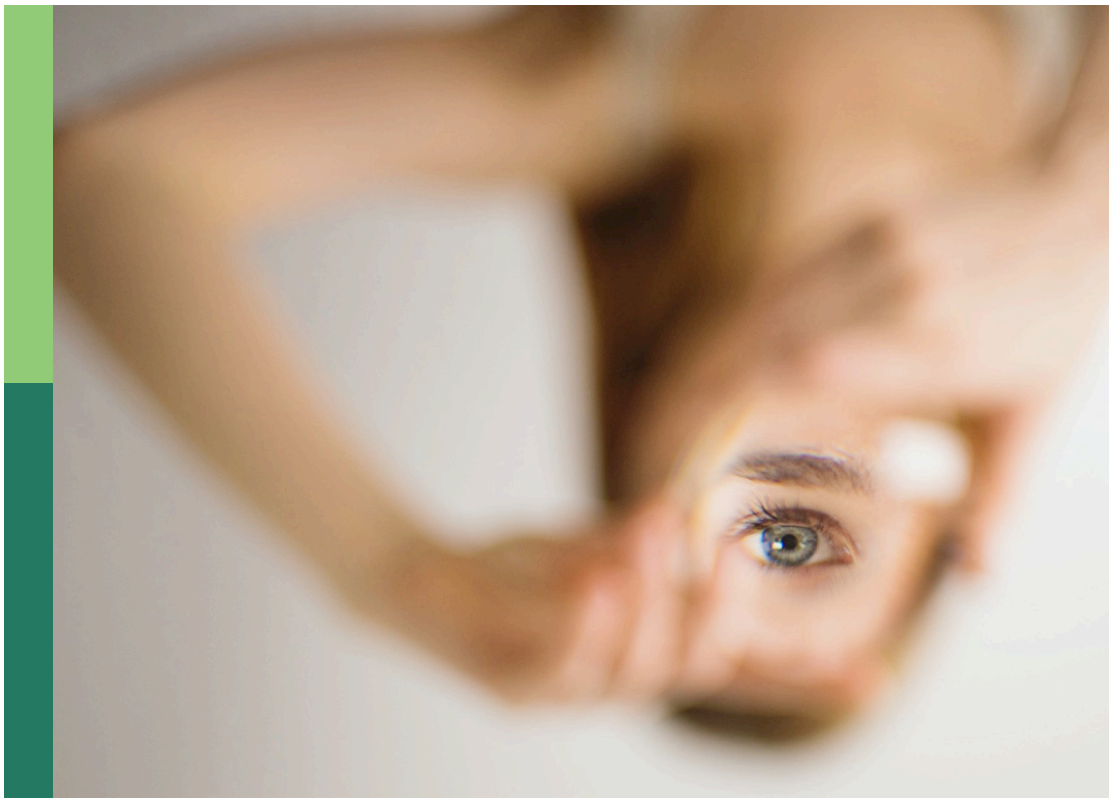
frontiers | Research Topics

Stem, steam, computational thinking and coding: Evidence-based research and practice in children's development

Edited by
Stamatios Papadakis, Michail Kalogiannakis and
Ali Ibrahim Can Gözüm

Published in
Frontiers in Psychology
Frontiers in Education



FRONTIERS EBOOK COPYRIGHT STATEMENT

The copyright in the text of individual articles in this ebook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this ebook is the property of Frontiers.

Each article within this ebook, and the ebook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this ebook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or ebook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-83251-168-8
DOI 10.3389/978-2-83251-168-8

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Stem, steam, computational thinking and coding: Evidence-based research and practice in children's development

Topic editors

Stamatios Papadakis — University of Crete, Greece
Michail Kalogiannakis — University of Crete, Greece
Ali Ibrahim Can Gözüm — Kafkas University, Türkiye

Citation

Papadakis, S., Kalogiannakis, M., Gözüm, A. I. C., eds. (2023). *Stem, steam, computational thinking and coding: Evidence-based research and practice in children's development*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83251-168-8

Table of contents

05 Editorial: STEM, STEAM, computational thinking, and coding: Evidence-based research and practice in children's development
Stamatios Papadakis, Michail Kalogiannakis and Ali İbrahim Can Gözüm

09 Applying Relatedness to Explain Learning Outcomes of STEM Maker Activities
Xiaojing Weng, Thomas K. F. Chiu and Morris S. Y. Jong

19 Children's Spatial Play With a Block Building Touchscreen Application
Naomi Polinsky, Breniel Lemley, Rachel M. Flynn, Ellen Wartella and David H. Uttal

34 Enhancing Digital Skills of Early Childhood Teachers Through Online Science, Technology, Engineering, Art, Math Training Programs in Estonia
Janika Leoste, Zsolt Lavicza, Kristof Fenyvesi, Maire Tuul and Tiia Õun

44 Effects of a Pair Programming Educational Robot-Based Approach on Students' Interdisciplinary Learning of Computational Thinking and Language Learning
Ting-Chia Hsu, Ching Chang, Long-Kai Wu and Chee-Kit Looi

59 Combined Effects of Block-Based Programming and Physical Computing on Primary Students' Computational Thinking Skills
Oliver Kastner-Hauler, Karin Tengler, Barbara Sabitzer and Zsolt Lavicza

71 Integrating Computational Thinking and Empowering Metacognitive Awareness in Stem Education
Nagalaxmy Markandan, Kamisah Osman and Lilia Halim

89 Developing Teaching Practice in Computational Thinking in Palestine
Abdel Ghani, David Griffiths, Soheil Salha, Saida Affouneh, Fakher Khalili, Zuheir N. Khlaif and Daniel Burgos

102 Evaluating Technology-Enhanced, STEAM-Based Remote Teaching With Parental Support in Luxembourgish Early Childhood Education
Ben Haas, Zsolt Lavicza, Tony Houghton and Yves Kreis

114 Educational Robotics Intervention to Foster Computational Thinking in Preschoolers: Effects of Children's Task Engagement
Anaclara Gerosa, Víctor Koleszar, Gonzalo Tejera, Leonel Gómez-Sena and Alejandra Carboni

127 How Might We Raise Interest in Robotics, Coding, Artificial Intelligence, STEAM and Sustainable Development in University and On-the-Job Teacher Training?
Jannik Henze, Carina Schatz, Shalina Malik and André Bresges

142 Retrieval Practices Enhance Computational and Scientific Thinking Skills
Osman Yaşar, Jose Maliekal, Peter Veronesi, Leigh Little, Michael Meise and Ibrahim H. Yeter

156 Promoting Secondary Students' Twenty-First Century Skills and STEM Career Interests Through a Crossover Program of STEM and Community Service Education
Biyun Huang, Morris Siu-Yung Jong, Ronnel B. King, Ching-Sing Chai and Michael Yi-Chao Jiang

167 Teaching Preschoolers Theory of Mind Skills With Mobile Games
Mariya Nikolayev, Anya S. Evmenova, Stephanie M. Reich, Kevin A. Clark and M. Susan Burns

185 Corrigendum: Teaching preschoolers theory of mind skills with mobile games
Mariya Nikolayev, Anya S. Evmenova, Stephanie M. Reich, Kevin A. Clark and M. Susan Burns

186 A Systematic Review of Technologies to Teach Control Structures in Preschool Education
Ewelina Bakala, Anaclara Gerosa, Juan Pablo Hourcade, Gonzalo Tejera, Kerry Peterman and Guillermo Trinidad

214 Exploring Gender Differences in Coding at the Beginning of Primary School
Chiara Montuori, Lucia Ronconi, Tullio Vardanega and Barbara Arfé

229 K–12 Science, Technology, Engineering, and Math characteristics and recommendations based on analyses of teaching cases in China
Yunxiang Zheng, Panpan Liu, Xinru Yang, Yidong Guo, Xinxin Qiu, Xiunan Jin, Xianfei Luo and Tianxiang Zheng

243 Preschool teachers' STEM pedagogical content knowledge: A comparative study of teachers in Greece and Turkey
Ali İbrahim Can Gözüm, Stamatios Papadakis and Michail Kalogiannakis

262 Comparing the psychometric properties of two primary school Computational Thinking (CT) assessments for grades 3 and 4: The Beginners' CT test (BCTt) and the competent CT test (cCTt)
Laila El-Hamamsy, María Zapata-Cáceres, Pedro Marcelino, Barbara Bruno, Jessica Dehler Zufferey, Estefanía Martín-Barroso and Marcos Román-González

frontiers | Frontiers in Psychology

TYPE Editorial
PUBLISHED 13 December 2022
DOI 10.3389/fpsyg.2022.1110476

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Dragan Lambić,
University of Novi Sad, Serbia

*CORRESPONDENCE
Ali İbrahim Can Gözüm
✉ a_ibrahimcan@hotmail.com

SPECIALTY SECTION
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

RECEIVED 28 November 2022
ACCEPTED 02 December 2022
PUBLISHED 13 December 2022

CITATION
Papadakis S, Kalogiannakis M and Gözüm AIC (2022) Editorial: STEM, STEAM, computational thinking, and coding: Evidence-based research and practice in children's development. *Front. Psychol.* 13:1110476. doi: 10.3389/fpsyg.2022.1110476

COPYRIGHT
© 2022 Papadakis, Kalogiannakis and Gözüm. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Editorial: STEM, STEAM, computational thinking, and coding: Evidence-based research and practice in children's development

Stamatios Papadakis[1], Michail Kalogiannakis[1] and Ali İbrahim Can Gözüm[2]*

[1]Faculty of Education, The University of Crete, Crete, Greece, [2]Dede Korkut Faculty of Education, Kafkas University, Kars, Turkey

KEYWORDS

Science Technology Engineering Mathematics (STEM), STEM enriched with arts (STEAM), computational thinking (CT), coding, education

Editorial on the Research Topic
STEM, STEAM, computational thinking, and coding: Evidence-based research and practice in children's development

First, we would like to congratulate the research authors and reviewers worldwide for their contributions to the literature with up-to-date scientific information. An overall view of the Research Topics shows that they present up-to-date methodology and results, generating new ideas and alternative perspectives that turn theoretical foundations into practical applications. Another aspect of this particular research compilation is that it contains articles from many different countries, including the United States (USA), Austria, China-Hong Kong, Germany, Greek, Italy, Luxembourg, Malaysia, Palestine, Portugal, Singapore, Taiwan, Turkey and Uruguay. Examination of the frequencies of the keywords used in the 18 different articles in this special issue (see Figure 1) reveals the collection's thematic paradigm.

It is clear from Figure 1 that the studies frequently use such keywords as Science Technology Engineering Mathematics (STEM), STEM enriched with arts (STEAM), STEM education, computational thinking (CT), robotics, coding, etc. Considering that the title of this special issue is *"STEM, STEAM, computational thinking, and coding: Evidence-based research and practice in children's development,"* we can see that the keywords reflect thematic approaches adopted by the studies. Moving forward, we chose the concept of STEM, one of the main keywords, as the critical keyword that best fits the purpose of the special edition and interpreted its relationship with other concepts. However, we also want to emphasize to our readers the importance of reviewing this special edition regarding the children, parents, and teachers that form the critical triangle in early childhood education, another of

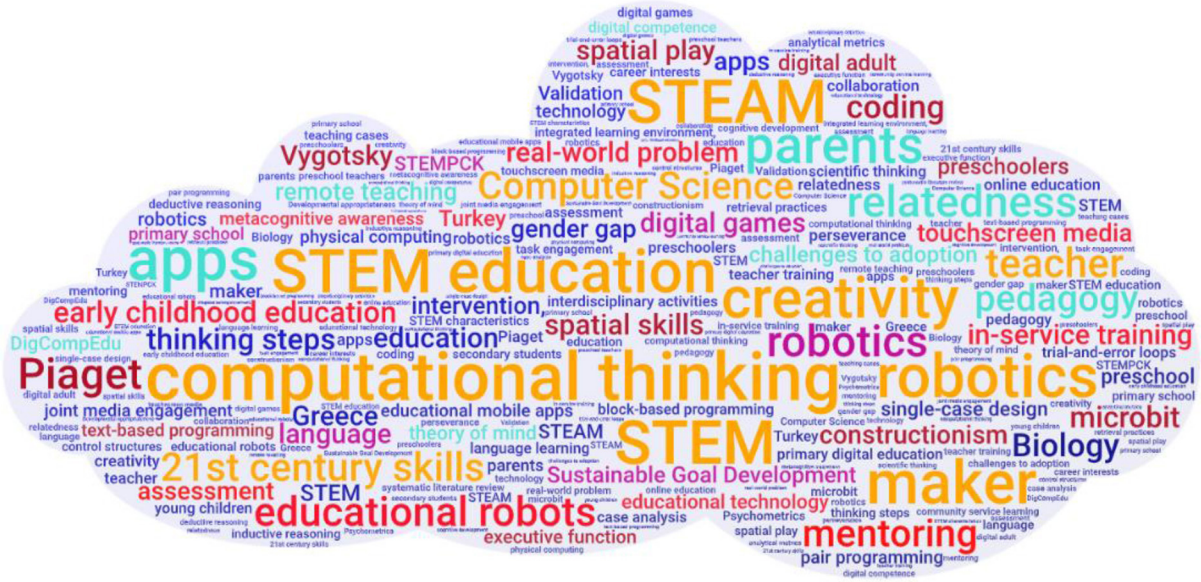


FIGURE 1
Cloud of research keywords.

the keywords in Figure 1. This being so, it would be helpful to briefly discuss the thematic paradigm of the research concerning children, parents, and teachers.

Regarding child development, STEM education emphasizes cognitive development (computational thinking, creativity, thinking steps, and metacognitive awareness) and twenty-first-century skills for the skills that children today need to have. However, we should keep sight of the fact that STEM activities involving real-world problems, another keyword seen in Figure 1, can support all areas of a child's development. This special edition presents new perspectives to researchers and readers by revealing new findings concerning the relationship between children and parents within the scope of STEM education. What kind of STEM education should children be given? The critical point in this question is the concept of teacher and education. Here, we can see that pedagogical content emerges with critical concepts in both pre-service and in-service teacher training and the teaching of pedagogical content knowledge (PCK) in line with STEM philosophy. As for the concept of teachers, the training programs and technology integration into them (educational robots, robotics, digital games, coding, etc.) also include central key concepts (see Figure 1). This special edition includes 18 research articles and the significant findings of different STEM-oriented research questions asked in these studies.

The first article in this special issue is the study called *"Applying Relatedness to Explain Learning Outcomes of STEM Maker Activities"* conducted by Weng et al.. This study was conducted with the participation of students and teachers in Hong Kong. The researchers examined the effects of STEM maker activities on learning outcomes. They found that maker activities affected the development of learners' non-cognitive characteristics, meaning their cognitive competencies such as STEM interest, STEM identity, and critical thinking by using the relationship between student and mentor and real-world problem (RWP) types.

The second article in this issue is the study *titled "Developing Teaching Practice in Computational Thinking in Palestine"* conducted by Ghani et al.. The researchers aimed to provide information about the challenges faced by CT teachers in K-12 schools in Palestine, the support they provided when incorporating CT strategies in their teaching, and the strategies they adopted when using CT approaches. They found that the most appropriate way to support teachers' CT presentations would be to provide peer exchange and expert coaching concerning the integration of CT into the curriculum.

The third study is an interesting one called *"Children's Spatial Play With a Block Building,"* conducted by Polinsky et al.. The researchers examined the interest of children aged 3–6 years living in the Chicago region of the United States in digital block games and the correlation between children's age, gender, and spatial skills in playing digital block games. Although they found significant differences by age and gender in children playing digital games, they found no correlation between spatial skills and digital games. They reported that physical and digital block games support similar play behavior in children.

Researchers in Luxembourg identified numerous problems in using educational technology in early childhood (STEAM) education. Haas et al. conducted a study called *"Evaluating Technology-Enhanced, Steam-Based Remote Teaching With Parental Support in Luxembourgish Early Childhood Education."* The researchers collected data from teachers, children, and parents *via* software during the parent-assisted distance learning process to examine the roles of parents in STEAM education and to determine child-parent interaction. They identified new roles in the parent-child relationship in distance STEAM teaching and new opportunities in the use of technology in early childhood education. They also proposed interesting ideas to provide technical knowledge and support for factors that affect children's and parent's perceptions and motivations in distance learning.

The fifth article is the study *"Integrating Computational Thinking and Empowering Metacognitive Awareness in Stem Education"* by Markandan et al.. The study, conducted to support the metacognitive skills of biology students in Malaysia, found significant differences in student achievement by examining their skills at programming the Me-Cot learning module based on four learning theories.

The study *called "How Might We Raise Interest in Robotics, Coding, Artificial Intelligence, STEAM and Sustainable Development in University and On-the-Job Teacher Training?"* by Henze et al. is the sixth article in this issue. The researchers researched the project known as "Qualitätsoffensive Lehrerbildung," a joint initiative between the Federal Government and the individual States in Germany to improve teacher education quality. The researchers determined the positive effects of the 5E approach on in-class STEM practices based on what pre-service teachers, teachers, and students did on the job according to the 5E approach, the Biological Sciences Curriculum Study, and the pedagogic concept based on the STEM paradigm.

Nikolayev et al. conducted a single case study with six female participants aged 46–52 months in the USA. The study examined the effect on children's theory-of-mind skills of interactive touch screen apps based on the theory of mind. The study called *"Improving Preschoolers' Theory of Mind Skills With Mobile Games"* found that when children are supported in the use of mobile apps by a conversation they are to have with the help of adults, this can support the development of theory-of-mind skills in preschoolers.

The eighth study, called *"Combined Effects of Block-Based Programming and Physical Computing on Primary Students' Computational Thinking Skills,"* was conducted by Kastner-Haule et al.. By conducting a longitudinal and multi-stage study, the researchers support the increasing importance of computational thinking skills in Austrian classrooms. The study found that block-based programming applications can yield positive results at an early age in support of computing skills.

The ninth article in this issue is the research called *"Exploring Gender Differences in Coding at the Beginning of Primary School,"* conducted by Montuori et al.. Children's coding skills, response blocking, and planning skills were evaluated following the coding training given to first-year students in Italy. The researchers found that the final test results for coding skills following the coding training favored boys. The researchers examined the mediating role of differential executive functions (EF) on gender in coding and found that gender difference had no mediating role effect on EF in coding.

The study titled "Effects of a Pair Programming Educational Robot-Based Approach on Students' Interdisciplinary Learning of Computational Thinking and Language Learning" conducted by Hsu et al. is the 10th article in this issue. This research was conducted with the participation of children in Singapore who learned Chinese as a Second Language (CSL) in Singapore and children in Taiwan who learned English as a Foreign Language (EFL). The study's main result was that the children in the EFL group showed better collaborative social skills in Computational Thinking (CT) skills and lower learning anxiety in learning the target language (TL). While the children in the CSL group had better problem-solving skills in CT, they were observed to exhibit more trial-and-error cycle behaviors.

The study called *"Retrieval Practices Enhance Computational and Scientific Thinking Skills,"* conducted by Yaşar et al. is the 11th article in this issue. Researchers conducted class activity research on teaching STEM concepts and CT to secondary school teachers in the USA and how they learned them. The research found that the majority of teachers (96%) understood the retrieval strategies well and how the in-class applications of the related ideas could be tested. The research results can make a significant contribution to the literature in terms of examples of retrieval practices. It is an exciting article that those looking for good practices, especially for developing CT and STE skills, will find helpful.

The study called *"Enhancing Digital Skills of Early Childhood Teachers Through Online Science, Technology, Engineering, Art, Math Training Programs in Estonia,"* conducted by Leoste et al. is the 12th article in this issue. The researchers examined the effects of online activities for early childhood teachers based on STEAM-integrated learning activities. They reported that early childhood educators would increase their digital competencies faster by receiving online training than face-to-face courses.

The 13th article in this issue is the study called *"Promoting Secondary Students' Twenty-First Century Skills and STEM Career Interests Through a Crossover Program of STEM and Community Service Education"* conducted by Huang et al.. After providing 8 weeks of STEM education to secondary school students as part of community service education. The researchers examined the improvement in Hong Kong mass housing residents' problem-solving skills using the information they gained from the real-world problems faced by disadvantaged groups. They observed positive development in the children's creative thinking, cooperation, and STEM career interests.

The study *"Educational Robotics Intervention to Foster Computational Thinking in Preschoolers: Effects of Children's Task Engagement"* by Gerosa et al. makes up the 14th article in this issue. The researchers conducted a quasi-experimental study based on an educational robotics app with children attending a public kindergarten in Uruguay. They examined the improvement in the task commitment, distraction, verbal participation, and goal realization skills of the children in the experimental group using an app called RoboTito. An overall view of the study shows that commitment to task supports children's learning and that robotic intervention supports the development of computational thinking.

The study called *"A Systematic Review of Technologies to Teach Control Structures in Preschool Education"* conducted by Bakala et al. is the 15th article in this issue. The researchers analyzed empirical evidence by examining the tools that constitute those programs that include the control structures of preschool children and by a systematic literature review of the roles played by the tools in question in teaching the control structures of early childhood children. The study is an exciting article for discovering the technologies for teaching children's control structures.

The study called *"Preschool teachers' STEM pedagogical content knowledge: A comparative study of teachers in Greece and Turkey"* by Gözüm et al. is the 16th article in this issue. The researchers compared the STEM pedagogical content knowledge (STEMPCK) levels of Greek and Turkish preschool teachers and determined no significant difference between the STEMPCK scores of Greek and Turkish teachers. The study also revealed the critical importance of STEM education in the STEMPCK differentiation of Greek and Turkish teachers.

The 17th article in this issue is the study called *"K-12 Science, Technology, Engineering, and Math characteristics and recommendations based on analyses of teaching cases in China"* conducted by Zheng et al.. The researchers examine STEM applications in different regions of China from the program's perspective and offer suggestions for practice. They suggest that STEM education in China should be localized for the integrated interdisciplinary application of STEM education and innovative STEM practices to be effective.

The study called *"Comparing the psychometric properties of two primary schools Computational Thinking (CT) assessments for grades 3 and 4: the Beginners' CT test (BCTt) and the competent CT test (cCTt)"* by El-Hamamsy et al. is the 18th and final article in this issue. The researchers compare psychometric characteristics based on the age validity of the data collection tools used to measure children's Computational Thinking skills. They report that the CT test (BCTt) was more beneficial for beginners in identifying children with low abilities.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Psychology

ORIGINAL RESEARCH
published: 17 January 2022
doi: 10.3389/fpsyg.2021.800569

Applying Relatedness to Explain Learning Outcomes of STEM Maker Activities

Xiaojing Weng[1,2], *Thomas K. F. Chiu*[1,2]* *and Morris S. Y. Jong*[1,2]

[1] Department of Curriculum and Instruction, Faculty of Education, The Chinese University of Hong Kong, Hong Kong, Hong Kong SAR, China, [2] Centre for Learning Sciences and Technologies, Faculty of Education, The Chinese University of Hong Kong, Hong Kong, Hong Kong SAR, China

A growing interest has been observed among K-12 school educators to incorporate maker pedagogy into science, technology, engineering, and mathematics (STEM) education to engage students in the design and making process. Both cognitive engagement and emotional engagement of students can be promoted through satisfying the psychological need of relatedness that concerns a sense of connection and belonging. How to support relatedness would influence the effective development of students' cognitive competencies, namely creativity and critical thinking, and non-cognitive characteristics, namely interest and identity. Therefore, the present study investigated how two relatedness support strategies—real-world problems (RWP) and mentoring influence the development of student's STEM-related cognitive competencies and non-cognitive characteristics in STEM marker activities. We implemented a 7-week intervention study with three classes of Grade 9 students (aged 13–15 years) in Hong Kong (n = 95). Three intervention conditions were designed in the experiment, comprising textbook problem (TBP), RWP, and RWP with mentoring (RWPM). Our analysis showed that (i) the differences in creativity among the three groups were non-significant, (ii) the RWP and RWPM groups showed stronger critical thinking than the TBP group, and (iii) the RWPM group exhibited stronger STEM interest and identity than the other two groups. This study revealed the effectiveness of adopting RWP strategy in developing secondary students' perceived cognitive competencies (e.g., creativity and critical thinking) and the feasibility of employing a mentoring mechanism for cultivating learners' perceived non-cognitive characteristics (e.g., STEM identity and interest). Hence, we also offered practical suggestions for teachers.

Keywords: STEM education, maker, relatedness, mentoring, real-world problem, 21st century skills

OPEN ACCESS

Edited by:
Stamatios Papadakis, University of Crete, Greece

Reviewed by:
Zhi Hong Wan, The Education University of Hong Kong, Hong Kong SAR, China
Michail Kalogiannakis, University of Crete, Greece

***Correspondence:**
Thomas K. F. Chiu
tchiu@cuhk.edu.hk

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

Received: *23 October 2021*
Accepted: *07 December 2021*
Published: *17 January 2022*

Citation:
Weng XJ, Chiu TKF and Jong MSY (2022) Applying Relatedness to Explain Learning Outcomes of STEM Maker Activities. Front. Psychol. 12:800569. doi: 10.3389/fpsyg.2021.800569

INTRODUCTION

After the release of the Make magazine in 2005 and the hosting of the first Maker Faire in 2006, the maker movement started to gain momentum into becoming a worldwide phenomenon (Sang and Simpson, 2019). This movement is a cultural trend focused on creating makers rather than consumers of products in the 21st century (Marshall and Harron, 2018) and advocates for creativity, excitement, and innovation (Bevan et al., 2015; Papadakis, 2021). Within this movement, individuals can use different tools and materials to present their ideas through physical products

such as prototypes and artifacts that they feel are relevant and interesting. This maker-centered approach has been applied in science, technology, engineering, and mathematics (STEM) learning and teaching (Honey et al., 2014; Godhe et al., 2019) primarily by means of after-school or extracurricular activities, and in libraries, museums, or playgrounds.

The incorporation of maker education in STEM learning and teaching is considered as contemporary pedagogy that embraces collaboration, experimentation, and open-endedness (Nemorin and Selwyn, 2017; Godhe et al., 2019); however, this pedagogy is less established in classrooms and relatively new to most schools and teachers (Honey et al., 2014). How to use maker approach to better engage in STEM learning warrants further exploration. Student engagement can be motivated by supporting a psychological need—relatedness (a sense of connection and belonging)—posited by Self-determination theory (SDT) that is a motivation theory (Ryan and Deci, 2020). To satisfy the need for relatedness, teachers could use authentic or real-world examples and assignments to establish students' perceived relevance and connection of the learning materials (Hung, 2016), and adopt mentorship programs to foster stronger students' belonging by developing student–expert relationships (Simões and Alarcão, 2014). Accordingly, real-world problems (RWPs) and mentoring would better motivate student engagement than textbook problem (TBP) in STEM maker activities.

Incorporating RWPs into STEM learning activities is more likely to expose students to authentic problems, which connects content with their daily life and lead to greater cognitive and emotional engagement. The inclusion of authentic problems will require students to solve ill-defined problems that are complex and cognitively challenging (English and Mousoulides, 2015; English, 2016; Woods and Hsu, 2020). The solving process requires students to process volumes of different information and brainstorm solutions. Students are expected to collaborate creatively, perform critical thinking, and communicate to propose, devise, and evaluate solutions to address the problems. The relevance of and connection with RWPs encourage students to be aware of the choice of solutions they arrive at and how their choices fit into a societal context in which they feel loved and care for, i.e., support the need for relatedness (Hung, 2016). Therefore, how maker pedagogy support relatedness could determine how effectively students' cognitive competencies such as creativity and critical thinking can be developed. Moreover, mentorship programs support the need for relatedness, allow students to avail themselves of student–expert relationships, and to connect with their mentors (Simões and Alarcão, 2014). As mentees, students engage in different interactions with mentors in course of their learning process (Tofel-Grehl et al., 2017). They receive mentor's guidance to shape their ideas, endorsement of their choices, and recognition of their efforts and works. This process could facilitate positive non-cognitive characteristics, such as interest and identity development in a social environment (Schlegel et al., 2019; Nganga et al., 2020). Relationships with mentors could influence the positive development of students' STEM interest and identity (Nganga et al., 2020). However, the adoption of mentoring that supports relatedness (Simões and Alarcão, 2014) for promoting these non-cognitive characteristics has been less discussed in K-12 education (Honey et al., 2014), because mentorship programs require numerous resources involving an enormous number of school students.

In sum, further research to understand how the pedagogical designs of maker education to use RWPs and mentoring to support relatedness (Simões and Alarcão, 2014; Hung, 2016) in K-12 formal schooling settings is required (Honey et al., 2014; Godhe et al., 2019). Creativity, critical thinking, interest, and identity are the four major outcomes of STEM education in K-12 STEM education (Honey et al., 2014; Johnson et al., 2020). These vital features are necessary for the identification of young children as STEM makers and for their adaptation to the future society (Chiu et al., 2020; Chiu and Lim, 2020). Accordingly, the present study aimed to explain how RWPs and mentoring influence the development of student's cognitive competencies and non-cognitive characteristics from a different perspective: relatedness.

First, we present our conceptual framework, and critically discussed previous studies on cognitive competencies and non-cognitive characteristics as learning outcomes of maker pedagogy. Then, we state the goal of this study, and its three research questions, following by describing research design and procedure. Final, we present the results answering the questions, and a discussion of the findings with limitations.

LITERATURE REVIEW

Maker Pedagogy in STEM Education

Making is an essential human activity because "we must make, create, and express ourselves to feel whole" (Hatch, 2014, p. 11). The concept of making is well-accepted as minds-on and hands-on learning in STEM education. Many schools have adopted making as a pedagogical approach with strong emphasis on designing, doing, and creating to nurture students to be creative STEM thinkers (Halverson and Sheridan, 2014; Godhe et al., 2019; Suh et al., 2020).

Creativity refers to the ability to generate multiple solutions to a problem through divergent thinking rather than one solution through convergent thinking (Daly et al., 2014). Maker pedagogy may allow students to explore a problem as a team, exchange ideas among their members, and build and rebuild their ideas. Furthermore, creativity acts as a vital tool that helps students put their new ideas into practice and develop their creative competence through innovation (Papadakis, 2021; Xia et al., 2021). Literature suggests that STEM maker projects can effectively cultivate students' creativity in the school context (Saorín et al., 2017; Searle et al., 2018). For example, students can work in groups and share their ideas with peers to make creatively designed dolls using 3D printing (Saorín et al., 2017). Similarly, high school students in a circuit creation workshop had new ideas for their artifacts of electronic textiles when they were given opportunities to brainstorm different solutions and receive ideas from others (Searle et al., 2018). In general, maker pedagogy fosters a creative culture and play a positive role in helping students imagine innovative new possibilities.

Maker pedagogy, including hands-on activities and inquiry-based learning, is fundamentally linked to experiential learning (Honey et al., 2014; Suh et al., 2020). This type of pedagogy is based on the constructivist paradigm that students learn through experiencing and reflecting on those experiences (Martin, 2015). Teachers set the stage for students with real-life challenges, and students identify problems, design different possible solutions, and propose evaluation methods related to the challenges (Gettings, 2016; Chiu et al., 2021b). The students engage in in-depth investigations with materials, objects, and ideas and draw meaning and understanding from their experiences in a fun manner, in order to solve challenges with originality and imagination. The entire solving process includes observations and reflections. They are required to recruit and coordinate personal, social, and material resources for meaningful learning participation, i.e., they should have abilities to engage with sophisticated practices (Brahms and Crowley, 2016). Therefore, making as learning process for young children requires their access and evolving relationship to teacher/mentor assistance and expertise. These suggest that relatedness satisfaction would become more crucial in solving RWPs in STEM making activities.

Textbook Problem, Real-World Problem, Mentoring, and Relatedness in Self-Determination Theory

Maker approach is less established pedagogy being adopted in STEM classrooms (Honey et al., 2014). How student engagement in maker learning activities affect the development of cognitive competencies such as collaboration, creativity, and critical thinking and non-cognitive characteristics such as enthusiasm, interest, and identity remains unclear (Honey et al., 2014; Martin, 2015). Student engagement can be motivated by satisfying relatedness, a psychological need that refers to a sense of connection and belonging suggested in Self-determination theory (SDT) (Ryan and Deci, 2020). In the course of learning, students' behavioral, cognitive and emotional engagement can be speculated by the relatedness of the learning context (Reeve, 2013; Chiu, 2021b,c; Chiu et al., 2021a). It is because desirable teacher-student relationships can promote students' active participation in maker activities (i.e., behavioral engagement), stimulate their positive feelings when pursuing the activities (i.e., emotional engagement), and make them more confident in accomplishing the challenging tasks therein (i.e., cognitive engagement) (Chiu, 2021b,c). In this paper, TBPs refer to the problem are not in real-life context. Accordingly, compared to TBPs, RWPs and mentoring are approaches to support need for relatedness: Students would perceive stronger relevance and connection of the learning materials when using RWPs, and would develop stronger belongings and student–expert relationships when adopting mentoring approach in learning activities (English and Mousoulides, 2015; Chiu, 2021a; Chiu et al., 2021a).

In TBPs students are often given specific ill-defined problems (Hanif et al., 2019). However, students may not be required to identify issues and formulate their own problems in real-life context. They may follow teachers' instruction, and replicate peers' ideas to solve the TBPs that are less complicated and authentic than RWPs. Therefore, students would find TBPs and their solutions less relevant, connected, and ownerships (Lee et al., 2013). In RWPs, students are afforded opportunities to make connections between STEM concepts and real-life applications (Achmetli et al., 2019), such as building own surgical masks and designing own smart home appliances. Their lived experiences and outcomes from authentic context-based activities and tasks provide some encouragement in working through these issues (Lee et al., 2013). Allowing students to identify their own issue in an RWPs and suggest solutions develop a strong sense of belonging and ownership to the activities and tasks (Lee et al., 2013; Achmetli et al., 2019), which support the need for relatedness. Accordingly, TBPs and RWPs are both problem-based learning, and enhance students' creativity (Hanif et al., 2019).

As discussed in last section, young children require their access and evolving relationship to teacher/mentor assistance and expertise for making activities, particularly for more complex problems, i.e., RWPS. Effective mentorship occurs when mentors and mentees develop trust, and identify with and authentically engage with one another (Stoeger et al., 2013; Tenenbaum et al., 2014). They can speak freely and express ideas without concerns for interpersonal comfort (Chiu, 2021b; Chiu et al., 2021a). The effective mentor-mentee relationships allow students to interact and be connected to their methods, which lead to an enhanced sense of relatedness and effectiveness in the mentee (Ryan and Deci, 2020). Mentees benefit from engaging with mentors who share expert knowledge and experiences.

Accordingly, RWPs and mentoring that support relatedness, develop better 21st century cognitive competencies such as creativity, and critical thinking, and more positive non-cognitive learning outcomes such as interest, and identity (Ryan and Deci, 2020; Chiu, 2021b).

Critical Thinking and Real-World Problems

Critical thinking is the ability to think clearly and rationally to form a judgment. It requires students to analyze complex problems, make connections across different disciplines, and evaluate solutions (Hu et al., 2020). STEM disciplines often work together seamlessly in the real world. Introducing students to RWPs enables them to see the connection between "what's inside" and "what's outside" their classroom; they see that their learning is more than giving correct answer to tests and getting good grades. The RWPs build bridges beyond the classroom by connecting discipline content with daily life, offering learning opportunities with local and global communities, showing everyday applications of learning, and engaging students in authentic learning that are meaningful. These result in greater cognitive engagement that facilitates their critical thinking advancements (Hu et al., 2020). For instance, Hollman et al. (2019) reported that 645 secondary school students engaged with real world problems that improved their critical thinking; English and Mousoulides (2015) real world problems facilitated the development of Grade 6 students' critical thinking. It may be because real world problems are cognitively challenging

(English and Mousoulides, 2015; Woods and Hsu, 2020) and require students to carefully consider all the decisions concerning their solutions and how the solutions fit into societal contexts (English, 2016). To solve these problems, students need to collaborate creatively and think critically to propose possible solutions. These RWPs are more relevant to students than textbook problems, for example, building a bridge to connect a highway. The relevance of RWPs can support students need for relatedness by connecting problems to their life and own understanding. i.e., it encourages students to perform comprehensive research, seek advice and feedback from others, and employ their critical thinking. In addition to cognitive competencies (e.g., critical thinking), maker pedagogy should aim to cultivate students' non-cognitive characteristics such as STEM interest and identity (Fasso and Knight, 2020).

STEM Interest and Identity and Mentoring

STEM interest and identity are two critical predictors of students' higher studies and career plans in STEM-related fields (Hurk et al., 2019). This is because they reflect individuals' self-images and enables them to derive personal meaning from their endeavors in the STEM field. Hence, fostering students' STEM interest and identity are two major learning outcomes of maker pedagogy in STEM classrooms (Cannady et al., 2014; Honey et al., 2014; Hurk et al., 2019). Interest is often characterized in terms of curiosity, persistence, and resourcefulness (Hidi and Renninger, 2006), and develops over time—it germinates when people's attention is triggered and develops through voluntary engagement or re-engagement (Renninger and Hidi, 2011). The development of students interest can be guided (Renninger, 2010). Interested students have stronger feelings of self-efficacy and can cultivate better self-regulated behaviors to persevere on with challenging tasks (Hidi and Ainley, 2008). Maker pedagogy that promotes student autonomy in terms of allowing them to decide which solutions to arrive at is postulated to increase student motivation for STEM learning because of the sense of personal ownership that students feel throughout the design, making, and testing process. Furthermore, maker pedagogy has been demonstrated to be effective in increasing students' interest and engagement in science and math classes (Gerber et al., 2012; Honey et al., 2014; Holmlund et al., 2018).

Identity refers to who one is and how one is recognized by others. A person's identity is shaped by how they are recognized in a given context, with particular interests, talents, and ways of being in particular social contexts (Honey et al., 2014; Goos and Bennison, 2019). Identity development is a matter of finding oneself by matching one's talents and potential with available social roles (Goos and Bennison, 2019). Recognition from other people, which determines the credibility and value of a person's performance, is a necessary component of identity formation. A few studies have suggested that integrated experience from project- and problem-based learning can promote identity development. For example, Hachey et al. (2021) suggested that the flexibility and openness of STEM learning activities that allow students to define their own designs are influential factors in fostering a STEM-related identity. However, this was a very preliminary finding (Honey et al., 2014). Maker pedagogy alone may not effectively foster STEM identity.

Introducing role models or mentors in STEM education can enhance students' STEM interest and identity development (Honey et al., 2014; Nelson et al., 2017; Huvard et al., 2020). In mentorship programs, students often engage in three types of interactions with mentors: shareability, tangibility, and aesthetic (Tofel-Grehl et al., 2017). Shareability enables students to obtain access to various resources, better understand problems, and make creative prototypes under their mentors' guidance (Kreider et al., 2018). Tangibility allows students to forge personal connections to internalize their interest and identity (Fasso and Knight, 2020). Aesthetic allows for students' efforts and/or works to be endorsed by mentors, and this acceptance further helps students to recognize their identity in a social environment. These interactions trigger and sustain interest and lead to the development of a stronger identity (Tofel-Grehl et al., 2017). Studies have suggested that the mentoring approach can better develop students' STEM identity. For example, Ladeji-Osias et al. (2018) and Musavi et al. (2018) revealed under-represented minority students developed a stronger STEM identity and a stronger desire to pursue STEM careers in a mentorship program; Pinkard et al. (2017) found that middle school female student-mentees developed a stronger interest and identity toward STEM. In summary, adding mentoring to maker pedagogy would lead to the development of stronger STEM interest and identity.

THE PRESENT STUDY

Supporting the need for relatedness that better cognitively and emotionally engages students in learning may better develop critical thinking, and enable the positive formation of interest and identity. This paper constitutes an interventional study that investigates whether RWPs and mentorships in maker pedagogy can influence the development of learners' cognitive competencies, namely creativity and critical thinking, and non-cognitive characteristics, namely interest and identity, in K-12 school classrooms. To achieve the research goal, testing the intervention is the main research task; therefore, we employed an experimental method with three interventional conditions, namely TBP, RWP, and RWP with mentoring (RWPM). The TBP group used an ill-defined problem in non-real-life context; the RWP group used an ill-defined problem in real-life context; the RWP group used an ill-defined problem in real-life context and mentors. Accordingly, the following research questions were proposed:

RQ1. Are there differences among the three groups in terms of perceived cognitive competencies?
RQ2. Are there differences among the three groups in terms of perceived non-cognitive characteristics?
RQ3. Do the three groups have stronger cognitive competencies and non-cognitive characteristics?

We hypothesized the following: no significant difference exists in terms of creativity among the three groups (H1); the RWP and

TABLE 1 | Intervention.

Time and theme	Activity examples
Week 1 Meet Arduino and my first Arduino program	• Teachers introduced Arduino I/O circuit board and computer programming interface to students. • Student groups ◦ Learned to connect the Arduino board to the laptop with the provided USB cable. ◦ Worked in groups to follow instructions on a worksheet to finish their first Arduino program (lighting up the LED). ◦ Explored other extension programs.
Week 2 Basic knowledge 1	• Teachers acted as facilitators. • Student groups ◦ Modified the program and observed the changes in the brightness of the LED lights. ◦ Created extension programs to control the LEDs at three different brightness levels by using the push buttons.
Week 3 Basic knowledge 2	• Teachers acted as facilitators. • Student groups ◦ Explored a way to connect the variable resistor. ◦ Adjusted the variable resistor to observe the number changes. ◦ Designed a variable by recording the input voltage. • Teachers guided students to review the mathematical knowledge about proportions. • Student groups ◦ Designed a new variable to control the brightness of the LEDs and use variable resistance to change it. ◦ Tried to write extension programs according to the instructions on the worksheet.
Week 4 Basic knowledge 3	• Teachers acted as facilitators. • Student groups ◦ Wrote/rewrote a new program to control the brightness of LEDs according to the brightness of the surrounding environment. ◦ Tried to write extension programs according to the instructions on the worksheet.
Week 5–7 Problem-solving with Arduino	• Teachers acted as facilitators. • In the TBP groups, teachers introduced and reintroduced different suggested traffic light systems as problems in Weeks 1 and 5. • In the RWP and TWPM groups, teachers explained an existing situation of public traffic light systems in the community. For example, the public traffic light is turned on by workers at 6 pm every evening with a fixed brightness level. Students introduced and reintroduced a real-world problem—propose solutions for specific areas or occasions in Weeks 1 and 5. • Students in TWPM groups were introduced to their mentors in Week 2.

RWPM groups develop stronger critical thinking skills (H2); the RWPM group develop stronger STEM interest (H3) and STEM identity (H4); and the three groups develop stronger cognitive competencies and non-cognitive characteristics (H5).

METHODOLOGY

Participants

The participants comprised 95 Secondary Three students, aged between 13 and 15 years, and three teachers from three different schools in Hong Kong with similar academic performance. The schools were randomly assigned to three interventional conditions: TBP (n = 32), RWP (n = 31), and RWPM (n = 32). The average teaching experience of the teachers was 5 years. Eight undergraduate student mentors pursuing STEM-related majors were recruited to facilitate the intervention.

Procedures

We got the ethical approval from university, and first obtained the consent of the students and their parents. Before the intervention, we conducted two 3-h workshops on Arduino kits for making pedagogy in STEM education for all the teachers, conducted two co-planning sessions with the teachers, recruited all the student mentors, and administered the pre-questionnaire among the students. The intervention was conducted in 14 lessons for 7 weeks. During the intervention, all the students learned in group of three to four. Student in TBP used a textbook-like problem—assembling Arduino kits to make different traffic light systems for their learning. The RWP group learned through a real-life problem—traffic light system for a specific area, they identified their own areas or occasions and proposed solutions. Students in RWPM used the same real-life problem for their project as students in RWP with mentoring from eight mentors, i.e., each student team was guided by one mentor. The students finished the post-questionnaire in the last lesson. **Table 1** outlines the pedagogies of the three groups. In Week 1, the teachers introduced Arduino I/O circuit board and computer programming interface to students. From Week 2 to Week 4, the teachers acted as facilitators to foster student understanding of STEM prerequisite knowledge by asking questions, giving feedback, and monitoring learning progress. The prerequisite knowledge of Science, Technology, Mathematics were electricity (Physics), coding and ratios respectively. Engineering design

process was adopted. From Week 5–7, different student groups learned with different problems—TBP, RWP and RWPM. The problem in RWP and RWPM groups is a design task. In RWPM, all the students worked collaboratively, guided and facilitated by their mentors, to design solutions for their problems. The mentors encouraged their mentees, helped with problem solving, and used active-listening techniques, and served as a guide for mentees' behavior, values, and attitudes. The teachers joined the discussions if necessary. Compared with the teachers, the mentors had much more intimate and frequent communications with students.

Measures

Questionnaires were used to measure four variables in the before and after the intervention, namely creativity, critical thinking, STEM interest and identity. Each of the variables was measured by three items using a 5-point Likert scale that solicited ordinal responses from 1 (strongly disagree) to 5 (strongly agree). All the questions were adapted from the previously published instruments.

Creativity and critical thinking were measured using the items from the study of Kelley et al. (2019) with good reliabilities (e.g., creativity $\alpha = 0.75$, critical thinking $\alpha = 0.88$). The participants of Kelly and colleagues were K-12 students, which fit the participant of this study. The items were modified to fit the research goals designed for creativity, for example, I am confident in my ability to understand how knowledge or insights might transfer to other situations or contexts, and I am confident in my ability to elaborate and improve on ideas. The examples of the items that evaluated critical thinking were: I am confident in my ability to evaluate reasoning and evidence that support an argument, and I am confident in my ability to justify choices of evaluation criteria.

STEM interest was measured using the items of STEM Semantics Survey Scales designed by Tyler-Wood et al. (2010) with acceptable reliability (e.g., α of Science, Math, Engineering, Technology, and STEM as a career interest ranged from 0.84 to 0.93). Their assessment target was participants' perceptions of scientific disciplines, which was similar to this study purpose. The items were revised to match our research context; for instance, I find STEM fascinating, I find STEM exciting, and I find STEM appealing.

STEM identity was evaluated using the three items of identification adopted in the study of Godwin (2016): the constructs of competence, self-recognition, and recognition by others. Godwin's study investigated students' identity in STEM-related domain. The original items were adapted to fit our research situations through statements such as I am able to do well in activities that involve STEM, I see myself as a STEM individual, and My best friends see me as a STEM individual.

Moreover, our analysis demonstrated that the reliability of all variables ranged from 0.78 to 0.95 in this study and was acceptable [Cronbach alpha (CA) ≥ 0.7, see **Table 2**; Taber, 2018].

Analysis Approaches

A one-way ANCOVA was utilized to examine how the three pedagogical designs affect participants' perceived cognitive competencies and non-cognitive characteristics while controlling the pre-questionnaire scores, to answer RQ1 and RQ2. The pedagogical design was the independent variable and the pre- and post-questionnaire scores were covariates and dependent variables, respectively. To answer RQ3, paired-sample *t* tests were conducted to examine whether perceived cognitive competencies and non-cognitive characteristics improved in the three groups by comparing the mean scores of their pre- and post-questionnaires.

RESULTS

Descriptive Statistics

Table 3 presents the descriptive statistics of creativity, critical thinking, STEM interest, and identity in both the two questionnaires. Our analyses revealed there was no significant difference among the pre-questionnaire scores of creativity, F (2, 94) = 0.10, $p = 0.91$; critical thinking, F (2, 94) = 0.57, $p = 0.57$; STEM interest, F (2, 94) = 0.38, $p = 0.69$; and STEM identity, F (2, 94) = 1.81, $p = 0.17$. These scores showed that the three experimental groups perceived cognitive competencies and non-cognitive characteristics in similar way before the intervention.

RQ1: Differences in Perceived Cognitive Competencies

Creativity

According to our Levene's test, homogeneity existed with $F = 1.17$, $p = 0.31$. One-way ANCOVA showed no significant differences in students' creativity among the three groups in the post-questionnaire by excluding the effect of their pre-questionnaire scores, F (2, 91) = 0.22 and $p = 0.80$ (see **Table 4**).

Critical Thinking

Levene's test was conducted and the assumption was validated, $F = 2.93$, $p = 0.06$. The analysis revealed significant differences in students' critical thinking after the intervention, with F (2, 91) = 8.41, $p < 0.001$. A *post hoc* comparison showed no significant improvements for the RWP group (adjusted mean = 3.78) over the TBP group (adjusted mean = 3.21, p = 0.04) and the RWPM group (adjusted mean = 4.10) over the TBP group ($p < 0.001$).

RQ2: Differences in Perceived Non-cognitive Characteristics

STEM Interest

Homogeneity existed with $F = 1.29$, $p = 0.28$. The analysis was significant, with F (2, 91) = 14.86, $p < 0.001$. A *post hoc*

TABLE 2 | Scale characteristics.

Scales	CA in pre-questionnaire	CA in post-questionnaire
Creativity	0.83	0.78
Critical thinking	0.88	0.95
STEM interest	0.93	0.88
STEM identity	0.90	0.92

TABLE 3 | Descriptive statistics and paired *t*-test for pre- and post-questionnaires.

		Pre-questionnaire		Post-questionnaire		Paired *t*-test
Groups	**Variables**	**Mean**	**SD**	**Mean**	**SD**	
TBP (*N* = 32)	Creativity	2.79	0.59	4.30	0.43	−12.15***
	Critical thinking	2.90	0.84	3.22	0.74	0.12
	STEM interest	2.85	0.80	3.84	0.56	−7.11***
	STEM identity	2.89	0.83	3.19	0.69	−3.26***
RWP (*N* = 31)	Creativity	2.85	0.72	4.27	0.49	−9.37***
	Critical thinking	3.08	0.65	3.77	0.82	−3.62***
	STEM interest	2.87	0.85	3.85	0.65	−6.95***
	STEM identity	2.67	0.56	3.18	0.38	−5.36***
RWPM (*N* = 32)	Creativity	2.78	0.64	4.22	0.57	−8.90***
	Critical thinking	2.98	0.46	4.10	1.03	−5.59***
	STEM interest	3.02	0.88	4.60	0.68	−8.50***
	STEM identity	3.02	0.80	4.60	0.45	−13.12***

*$p < 0.05$; **$p < 0.01$; and ***$p < 0.001$.

comparison indicated that the RWPM group (adjusted mean = 4.58) significantly outperformed the TBP group in the post-questionnaire (adjusted mean = 3.86, $p < 0.001$), and the RWP group (adjusted mean = 3.86, $p < 0.001$).

STEM Identity

Homogeneity was proved by Levene's test, $F = 2.68$, $p = 0.07$. The analysis was significant with F (2, 91) = 104.12, $p < 0.001$. Significant differences were observed in the three groups' STEM identity after the intervention. A *post hoc* comparison revealed that the post-questionnaire scores of the RWPM group (adjusted mean = 4.53) were significantly higher than that of the TBP group (adjusted mean = 3.18, $p < 0.001$), and the RWP group (adjusted mean = 3.27, $p < 0.001$).

RQ3: Improvement of Perceived Cognitive Competencies and Non-cognitive Characteristics

In response to RQ3, this study further examined the perceived improvement of the three groups before and after the intervention by conducting paired-sample *t* tests, see **Table 3**. All the participants showed significant improvement in creativity, with TBP group ($t = -12.15$, $p < 0.001$), RWP group ($t = -9.37$, $p < 0.001$), and RWPM group ($t = -8.90$, $p < 0.001$). Furthermore, the RWP and RWPM groups achieved significant improvement in critical thinking, with $t = -3.62$, $p < 0.001$ and $t = -5.59$, $p < 0.001$, respectively. However, the TBP group showed no significant improvement in this dimension with $t = -1.60$, $p = 0.12$.

STEM interest and identity of the three groups were significantly different before and after the intervention. Students of the TBP group ($t = -7.11$, $p < 0.001$), RWP group ($t = -6.95$, $p < 0.001$) and RWPM groups ($t = -8.50$, $p < 0.001$) showed perceived stronger STEM interest. Similarly, the TBP group ($t = -3.26$, $p < 0.001$), RWP group ($t = -5.36$, $p < 0.001$) and RWPM group ($t = -13.12$, $p < 0.001$) reported significantly stronger STEM identity.

DISCUSSION

The intervention conducted in this paper was designed to examine how different pedagogical designs supporting the need for relatedness, i.e., real-life problems and mentoring, influence the development of cognitive competencies and non-cognitive characteristics. This study presents four major findings; accordingly, three practical suggestions are presented for practitioners and instructional designers.

First, as predicted, maker pedagogy exerted an impact on nurturing students' creative competence. In line with the studies of Saorín et al. (2017) and Searle et al. (2018), our results highlighted maker activities improved creativity for all three groups (H1 and H5). These results imply that (i) maker activities provided students with learning opportunities that facilitated their creative competence as they implemented their ideas through creating artifacts and (ii) the relevance of problems pertaining to maker pedagogy and the relationships with and guidance from a more experienced or knowledgeable person

TABLE 4 | ANCOVA results of post-questionnaire for the three groups.

Scales	**Groups**	***N***	**Adjusted mean**	**SE**	***F***	***Post hoc***
Creativity	TBP	32	4.30	0.09	0.22	
	RWP	31	4.27	0.09		
	RWPM	32	4.22	0.09		
Critical thinking	TBP	32	3.21	0.16	8.41***	(1) < (2)
	RWP	31	3.78	0.16		
	RWPM	32	4.10	0.16		(1) < (3)
STEM interest	TBP	32	3.86	0.11	14.86***	(1) < (3)
	RWP	31	3.86	0.11		(2) < (3)
	RWPM	32	4.58	0.11		
STEM identity	TBP	32	3.18	0.07	104.12***	(1) < (3)
	RWP	31	3.27	0.08		(2) < (3)
	RWPM	32	4.53	0.07		

*$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

did not better foster the competence of students. Furthermore, students were found to better develop their creative skills through innovation by putting their new ideas into practice that encouraged divergent thinking when students are encouraged to envision and suggest multiple solutions (Xia et al., 2021).

Second, the critical thinking of all the three groups showed improvement and the students who solved RWPs outperformed those who solved TBPs (H2 and H5). This is in alignment with earlier studies that advocated the use of RWPs in design and maker activities (English and Mousoulides, 2015; Hollman et al., 2019). A plausible explanation is that students felt connected and relevant, and were motivated to explore and brainstorm more solutions for the problems (Ryan and Deci, 2020). The processes of coming up solutions encouraged students to evaluate their ideas for developing better critical thinking skills (Hu et al., 2020). The connection also encourages students to carefully and thoughtfully consider and decide their choices fit into a societal context in which they feel loved and care for. Another explanation is about, when addressing RWPs, students process a large amount of information that is contradictory and potentially unreliable and have to make connections across different disciplines to arrive at solutions. In this learning process, they are needed to carefully read and critically process information to consolidate their ideas (English and Mousoulides, 2015). Therefore, the relevance and complexity of the problems influence how students can think clearly, rationally, and critically to form a decision.

Third, students in all three groups developed more positive interest and identity toward STEM, and students in the mentorship group particularly developed much stronger interest and identity than the other two groups that did not involve mentors (H3, H4, and H5). These results are supported by studies that suggest that engaging role models or mentors in STEM learning can better cultivate students' STEM interest and identity (Honey et al., 2014; Pinkard et al., 2017; Ladeji-Osias et al., 2018; Musavi et al., 2018; Huvard et al., 2020). These findings also imply that positive interest and identity toward STEM are developed in a social context (Renninger and Hidi, 2011; Goos and Bennison, 2019). Both interest and identity reflect how students view themselves (self-images and endeavors) and how they feel others perceive them (recognitions from others) in the context of STEM. Therefore, their development requires the internalization process of need satisfaction for relatedness (Ryan and Deci, 2020). This internalization process is catalyzed by engaging mentors or role models in maker learning due to the different interactions with mentors, namely shareability, tangibility, and aesthetic (Tofel-Grehl et al., 2017). In this study, the mentors used active-listening techniques to establish and maintain positive mentor-mentee relationships by encouraging their mentees and helping with problem solving. The students can speak freely and express ideas, resulting in interpersonal comfort and more intimate and frequent communications. Accordingly, offering positive student–expert relationships in maker pedagogy can more effectively foster students' STEM interest and identity.

The final finding suggests that in maker pedagogy, supporting relatedness by using student–expert relationships and RWPS types can determine the effectiveness of the development of learners' non-cognitive characteristics, namely STEM interest and identity (third finding), and cognitive competency, that is critical thinking. Therefore, non-cognitive characteristics in STEM are influenced by interaction with role models but not by analyzing problems and building artifacts. Cognitive competencies that involve extensive thinking require deeper learning, and these attributes can be affected only through the processing, application, and emulation of knowledge.

PRACTICAL SUGGESTIONS

Creativity, critical thinking, interest, and identity are the significant outcomes of STEM education (Honey et al., 2014). Our study shows how STEM learning can be designed using relatedness support to foster the four outcomes in a formal curriculum. Our first suggestion is to encourage STEM curriculum leaders should emphasize the development of student STEM interest and identity by offering a promising pedagogical approach—satisfying the need for relatedness. We suggest that teachers should use learning strategies that make students feel connected and loved, such as students' favorable raw materials for making, community- and school-based problems and involvement of more knowledgeable persons (e.g., role models or mentors). Our second suggestion is that teachers should consider student-mentor relationships when planning and designing their STEM maker lessons or projects. To achieve that, teacher professional programs on how to co-teach with mentors should be offered because this approach is new to most STEM teachers. Finally, we recommend that teachers should use scaffolding ideas to help students achieve the four learning outcomes (Chiu et al., 2021b). Because mentorship programs require numerous resources, schools may not be able to sustainably provide such programs. Teachers can use TBPs to build students' basic making skills, followed by RWPs and mentorship programs.

CONCLUSION AND LIMITATIONS

Overall, this study suggested supporting relatedness can influence students' cognitive competencies and non-cognitive characteristics in STEM learning. This study had limitations and five of them are noted here. First, while this study appears to support how to satisfy the need for relatedness affect the effectiveness of learning outcomes (Nganga et al., 2020), additional studies validating these findings are required. The results of the present intervention could also be extended by additional studies by including other problem types (local and global), other experts such as STEM professionals, and other cognitive and non-cognitive skills. Second, this study did not consider how long interest and identity can persist. Further research should be conducted using longitudinal design to examine the long-term effects of different maker pedagogies. Third, how students collaborate with peers or mentors in making may have impact on the outcomes of this study. Future studies are suggested to examine how different collaborative learning affect the outcomes of the students. The fourth limitation are the definition of creativity. This study adopted the definition from an engineering education (Daly et al., 2014); however, creativity

can be defined in terms of three key aspects—fluency, flexibility, and originality. Future studies should adopt these aspects in measuring student creativity with comprehensive approach. The final limitation of the study is that the invention was conducted over different sessions. Environmental factors may influence student motivation for learning, leading to differences between conditions. The intervention should be conducted in parallel sessions by future studies.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Faculty of Education, The Chinese University of Hong Kong. Written informed consent to participate in this study was provided by the participants' legal guardian/ next of kin.

AUTHOR CONTRIBUTIONS

XW analyzed the data, drafted and revised the manuscript. TC conceptualized this study, managed the research project, and revised the work. MJ revised and proofread the article. All authors approved the submission.

FUNDING

This work was fully supported by Faculty research supportive fund, Faculty of Education, The Chinese University of Hong Kong.

REFERENCES

Achmetli, K., Schukajlow, S., and Rakoczy, K. (2019). Multiple solutions for real-world problems, experience of competence and students' procedural and conceptual knowledge. *Int. J. Sci. Math. Educ.* 17, 1605–1625. doi: 10.1007/s10763-018-9936-5

Bevan, B., Gutwill, J. P., Petrich, M., and Wilkinson, K. (2015). Learning through STEM-rich tinkering: Findings from a jointly negotiated research project taken up in practice. *Sci. Educ.* 99, 98–120. doi: 10.1002/sce.21151

Brahms, L., and Crowley, K. (2016). "Learning to make in the museum: The role of maker educators," in *Makeology: Makerspaces as learning environments*, eds K. Peppler, E. R. Halverson, and Y. B. Kafai (New York, NY: Routledge), 15–29. doi: 10.4324/9781315726519

Cannady, M. A., Greenwald, E., and Harris, K. N. (2014). Problematizing the STEM pipeline metaphor: is the STEM pipeline metaphor serving our students and the STEM workforce? *Sci. Educ.* 98, 443–460. doi: 10.1002/sce.21108

Chiu, T. K. F. (2021a). A holistic approach to Artificial Intelligence (AI) curriculum for K-12 schools. *TechTrends* 65, 796–807. doi: 10.1007/s11528-021-00637-1

Chiu, T. K. F. (2021b). Applying the self-determination theory (SDT) to explain student engagement in online learning during the COVID-19 pandemic. *J. Res. Technol. Educ.* 2021, 1–17. doi: 10.1080/15391523.2021.1891998

Chiu, T. K. F. (2021c). Digital support for student engagement in blended learning based on self-determination theory. *Comput. Hum. Behav.* 2021:106909. doi: 10.1016/j.chb.2021.106909

Chiu, T. K. F., Chai, C. S., Williams, J., and Lin, T. J. (2021a). Teacher professional development on Self-determination Theory-based design thinking in STEM education. *Educ. Technol. Soc.* 24, 153–165.

Chiu, T. K. F., Meng, H., Chai, C. S., King, I., Wong, S., and Yam, Y. (2021b). Creation and Evaluation of a Pretertiary Artificial Intelligence (AI) Curriculum. *IEEE Trans. Educ.* 2021:3085878. doi: 10.1109/te.2021.3085878

Chiu, T. K. F., Jong, M. S. Y., and Mok, I. A. C. (2020). Does learner expertise matter when designing emotional multimedia for learners of primary school mathematics? *Educ. Tech. Res. Dev.* 68, 2305–2320. doi: 10.1007/s11423-020-09775-4

Chiu, T. K. F., and Lim, C. P. (2020). Strategic use of technology for inclusive education in Hong Kong: A content-level perspective. *ECNU Rev. Educ.* 3, 715–734. doi: 10.1177/2096531120930861

Daly, S. R., Mosyjowski, E. A., and Seifert, C. M. (2014). Teaching creativity in engineering courses. *J. Eng. Educ.* 103, 417–449. doi: 10.1002/jee.20048

English, L. D. (2016). STEM education K-12: Perspectives on integration. *Int. J. STEM Educ.* 3, 1–8. doi: 10.1186/s40594-016-0036-1

English, L. D., and Mousoulides, N. G. (2015). Bridging STEM in a real-world problem. *Math. Teach. Middle Sch.* 20, 532–539. doi: 10.5951/mathteacmiddscho.20.9.0532

Fasso, W., and Knight, B. A. (2020). Identity development in school makerspaces: Intentional design. *Int. J. Technol. Des. Educ.* 30, 275–294. doi: 10.1007/s10798-019-09501-z

Gerber, E. M., Marie Olson, J., and Komarek, R. L. (2012). Extracurricular design-based learning: Preparing students for careers in innovation. *Int. J. Sci. Educ.* 28, 317–324.

Gettings, M. (2016). Putting it all together: STEAM, PBL, scientific method, and the studio habits of mind. *Art Educ.* 69, 10–11. doi: 10.1080/00043125.2016.1176472

Godhe, A. L., Lilja, P., and Selwyn, N. (2019). Making sense of making: critical issues in the integration of maker education into schools. *Technol. Pedagag. Educ.* 28, 317–328. doi: 10.1080/1475939x.2019.1610040

Godwin, A. (2016). "The development of a measure of engineering identity," in *Proceedings of the ASEE Annual Conference and Exposition*, (New Orleans, LA), doi: 10.18260/p.26122

Goos, M., and Bennison, A. (2019). A zone theory approach to analysing identity formation in mathematics education. *ZDM-Math. Educ.* 51, 405–418. doi: 10.1007/s11858-018-1011-8

Hachey, A. C., An, S. A., and Golding, D. E. (2021). Nurturing Kindergarteners' early STEM academic identity through makerspace pedagogy. *Early Child. Educ. J.* 2021:1154. doi: 10.1007/s10643-021-01154-9

Halverson, E. R., and Sheridan, K. (2014). The maker movement in education. *Harv. Educ. Rev.* 84, 495–504. doi: 10.1093/acrefore/9780190264093.013.312

Hanif, S., Wijaya, A. F. C., and Winarno, N. (2019). Enhancing students' creativity through STEM project-based learning. *J. Sci. Learn.* 2, 50–57. doi: 10.17509/jsl.v2i2.13271

Hatch, M. (2014). *The maker movement manifesto: Rules for innovation in the new world of crafters, hackers, and tinkerers*. New York, NY: McGraw-Hill Education, doi: 10.5860/choice.51-3798

Hidi, S., and Ainley, M. (2008). "Interest and self-regulation: Relationships between two variables that influence learning," in *Motivation and self- regulated learning: Theory, research, and applications*, eds D. H. Schunk and B. J. Zimmerman (Mahwah, NJ: Lawrence Erlbaum Associates Publishers), 77–109. doi: 10.4324/9780203831076

Hidi, S., and Renninger, K. A. (2006). The four-phase model of interest development. *Educ. Psychol.* 41, 111–127. doi: 10.1207/s15326985ep4102_4

Hollman, A. K., Hollman, T. J., Shimerdla, F., Bice, M. R., and Adkins, M. (2019). Information technology pathways in education: Interventions with middle school students. *Comput. Educ.* 135, 49–60. doi: 10.1016/j.compedu.2019.02.019

Holmlund, T. D., Lesseig, K., and Slavit, D. (2018). Making sense of "STEM education" in K-12 contexts. *Int. J. STEM Educ.* 5, 1–18. doi: 10.1186/s40594-018-0127-2

Honey, M., Pearson, G., and Schweingruber, H. A. (eds) (2014). *STEM integration in K-12 education: Status, prospects, and an agenda for research*. Washington, D.C: National Academies Press, doi: 10.17226/18612

Hu, C. C., Yeh, H. C., and Chen, N. S. (2020). Enhancing STEM competence by making electronic musical pencil for non-engineering students. *Comput. Educ.* 150:103840. doi: 10.1016/j.compedu.2020.103840

Hung, W. (2016). All PBL starts here: The problem. *Interdiscip. J. Probl.-Based Learn.* 10:2. doi: 10.7771/1541-5015.1604

Hurk, A., Van Den, Meelissen, M., and van Langen, A. (2019). Interventions in education to prevent STEM pipeline leakage. *Int. J. Sci. Educ.* 41, 150–164. doi: 10.1080/09500693.2018.1540897

Huvard, H., Talbot, R. M., Mason, H., Thompson, A. N., Ferrara, M., and Wee, B. (2020). Science identity and metacognitive development in undergraduate mentor-teachers. *Int. J. STEM Educ.* 7, 1–17. doi: 10.1186/s40594-020-00231-6

Johnson, C. C., Mohr-Schroeder, M. J., Moore, T. J., and English, L. D. (eds) (2020). *Handbook of Research on STEM Education*. Milton Park: Routledge, doi: 10.4324/9780429021381

Kelley, T. R., Knowles, J. G., Han, J., and Sung, E. (2019). Creating a 21st Century Skills Survey Instrument for High School Students. *Amer. J. Educ. Res.* 7, 583–590. doi: 10.12691/education-7-8-7

Kreider, C. M., Medina, S., Lan, M. F., Wu, C. Y., Percival, S. S., Byrd, C. E., et al. (2018). Beyond academics: A model for simultaneously advancing campus-based supports for learning disabilities, STEM students' skills for self-regulation, and mentors' knowledge for co-regulating and guiding. *Front. Psychol.* 9:1466. doi: 10.3389/fpsyg.2018.01466

Ladeji-Osias, J. O., Partlow, L. E., and Dillon, E. C. (2018). Using Mobile Application Development and 3-D Modeling to Encourage Minority Male Interest in Computing and Engineering. *IEEE Trans. Educ.* 61, 274–280. doi: 10.1109/te.2018.2826466

Lee, J. J., Ceyhan, P., Jordan-Cooley, W., and Sung, W. (2013). GREENIFY: A real-world action game for climate change education. *Simul. Gaming.* 44, 349–365. doi: 10.1177/1046878112470539

Marshall, J. A., and Harron, J. R. (2018). Making learners: A framework for evaluating making in STEM education. *Interdiscip. J. Probl.-Based Learn.* 12:3. doi: 10.7771/1541-5015.1749

Martin, L. (2015). The promise of the maker movement for education. *J. Pre-Coll. Eng. Educ. Res.* 5, 30–39. doi: 10.7771/2157-9288.1099

Musavi, M., Friess, W. A., James, C., and Isherwood, J. C. (2018). Changing the face of STEM with stormwater research. *Int. J. STEM Educ.* 5, 1–12. doi: 10.1186/s40594-018-0099-2

Nelson, K., Sabel, J., Forbes, C., Grandgenett, N., Tapprich, W., and Cutucache, C. (2017). How do undergraduate STEM mentors reflect upon their mentoring experiences in an outreach program engaging K-8 youth? *Int. J. STEM Educ.* 4, 1–13. doi: 10.1186/s40594-017-0057-4

Nemorin, S., and Selwyn, N. (2017). Making the best of it? Exploring the realities of 3D printing in school. *Res. Pap. Educ.* 32, 578–595. doi: 10.1080/02671522.2016.1225802

Nganga, C., Bowne, M., and Stremmel, A. (2020). Mentoring as a developmental identity process. *Mentor. Tutor.: Partner. Learn.* 28, 259–277. doi: 10.1080/13611267.2020.1783498

Papadakis, S. (2021). The Impact of Coding Apps to Support Young Children in Computational Thinking and Computational Fluency. A Literature Review. *Front. Educ.* 6:657895. doi: 10.3389/feduc.2021.657895

Pinkard, N., Erete, S., Martin, C. K., and McKinney de Royston, M. (2017). Digital youth divas: Exploring narrative-driven curriculum to spark middle school girls' interest in computational activities. *J. Learn. Sci.* 26, 477–516. doi: 10.1080/10508406.2017.1307199

Reeve, J. (2013). How students create motivationally supportive learning environments for themselves: the concept of agentic engagement. *J. Educ. Psychol.* 105, 579–595. doi: 10.1037/a0032690

Renninger, K. A. (2010). "Working with and cultivating interest, self-efficacy, and self-regulation," in *Innovations in educational psychology: Perspectives on learning, teaching and human development*, eds D. Preiss and R. Sternberg (New York, NY: Springer), 107–138.

Renninger, K. A., and Hidi, S. (2011). Revisiting the conceptualization, measurement, and generation of interest. *Educ. Psychol.* 46, 168–184. doi: 10.1080/00461520.2011.587723

Ryan, R. M., and Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemp. Educ. Psychol.* 61:101860. doi: 10.1016/j.cedpsych.2020.101860

Sang, W., and Simpson, A. (2019). The Maker Movement: A global movement for educational change. *Int. J. Sci. Math. Educ.* 17, 65–83. doi: 10.1007/s10763-019-09960-9

Saorín, J. L., Melian-Diaz, D., Bonnet, A., Carrera, C. C., Meier, C., and De La Torre-Cantero, J. (2017). Makerspace teaching-learning environment to enhance creative competence in engineering students. *Think. Skills Creat.* 23, 188–198. doi: 10.1016/j.tsc.2017.01.004

Schlegel, R. J., Chu, S. L., Chen, K., Deuermeyer, E., Christy, A. G., and Quek, F. (2019). Making in the classroom: Longitudinal evidence of increases in self-efficacy and STEM possible selves over time. *Comput. Educ.* 142:103637. doi: 10.1016/j.compedu.2019.103637

Searle, K. A., Litts, B. K., and Kafai, Y. B. (2018). Debugging open-ended designs: High school students' perceptions of failure and success in an electronic textiles design activity. *Think. Skills Creat.* 30, 125–134. doi: 10.1016/j.tsc.2018.03.004

Simões, F., and Alarcão, M. (2014). Promoting well-being in school-based mentoring through basic psychological needs support: Does it really count? *J. Happiness Stud.* 15, 407–424. doi: 10.1007/s10902-013-9428-9

Stoeger, H., Duan, X., Schirner, S., Greindl, T., and Ziegler, A. (2013). The effectiveness of a one-year online mentoring program for girls in STEM. *Comput. Educ.* 69, 408–418. doi: 10.1016/j.compedu.2013.07.032

Suh, E. K., Hoffman, L., and Zollman, A. (2020). "Making STEM accessible to all," in *Handbook of research on STEM education*, eds C. C. Johnson, M. J. Mohr-Schroeder, T. J. Moore, and L. D. English (Milton Park: Routledge), 311–322.

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Res. Sci. Educ.* 48, 1273–1296. doi: 10.1007/s11165-016-9602-2

Tenenbaum, L. S., Anderson, M. K., Jett, M., and Yourick, D. L. (2014). An innovative near-peer mentoring model for undergraduate and secondary students: STEM focus. *Innov. High. Educ.* 39, 375–385. doi: 10.1007/s10755-014-9286-3

Tofel-Grehl, C., Fields, D., Searle, K., Maahs-Fladung, C., Feldon, D., Gu, G., et al. (2017). Electrifying engagement in middle school science class: Improving student interest through e-textiles. *J. Sci. Educ. Technol.* 26, 406–417. doi: 10.1007/s10956-017-9688-y

Tyler-Wood, T., Knezek, G., and Christensen, R. (2010). Instruments for assessing interest in STEM content and careers. *J. Technol. Teach. Educ.* 18, 345–368.

Woods, S., and Hsu, Y. C. (2020). Making spaces for STEM in the school library. *TechTrends* 64, 388–394. doi: 10.1007/s11528-019-00460-9

Xia, T., Kang, M., Chen, M., Ouyang, J., and Hu, F. (2021). Design training and creativity: Students develop stronger divergent but not convergent thinking. *Front. Psychol.* 12:695002. doi: 10.3389/fpsyg.2021.695002

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Weng, Chiu and Jong. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Education

ORIGINAL RESEARCH
published: 31 March 2022
doi: 10.3389/feduc.2022.871895

Children's Spatial Play With a Block Building Touchscreen Application

Naomi Polinsky[1], Breniel Lemley[2], Rachel M. Flynn[3], Ellen Wartella[2] and David H. Uttal[1]*

[1] Department of Psychology, Northwestern University, Evanston, IL, United States, [2] Department of Communication Studies, Northwestern University, Evanston, IL, United States, [3] Department of Child and Adolescent Development, San Francisco State University, San Francisco, CA, United States

Spatial play contributes to children's early development of spatial skills, which are foundational for STEM achievement. A growing genre of spatial play for young children is digital block play. We asked how 3- to 6-year-old children (*N* = 117) engaged in digital block play and whether children's age, gender, and spatial skills were correlated with this play. Children completed a spatial skills assessment and played a popular digital block play app, *Toca Blocks*. We developed a coding scheme that measured children's play behaviors in the app, and reliably detected individual differences in this play. Children actively manipulated the digital blocks, and there were differences in their block play by age and gender. However, children's spatial skills were not associated with their play in the app. The present work shows that digital block play supports play behaviors similar to those supported by physical blocks, but whether and how digital block play facilitates spatial learning is still unknown. The results are discussed in terms of potential ways to implement digital spatial play apps that might engage children's spatial skills and support their spatial and STEM learning.

Keywords: apps, spatial play, touchscreen media, spatial skills, digital games

OPEN ACCESS

Edited by:
Stamatios Papadakis, University of Crete, Greece

Reviewed by:
Hari Palani, Others, Portland, OR, United States
H. Özlen Demircan, Middle East Technical University, Turkey

****Correspondence:***
Naomi Polinsky
naomipolinsky2022@u.northwestern.edu

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Education

Received: *08 February 2022*
Accepted: *07 March 2022*
Published: *31 March 2022*

Citation:
Polinsky N, Lemley B, Flynn RM, Wartella E and Uttal DH (2022) Children's Spatial Play With a Block Building Touchscreen Application. Front. Educ. 7:871895. doi: 10.3389/feduc.2022.871895

INTRODUCTION

Spatial skills are important for STEM learning, even from a young age. For example, there is a strong association between spatial skills and mathematics performance, a critical foundation of many STEM topics (e.g., Lubinski and Benbow, 1992; Casey et al., 2008; Mix and Cheng, 2012; Verdine et al., 2014a,c). The connection between spatial skills and mathematics potentially begins as early as infancy (Gallistel and Gelman, 1992), perhaps due to a shared common neural code between space and number within the intraparietal sulcus (McCrink and Opfer, 2014; Hawes and Ansari, 2020). Spatial skills also play a specific role in intellectual creativity in STEM (Kell et al., 2013) and if and how STEM learners use external spatial representations, such as maps, models, diagrams, graphs, and sketches, during problem solving (Uttal et al., 2013; Mix, 2019). Finally, there is emerging evidence (although still preliminary), including well-controlled experiments, that training spatial skills may lead to gains in STEM interest, achievement, and retention (e.g., Sorby et al., 2013; Cheng and Mix, 2014).

One experience that contributes to children's spatial skill development is their *spatial play,* such as play with blocks, puzzles, and board games (Caldera et al., 1999; Siegler and Ramani, 2008; Levine et al., 2012; Verdine et al., 2014c; Jirout and Newcombe, 2015). Spatial play facilitates

children's spatial thinking and creates opportunities for them to practice their spatial skills. When children engage in spatial play, they often must rotate and rearrange toys, which results in visual changes to object orientations. The opportunity for children to notice the spatial changes that they physically create promotes spatial thinking (Wakefield et al., 2019). Moreover, the frequency of children's spatial play is predictive of their spatial skills (Levine et al., 2012; Verdine et al., 2014a; Jirout and Newcombe, 2015). There is now substantial evidence that spatial play can contribute both to spatial development and to the development of relevant STEM skills (e.g., Gunderson et al., 2012).

In recent years the opportunities for children's spatial play have been transformed through the tremendous growth in digital apps. Ten years ago, touchscreens and tablets were novelties; now they are nearly ubiquitous and are available for very young children (Rideout, 2017). Consequently, spatial games that were once reserved for children's play with physical materials, such as blocks and puzzles, now occur on touchscreen devices in the form of *digital spatial play*. Early evidence suggests a positive association between children's spatial skills and their digital spatial play (Polinsky et al., 2021) and that these skills can be strengthened through children's play on touchscreen devices (Bower et al., 2021). Additionally, digital spatial play is also a culturally relevant part of young children's everyday lives and is a growing context for children's play (Gee, 2003; Flynn et al., 2019). Therefore, we must also examine *how* children play with these apps.

In this paper, we focus on how children engage in one form of digital spatial play, digital block play. There is increasing excitement for and growth of digital block building games, and some are designed for young children, including 3-year-olds. One example of a digital block play touchscreen game, or app, is Minecraft by Microsoft, which has sold over 100 million copies (Sarkar, 2017) and is most often played by children who are five and older. Another example is Toca Blocks by Toca Boca, which is currently ranked 40th in paid education apps in the Apple App Store and is designed for children ages four plus. Most block play apps are completely open-ended and unstructured; children can build whatever they choose without the constraints of defined levels or time limits.

We focus on digital block play because physical block play has been studied extensively (Reifel and Greenfield, 1982; Brosnan, 1998; Caldera et al., 1999; Stiles and Stern, 2001; Casey et al., 2008; Bower et al., 2020). Blocks are appropriate for children across the preschool years (Casey et al., 2008) and facilitate several different types of play that include both structured and unstructured building activities (Bower et al., 2020). When children play with physical blocks, they can stack, rotate, and arrange them in a variety of different ways. Some children use similarly shaped blocks to create the same relation exclusively and consecutively, for example, by only stacking blocks vertically on top of each other to create a tall tower (Stiles and Stern, 2001). Other children may use a wide range of blocks and produce several different types of spatial relations by building tall towers upward and using blocks of different dimensions to create bridges and structures with more depth (Stiles and Stern, 2001). These different ways of building with blocks allow children to notice and experience the outcomes of different types of spatial manipulations that they physically create. Because research on physical block play shows an association between children's block building behaviors and their characteristics (Goodson, 1982; Reifel, 1984; Caldera et al., 1999; Casey et al., 2008; Verdine et al., 2017), in this study we consider how children engage in digital block play and the role of children's age, gender, and spatial skills in this play.

First, children's age is associated with their block play behaviors. In general, the complexity and sophistication of the structures that children build during block play develops during early childhood. Children begin their block building by using a single block that represents one object and then transition toward building horizontal, floor-like, two-dimensional structures. By 3 and 4 years of age, children build three-dimensional structures, which primarily include vertical towers, and between four and seven, begin to incorporate complex structures, such as bridges and arches, into their building (Goodson, 1982; Reifel, 1984; Casey et al., 2008). This development is associated with children's evolving motor abilities, their emergent understanding of part-whole relationships, and their growing spatial abilities (Reifel and Greenfield, 1982, 1983; Gura, 1992; Caldera et al., 1999; Casey and Bobb, 2003).

Additionally, boys and girls may build with blocks in different ways, with boys tending to build more complex structures than girls (Goodfader, 1982; Sluss, 2002). This difference may be due to early socialized gendered toy preferences (Caldera et al., 1989; Campbell et al., 2002; Cherney, 2018; Coyle and Liben, 2020), and the fact that construction toys are often marketed toward and considered made for boys (Cherney et al., 2003; Cherney and Voyer, 2010; LoBue and DeLoache, 2011; Coyle and Liben, 2020). Consequently, male children may have more block play experiences than their female peers, which supports them in building complex structures (Doyle et al., 2012; Nazareth et al., 2013; Pruden et al., 2019). Beyond complexity, gender differences sometimes emerge in the types of structures children build. For example, Ramani et al. (2014) found that girls included more symbolic features of buildings, such as windows and doors, in their constructions than boys did. However, it is important to note that not all studies of children's block play consistently show gender differences (e.g., Verdine et al., 2014c).

Finally, children's spatial skills are associated with what and how they build with physical blocks. By the time children are in preschool there is variability in the strength of children's spatial skills (Newcombe and Frick, 2010; Levine et al., 2012). Moreover, those children with stronger spatial skills tend to build more complex structures than their peers (Brosnan, 1998; Verdine et al., 2014c). For example, Caldera et al. (1999) observed children's free play with blocks, and measured the complexity of their building process, such as the types of block placements, the adjustments children made to their constructions, and how frequently they rotated their blocks. The results revealed a significant association between the complexity of children's building approach and their spatial visualization skills. Additionally, in a more recent study, in which children were asked to replicate a model of a block

construction, Bower et al. (2020) demonstrated an association between 3-year-olds block building behaviors and strategies and their spatial skills.

As age, gender, and spatial skills play a role in what and how children build with physical blocks, they may also play an important role in children's digital block play. However, a question remains about the unique features of the digital technology and whether children have different experiences when they engage with digital blocks in comparison to those they have when playing with physical blocks (Lee et al., 2018; Worsley and Bar-El, 2020).

In some ways, digital features may augment children's block play. For example, digital block play apps provide children with an endless number of blocks, allowing children to build without the limitation of quantity. Additionally, during digital block play children have more options for the types of block structures, arrangements, and patterns they can create (Lee et al., 2018) because their built structures do not have to comply with the rules of physics. For example, when playing with digital blocks children can defy gravity and build structures that lack solid foundations but will not crumble. Finally, in digital block play games, children can easily navigate through their environment and change their perspectives on the screen. Sometimes this navigation is carried out with the use of avatars, digital characters, who can stand on top of blocks, explore the digital world, and easily spin around. Other apps simply make it possible for children to use their fingers on the screen to rotate the digital environment and even to take on a birds-eye point of view (Worsley and Bar-El, 2020).

However, some of the unique features of digital technology may also hinder children's block play. Specifically, children may be limited in the ways they can manipulate the blocks as the blocks are within a two-dimensional medium. The design of some apps simulates a three-dimensional depth, which allows children to change their perspectives to see all sides of the blocks or environment. However, even with that simulation of depth the blocks are being manipulated on a two-dimensional screen. Without the physical three-dimensionality, digital blocks may not create realistic depth cues when stacked and may not give children the same tactical cues needed to develop spatial relations.

These differences across the physical and digital mediums lead to the two aims of this paper. The first aim was to examine how children engaged in digital block play and to investigate the different patterns of this play. The second aim was to explore whether and how children's age, gender, and spatial skills were associated with their digital block play. In addition, we explored the influence of the potential covariate of prior media experience on children's digital block play because it has been shown to impact children's touchscreen play in past research (Aladé and Nathanson, 2016). Children between the ages of three and six, were asked to play with a commercially available block building touchscreen application called Toca Blocks, by Toca Boca (Toca Blocks, 2017). This app was chosen due to its popularity (it is the first block building app, on the Top Paid iPad Kids Apps list created by Apple), making it a prototypical example of a block building app for children of this age group. Therefore, we believe a focus on how children play with this app could provide insight into the spatial skills children practice during block play.

MATERIALS AND METHODS

Participants

The participants were 117, 3- to 6-year-old children (M = 4.88 years, SD = 1.19, 50.4% females)[1]. Children were recruited from a large laboratory database of families from the greater Chicago area who had expressed interest in participating in research in response to emails, phone calls, and other advertisements. Parents indicated their children's race and ethnic background: 71% Caucasian, non-Hispanic, 6% Asian, 4% Black or African American, and 13% indicated multiracial. Additionally, parents recorded their highest level of education: 2.7% completed some college, 33% held a college degree, 3% had completed some of their post graduate education, 45% held a master's degree, and 12% held a Ph.D. or a professional degree. Children received a book and a tee-shirt as a thank you gift. Additionally, data from 14 children (M_{age} = 4.4, 43% females) was discarded due to refusal to complete the task, and data from 2 children (6.9 and 6.2 years, both male) was discarded due to a screen recording failure.

Spatial Skills Measures

To our knowledge, there is currently not a spatial skills assessment that spans the 3- to 6-year-old age range; therefore, the assessments we used to measure children's spatial skills differed for younger and older children. The 3- and 4-year-old participants completed the 2D and 3D Test of Spatial Assembly (TOSA) (Verdine et al., 2014c), whereas the 5- and 6-year-old participants completed the Children's Mental Transformation Task (CMTT) (Ehrlich et al., 2006).

Test of Spatial Assembly

The TOSA is a match-to-sample spatial assembly task (Verdine et al., 2014c). Children completed 14 trials, including two practice trials, in which they created a copy of a sample arrangement of geometric shapes for the 2D trials and a copy of interlocking blocks for the 3D trials (**Figure 1**). The stimuli for this assessment were created based on the Test of Spatial Assembly (TOSA) Instruction Manual (Verdine and Golinkoff, 2017). For each 2D trial, children are provided an 8.5-inch by 10.5-inch magnetic white board, on which there is a 2.25-inch by 2.25-inch laminated picture of the sample geometric arrangement placed on the left side of the board, and the corresponding cut out magnetic foam shapes, randomly dispersed behind a black line drawn down the right side of the board. Children used these magnetic foam pieces to create a copy of the sample arrangement displayed in the picture on the white board. Similarly, for each 3D TOSA trial, children are provided a glued together sample arrangement of interlocking LEGO Duplo Blocks, and the corresponding free LEGO Duplo Blocks that they could use to recreate the sample arrangement. The order of both the 2D and 3D TOSA trials are fixed beginning with a training trial and become progressively more difficult with each trial.

[1]The spatial abilities data of the 3- and 4-year-old children (n = 51), was previously published in Polinsky et al. (2021) paper. However, this publication did not include data on the 5- and 6-year-old children or on any children's play with Toca Blocks.

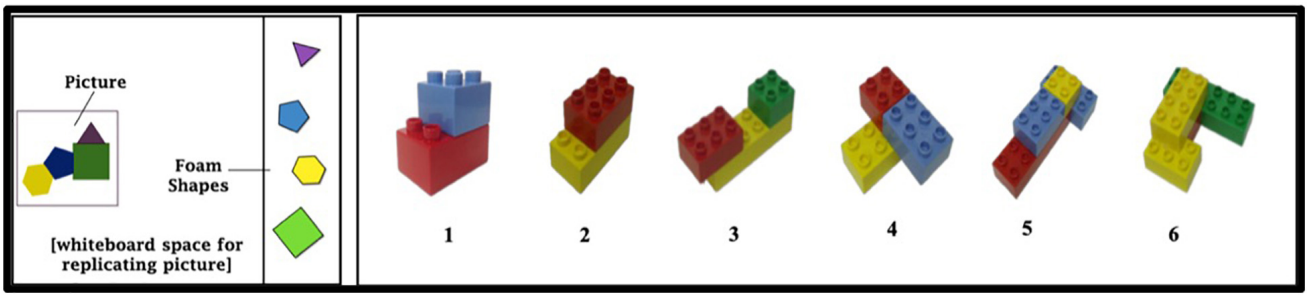


FIGURE 1 | The left side of the figure is a schematic of one trial of the 2D TOSA. The right side of the figure includes the block constructions used for each 3D TOSA trial. The number below the construction indicates on which trial each construction was used (Verdine and Golinkoff, 2017).

Both the 2D and 3D TOSAs begin with a practice trial, in which the experimenter showed the child a target arrangement and said, "See my model/picture and see my blocks/pieces. I am going to make my blocks/pieces look just like my model/picture." Then the experimenter moved their blocks or pieces into an incorrect arrangement, and asked the child, "Does this look like my model/picture?" When the child declared that the arrangement was incorrect, the experimenter repeated this routine a second time, again placing the blocks/pieces into the incorrect arrangement. After the child declared for a second time that the arrangement was incorrect, the experimenter finally created an accurate arrangement. Then, the experimenter asked the child, "Can you make your blocks/pieces look just like the model/picture?" and passed them the practice blocks or pieces and the practice model or picture. Once the child correctly completed the practice trial, the experimenter moved on to the latter trials. At the beginning of each trial the experimenter said, "Can you make these blocks/pieces look just like this model/picture," and then passed the model and blocks or whiteboard to the child. The target designs remained visible throughout each respective trial, and feedback was not provided. Trials were not timed; they ended when children indicated to the researcher that they had completed their geometric arrangement or block construction. Children's answers were recorded by the experimenter, who took a photograph of each arrangement created by the child at the end of the experimental session. Each photograph was taken with a digital camera from the same height and angle for each trial. These photographs were used for later coding by trained researchers.

Scoring

The 2D and 3D TOSA were scored using the coding scheme used by Verdine et al. (2014b,c). The 2D TOSA trials received a score based on how closely the child's geometric arrangement matched the model. For every trial, each shape was compared to the model and was coded based on its accuracy (0 or 1) of three spatial relations: *horizontal and vertical direction*, the placement of the shape on the whiteboard, *adjacent pieces*, next to which other shapes the shape being scored was placed, and *relative position*, the placement of the shape being scored in relation to the central shape of the arrangement. The points for each shape across all 2D trials were summed for a maximum of 73 points, which was then transformed into a *z*-score. One researcher independently coded all the 2D TOSAs. To check reliability, a second researcher coded 20% of the 2D TOSAs, resulting in a Cohen's kappa of 0.97. In the current study, the 2D TOSA was reliable, $a = 0.79$.

The 3D TOSA trials were scored in a similar manner to the 2D trials using the Verdine et al. (2014b,c) coding scheme. Each block of every trial was compared to the model and received an accuracy score (0 or 1) for its placement regarding three spatial relations: *vertical location*, a block's placement above or below other blocks, *rotation*, the orientation of each block, and *translation*, the placement of the block in the correct location within the model. The points for each block across all 3D trials were summed for a maximum of 46 points, which was then transformed into a *z*-score. One researcher independently coded all the 3D TOSAs. To check reliability, a second researcher coded 20% of the 2D TOSAs, resulting in a Cohen's kappa of 0.90. In the current study, the 3D TOSA did not have high reliability, $a = 0.552$[2]. Younger children's spatial abilities scores were created by averaging each child's z-score for the 2D and 3D trials.

Children's Mental Transformation Task

Five- and six-year-old participants completed the CMTT, a multiple-choice spatial skills assessment that has been used to assess spatial skills in young children in past research (Ehrlich et al., 2006). Children are shown a target picture of a 2-dimensional shape that is divided and separated into two halves on the vertical line of symmetry (Ehrlich et al., 2006). These halves are rotated and translated apart and placed on their own full piece of paper in a flip book. While the target is still visible, children are shown four whole shape choices that the two target pieces could make if put together (**Figure 2**). The four shape choices are all on the same piece of paper and contain three foils and one correct answer. Children are then asked to choose which shape would be created by putting together the two target pieces. The CMTT has 32 trials, and each trial is one of four different types: horizontal translation, diagonal translation, horizontal rotation, and diagonal rotation. On the first trial of the CMTT the experimenter tells the child, "Look at these pieces. Look at these pictures. If you put the pieces together, they will make one of the

[2]Despite the low reliability of the 3D TOSA, the findings do not change when we remove it from the analysis. Therefore, we maintain children's 3D TOSA score as an element of our measure of younger children's spatial abilities.

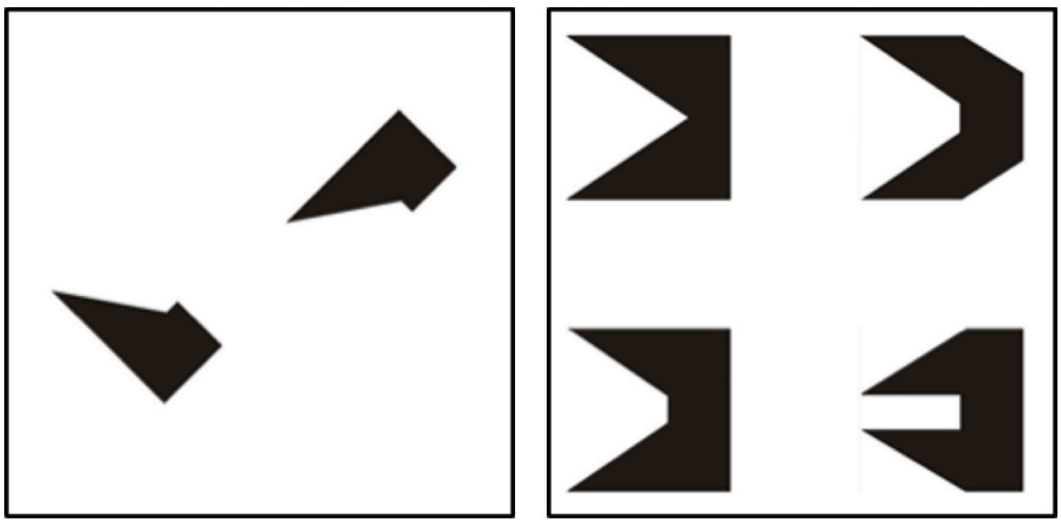


FIGURE 2 | A sample item of the Children's Mental Transformation Task (Ehrlich et al., 2006).

pictures. Point to the picture the pieces make." For all subsequent trials, the experimenter asks the child to, "Point to the picture the pieces make." Children's choices for each trial were recorded by the experimenter. The number of trials in which children chose the correct shape were summed and divided by the total number of trials to create a proportion correct score for each child. In the current study the CMTT was reliable, $a = 0.74$.

Prior Media Experience Survey

Parents completed a survey on their children's prior media experience on an iPad. The media usage and attitudes questions were based on a survey used by Sheehan et al. (2018). This survey yielded two measures of media use: (1) amount of time the child spent using media yesterday and (2) children's age of first exposure to media. First, parents reported the amount of time that their child spent using the computer, internet, video game devices, smartphones, tablets, eReaders, and voice control systems the previous day. Second, parents indicated children's earliest age of exposure to smartphones/tablets, video calling, videos, and smartphone and tablet applications. Parents chose the age range, below 9 months to between 5- and 6-years of age, in which their child's earliest exposure to each media occurred, resulting in an age of first exposure to tablets variable.

Touchscreen Game

Toca Blocks (2017) is a touchscreen game for children 4-years-old and older. The game can be downloaded from the Apple App Store or from Google Play. Toca Blocks does not include any levels or stated goals and instead allows children to build and explore freely using a variety of blocks. Toca Blocks is set in a landscape that includes green grass and an endless sky, which sometimes displays features resembling daytime and sometimes displays features resembling nighttime. The virtual environment includes a wide range of different colored blocks, which children can place throughout the landscape. Children can use the blocks in several ways. First, they can place blocks throughout the landscapes to create patterns and/or build two dimensional structures. Second, they can change the colors of the blocks by layering them on top of one another. Third, children can remove blocks from structures that were pre-designed to create pathways or rooms. In addition, the game had a variety of other props that children could use with the blocks, such as furniture, balloons, and fruits and vegetables. Children could use three avatars to engage with the structures they created and the objects in the environment. These avatars could run, jump, and dance, and could also destroy and remove blocks that were on their path. Beyond avatars, the game provided children with a variety of tools they could use to explore and manipulate the world. These tools included a digital scroll wheel that children could use to change their view of the landscape, a pencil that could color large spaces with blocks, and a chomping avatar that destroyed unwanted blocks. See **Figure 3** for a visual of the game's features. Finally, there are no rules in Toca Blocks, so children are free to engage with the game as they choose.

Procedure

The experiment was video recorded with consent from the children's parents and assent from the child. Children participated in the study in a quiet testing room in a lab space on a Midwestern campus with a trained researcher. First, children completed their age group's respective spatial skills assessment. Next, all children played Toca Blocks on an iPad for 7 min, as this time period was long enough to capture variation in game play while maintaining children's attention within the longer study session and was similar to the time allotted for game play in other studies (Polinsky et al., 2021). A timer was set at the beginning of the game. Before beginning the game, every child was provided the same instructions developed for the study.

> "Let's play the game Toca Blocks. Have you ever played Toca Blocks before? Ok great! I can show you how. In this game you can

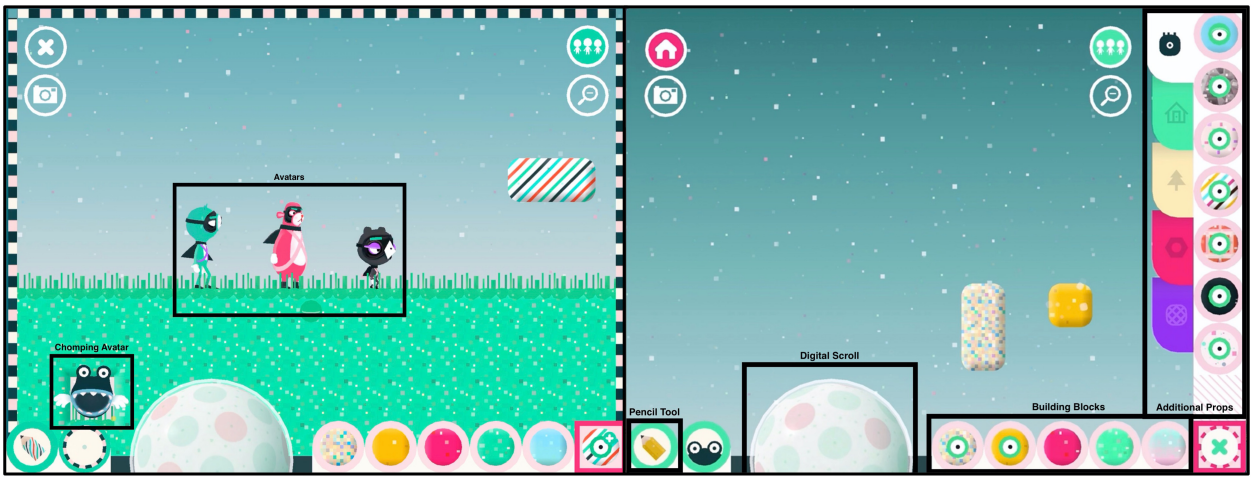


FIGURE 3 | The left side of the figure showcases the avatars with which children can play. The right side of the figure showcases the tools available for children to use to build structures and explore the environment.

build and explore, there are no right or wrong answers. See down here (point to blocks on bottom right of screen), there are blocks. You can drag the blocks up to the grass and build with them. To make the blocks change you can put one on top of the other. You can build anything you want, sort of like regular blocks. Then you see down here (point to avatars), there are all of these characters. You can drag them up here so they explore the world you build. If you press this button (point to circle triangle button) the character will move. Do you want to give it a try?"

During the game children's performance was recorded using the Softin Technology Co., Ltd. Screen Recorder application on the iPad.

Touchscreen Game Coding

We created a coding scheme to capture children's Toca Blocks play behaviors from screen recordings of them playing the game (see **Table 1** for more description). This coding scheme was based on research that coded children's block play behaviors (Ramani et al., 2014) and play during digital games (Marsh et al., 2016).

To adapt these coding schemes for our purposes we held a series of consensus meetings between the first three authors to establish codes that captured the observed behaviors. At the first meeting the researchers reviewed the screen recordings together and discussed and identified global play behaviors. Through these discussions the group came to a consensus on three main categories of game play behaviors: (1) object play, (2) avatar play, and (3) perspective changes. These behaviors are described in detail below. Next, we coded a set of videos using those codes. In a second coding meeting we analyzed differences in the three main types of behaviors and created a coding scheme to better capture observations. This analysis resulted in classifying the three main types of behaviors into subcategories: *strategic*, appearing direct and systematic, or *exploratory*, appearing unstructured. These behaviors were notable because they appeared to parallel pre-existing hierarchies of the types of children's play ranging from entirely goal-directed and systematic to completely unstructured and open-ended (Zosh et al., 2018). In the third step, we coded a subset using the more detailed coding reaching consensus that all observed behaviors were captured by our coding scheme. In the final phase we conducted moment-by-moment coding of the screen recordings using Datavyu (Datavyu Team, 2014).

To establish inter-rater reliability, two researchers coded 20% of the touchscreen play recordings, which resulted in an ICC > 0.82, indicating good reliability. Intraclass correlations are the most appropriate reliability measure for coding continuous data because they account for similarity and proximity (Shrout and Fleiss, 1979; Syed and Nelson, 2015). ICCs greater than or equal to 0.70 are considered acceptable reliability levels (Ostrov and Hart, 2012). The same two trained researchers split the remaining screen recordings and coded every child's behavior, resulting in a set of continuous codes for each child. From this coding, we created a variable that represents proporion of time that each type of behavior occurred. This variable was calculated by dividing the total amount of time in seconds for each behavior by the total amount of time children played. See **Figure 4** for proportion of time children spent in each play behavior.

Object Play

Object play was coded when children engaged with the on-screen building blocks or other available props, such as miniature furniture or balloons. We coded two subcategories of object play, a strategic subcategory and an exploratory subcategory. The strategic subcategory was *Object Play Construction* and was coded when children used the building blocks to create structures. These structures needed to include at least two blocks to be coded as construction because one block can be randomly placed without revealing intention to build. This category of play is the most like play with tangible building blocks, even though the virtual blocks can be used in ways that defy physical properties (e.g., floating in space or removing blocks from foundations without consequence). On average, children engaged in Object Play Construction for 21% of their play time. The exploratory

TABLE 1 | The coding scheme used to code children's Toca Blocks play.

	Code	Code subcategory	Description
	No move	Not applicable	No changes are made to the screen for a period of 6 s or longer.
Object play	Object play exploration	Exploratory	The child tests out different block and object functions by layering several blocks on top of each other, exploring how the side panel objects interact with the blocks or avatar, and sporadically placing individual blocks.
	Object play construction	Strategic	The child builds a tower or creates a pattern by placing two or more blocks on, next to, on top of, or under each other.
Avatar play	Avatar exploration	Exploratory	The child moves their avatar through the environment. This includes the avatar hopping, walking, sliding, skipping, or dancing on any material without creating change to the space.
	Avatar manipulation	Strategic	The child uses an avatar to move or destroy any object.
Perspective changes	Exploratory perspective changes	Exploratory	The child changes the view of the environment by scrolling in a non-linear or circular direction.
	Strategic perspective changes	Strategic	The child changes the view of the environment and immediately engages with the available materials in that new location.

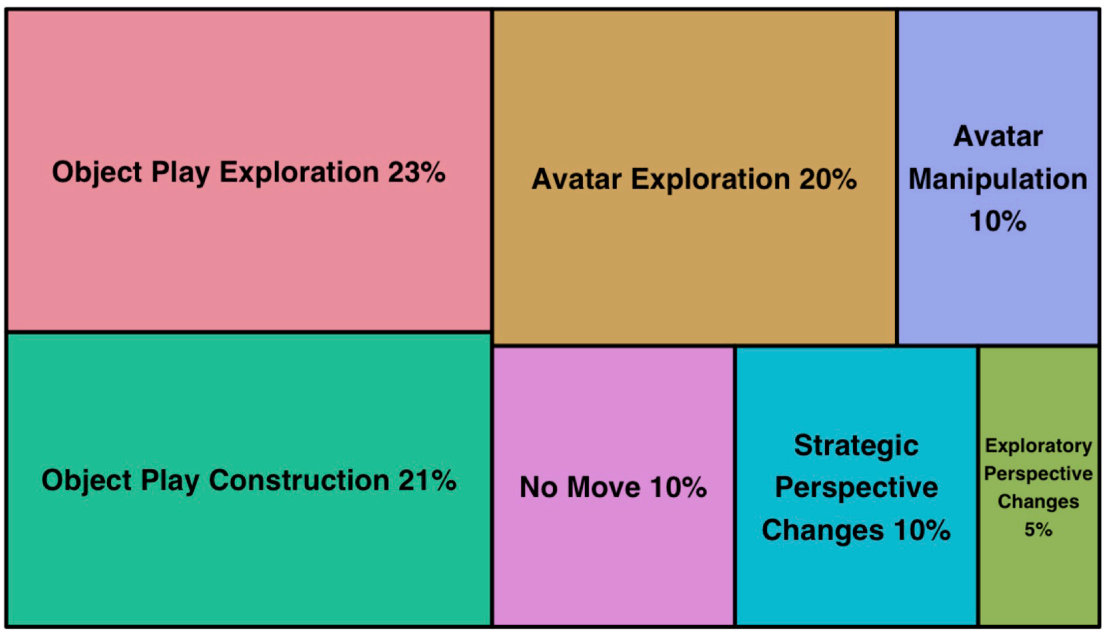


FIGURE 4 | Treemap illustrating the percentage of Toca Blocks play time children on average spent engaging in each play behavior.

subcategory was *Object Play Exploration* and was coded when children engaged with the blocks in ways that weren't building. For example, children could change block colors and features by layering the blocks on top of one another without creating a visible structure or pattern. Object Play Exploration could also include singular blocks being placed on the screen not in relation to a structure or block pieces, for example, blocks placed in mid-air not near other blocks. This category of play was unique to the digital medium as children cannot easily change the color or features of tangible blocks. On average, children spent 23% of their play time engaging in Object Play Exploration.

Avatar Play

Avatar play was coded when children engaged with any of the on-screen characters. This code included when children moved the avatar through the game or used the avatar for deleting objects from the space. A strategic subcategory of avatar play was coded as well as an exploratory subcategory of avatar play. The strategic subcategory was *Avatar Manipulation* and was coded when the child used the avatars to remove, destroy, or erase objects in the environment. For example, the game includes an avatar who removes blocks by chomping them with their mouth and a child might manipulate this avatar to "erase" an existing building or even part of the ground in order to create an underground room. This code was subcategorized as strategic as the child manipulated the avatar to do something purposeful in the environment, often related to building or destroying structures. On average, children spent 10% of their play time engaging in Avatar Manipulation. The exploratory subcategory was *Avatar Exploration* and was coded when the child used the avatar to move through the game space seemingly without purpose besides exploring. For example, when we observed this behavior, a child

might have their avatar walk around the environment, jumping over encountered objects, but not interacting with them in any way. On average, children engaged in Avatar Exploration for 20% of their play time.

Perspective Changes

Perspective changes were play behaviors in which children adjusted their view on the screen. For example, this behavior would occur when a child zoomed in or out, scrolled up or down, and navigated horizontally. We coded a strategic subcategory and an exploratory subcategory of perspective changes. The strategic subcategory was *Strategic Perspective Changes*, which were coded when a child manipulated the screen to have a new point of view and then moved or placed a block or an avatar within this new perspective. For a behavior to be coded this way it must have been immediately followed by either an object play or avatar play behavior. This code also included behaviors where children adjusted their view momentarily and returned immediately to a landmark on the screen, or a previously seen location or structure. For example, a child might have a view of the on-screen environment that includes a tower that they just built. Then they might manipulate their view of the screen by scrolling toward the right, such that the tower is out of view, and then immediately scroll back to the tower. This code was strategic because the changes to the on-screen environment appeared precise and predictable. On average, children made Strategic Perspective Changes for 10% of their play time. The exploratory subcategory was *Exploratory Perspective Changes*, which were behaviors that manipulated the screen to have a new point of view, but then did not result in an immediate object play or avatar play behavior. For example, having the block or avatar spinning around in a circle or continuously scrolling through the screen. On average, children spent 5% of their play time making Exploratory Perspective Changes.

No Move

We also included a No Move code, with which we marked all time periods greater than 6 seconds in which children were not engaging with the on-screen environment. This lapse in engagement with the game may have been a result of the child considering their next move, speaking with the researcher, or focusing their attention elsewhere. Coding this category allowed us to determine the time children spent actively engaged with the game. On average children's play behaviors were coded as No Move for 10% of play time.

RESULTS

Data were analyzed using the Psych Package in R (Revelle, 2019). In line with our first aim, to examine how children engaged in digital block play, we present descriptive statistics of behaviors and results from a One-Way ANOVA that examined differences in the proportion of time children spent engaging in each play behavior. Then, we report findings from a cluster analysis that grouped children based on their play behaviors. Finally, to examine our second aim, to explore whether and how children's age, gender, and spatial abilities were associated with their digital block play, we tested our potential covariates of prior media experience and used correlational analyses and t-tests to test the relationships between children's characteristics and their game play.

TABLE 2 | Means and standard deviations of the proportion of Toca Blocks play time children spent engaging in each play behavior.

Play behavior	*M*	*SD*	Minimum	Maximum
Object play construction	0.21	0.17	0	0.77
Object play exploration	0.23	0.17	0.02	0.83
Avatar exploration	0.20	0.17	0	0.67
Avatar manipulation	0.10	0.14	0	0.80
Strategic perspective changes	0.10	0.06	0	0.28
Exploratory perspective changes	0.05	0.07	0	0.28
No move	0.10	0.11	0	0.53

TABLE 3 | Means and standard deviations of the proportion of time children spent in each play behavior by cluster.

Play behavior	Object play group		Avatar play group	
	M	*SD*	*M*	*SD*
Object play construction	0.27	0.19	0.13	0.09
Object play exploration	0.33	0.17	0.11	0.06
Avatar exploration	0.11	0.10	0.31	0.18
Avatar manipulation	0.06	0.10	0.15	0.17
Strategic perspective changes	0.09	0.06	0.11	0.05
Exploratory perspective changes	0.05	0.07	0.06	0.07
No move	0.08	0.08	0.14	0.13

Children's Play With Toca Blocks

To address our first aim of investigating how children engaged in digital block play, we examined the proportion of play time children spent in each play behavior (see **Table 2** for means and standard deviations). We used a one-way ANOVA to examine differences between the proportion of play time children spent engaging in each play behavior. The independent variable was play behavior, such as Avatar Exploration or Object Play Construction, and the dependent variable was proportion of play time spent engaging in each behavior. There were significant differences in the overall proportion of time children spent engaging in each play behavior, $F(700) = 26.65$, $p < 0001$. Exploratory *post hoc* analysis with Tukey adjustments revealed that children spent a greater proportion of their time engaging in Object Play Construction, Object Play Exploration, and Avatar Play Exploration than they spent engaging in any other type of play (all p's < 0.001). However, there were no significant differences between the proportion of time children spent engaging in these three behaviors (Object Play Construction, Object Play Exploration, and Avatar Play Exploration) (all p's > 0.5).

To further understand the patterns of children's Toca Blocks play, we investigated if different children tended to use a combination of certain play behaviors. For example, did some

TABLE 4 | Means (and standard deviations) of and age differences in proportion of Toca Blocks play time children spent engaging in each type of play behavior.

	Younger children		Older children		Age	Cohen's
	M	*SD*	*M*	*SD*	Differences	*d*
1. Object play construction	0.23	(0.18)	0.19	(0.15)	$t(97) = 1.39$	0.28
2. Object play exploration	0.23	(0.17)	0.24	(0.17)	$t(97) = 0.14$	0.03
3. Avatar exploration	0.13	(0.12)	0.26	(0.19)	$t(97) = 4.09^{**}$	0.82
4. Avatar manipulation	0.09	(0.12)	0.11	(0.19)	$t(97) = 0.30$	0.82
5. Strategic perspective changes	0.09	(0.05)	0.11	(0.06)	$t(97) = 1.54$	0.31
6. Exploratory perspective changes	0.08	0.08	0.03	0.05	$t(97) = 3.24^{*}$	0.64
7. No move	0.14	(0.12)	0.07	(0.08)	$t(97) = 3.43^{**}$	0.68

*$p < 0.05$, **$p < 0.001$.

TABLE 5 | Means (and standard deviations) of and gender differences in proportion of Toca Blocks play time children spent engaging in each type of play behavior.

	Girls		Boys		Gender	Cohen's
	M	*SD*	*M*	*SD*	Differences	*d*
1. Object play construction	0.24	(0.18)	0.16	(0.15)	$t(97) = 2.55^{*}$	0.52
2. Object play exploration	0.25	(0.17)	0.22	(0.17)	$t(97) = 0.14$	0.74
3. Avatar exploration	0.17	(0.15)	0.23	(0.18)	$t(97) = 1.95$	0.39
4. Avatar manipulation	0.09	(0.12)	0.12	(0.16)	$t(97) = 1.31$	0.26
5. Strategic perspective changes	0.10	(0.05)	0.10	(0.06)	$t(97) = 0.50$	0.10
6. Exploratory perspective changes	0.05	0.06	0.06	0.08	$t(97) = 0.91$	0.08
7. No move	0.11	(0.10)	0.10	(0.11)	$t(97) = 0.47$	0.09

*$p < 0.05$.

children spend most of their time exploring without strategic play? Therefore, we conducted an exploratory cluster analysis using Ward's method to examine the process of children's Toca Blocks play more deeply. Ward's method is most appropriate for quantitative variables, and using a cluster analysis could reveal different styles of children's Toca Blocks play behaviors (Morey et al., 1983; Milligan and Cooper, 1987; Breckenridge, 2000; Phillips and Lonigan, 2009). We entered the proportion of time children spent engaging in each play behavior into our cluster analysis. The gap statistic method (Tibshirani et al., 2001), a technique used for estimating the number of clusters in a data set, yielded three distinct groups. To understand these groups, we first examined the descriptive statistics by cluster of the proportion of time children spent engaging in each play behavior. We found that children in Cluster 1 spent the greatest proportion of play time engaging in Object Play Exploratory, children in Cluster 2 spent the greatest proportion of play time engaging in Object Play Construction, and children in Cluster 3 spent the greatest proportion of their play time engaging in Avatar Play Exploratory. Next, we examined the number of children in each group (Cluster 1: $n = 34$; Cluster 2: $n = 22$; Cluster 3: $n = 45$). Given the large size discrepancy between Cluster 2 and Cluster 3, and that children in both Cluster 1 and Cluster 2 spent the greatest proportion of their play time engaging in a form of object play, we combined Cluster 1 and Cluster 2 into one group of children. Therefore, our cluster analysis led to the creation of two groups of children, an Object Play group ($n = 56$) and an Avatar Play group ($n = 45$). **Table 3** shows the proportion of time children spent engaging in each play behavior as a function of these two groups.

Although these clusters provide a picture of how different children played Toca Blocks, t-tests and chi-squared analyses did not reveal any differences in age, gender, and spatial skills between these two groups (all p's > 0.09). Thus, we do not use the clusters in the remaining analyses, which examined our second research aim.

Children's Characteristics and Toca Blocks Play

To address our second aim, we conducted a second set of analyses examining differences in children's play behaviors depending on several different characteristics including age, gender, and spatial skills. We begin this section with preliminary analyses focused on children's prior media experiences as a potential covariate. On average children used digital media for 123.28 min ($SD = 153.8$) the day prior to the study, and had their first exposure to tablets between 2- and 3-years-old ($SD = 1.52$). The amount of time children spent engaging with digital media the day prior to the study was not associated with the proportion of play time children spent engaging in each Toca Blocks play behavior (p's > 0.16). Similarly, there were no differences in the proportion of time children spent engaging in any play behaviors based on the age range of their first exposure to tablet devices (p's > 0.10). Therefore, we did not include either measure of prior experience in our analyses.

Developmental Differences in Toca Blocks Play

In our first analysis, we focused on developmental change in children's play behaviors. We used a series of Pearson's correlations to examine the association between the proportion of time children spent in each play behavior and their age in months (see **Table 4** for *r*-values). We found that on average, as age increased, the proportion of time children spent engaging in Exploratory Perspective changes decreased, $r(94) = -0.25$, $p = 0.02$. Additionally, on average, as age increased the proportion of time children spent engaging in Avatar Exploration increased as well, $r(94) = 0.35$, $p < 0.001$. Finally, as age increased the proportion of Toca Blocks play time in which children did not interact with the screen decreased, $t(94) = -0.32$, $p = 0.002$. There were no other significant associations between children's age and their Toca Blocks play behaviors.

Gender Differences in Toca Blocks Play

For our second analysis, we examined if there were gender differences in children's Toca Blocks play. For each play behavior, we ran a *t*-test with gender as the between group variable and the proportion of time spent engaging in that behavior as the dependent variable (see **Table 5** for means by gender and for *t*-tests and respective effect sizes). Results revealed that girls ($M = 0.25$, $SD = 0.17$) spent a greater proportion of their time engaging in Object Play Construction than boys did ($M = 0.16$, $SD = 0.14$), $t(97) = 2.55$, $p = 0.01$.

Spatial Skills Differences in Toca Blocks Play

In our final analysis, we examined the connection between children's spatial skills and their Toca Blocks play. Since younger and older children completed different spatial skills assessments (because of the lack of an appropriate test that spanned the entire age range), we conducted these analyses by age group. For both groups of children, we ran a series of correlations between scores on the spatial skills assessments and the proportion of time children spent engaging in each play behavior (see **Table 6** for all correlations). Our analyses revealed that there was not an association between younger children's spatial skills as measured by the TOSA and the proportion of time they spent engaging in any of the play behaviors, all *r*'s < 0.18, all *p*'s > 0.1. Similarly, we did not find a significant correlation between older children's spatial skills as measured by the CMTT and the proportion of time they spent engaging in any of the play behaviors, all *r*'s < 0.12, all *p*'s > 0.2.

TABLE 6 | Correlations between spatial skill assessment scores and the proportion of play time younger (TOSA scores) and older (CMTT scores) children spent engaging in each type of play behavior.

	TOSA scores	CMTT scores
	r-values	*r*-values
1. Object play construction	0.10	−0.09
2. Object play exploration	0.06	0.16
3. Avatar exploration	0.02	−0.08
4. Avatar manipulation	0.09	−0.08
5. Strategic perspective changes	−0.13	0.26
6. Exploratory perspective changes	−0.08	−0.11
7. No move	−0.25	0.05

DISCUSSION

This study aimed to examine how young children played with a digital block play app, Toca Blocks, and to explore the role of children's characteristics in this play. In pursuit of these aims, we developed a coding scheme that could reliably measure how children played with Toca Blocks. Using this coding scheme, we discovered differences in the proportion of time children engaged in various play behaviors. Additionally, we found that age and gender were associated with some play behaviors, but that spatial skills were not. These findings provide insight into the potential role of digital block play apps in children's spatial skill development, which could have implications for STEM learning.

Measuring Digital Block Play

As digital block play is a growing component of children's early spatial experiences, an important contribution of this study was the development of a coding scheme that provided insight into how children play with these apps. This coding scheme built on well-established methodologies for measuring children's play with physical blocks (e.g., Caldera et al., 1999; Kamii et al., 2004; Casey et al., 2008; Ramani et al., 2014; Bower et al., 2020) and simultaneously accounted for the unique elements of digital technology. For example, in addition to block play behaviors our coding scheme captured children's engagement with the Toca Blocks avatars and with the tools provided for changing perspectives on the screen. Moreover, our coding scheme measured two types of play for each main play behavior category. These subcategories of play behaviors were broadly associated with actions that appeared more exploratory, they did not end with a change to the objects in the environment, or more strategic, the behavior ended with a change to objects in the environment. This breakdown reflects that digital block play apps present children with opportunities to explore the game environment and experiment with objects, in addition to building definitive structures. Therefore, with our coding scheme we could examine how children engaged in an environment designed for block play in ways that may only be supported by the unique digital elements of digital block play apps.

Digital Block Play Behaviors

We found that in comparison to all other play behaviors children spent a greater proportion of their time engaging in Object Play Construction, Object Play Exploration, and Avatar Exploration. Although Toca Blocks is designed for these play behaviors, when children play with the app they are not instructed to play with the blocks or the avatars. Thus, these findings indicate that children interact with blocks in a digital environment, even without instructions.

Children's natural engagement with the blocks and avatars may reflect the strength of the cultural forms of tangible blocks

and figurines in engaging children in play regardless of the medium. Cultural forms are elements of a design that tap into a user's pre-existing conventions of how to interact with objects or others in a given situation (Horn, 2018). As the avatars resemble the physical figurines and characters children play with outside of apps, avatars may act as a cultural form that engage children in playing with their character. Additionally, children's engagement in Object Play could be a result of the strength of the cultural form of blocks, even those on a digital device, to involve children in object manipulation. Our finding that digital blocks engage children in block play behaviors without explicit instructions indicates that at a minimum digital block play apps can immerse children in the types of spatial activities, such as block play, that support spatial learning (e.g., Casey et al., 2008).

Additionally, our cluster analysis, intended to understand patterns in children's digital block play, revealed that children tended to spend most of their play time engaging in either Object Play or Avatar Play, as opposed to a combination of them both. Those children who primarily engaged in Object Play tended to stack blocks into towers, place blocks in midair, and layer blocks on top of each other to change their colors. Those children who primarily engaged in Avatar Play tended to use the avatars to move around the world and navigate through the created block structures. However, children did not seem to build structures for the avatars to play in, suggesting that children's play with either objects or avatars does not necessarily lead to play with the other. These findings parallel research on children's block play in the classroom, which demonstrates that when replica play toys, such as cars, trucks, and figurines, are placed in the block building center, children engage in less block building. Trawick-Smith et al. (2017) suggest that replica toys in block building centers simply provide an alternative activity to the block play. Consequently, when children choose to play with replica toys in building centers they are limited in the quantity and quality of building they can simultaneously complete. The current study furthers this past research on physical block play by demonstrating that in the digital medium children do not integrate construction play with their play with figurines.

Although children may spend more time engaging in either Object or Avatar Play, both types of play may individually provide children unique spatial opportunities. Object Play and Avatar Play emphasize one or the other of two kinds of engagement by children that parallel the main categories of spatial skills – those that support object manipulation and those that support representing and navigating environments (Chatterjee, 2008; Newcombe et al., 2013). Object Play more closely resembles object manipulation, as during this play children can manipulate and arrange the blocks and visualize the outcomes of these spatial transformations. Avatar Play more closely resembles navigation, as children use their avatars to explore the world. An important element of children's early spatial skill development is having play experiences that involve object manipulation and having play experiences that involve navigation (Pritulsky et al., 2020). The current study shows that digital block play may be a platform that can engage children in object manipulation and navigation, even if those activities do not occur simultaneously.

Children's Play Behaviors and Their Age, Gender, and Spatial Skills

The second aim of this study was to examine linkages between children's individual characteristics and how they played with digital blocks. First, we found age differences in children's play. Younger children tended to spend a greater proportion of their time engaging in Perspective Change Exploratory behaviors than older children, and older children spent a greater proportion of their play time engaging in Avatar Exploration behaviors than younger children. This increased engagement in avatar play with age reflects the developmental trajectory of children's symbolic understanding of dolls, especially when being used as representations of themselves (DeLoache et al., 1995; Uttal et al., 1998). Children's ability to use dolls as representations of themselves, for example to demonstrate where a sticker was placed on their own body, develops between the ages of two and five (Lytle et al., 2015). This symbolic representation of dolls is challenging for children younger than 5, particularly when the doll is in the two-dimensional form, such as a paper doll. During digital block play, avatars act as two-dimensional dolls that represent the child in the game. Given the developmental trajectory of children's symbolic understanding of dolls, it could be that younger children do not fully comprehend what avatars represent, preventing the young children in our study from engaging in Avatar Exploration as much as the older children.

Moreover, these developmental differences could also be attributed to the types of apps different aged children play with at home, and subsequently the digital play experiences most familiar to them. Research shows that between 4 and 5 years of age there is a sharp increase in the number of children who begin playing the extremely popular game, Minecraft (Mavoa et al., 2018). During Minecraft children explore the game world and build with blocks using an avatar. Accordingly, the older children in our study may have had more experiences using avatars in their play at home, and consequently may have been more drawn to engage with them for exploration during Toca Blocks. However, we did not measure the types of apps with which children play at home.

Additionally, we found gender differences in children's Object Play. Overall, girls tended to engage in Object Play Construction, meaning they built structures, more than boys. This finding parallels past research on gender differences in the types of structures children build during block play. As an example, Ramani et al. (2014) found that while boys and girls spent the same amount of time engaging in block play and built equivalently complex structures, girls included more symbolic features (e.g., a window) of the house structure they were building. Similarly, in the digital medium, girls tend to play Minecraft in Creative mode, where the focus is on creating structures using freely available resources, more than boys, who often play in Survival mode, where they must avoid hostile creatures and spend time collecting resources for building (Mavoa et al., 2018). Together, these findings suggest that during either physical or digital block play girls may focus more on what they are building than boys, but more research is still needed.

Finally, there were no associations between the proportion of time children spent engaging in any Toca Blocks play behaviors

and their spatial skills. This null result contrasts with recent research demonstrating a positive connection between children's spatial skills and their performance on other digital spatial play activities (Bower et al., 2021; Polinsky et al., 2021). For example, using a portion of this same sample of children, Polinsky et al. (2021) found a positive correlation between children's performance on the TOSA and their play with two digital puzzle play apps. However, the digital spatial play activities used in prior research differ from Toca Blocks in that they provide children with very specific goals to complete. As play goals can be important for promoting children's spatial thinking (Ferrara et al., 2011), the open-ended and creative nature of Toca Blocks may contribute to our finding that children's spatial skills are not associated with their play behaviors in this app. Additionally, this null finding could also be due to the spatial assessments used in this study. Given the unique affordances of digital block play, such as the ability to defy gravity, it is possible that the TOSA and CMTT, physical spatial skill assessments, could not capture the types of spatial skills children used when engaging in play with Toca Blocks. While this game is unrelated to spatial skills (in this sample) there is still more research needed to understand the role of digital block building apps for young children.

Limitations

A few limitations of this study should be noted. First, children's play was measured by the time spent engaging in certain types of play behaviors as opposed to a more in-depth coding of how children were manipulating objects on the screen. Attention toward how children specifically manipulated the on-screen environment will help reveal children's spatial practices during digital block play. Second, we faced a challenge in finding a pre-existing spatial skill assessment that could reliably measure the spatial skills of children across all age groups in our study. This challenge prevented us from comparing the role of spatial skills in children's Toca Blocks play across all age groups. Moreover, this challenge highlights a need for the development of spatial skills assessments that can be more readily conducted with children of a wide age range. Third, although we accounted for some of children's prior media experiences in our analyses, we did not specifically measure the types of apps with which children play at home. In turn, we cannot make claims about the potential role of the types of children's prior media experiences in their digital block play. Fourth, the study's correlational design prevents us from assessing the directionality and causality between children's individual characteristics and their Toca Blocks play. Fifth, children only played with one digital block play app in this study, even though this is a growing genre of apps for young children. To understand the generalizability of our findings to other apps from the same genre future research must examine children's play with other digital block play apps. Finally, our findings are constrained by the diversity of our sample. Most of the participating children were Caucasian and from relatively high socioeconomic backgrounds. This sample of children may have greater access to touchscreens and apps than other children (Rideout, 2017). These potential differences in experiences could impact how children played Toca Blocks. Therefore, to understand children's digital block building sandbox games, future research must include a broad sample.

Implications for STEM Education and Future Directions

Despite these limitations, our work highlights the potential role of digital block play apps in children's spatial skill development and STEM learning. This study demonstrates a promising finding that apps can successfully engage children in playing with blocks. However, how children engage in this play varies by individual characteristics and a short amount of open-ended play time may not engage children's spatial skills. Nonetheless, playing with digital blocks in an educational setting could capitalize on the opportunities for object manipulation these apps provide and might engage children in using their spatial skills in ways that may lead to spatial learning. We would suggest that when using digital block play apps in classrooms, educators may need to provide students with a clear goal that requires students to build structures that could facilitate their noticing of and considering spatial relations. Second, educators may need to provide play instructions that vary by student, such that boys may need more encouragement to build structures with these blocks, and slightly older children may need support to build with blocks in addition to exploring with their avatars. Finally, it could be that repeated and prolonged Toca Blocks play may be necessary to engage children in using their spatial skills while using this digital block play app. Future research should examine whether Toca Blocks can successfully engage children in using their spatial skills when these suggested conditions are met. This continued research can provide insight into whether digital block play apps can support children's spatial skill development and ultimately their STEM learning.

Conclusion

In conclusion, this study provides a starting point for continued research on digital block play. These digital block play apps are important because they extend the spatial play opportunities available to young children. Spatial play opportunities are one main contributor to children's early spatial skill development. As we found that these digital block play apps can engage children in object manipulation with digital blocks and objects, our research demonstrates that widely available digital block play apps may be an important and fun source for children's spatial learning. However, in the current study we found that children varied in the amount they engaged in block play based on their characteristics and we did not find that this play engaged children's spatial skills. Therefore, continued research must investigate how to design these apps in ways that best support all children's block play and spatial learning. Given the popularity of digital block play apps and the connection between spatial skills and STEM achievement (Lubinski and Benbow, 1992; Casey et al., 2008; Wai et al., 2009; Mix and Cheng, 2012; Uttal and Cohen, 2012; Uttal et al., 2013; Verdine et al., 2014a,c) these apps have the potential to support children's 21st century skills.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Northwestern University Institutional Review Board. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

NP contributed to the conceptualization, study design, data collection, data analysis, interpretation, and writing editing, and revising the manuscript. BL contributed to the data analysis, interpretation, writing, editing, and revising the manuscript. RF contributed to the conceptualization, study design, data analysis, interpretation, editing, and revising the manuscript. EW and DU contributed to the conceptualization, study design, interpretation, and review of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This project was supported by a grant from Northwestern University Institute for Innovations in Developmental Science. This work was also supported by a National Science Foundation Graduate Research Fellowship to BL (DGE-1842165).

ACKNOWLEDGMENTS

We extend our thanks to the children and families who participated, to Sarah Pila for her thoughtful feedback on a prior version of this manuscript, and to Camille Msall, Ryan Millman, Maya Davis, Elena Feigen, Stephanie McCarty, and Anjelique Bomar for their help with data collection.

REFERENCES

Aladé, F., and Nathanson, A. I. (2016). What preschoolers bring to the show: the relation between viewer characteristics and children's learning from educational television. *Media Psychol.* 19, 406–430. doi: 10.1080/15213269.2015.1054945

Bower, C., Odean, R., Verdine, B. N., Medford, J. R., Marzouk, M., Golinkoff, R. M., et al. (2020). Associations of 3-year-olds' block-building complexity with later spatial and mathematical skills. *J. Cogn. Dev.* 21, 383–405. doi: 10.1080/15248372.2020.1741363

Bower, C., Zimmermann, L., Verdine, B. N., Pritulsky, C., Golinkoff, R. M., and Hirsh-Pasek, K. (2021). Enhancing spatial skills of preschoolers from under-resourced backgrounds: a comparison of digital app vs. concrete materials. *Dev. Sci.* 25:e13148. doi: 10.1111/desc.13148

Breckenridge, J. N. (2000). Validating cluster analysis: consistent replication and symmetry. *Multiv. Behav. Res.* 35, 261–285.

Brosnan, M. J. (1998). Spatial ability in children's play with Lego blocks. *Percept. Mot. Skills* 87, 19–28. doi: 10.2466/pms.1998.87.1.19

Caldera, Y. M., Culp, A. M., O'Brien, M., Truglio, R. T., Alvarez, M., and Huston, A. C. (1999). Children's play preferences, construction play with blocks, and visual-spatial skills: are they related? *Int. J. Behav. Dev.* 23, 855–872. doi: 10.1080/016502599383577

Caldera, Y. M., Huston, A. C., and O'Brien, M. (1989). Social interactions and play patterns of parents and toddlers with feminine, masculine, and neutral toys. *Child Dev.* 60, 70–76. doi: 10.2307/1131072

Campbell, A., Shirley, L., and Caygill, L. (2002). Sex-typed preferences in three domains: do two-year-olds need cognitive variables? *Br. J. Psychol.* 93, 203–217. doi: 10.1348/000712602162544

Casey, B. M., Andrews, N., Schindler, H., Kersh, J. E., Samper, A., and Copley, J. (2008). The development of spatial skills through interventions involving block building activities. *Cogn. Instr.* 26, 269–309. doi: 10.1080/07370000802177177

Casey, B., and Bobb, B. (2003). The power of block building. *Teach. Child. Math.* 10, 98–103.

Chatterjee, A. (2008). The neural organization of spatial thought and language. *Semin. Speech Lang.* 29, 226–238. doi: 10.1055/s-0028-1082886

Cheng, Y.-L., and Mix, K. S. (2014). Spatial training improves children's mathematics ability. *J. Cogn. Dev.* 15, 2–11. doi: 10.1080/15248372.2012.725186

Cherney, I. D. (2018). "Characteristics of masculine and feminine toys and gender-differentiated play," in *Gender Typing of Children's Toys: How Early Play Experiences Impact Development*, eds E. S. Weisgram and L. M. Dinella (Washington, DC: American Psychological Association), 73–93. doi: 10.1037/0000077-005

Cherney, I. D., and Voyer, D. (2010). Development of a spatial activity questionnaire i: items identification. *Sex Roles* 62, 89–99. doi: 10.1007/s11199-009-9710-9

Cherney, I. D., Kelly-Vance, L., Gill Glover, K., Ruane, A. M. Y., and Oliver Ryalls, B. (2003). The effects of stereotyped toys and gender on play assessment in children aged 18-47 months. *Educ. Psychol.* 23, 95–106. doi: 10.1080/01443410303222

Coyle, E. F., and Liben, L. S. (2020). Gendered packaging of a STEM toy influences children's play, mechanical learning, and mothers' play guidance. *Child Dev.* 91, 43–62. doi: 10.1111/cdev.13139

Datavyu Team (2014). *Datavyu: A Video Coding Tool. Databrary Project*. New York University. Available online at: http://datavyu.Org (accessed January 25, 2020).

DeLoache, J. S., Anderson, K., and Smith, C. M. (1995). "Interviewing children about real-life events," in *Proceedings of the Biennial Meeting of the Society for Research in Child Development*, Indianapolis, IN, 91–112.

Doyle, R. A., Voyer, D., and Cherney, I. D. (2012). The relation between childhood spatial activities and spatial abilities in adulthood. *J. Appl. Dev. Psychol.* 33, 112–120. doi: 10.1016/j.appdev.2012.01.002

Ehrlich, S. B., Levine, S., and Goldin-Meadow, S. (2006). The importance of gesture in children's spatial reasoning. *Dev. Psychol.* 42:1259. doi: 10.1037/0012-1649.42.6.1259

Ferrara, K., Hirsh-Pasek, K., Newcombe, N., Golinkoff, R. M., and Lam, W. S. (2011). Block talk: spatial language during block play. *Mind Brain Educ.* 5, 143–151. doi: 10.1111/j.1751-228X.2011.01122.x

Flynn, R. M., Richert, R. A., and Wartella, E. (2019). Play in a digital world: how interactive digital games shape the lives of children. *Am. J. Play* 12, 54–73.

Gallistel, C. R., and Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition* 44, 43–74. doi: 10.1016/0010-0277(92)90050-r

Gee, J. P. (2003). *What Video Games Have to Teach us About Learning and Literacy*, 1st Edn. New York, NY: Palgrace Macmillan.

Goodfader, R. A. (1982). Sex differences in the play constructions of pre-school children. *Smith Coll. Stud. Soc. Work* 52, 129–144. doi: 10.1080/00377318209516556

Goodson, B. D. (1982). "The development of hierarchic organization: the reproduction, planning, and perception of multiarch block structures," in *Action and Thought: From Sensorimotor Schemes to Symbolic Operations*, ed. G. E. Forman (New York, NY: Academic Press), 167–201.

Gunderson, E. A., Ramirez, G., Beilock, S. L., and Levine, S. (2012). The relation between spatial skill and early number knowledge: the role of the linear number line. *Dev. Psychol.* 48, 1229–1241. doi: 10.1037/a0027433

Gura, P. (1992). "Developmental aspects of blockplay," in *Exploring Learning: Young Children and Block Play*, ed. P. Gura (London: Paul Chapman Publishing Ltd.), 48–74.

Hawes, Z., and Ansari, D. (2020). What explains the relationship between spatial and mathematical skills? A review of evidence from brain and behavior. *Psychon. Bull. Rev.* 27, 465–482. doi: 10.3758/s13423-019-01694-7

Horn, M. S. (2018). Tangible interaction and cultural forms: supporting learning in informal environments. *J. Learn. Sci.* 27, 632–665. doi: 10.1080/10508406.2018.1468259

Jirout, J. J., and Newcombe, N. S. (2015). Building blocks for developing spatial skills: evidence from a large, representative U.S. sample. *Psychol. Sci.* 26, 302–310. doi: 10.1177/0956797614563338

Kamii, C., Miyakawa, Y., and Kato, Y. (2004). The development of logico-mathematical knowledge in a block-building activity at ages 1–4. *J. Res. Child. Educ.* 19, 44–57. doi: 10.1080/02568540409595053

Kell, H. J., Lubinski, D., Benbow, C. P., and Steiger, J. H. (2013). Creativity and technical innovation: spatial ability's unique role. *Psychol. Sci.* 24, 1831–1836. doi: 10.1177/0956797613478615

Lee, J., Ho, A., and Wood, E. (2018). "Harnessing early spatial learning using technological and traditional tools at home," in *Creativity and Technology in Mathematics Education* Mathematics Education in the Digital Era, eds V. Freiman and J. L. Tassell (Cham: Springer International Publishing), 279–302. doi: 10.1007/978-3-319-72381-5_11

Levine, S., Ratliff, K. R., Huttenlocher, J., and Cannon, J. (2012). Early Puzzle Play: a predictor of preschoolers' spatial transformation skill. *Dev. Psychol.* 48, 530–542. doi: 10.1037/a0025913

LoBue, V., and DeLoache, J. S. (2011). Pretty in pink: the early development of gender-stereotyped colour preferences. *Br. J. Dev. Psychol.* 29, 656–667. doi: 10.1111/j.2044-835X.2011.02027.x

Lubinski, D., and Benbow, C. P. (1992). Gender differences in abilities and preferences among the gifted: implications for the math-science pipeline. *Curr. Dir. Psychol. Sci.* 1, 61–66. doi: 10.1111/1467-8721.ep11509746

Lytle, N., London, K., and Bruck, M. (2015). Young children's ability to use two-dimensional and three-dimensional symbols to show placements of body touches and hidden objects. *J. Exp. Child Psychol.* 134, 30–42. doi: 10.1016/j.jecp.2015.01.010

Marsh, J., Plowman, L., Yamada-Rice, D., Bishop, J., and Scott, F. (2016). Digital play: a new classification. *Early Years* 36, 242–253. doi: 10.1080/09575146.2016.1167675

Mavoa, J., Carter, M., and Gibbs, M. (2018). Children and Minecraft: a survey of children's digital play. *New Media Soc.* 20, 3283–3303. doi: 10.1177/1461444817745320

McCrink, K., and Opfer, J. E. (2014). Development of spatial-numerical associations. *Curr. Dir. Psychol. Sci.* 23, 439–445. doi: 10.1177/0963721414549751

Milligan, G. W., and Cooper, M. C. (1987). Methodology review: clustering methods. *Appl. Psychol. Meas.* 11, 329–354. doi: 10.1177/014662168701100401

Mix, K. S. (2019). Why are spatial skill and mathematics related? *Child Dev. Perspect.* 13, 121–126. doi: 10.1111/cdep.12323

Mix, K. S., and Cheng, Y.-L. (2012). "The relation between space and math: developmental and educational implications," in *Advances in Child Development and Behavior*, ed. J. B. Benson (Amsterdam: Elsevier), 197–243. doi: 10.1016/b978-0-12-394388-0.00006-x

Morey, L. C., Blashfield, R. K., and Skinner, H. A. (1983). A comparison of cluster analysis techniques withing a sequential validation framework. *Multiv. Behav. Res.* 18, 309–329. doi: 10.1207/s15327906mbr1803_4

Nazareth, A., Herrera, A., and Pruden, S. M. (2013). Explaining sex differences in mental rotation: role of spatial activity experience. *Cogn. Process.* 14, 201–204. doi: 10.1007/s10339-013-0542-8

Newcombe, N., and Frick, A. (2010). Early education for spatial intelligence: why, what, and how. *Mind Brain Educ.* 4, 102–111.

Newcombe, N., Uttal, D. H., and Sauter, M. (2013). "Spatial development," in *The Oxford Handbook of Developmental Psychology*, Vol. 1, ed. P. D. Zelazo (Oxford: Oxford University Press), 563–590.

Ostrov, J. M., and Hart, E. J. (2012). "Observational methods," in *Oxford Handbook of Quantitative Methods*, Vol. 1, ed. T. D. Little (New York, NY: Oxford University Press), 285–303.

Phillips, B. M., and Lonigan, C. J. (2009). Variations in the home literacy environment of preschool children: a cluster analytic approach. *Sci. Stud. Read.* 13, 146–174. doi: 10.1080/10888430902769533

Polinsky, N., Flynn, R., Wartella, E. A., and Uttal, D. H. (2021). The role of spatial abilities in young children's spatially-focused touchscreen game play. *Cogn. Dev.* 57:100970. doi: 10.1016/j.cogdev.2020.100970

Pritulsky, C., Morano, C., Odean, R., Bower, C., Hirsh-Pasek, K., and Michnick Golinkoff, R. (2020). Spatial thinking: why it belongs in the preschool classroom. *Transl. Issues Psychol. Sci.* 6:271. doi: 10.1037/tps0000254

Pruden, S. M., Nazareth, A., Odean, R., Abad, C., Bravo, E., and Garcia, N. (2019). "Movement, space, and the development of spatial thinking," in *The Encyclopedia of Child and Adolescent Development* (Hoboken, NJ: John Wiley & Sons, Inc.), 1–15.

Ramani, G. B., Zippert, E., Schweitzer, S., and Pan, S. (2014). Preschool children's joint block building during a guided play activity. *J. Appl. Dev. Psychol.* 35, 326–336. doi: 10.1016/j.appdev.2014.05.005

Reifel, S. (1984). Block construction: children's developmental landmarks in representation of space. *Young Child.* 40, 61–67.

Reifel, S., and Greenfield, P. M. (1982). "Structural development in a symbolic medium: the representational use of block constructions," in *Action and Thought: From Sensorimotor Schemes to Symbolic Operations*, ed. G. Forman (New York, NY: Academic Press), 203–235.

Reifel, S., and Greenfield, P. M. (1983). Part-whole relations: some structural features of children's representational block play. *Child Care Q.* 12, 144–151. doi: 10.1007/bf01151602

Revelle, W. (2019). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Evanston, IL: Northwestern University.

Rideout, V. (2017). *The Common Sense Census: Media Use by Kids Age Zero to Eight*. San Francisco, CA: Common Sense Media, 263–283.

Sarkar, S. (2017). *Minecraft Sales Hit 122M Copies. Polygon*. Available online at: https://www.polygon.com/2017/2/27/14755644/minecraft-sales-122m-copies (accessed December 15, 2021).

Sheehan, K., Hightower, B., Lauricella, A. R., and Wartella, E. (2018). STEM media in the family context: the effect of STEM career and media use on preschoolers' science and math skills. *Eur. J. STEM Educ.* 3:17. doi: 10.20897/ejsteme/3877

Shrout, P. E., and Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychol. Bull.* 86:420. doi: 10.1037//0033-2909.86.2.420

Siegler, R. S., and Ramani, G. B. (2008). Playing linear numerical board games promotes low-income children's numerical development. *Dev. Sci.* 11, 655–661. doi: 10.1111/j.1467-7687.2008.00714.x

Sluss, D. J. (2002). Block play complexity in same-sex dyads of preschool children. *Play Cult. Stud.* 4, 77–92.

Sorby, S., Casey, B., Veurink, N., and Dulaney, A. (2013). The role of spatial training in improving spatial and calculus performance in engineering students. *Learn. Individ. Differ.* 26, 20–29. doi: 10.1016/j.lindif.2013.03.010

Stiles, J., and Stern, C. (2001). Developmental change in spatial cognitive processing: complexity effects and block construction performance in preschool children. *J. Cogn. Dev.* 2, 157–187. doi: 10.1207/s15327647jcd0202_3

Syed, M., and Nelson, S. C. (2015). Guidelines for establishing reliability when coding narrative data. *Emerg. Adulthood* 3, 375–387. doi: 10.1177/2167696815587648

Tibshirani, R., Walther, G., and Hastie, T. (2001). Estimating the number of clusters in a data set *via* the gap statistic. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 63, 411–423. doi: 10.1111/1467-9868.00293

Toca Blocks (2017). *Toca Boca*. Available online at: https://tocaboca.com/app/toca-blocks/ (accessed September 15, 2018).

Trawick-Smith, J., Swaminathan, S., Baton, B., Danieluk, C., Marsh, S., and Szarwacki, M. (2017). Block play and mathematics learning in preschool: the effects of building complexity, peer and teacher interactions in the block area, and replica play materials. *J. Early Child. Res.* 15, 433–448. doi: 10.1177/1476718X16664557

Uttal, D. H., and Cohen, C. A. (2012). "Spatial thinking and STEM education: when, why, and how?," in *Psychology of Learning and Motivation*, ed. B. H. Ross (Amsterdam: Elsevier), 147–181. doi: 10.4135/9781452229379.n10

Uttal, D. H., Marzolk, S. L., Pierroutsakos, S. L., Smith, C. M., Troseth, G. L., and Scudder, K. V. (1998). "Seeing through symbols: the developmental of children's understanding of symbolic relations," in *Multiple Perspectives on Play in Early*

Childhood Education, eds O. N. Saracho and B. Spodek (Albany, NY: State University of New York Press), 59–79.

Uttal, D. H., Miller, D. I., and Newcombe, N. (2013). Exploring and enhancing spatial thinking: links to achievement in science, technology, engineering, and mathematics? *Curr. Dir. Psychol. Sci.* 22, 367–373. doi: 10.1177/0963721413484756

Verdine, B. N., and Golinkoff, R. M. (2017). *Test of Spatial Assembly (TOSA) Instruction Manual*. Newark, DE: University of Delaware.

Verdine, B. N., Golinkoff, R. M., Hirsh-Pasek, K., and Newcombe, N. (2014a). Finding the missing piece: blocks, puzzles, and shapes fuel school readiness. *Trends Neurosci. Educ.* 3, 7–13. doi: 10.1016/j.tine.2014.02.005

Verdine, B. N., Golinkoff, R. M., Hirsh-Pasek, K., Newcombe, N. S., Filipowicz, A. T., and Chang, A. (2014b). Deconstructing building blocks: preschoolers' spatial assembly performance relates to early mathematical skills. *Child Dev.* 85, 1062–1076. doi: 10.1111/cdev.12165

Verdine, B. N., Irwin, C. M., Golinkoff, R. M., and Hirsh-Pasek, K. (2014c). Contributions of executive function and spatial skills to preschool mathematics achievement. *J. Exp. Child Psychol.* 126, 37–51. doi: 10.1016/j.jecp.2014.02.012

Verdine, B. N., Golinkoff, R. M., Hirsh-Pasek, K., and Newcombe, N. (2017). *Links Between Spatial and Mathematical Skills Across the Preschool Years*. Hoboken, NJ: Wiley.

Wai, J., Lubinski, D., and Benbow, C. P. (2009). Spatial ability for STEM domains: aligning over 50 years of cumulative psychological knowledge solidifies its importance. *J. Educ. Psychol.* 101, 817–835. doi: 10.1037/a0016127

Wakefield, E. M., Foley, A. E., Ping, R., Villarreal, J. N., Goldin-Meadow, S., and Levine, S. (2019). Breaking down gesture and action in mental rotation: understanding the components of movement that promote learning. *Dev. Psychol.* 55, 981–993. doi: 10.1037/dev0000697

Worsley, M., and Bar-El, D. (2020). Inclusive Making: designing tools and experiences to promote accessibility and redefine making. *Comput. Sci. Educ.* 30, 1–33. doi: 10.1080/08993408.2020.1863705

Zosh, J. M., Hirsh-Pasek, K., Hopkins, E. J., Jensen, H., Liu, C., Neale, D., et al. (2018). Accessing the inaccessible: redefining play as a spectrum. *Front. Psychol.* 9:1124. doi: 10.3389/fpsyg.2018.01124

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Polinsky, Lemley, Flynn, Wartella and Uttal. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Education

EMPIRICAL STUDY
published: 06 May 2022
doi: 10.3389/feduc.2022.894142

Enhancing Digital Skills of Early Childhood Teachers Through Online Science, Technology, Engineering, Art, Math Training Programs in Estonia

Janika Leoste[1], Zsolt Lavicza[2], Kristof Fenyvesi[3], Maire Tuul[1] and Tiia Õun[1]*

[1] School of Educational Sciences, Tallinn University, Tallinn, Estonia, [2] Linz School of Education, Johannes Kepler University, Linz, Austria, [3] Finnish Institute for Educational Research, University of Jyväskylä, Jyväskylä, Finland

OPEN ACCESS

Edited by:
Stamatios Papadakis, University of Crete, Greece

Reviewed by:
Sara Dias-Trindade, University of Coimbra, Portugal
Daisy Volmer, University of Tartu, Estonia
Fuat Balcı, University of Manitoba, Canada

***Correspondence:**
Janika Leoste
leoste@tlu.ee

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Education

Received: *11 March 2022*
Accepted: *12 April 2022*
Published: *06 May 2022*

Citation:
Leoste J, Lavicza Z, Fenyvesi K, Tuul M and Õun T (2022) Enhancing Digital Skills of Early Childhood Teachers Through Online Science, Technology, Engineering, Art, Math Training Programs in Estonia. Front. Educ. 7:894142. doi: 10.3389/feduc.2022.894142

Teacher professional development programs, including mid- and long-term Science, Technology, Engineering, Art, Math (STEAM) courses, have recently moved from in person learning at university premises to an online environment. Whether it is a temporary change in learning methods caused by the COVID-19 restrictions or whether it will become a new normal is currently under discussion in many teacher training institutions around the world. The aim of this study was to design and implement time- and money-saving synchronous online teacher training format for conducting co-design courses for early childhood teachers in the theme of STEAM integrated learning activities. Based on Tallinn University's curriculum of in-person training courses on the same topic, with the volume of 40 contact hours, we delivered the content in two different formats: in 11-months (as it used to be in pre-COVID period) and in 4-months, adapted to participants' needs. We used a self-assessment survey, based on DigCompEdu framework, to assess the increase of digital competences in the two formats. The long-format course had 31 participants and the short-format course had 50 participants. The assessment was based on pre- and post-test and we used structured live video presentations to let participants retrospectively describe their learning experiences. Results indicate that the participants of both courses had improved their digital competences and achieved the learning outcomes set by course content. There was no significant difference in increase of digital competences or the way the course was perceived between participants of both courses. This brings us to the cautious consideration that it is possible

to achieve desired outcomes of STEAM courses even in a shorter period when conducting them online compared to the in-person courses. There is a need for further research where results from participants of in-person and online teaching courses are compared.

Keywords: digital competence, teacher training, STEAM, DigCompEdu, online education

INTRODUCTION

In the near future at least 90% of professional roles in Europe will require basic digital skills, similarly to the skills of basic literacy or numeracy (European Commission, 2017a). However, the improvements in this area have been slow. According to the DESI 2021 report (European Commission, 2021a), by 2020 about 42% of the adult population still lacked basic digital skills, a 1% improvement compared to 2015. The situation is further aggravated by the fact that 88% of workplaces have done nothing to increase their employees' digital skills, often citing high costs as the main barrier to actions to deal with digital skills caps (European Commission, 2017b). Combined with the ambitious goal of the EU to ensure that at least 80% of adults would have basic digital skills by 2030 (European Commission, 2021b), it becomes clear that, together with other actors, the educational systems of EU countries are expected to contribute to achieving this goal.

According to the United Nations Development Programme's Global Knowledge Index (GKI), Estonia is a leading performer in knowledge infrastructure. It ranks 15th out of 154 countries in the GKI 2021 and 15th out of the 61 countries with very high human development. Estonia's areas of strength include E-participation, and the country's areas of improvement show an increase in teaching staff compensation (% tertiary expenditure). Estonia's recent global success in the PISA assessment has been widely discussed. The country's "Learning-adjusted years of schooling index" is also ranking Estonia in 5th place according to GKI (based on World Bank's Human Capital Index). It is worth mentioning that in GKI's global comparison, Estonia ranks 1st both in "Schools with access to computers in primary education (%)" and "Schools with access to computers in secondary education (%)" (based on UNESCO Institute for Statistics' Database). According to GKI, "ICT Employment" and "Internet activities by individuals (%)" are also very high in Estonia; in both categories, the country is ranking 5th. However, at the same time, the "Gross enrolment ratio in early childhood education" is not necessarily high: Estonia ranks in 49th place in this category according to GKI (UNDP & Mohammed Bin Rashid Al Maktoum Knowledge Foundation, 2021).

The task of educational systems to prepare the youth for the requirements of the modern labor market, including providing them with adequate digital skills, has been discussed frequently for several decades already. Integrating Science, Technology, Engineering, Art, Math (STEAM) principles into regular curricula has been offered as a viable solution for making students become interested in jobs that require good digital skills (i.e., the majority of jobs in the coming decades). However, integrating STEAM to curricula requires teachers to en masse accept the principles of technology-enhanced learning (TEL) – a task that according to various studies has not been entirely successful (Niederhauser et al., 2018; Leoste, 2021) for various reasons. One of these reasons being the low level of existing digital competences (Sánchez-Cruzado et al., 2021). As teachers are expected to use technology for both making their work more efficient and for providing students with relevant subject-related technological skills, it has become crucial to provide teachers with training programs with teachers' digital competences as one of the key aims (Usart Rodríguez et al., 2021).

Digital technologies are recognized not only as working and learning tools but also tools of participation in society. Because of the deep embeddedness of digital technology in our society and to prevent the further growth of the digital gap, early childhood educators need to develop their own and their students' digital competencies (Galindo-Domínguez and Bezanilla, 2021). At the same time, early childhood education's vulnerability in these aspects has been recognized in connection with the worrying situation of the lower technological competence at lower educational levels (Portillo et al., 2020). The complex needs and existing shortcomings in this field have become clearly visible in the COVID-19 era.

The COVID-19 pandemic affected all levels and areas of education, including teacher training and professional development of teachers. To cope with the first shocks of the forced transition to "emergency remote teaching" (Bozkurt and Sharma, 2020), both educators and learners needed to solve complex technical, pedagogical, social, cognitive, and practical challenges. Teacher training institutes, including the Tallinn University were required to develop new infrastructure and learning environments and adapt to new teaching methods, learning scenarios, and study materials. Moreover, both the instructors and the participants of the teacher education and professional development programs had to develop their digital competencies to respond to all the new kinds of challenges.

In Europe the demand for a workforce with meaningful basic digital skills is growing rapidly while the educational systems have difficulties in keeping pace with the demand, partly due to low digital competence of teachers. In order to increase the educational systems' capacity of improving teachers' digital competence, relevant teacher training courses need to become shorter while at least retaining the existing efficiency – or even improving it. In this article, we summarize the main outcomes of a case study, which provided the scientific background to Tallinn University's synchronous online teacher training format for conducting co-design courses for early childhood teachers in the theme of STEAM integrated learning activities. We start by opening the theoretical background on teacher digital competence, its impact on student digital skills development,

ways of training it, and stating the research aim and research questions. Next, we describe the background of the case study that examines the results of two different teacher training courses, followed by the description of the data collection and analysis methods and of the sample. We then answer our two research questions, and discuss the findings.

Theoretical Background

Building Digital Literacy in Education

Digital skills make it possible for people to participate in modern learning, working and social activities by allowing them to manipulate digital content, use digital communication and collaboration tools for solving various problems in their lives (UNESCO, 2018; European Commission, 2020). Of these, basic digital skills are required for basic use of digital and online technologies and, together with reading, writing and numeracy, are considered an important component of the modern literacy skill set (UNESCO, 2018). Teaching digital literacy at schools requires schools to accept several TEL innovations and to appropriate relevant novel classroom practices (Heidmets and Eisenschmidt, 2020), integrating these technologies and related practices to various subject disciplines. Students need their teachers to act as guides and facilitators (Pérez-Jorge et al., 2020) when constructing their digital skills, as their learning outcomes are significantly related to the teachers' mastery of teaching subject discipline (Bakar, 2018; Fauth et al., 2019). The success of integrating digital technologies in classroom practices depends largely on teachers' digital competences (Pérez-Jorge et al., 2020; Sabalete Suárez and Roblizo Colmenero, 2021) and attitudes (Papadakis et al., 2021a). However, the reality of technology integration can be different from theoretical approaches, as teachers are often unable to integrate technology and relevant methodologies with their subject content (Franzoni Velázquez et al., 2020). This deficiency suggests that teachers would need proper training and support when adopting technology in their classrooms (Franzoni Velázquez et al., 2020). Besides providing teachers with technology-related knowledge, these training courses should help them to develop their digital competence, needed to teach their students (Pérez-Jorge et al., 2020). Providing effective and attractive learning environments requires a complex approach on every level, considering a wide range of design principles to support the implementation of multiple pedagogies and developing both subject-based and cross-curricular knowledge, skills, attitudes, values, and ethics (Mäkelä et al., 2020).

Early Childhood Educators' Digital Competences

Findings in teacher training, and professional development for early childhood education (Ananiadou and Rizza, 2010; Tondeur et al., 2017; Casillas et al., 2020) show future teachers' perception of medium digital competence. While the COVID-19 crisis increased the emphasis on training and professional development in ICT and digital competences for early childhood education. In some cases, however, training programs were limited to using some online platforms, and there were no pedagogical practices and digital competences built for effectively engaging in online or distance education with young children (Atiles et al., 2021). A recent study (Galindo-Domínguez and Bezanilla, 2021) points out that pre-service teachers of early childhood education scored significantly lower regarding content creation compared to other digital competence dimensions. In line with Galindo-Domínguez and Bezanilla's (2021) discussion, the reinforcement of digital content creation is crucial. Both teachers and children already from an earlier age are expected to be not only end-users of technology but to become content creators (Drotner, 2020). More digital creativity in the learning environment also leads to more effective implementation of active and innovative methodologies (López Belmente et al., 2019).

Teachers' Professional Development in Digital Competences Through Science, Technology, Engineering, Art, Math Approaches

Teachers' digital competences are often developed through the means of teacher professional development (TPD) programs that utilize co-creation related social processes (Prieto-Alvarez et al., 2018) while providing teachers with knowledge and skills about certain STEAM approaches or technologies (Herranen et al., 2021). In such training programs, experts of various domains (e.g., technology, pedagogy, and learning content) combine their knowledge in order to provide participants with in-depth knowledge about using the selected technologies in the context of teachers' everyday teaching routines (Leoste, 2021). The co-creation approach relies on the ideas about user innovation, where end-users are involved in co-creation of innovation-related artifacts (in the context of education: teaching methods and relevant materials, related to various educational digital technologies), leading to higher user motivation and better innovation adoption rates (Von Hippel, 2017; Bradonjic et al., 2019).

Previous research indicates that participant learning can be better supported via iteratively arranged longitudinal training courses, which consist of cycles of contact days followed by participant activities of co-creation and implementation (Botha and Herselman, 2018; Leoste et al., 2020). Similarly, Henriques et al. (2021) argue for longer duration training courses, as these would help participants to better associate the learned content with their real-life practices. Such a daily teaching practice-oriented, continuous training model is argued positively by further studies as well (Papadakis et al., 2021b).

However, the question of duration is not an easy one as different authors define it differently. Burgess and McGregor (2018), in their literature review about formal teacher training programs, describe brief programs with the duration of several hours to some weeks, whereas in their classification there are also longitudinal programs that stretch over multiple years, while delivering tens of contact hours combined with a few online modules. In general, it seems that a course that lasts less than a semester is considered a short-term program, while a program with longer duration is considered a long-term program. While the short-term training courses might not see the same maturation of participant knowledge compared to long-term courses, they still have some important benefits. For example, short-term courses could lead to savings in financial

resources, and are also less demanding on teachers' already intensive schedules (Leoste et al., 2019).

Research Aims

Our goal in this paper is to examine two teacher-training courses with different durations that aim at increasing teachers' digital competences, using the STEAM technology of educational robotics as the learning focus of the courses. In particular, we are examining how the course duration influences the teachers' post-course digital competences. The underlying rationale for the study is based on teachers' overbooked schedules. The lack of time makes it more difficult for teachers to participate in long-term courses, forcing them often to choose for shorter-term courses.

We are using the European Commission's DigCompEdu framework (European Commission, 2022) to evaluate participants' digital competences before and after the courses, and, by using participant self-reflection, we will gather their more general feedback about the courses. To keep our study in focus, we have formulated the following research questions (RQs):

1. How did teachers assess their digital competences before and after the STEAM course?
2. Are there any differences in development of digital competences between long and mid-term course participants?
3. How did teachers describe their learning experience they got in the STEAM course?

MATERIALS AND METHODS

Case Descriptions

During the year 2021, two online in-service teacher-training courses were conducted at Tallinn University. The participants of both courses were the teachers of early childhood and primary school education levels. The goal of the courses was to provide teachers with necessary skills and knowledge for integrating digital STEAM tools to their teaching practices. Both of the courses included online contact lessons (32 academic hours) and independent work (20 academic hours). One academic hour in Estonia is 45 min. The major difference between the courses was their longitudinal structure: one of them (Case 1) took place from January to November 2021, having in total 13 online gatherings (8 of those had a duration of 1.5 academic hours and 5 had a duration of 4 academic hours). The other course (Case 2) took place from August to December 2021, having in total 8 online gatherings with the duration of 4 academic hours each. Considering that most of the learners were kindergarten teachers and took part from their workplaces, the courses started at 1 PM when in the Estonian kindergartens there is a sleeping hour for children.

The gatherings of both courses included the following content:

(a) lectures about digital competences of teachers and students, the influence of digital devices on the different aspects of child development, appropriate teaching practices for digital devices, and special teaching cases such as with students with special educational needs;
(b) introduction of STEAM teaching approaches and devices, such as digital educational games, simple educational robots, digital educational toys, various novel educational technologies (VR, AR, interactive flat panels, etc.), and digital environments for creating simple educational games and apps;
(c) additional activities, such as co-creation of teaching activities, individual and group reflections, presentations of final assessments.

The actual content of courses was similar and the lecturers were the same, to ensure that besides the difference in overall duration the training course experience would be the same for the participants of both courses.

The participants were pre-informed about the type of STEAM kits introduced and used in the dedicated sessions, so they could equip themselves with the needed kits or similar substitutions. In case the participants did not have any kits available at their workplace, they were able to lend them from the university's lab. The participants had to create teaching activities and test them with their students during the periods between the online gatherings, and reflect on their experience of conducting activities.

Data Collection and Analysis

RQ 1 and 2: Our interest is to understand the differences in growth of participant digital competences between an online training course with its duration stretched on a longer period (e.g., 11 months) and a greater number of online gatherings (e.g., 13 gatherings), and a course with a shorter duration (e.g., 4 months) and less gatherings (e.g., 8 gatherings).

To this end, we measured the digital competences of course participants before and after the course, using a shortened online query tool that is based on the Estonian translation (HARNO, 2022) of the European Commission's DigCompEdu framework (European Commission, 2022). We left out the areas "personal learning and development" and "student assessment" as the content of the courses did not especially target these areas. The final questionnaire in Estonian consisted of 32 indicators, which are divided into four areas of competence:

- Area 2, "Digital Resources": sourcing, creating and sharing digital resources.
- Area 3, "Teaching and Learning": managing and orchestrating the use of digital technologies in teaching and learning.
- Area 5, "Empowering Learners": using digital technologies to enhance inclusion, personalization and learners' active engagement;
- Area 6, "Facilitating Learners' Digital Competence": enabling learners to use digital technologies creatively and responsibly for managing information, communication, content creation, wellbeing and problem solving.

The questionnaire used Likert-type scales with six levels as follows:

- 0: No previous experience. I have not yet developed this competence.
- 1: Beginner. I can explain what it is, and I have tried it in my work.
- 2. I'm halfway to becoming an expert.
- 3: Expert. I am routinely using it.
- 4: I'm halfway to being a leader.
- 5: Leader. I am an advisor and expert in this field both inside and outside my organization.

The levels 0 and 1 indicate little to no previous contact with digital technologies. The levels 2 and 3 refer to educators' ability to appropriate and adapt basic digital practices, levels 4 and 5 refer to educators' ability to effectively manage their digital practices, with the level 5 indicating the educators' ability to critically assess existing practices, develop new ones and share their knowledge with peers. The questionnaire in its adapted form was previously used and validated in a study by Heinmäe et al. (2022) where university experts confirmed that the meaning of the questionnaire items, when compared to its original form, was retained as in the sources. We used Google Forms as the media to deliver the questionnaire to the participants. The questionnaire was asked to be filled in before and after the training course that the participant took part in. Answering the questionnaire was personalized to eliminate entries that were missing either the pre or post-test answers. However, personal information was removed before data analysis.

For analyzing the Likert-type scale data (see also Pimentel, 2019) we first summed the responses by each scale level, and by merging the scale levels within every DigCompEdu area by its subcategories. Next, we reduced the number of Likert items' response levels by merging the scale level responses 0 and 1 (into the level "Beginner"), the scale level responses 2 and 3 (into the level "Expert") and the scale level responses 4 and 5 (into the level "Leader"). Subsequently we counted the number of responses in each of the new levels in every DigCompEdu area and calculated the share of each level (in percentages) per DigCompEdu area. That way we found the share of beginners, experts and leaders among the participants of each course before and after the course (**Tables 1** and **2**).

RQ 3: We also wanted to understand how the participants appropriated the innovative method, introduced them during the training course, and how they planned to start using it in their organizations. To this end, we asked the participants to prepare a short presentation as part of their final assessment, based on the following two questions: "How did the participant assess their digital competences before and after the training course?" and "How did the participant describe the learning experience they got in the training course?" The presentations were video recorded during an online session and transcribed by two researchers. We used open coding when looking for emergence of meaning of clusters (Williams and Moser, 2019). The differences in coding results were removed through consensus-seeking discussions. The occurrence of codes was not quantified due to the different sample sizes.

Sampling

With both cases, participation in the study was voluntary for the training course participants. They were informed that their individual personal data would be removed from their input and that their decision to participate would not affect their chances to pass the course. Case 1 had in total 56 participants ($N1 = 56$), while Case 2 had in total 78 participants ($N2 = 78$). They all submitted their final assessment presentations (that we used to answer RQ3). All participants were female, their age was not recorded.

Of Case 1 participants, 31 ($n1 = 31$) filled in both the pre and posttest questionnaires (data for answering Research Questions 1 and 2). All of Case 1 survey respondents were female, with the average age of 42.7 years (at the moment of filling in the post test survey). Of Case 2 participants 50 ($n2 = 50$) filled in both the pre and posttest surveys. All of Case 2 survey respondents were female, with the average age of 43.1 years (at the moment of filling in the post test survey).

RESULTS

Teachers' Assessment of Their Digital Competences Before and After the Science, Technology, Engineering, Art, Math Course, and Differences in Digital Competence Development Between the Long and Mid-Term Course Participants

Before the long-term training course of Case 1, at least a third of Case 1 participants considered themselves possessing beginner-level digital competences, while slightly below 2/3rds viewed themselves as experts (**Table 1** and **Figure 1**). However, their competences when teaching their own students were assessed more critically – roughly half of the participants (48.4%) admitted that they had only limited knowledge about developing the digital competences of their own students, while less than half of the participants (47.3%) considered themselves experts in this area. In all DigCompEdu areas, only 4.9% of the participants considered to possess the leader-level digital competences. After the training course, the share of the participants who considered themselves to possess leader-level digital competences rose significantly in all DigCompEdu areas. The participants considered themselves especially competent in managing and using Digital Resources (57.3% on the expert level and 24.2 on the leader level), while the increase was the smallest in Facilitating Learner's Digital Competences (the majority of participants considered themselves as experts (54.3%) or beginners (31.7%) in this area).

According to the assessments of Case 2 participants, their digital competences before Case 2 training course were lower, compared to Case 1 participants (**Table 2** vs **Table 1**, **Figure 2** vs **Figure 1**). At least half of the participants considered themselves beginners, and about 40% considered themselves experts in most of the DigCompEdu areas, with the exception of the "Facilitating Learners' Digital Competence" area. In this latter area, about two thirds of the participants believed themselves to be at the

TABLE 1 | The level of Case 1 participants' digital competences, by percentage.

Competence level DigCompEdu Area	Beginner (levels 0 and 1)	Expert (levels 2 and 4)	Leader (levels 4 and 5)
Digital resources			
Pre-test	31.5	59.7	8.9
Post-test	9.7	57.3	33.1
Δ	*−21.8*	*−2.4*	*24.2*
Teaching and learning			
Pre-test	33.1	59.7	7.3
Post-test	6.5	71.8	21.8
Δ	*−26.6*	*12.1*	*14.5*
Empowering learners			
Pre-test	34.4	61.3	4.3
Post-test	11.8	67.7	20.4
Δ	*−22.6*	*6.5*	*16.1*
Facilitating learners' digital competence			
Pre-test	48.4	47.3	4.3
Post-test	31.7	54.3	14.0
Δ	*−16.7*	*7.0*	*9.7*

TABLE 2 | The level of Case 2 participants' digital competences, by percentage.

Competence level DigCompEdu Area	Beginner (levels 0 and 1)	Expert (levels 2 and 4)	Leader (levels 4 and 5)
Digital resources			
Pre-test	50.0	39.0	11.0
Post-test	14.0	65.0	21.0
Δ	*−36.0*	*26.0*	*10.0*
Teaching and learning			
Pre-test	50.0	45.0	5.0
Post-test	19.0	68.0	13.0
Δ	*−31.0*	*23.0*	*8.0*
Empowering learners			
Pre-test	53.3	39.3	7.3
Post-test	19.3	64.7	16.0
Δ	*−34.0*	*25.3*	*8.7*
Facilitating learners' digital competence			
Pre-test	64.3	32.3	3.3
Post-test	28.7	57.0	14.3
Δ	*−35.7*	*24.7*	*11.0*

beginner level and about a third considered themselves experts. This exception was similarly present also with Case 1 participants. However, after the training course the self-assessed competence levels improved significantly.

In most DigCompEdu areas, the share of expert level Case 2 participants became similar or even exceeded that of Case 1 participants, although the share of beginners remained higher and the share of leaders lower, compared to Case 1. Remarkably though, in the area of "Facilitating Learners' Digital Competence" the Case 2 participants' post-test confidence about their competences exceeded that of the Case 1 participants in all areas: there were less beginners (28.7 vs 31.7%), more experts (57 vs 54.3%) and slightly more leaders (14.3 vs 14.0%).

When comparing the dynamics of the digital competences of the participants of the both cases (**Figures 1** and **2**), the following suggestions can be made. First, with the short-term course (Case 2) the share of beginner-level participants reduced significantly more in all DigCompEdu areas, compared to that of the long-term course (Case 1). Second, the share of expert level participants increased remarkably more in Case 2, compared to Case 1. Third, the share of leader-level participants did increase more in the long-term course (with the exception of "Facilitating Learners' Digital Competences," but this can be explained by the higher initial digital competence levels of its participants. These observations lead to the fourth suggestion that specifically concerns the "Facilitating Learners' Digital Competences" area. This area is about one of the key competences of the teachers – their ability to enable their students to use digital technologies creatively and responsibly. With Case 2, the initial digital competence levels in this area were remarkably higher compared to Case 1. However, after the course, this was the area where Case 2 participants' digital competence levels exceeded those of Case 1

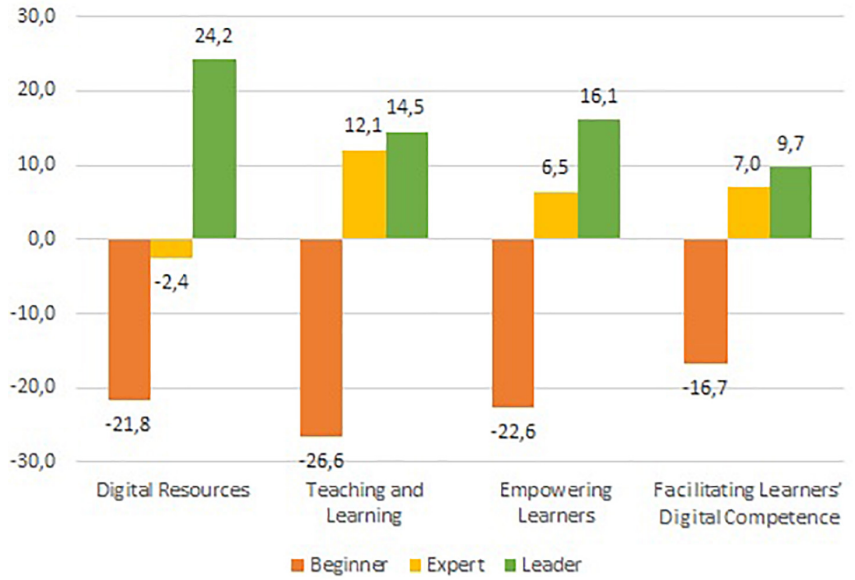


FIGURE 1 | Dynamics of Case 1 participants' digital competences (increase in competence-level group sizes, by DigCompEdu area, in percentage points).

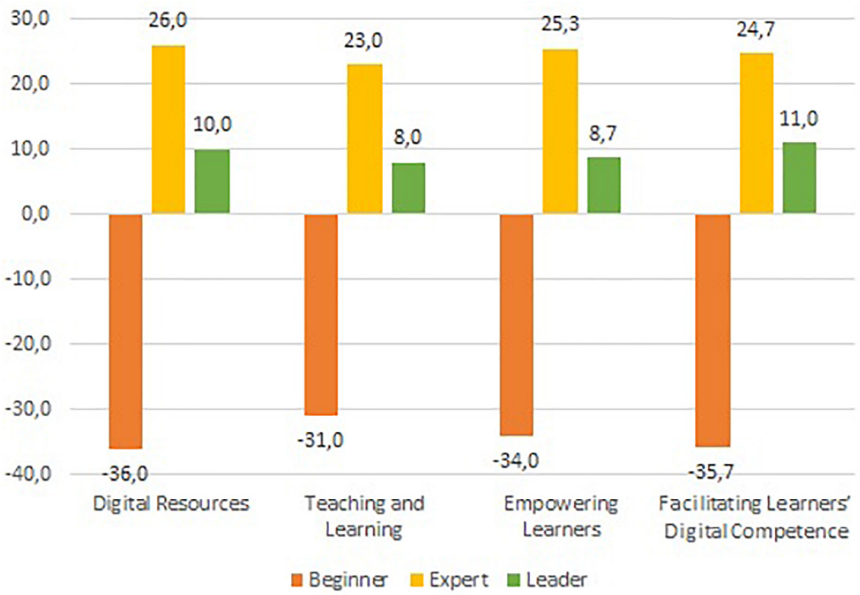


FIGURE 2 | Dynamics of Case 2 participants' digital competences (increase in competence-level group sizes, by DigCompEdu area, in percentage points).

participants. In other areas, the growth of competences was not so conclusively expressed.

Teachers' Description of Their Learning Experience They Got in the Science, Technology, Engineering, Art, Math Course

When reflecting on their learning experiences, all of the participating teachers of both training courses described how the training contributed to the growth of their knowledge and courage of using digital tools. The participants were able to become familiar with new tools and environments and learned to use them in their teaching practices. The participants found it important that the training courses used a variety of educational tools and good guiding materials, its instructors were supportive and that it was possible to learn from the experiences of their peers and feel their support. Many participants highlighted the possibility, considered as one of the biggest values of the courses, to learn from the activities, to use shared teaching materials

and practical experiences of their fellow teachers. For example, a kindergarten teacher said: "*Thanks to the course I became encouraged to start using (educational) robotics tools; regularity, consistency, courage to use developed in my activities. Whereas earlier I only used the tools I was familiar with, now I also have the courage to start using new (unfamiliar) tools.*" On the topic of substantive planning of learning activities, many participants pointed out the importance of understanding the meaning of the integration of subject areas. In their opinion, the training courses helped them to recognize better how to integrate subject disciplines while using digital tools to fulfill learning objectives.

Teachers who participated in the long-term training course repeatedly mentioned the increase in motivation, the development of self-analysis skills, and the systematic and consistent use of digital tools in their everyday teaching activities as an effect of the training on their professional development. In most cases, the reflection of teachers who had completed the shorter training was limited to mentioning the development of new knowledge, ideas and practical skills. In the case of supporting children's development, teachers of the shorter training highlighted their new experience in working with children with special needs, where digital tools have helped the child to learn actively and supported co-operation between teacher and child. For example, a comment from a schoolteacher with long-term teaching experience states: "*Using a digital tool makes it easier to get in touch with a student with SEN, makes it easier for the teacher to communicate with the child and the child is more in contact with the activity.*" Participants were asked to describe how they see the further development of the field of learning robotics in their educational institution.

A large number of participants mentioned that their educational institutions had purchased new digital educational tools. Teachers of the shorter training course saw further development mainly in the context of their own classroom: they mentioned how often and what different tools and activities they plan to use in their teaching practices in the future. Some teachers also mentioned training a teaching partner or conducting an in-house training course. In many cases, teachers who had completed the longer training course, mentioned training colleagues, organizing regular workshops and mentoring co-teachers as a part of integrating digital tools into teaching.

DISCUSSION

We aimed at clarifying how a long-term STEAM training course for teachers would influence their digital competences, compared to a shorter-term STEAM training course. To this end we conducted two similar training courses, both focused on providing teachers with the skills and knowledge needed for integrating the STEAM technology of educational robotics to their teaching practices. The main difference between these two courses was their duration: the length of Case 1 was 11 months whereas the duration of Case 2 was 4 months. There were no other meaningful differences between the two cases. The increase in digital competences was relatively similar with both the long-term and short-term cases. However, it seems that with the short-term training course the share of beginners decreased more vigorously and the increase of experts was more vigorous than with the longer training course.

The DigCompEdu area of "Facilitating Learners' Digital Competences" remains a question with particular importance as it reflects the ability of teachers to prepare their students for the digital future. With the short-term training course, the competence growth was uniform in all DigCompEdu areas (**Figure 2**), while with the long-term training course the growth was significantly higher in all other areas (both when compared to the "Facilitating Learners' Digital Competences" area or to the results of the short-term training course). Our experiment did not give full clarity about the reasons for this peculiarity, suggesting that further studies are needed to understand this question. It would be important to determine if a long-term training course would help teachers to become more realistic about the difficulties in enabling their students to creatively and responsibly use digital technologies, or if, considering that the final digital competence levels in this area were similar in both cases (e.g., 14.0% leaders in Case 1 vs 14.3% leaders in Case 2), it is possible for a teacher to understand relatively quickly these difficulties and therefore the problem is not solvable by allocating more training time but instead more effective teaching practices are required. However, it is also possible that the underlying reasons for this phenomenon are tied to the reasons teachers use digital technology in their lessons. In most countries, the curriculum goals have remained unchanged despite the pressure from stakeholders toward integrating TEL and STEAM (UNESCO, 2016). Thus, the primary focus of using technology in their lessons is on delivering daily teaching (Perifanou et al., 2021). In this case, facilitating the digital learning of their students would require national level curriculum change.

The qualitative results imply that both courses were beneficial for growing participants' digital knowledge and related confidence. With both approaches, long-term and short-term, it is important to cover a wide range of usable technologies, to have good guiding materials, supportive instructors, and to allow peer experience and support to be used as a part of the course. However, based purely on participant feedback, it would seem that the long-term courses could have better results in increasing participant motivation, in developing their self-analysis skills and changing their teaching practices – similarly to what is implied by Henriques et al. (2021). Then again, it is possible to argue that at least some of these developments can be achieved outside the training format naturally, as this maturation of knowledge happens as a result of teachers applying their newly-found knowledge and skills in their classrooms. In this case, it could save resources if instead of long-term training courses shorter ones were conducted with reasonable follow-up support, as competence decay (see Gawad et al., 2019) has to be addressed with both scenarios. For better understanding about this matter, in the similar future studies additional measurements need to be conducted after short-time training courses to get similar longitudinal data to the long-term courses.

The results of our study seem to suggest that a short-term training course would be useful for guiding teachers with beginner-level digital competences to the expert level, whereas a

long-term training course could result in higher share of teachers with top-level digital competences. While this suggestion seems to be confirmed by the data in **Figures 1** and **2**, there can also be other factors in play, for example, the different initial levels of the participants' digital competences of the compared cases. There are studies (e.g., Henriques et al., 2021) that support the idea that a long-term training course could lead to more beneficial outcomes, as the participants could better understand the impact of what they have learned on their classroom practices, and that changing teachers' teaching practices is a slow process. However, to fully understand the dynamics of both long-term and short-term training courses, it is important to provide clear definitions of what counts as short-term or what counts as long-term.

In this paper we examined if shortening the duration of STEAM training courses would have meaningful negative impact on various teachers' digital competences, measured by a survey that was based on the European DigCompEdu framework. We found out that while both of the observed STEAM courses had overall positive impact on the growth of teachers' digital competences (RQ1), the short-term course seemed to have better results in providing teachers with intermediate digital competences, and the long-term course seemed to provide more teachers with high-level digital competences.

Our study has some limitations that could have had an influence on its results. For example, the initial level of the participants' digital competences is not the same with the two cases compared. For better comparability, populations with similar characteristics are recommended. We only measured the participants' self-evaluations before and after the training courses. While self-evaluation of digital competence tends to correlate with evaluator's actual competence, some studies indicate that evaluators tend to overestimate their abilities (Tomczyk, 2021). The objectivity of results could be improved by using alternative methods that would measure participants' actual digital competences (instead of using their self-assessment). In addition, a third measurement, after some months or a year after the training, should be conducted in order to get an understanding about the long-term influence of the training-course.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The name of the repository and accession number can be found below: Google Drive, https://drive.google.com/drive/folders/15RBbRoaxFMuY7v0oEe27EyxusiBLfUY-?usp=sharing.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

JL, KF, and ZL: conceptualization and supervision. JL and KF: methodology. ZL and MT: validation. JL, MT, and TÕ: formal analysis and writing – original draft preparation. JL: investigation and project administration. JL and MT: resources, data curation, and visualization. JL, MT, TÕ, KF, and ZL: writing. TÕ: funding acquisition. All authors have read and agreed to the published version of the manuscript.

FUNDING

Open Access was supported by Tallinn University School of Educational Sciences Research Fund.

REFERENCES

Ananiadou, K., and Rizza, C. (2010). "ICT in initial teacher training: first findings and conclusions of an OECD study," in *Proceedings of the EDULEARN10 Conference. International Association of Technology, Education and Development*, Barcelona, 5621–5632.

Atiles, J. T., Almodóvar, M., Chavarría Vargas, A., Dias, M. J. A., and Zúñiga León, I. M. (2021). International responses to COVID-19: challenges faced by early childhood professionals. *Eur. Early Childh. Educ. Res. J.* 29, 66–78. doi: 10.1080/1350293X.2021.1872674

Bakar, R. (2018). The influence of professional teachers on Padang vocational school students' achievement. *Kasetsart J. Soc. Sci.* 39, 67–72.

Botha, A., and Herselman, M. (2018). Teachers become co-creators through participation in a teacher professional development (TPD) course in a resource constrained environment in South Africa. *Electron. J. Inf. Syst. Dev. Ctries.* 84:e12007. doi: 10.1002/isd2.12007

Bozkurt, A., and Sharma, R. C. (2020). Emergency remote teaching in a time of global crisis due to coronavirus pandemic. *Asian J. Dist. Educ.* 15, i–vi.

Bradonjic, P., Franke, N., and Lüthje, C. (2019). Decision-makers' underestimation of user innovation. *Res. Policy* 48, 1354–1361. doi: 10.1016/j.respol.2019.01.020

Burgess, A., and McGregor, D. (2018). Peer teacher training for health professional students: a systematic review of formal programs. *BMC Med. Educ.* 18:263. doi: 10.1186/s12909-018-1356-2

Casillas, S., Cabezas, M., and García, F. J. (2020). Digital competence of early childhood education teachers: attitude, knowledge and use of ICT. *Eur. J. Teach. Educ.* 43, 210–223. doi: 10.1080/02619768.2019.1681393

Drotner, K. (2020). Children's digital content creation: towards a processual understanding of media production among Danish children. *J. Child. Media* 14, 221–236. doi: 10.1080/17482798.2019.1701056

European Commission (2017a). *ICT for Work : Digital Skills in the Workplace : Final Report*. Publications Office. Available online at: https://data.europa.eu/doi/10.2759/498467 (accessed March 9, 2020).

European Commission (2017b). *European Framework for the Digital Competence of Educators (DigCompEdu)*. Available online at: https://ec.europa.eu/jrc/en/publication/eur-scienti?c-and-technical-research-reports/european-framework-digital-competence-educators-digcompeduColbert (accessed March 9, 2020).

European Commission (2020). *Digital Education Action Plan 2021–2027*. Available online at: https://ec.europa.eu/education/education-in-the-eu/digital-education-action-plan_en (accessed March 9, 2020).

European Commission (2021a). *Digital Economy and Society Index (DESI) 2021*. Publications Office. Available online at: https://digital-strategy.ec.europa.eu/en/policies/desi (accessed March 9, 2020).

European Commission (2021b). *Decision of the European Parliament and of the Council Establishing the 2030 Policy Programme "Path to the Digital Decade". COM/2021/574 Final*. Available online at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0574 (accessed March 9, 2020).

European Commission (2022). *SELFIEforTEACHERS – Discover your Digital Potential*. Available online at: https://educators-go-digital.jrc.ec.europa.eu/ (accessed March 9, 2020).

Fauth, B., Decristan, J., Decker, A., Büttner, G., Hardy, I., Klieme, E., et al. (2019). The effects of teacher competence on student outcomes in elementary science education: the mediating role of teaching quality. *Teach. Teach. Educ.* 86:102882. doi: 10.1016/j.tate.2019.102882

Franzoni Velázquez, A. L., Cardenas Peralta, M. C., and Mandujano Canto, J. A. (2020). Lessons from the training and support of teachers in the development of digital skills: a case study of @prende 2.0. *Dig. Educ. Rev.* 37, 154–171. doi: 10.1344/der.2020.37.154-171

Galindo-Domínguez, H., and Bezanilla, M. J. (2021). Digital competence in the training of pre-service teachers: perceptions of students in the degrees of early childhood education and primary education. *J. Dig. Learn. Teach. Educ.* 37, 262–278. doi: 10.1080/21532974.2021.1934757

Gawad, N., Allen, M., and Fowler, A. (2019). Decay of competence with extended research absences during residency training: a scoping review. *Cureus* 11:e5971. doi: 10.7759/cureus.5971

HARNO (2022). *Selfie for Teachers*. Available online at: https://digipadevus.ee/opetaja-digipadevusmudel/selfie-for-teachers/ (accessed 11 March, 2022).

Heidmets, M., and Eisenschmidt, E. (2020). "Kooliuuendus: välised mõjurid ja ühiskondlik taust," in *Haridusmõte 2020*, ed. M. Heidmets (Tallinn: Tallinna Ülikooli Kirjastus).

Heinmäe, E., Leoste, J., Kori, K., and Mettis, K. (2022). "Enhancing teacher-students' digital competence with educational robots," in *Advances in Intelligent Systems and Computing: RiE 2021*, eds M. Merdan, W. Lepuschitz, G. Koppensteiner, R. Balogh, and D. Obdržálek (Cham: Springer), doi: 10.1007/978-3-030-82544-7_15

Henriques, S., Correia, J. D., and Dias-Trindade, S. (2021). Portuguese primary and secondary education in times of COVID-19 pandemic: an exploratory study on teacher training and challenges. *Educ. Sci.* 11:542. doi: 10.3390/educsci11090542

Herranen, J. K., Aksela, M. K., Kaul, M., and Lehto, S. (2021). Teachers' expectations and perceptions of the relevance of professional development MOOCs. *Educ. Sci.* 11:240. doi: 10.3390/educsci11050240

Leoste, J. (2021). *Adopting and Sustaining Technological Innovations in Teachers' Classroom Practices – The Case of Integrating Educational Robots into Math Classes. Ph.D. thesis*. Tallinn: Tallinn University.

Leoste, J., Heidmets, M., Reis, L., Sanfeliu, A., and Tardioli, D. (2019). "Factors influencing the sustainability of robot supported math learning in basic school," in *Advances in Intelligent Systems and Computing: ROBOT'2019*, eds M. Silva and J. Luís Lima (Cham: Springer), doi: 10.1007/978-3-030-35990-4_36

Leoste, J., Tammets, K., and Ley, T. (2020). Co-Creating learning designs in professional teacher education: knowledge appropriation in the teacher's innovation laboratory. *IxD&A Interact. Design Archit.* 42, 131–163. doi: 10.55612/s-5002-042-007

López Belmente, J., Pozo Sanchez, S., Fuentes Cabrera, A., Lopez Nuñez, J. A. (2019). Content creation and flipped learning: a necessary pairing for education in the new millennium. *Rev. Española Pedag.* 77, 535–555. doi: 10.22550/REP77-3-2019-07

Mäkelä, T., Fenyvesi, K., and Mäki-Kuutti, M. (2020). Developing a pedagogical framework and design principles for STEM learning environment design. *J. Res. STEM Educ.* 6, 1–17. doi: 10.51355/jstem.2020.74

Niederhauser, D. S., Howard, S. K., Voogt, J., Agyei, D. D., Laferriere, T., Tondeur, J., et al. (2018). Sustainability and scalability in educational technology. *Technol. Knowl. Learn.* 23, 507–523. doi: 10.1007/s10758-018-9382-z

Papadakis, S., Vaiopoulou, J., Sifaki, E., Stamovlasis, D., and Kalogiannakis, M. (2021a). Attitudes towards the use of educational robotics: exploring pre-service and in-service early childhood teacher profiles. *Educ. Sci.* 11:204. doi: 10.3390/educsci11050204

Papadakis, S., Vaiopoulou, J., Sifaki, E., Stamovlasis, D., Kalogiannakis, M., and Vassilakis, K. (2021b). "Factors that hinder in-service teachers from incorporating educational robotics into their daily or future teaching practice," in *Proceedings of the 13th International Conference on Computer Supported Education (CSEDU 2021)* (online), Vol. 2, 55–63. doi: 10.5220/0010413900550063

Pérez-Jorge, D., Rodríguez-Jiménez, M., Gutiérrez-Barroso, J., and Castro-León, F. (2020). Training in digital skills in early childhood education teachers: the case of the University of La Laguna. *Interact. J. Interact. Mobile Technol.* 14, 35–49.

Perifanou, M., Economides, A. A., and Tzafilkou, K. (2021). Teachers' Digital Skills Readiness During COVID-19 Pandemic. *Int. J. Emerg. Technol. Learn.* 16, 238–251. doi: 10.3991/ijet.v16i08.21011

Pimentel, J. L. (2019). Some biases in likert scaling usage and its correction. *Int. J. Sci. Basic Appl. Res.* 45, 183–191.

Portillo, J., Garay, U., Tejada, E., and Bilbao, N. (2020). Self-perception of the digital competence of educators during the COVID-19 pandemic: a cross-analysis of different educational stages. *Sustainability* 12:10128. doi: 10.3390/su122310128

Prieto-Alvarez, C. G., Martinez-Maldonado, R., and Dirndorfer Anderson, T. (2018). "Codesigning learning analytics tools with learners," in *Analytics in the Classroom. Translating Learning Analytics Research for Teachers*, eds J. M. Lodge, J. Cooney Horvath, and L. Corrin (Abingdon: Routledge).

Sabalete Suárez, A., and Roblizo Colmenero, M. (2021). The challenge of incorporating digital skills in the classroom: perceptions and attitudes of Spanish Salesian teachers. *Int. Stud. Catholic Educ.*

Sánchez-Cruzado, C., Santiago Campión, R., and Sánchez-Compaña, M. T. (2021). Teacher digital literacy: the indisputable challenge after COVID-19. *Sustainability* 13:1858. doi: 10.3390/su13041858

Tomczyk, Ł. (2021). Declared and real level of digital skills of future teaching staff. *Educ. Sci.* 11:619. doi: 10.3390/educsci11100619

Tondeur, J., Roblin, N. P., van Braak, J., Voogt, J., and Prestridge, S. (2017). Preparing beginning teachers for technology integration in education: ready for take-off? *Technol. Pedag. Educ.* 26, 157–177.

UNDP & Mohammed Bin Rashid Al Maktoum Knowledge Foundation (2021). *Global Knowledge Index*. Available online at: http://www.knowledge4all.org/ (accessed 10 March, 2022).

UNESCO (2016). *School and Teaching Practices for Twenty-First Century Challenges. Lessons from the Asia-Pacific Region – Regional Synthesis Report*. Bangkok: UNESCO.

UNESCO (2018). *Digital Skills Critical for Jobs and Social Inclusion*. Available online at: https://en.unesco.org/news/digital-skills-critical-jobs-and-social-inclusion (Accessed March 9, 2020).

Usart Rodríguez, M., Lázaro Cantabrana, J. L., and Gisbert Cervera, M. (2021). Validation of a tool for self-evaluating teacher digital competence. *Educac. XX1* 24, 353–373. doi: 10.5944/educXX1.27080

Von Hippel, E. (2017). *Free innovation*. Cambridge, MA: MIT Press.

Williams, M., and Moser, T. (2019). The art of coding and thematic exploration in qualitative Research. *Int. Manag. Rev.* 15, 45–55.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Leoste, Lavicza, Fenyvesi, Tuul and Õun. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Psychology

ORIGINAL RESEARCH
published: 24 May 2022
doi: 10.3389/fpsyg.2022.888215

Effects of a Pair Programming Educational Robot-Based Approach on Students' Interdisciplinary Learning of Computational Thinking and Language Learning

Ting-Chia Hsu[1†], Ching Chang[1†], Long-Kai Wu[2] and Chee-Kit Looi[3†]*

[1]Department of Technology Application and Human Resource Development, National Taiwan Normal University, Taipei City, Taiwan, [2]Faculty of Artificial Intelligence in Education, Central China Normal University, Wuhan, China, [3]National Institute of Education, Nanyang Technological University, Singapore, Singapore

OPEN ACCESS

Edited by:
Stamatios Papadakis, University of Crete, Greece

Reviewed by:
Rustam Shadiev, Nanjing Normal University, China
Golnar Mazdayasna, Yazd University, Iran
Hossein Keshmirshekan, Kharazmi University, Iran

***Correspondence:**
Ching Chang
chingtw2006@gmail.com

†ORCID:
Ting-Chia Hsu
orcid.org/0000-0001-6504-9540
Ching Chang
orcid.org/0000-0001-8032-247X
Chee Kit Looi
orcid.org/0000-0001-9905-2713

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

Received: *02 March 2022*
Accepted: *08 April 2022*
Published: *24 May 2022*

Citation:
Hsu T-C, Chang C, Wu L-K and Looi C-K (2022) Effects of a Pair Programming Educational Robot-Based Approach on Students' Interdisciplinary Learning of Computational Thinking and Language Learning. Front. Psychol. 13:888215. doi: 10.3389/fpsyg.2022.888215

Using educational robots (ERs) to integrate computational thinking (CT) with cross-disciplinary content has gone beyond Science, Technology, Engineering, and Mathematics (STEM), to include foreign-language learning (FL) and further cross-context target-language (TL) acquisition. Such integration must not solely emphasise CT problem-solving skills. Rather, it must provide students with interactive learning to support their target-language (TL) interaction while reducing potential TL anxiety. This study aimed to validate the effects of the proposed method of pair programming (PP) along with question-and-response interaction in a board-game activity on young learners' CT skills and TL learning across contexts. Two Grade 6 classes, one with 16 students who were studying Chinese as a Second Language (CSL) and the other with 16 students who were studying English as a Foreign Language (EFL), participated in the activity. A series of instruments on achievement assessment, questionnaires on CT skills and TL anxiety, and sequential learning behaviour analysis were used to critically examine the results. The main conclusion is that the EFL group showed better social skills of cooperation on CT and lower TL learning anxiety, while the CSL group demonstrated better problem-solving skills in CT, but presented more behaviours of trial-and-error loops. Results not only contribute suggestions for cross-disciplinary learning but also provide support for cross-context instruction beyond educational coursework.

Keywords: interdisciplinary activities, educational robots, pair programming, language learning, trial-and-error loops

INTRODUCTION

Educational robots (ERs) have gained popularity in classrooms, as they are considered as effective tools for fostering students' CT skills. The typical goals of ERs range from a low threshold of generating students' interest and learning with abstract concepts to a central CT development of problem decomposition, algorithm design, iteration, and debugging (Shute

et al., 2017). Many educators have seen their potential and have designed ER activities beyond Science, Technology, Engineering, and Mathematics (STEM) for interdisciplinary activities, such as music (Chung, 2014), arts (Burhans and Dantu, 2017), and foreign language learning (FL; e.g., Hsu and Liang, 2021). According to the review of Papadakis (2021), CT skills can be studied as a problem-solving mechanism and a way that allows users to identify problems or organise the situation by expressing their thoughts, and thus can support cognitive development. Although cross-discipline development was not highlighted in the analysis of Papadakis (2021), being able to express oneself, development of computational fluency, and language development are all essential skills included in CT development (Papadakis, 2021). The current study extrapolated the view of Papadakis (2021) beyond educational coursework on communication development in computational fluency. It aimed to tailor ER activities which integrated CT and target-language (TL) learning, as ERs can be programmed to not only be a medium for CT development, but also to offer unique ways of engaging students in problem-solving tasks while cultivating peer-to-peer communication and interactions in TL learning.

Simply putting these components (ERs, CT, and FL) together, however, does not guarantee the development of the anticipated competences. Indeed, interdisciplinary activities cannot be implemented without carefully designing meaningful ways to develop CT and TL, along with using an appropriate approach. Interdisciplinary activities designed for CT and TL development do not only consider the problem-solving skills involved in coding; they also need to consider allowing students to express themselves and state their thoughts in programming, and to support their TL interaction while lowering their potential TL learning anxiety.

One way to build problem-solving innovators is through establishing collaborative learning settings, as it has been evidenced as an effective approach to help students obtain CT skills while eliciting discussion in programming (Lewis, 2011; Wei et al., 2021). Pair programming (PP) is recommended by researchers as an intervention method that offers collaborative explicit guidelines to instructors on how to integrate ERs and classroom practice (Zhong et al., 2017; Wei et al., 2021). The rationale of PP stems from collaborative learning approaches that require two people to work together and switch roles during the coding process, where one (the driver) operates the device and writes the code while the other (the navigator) offers information by watching for possible defects or directing design decisions. Those who advocate the use of PP argue that it leads to better learning results (e.g., CT), intensive involvement, and communication in coding, and increases students' satisfaction (e.g., Zhong et al., 2017; Wei et al., 2021). However, results regarding PP are inconsistent, and studies using PP in interdisciplinary activities remain scarce.

Thus, tangible manipulation of ERs was integrated into the PP approach (ER-integrated-PP) to allow children to identify a situation, define a problem, and come up with a solution (Chevalier et al., 2020). Such collaborative activities afford students the opportunity to express their ideas and develop their abilities of problem decomposition, abstraction, algorithm design, debugging, iteration, and generalisation, corresponding to Brennan and Resnick's framework of the description of computational practice (Brennan and Resnick, 2012; Shute et al., 2017). While engaging in collaboration, the ER-integrated-PP approach appears to be feasible for engaging students in TL interaction and communication in the process of problem-solving learning using target languages.

The premise of acquiring target-language acquisition is a way of activity construction that affords students opportunities for target-language practice (Sato and Storch, 2020). This activity construction also needs to support an open environment that offers students multiple ways to solve a problem in their CT development. A board-game activity adopting a gamified mechanism is a type of communicative task which actively engages students in group talk. The question-and-response interaction tailored in the board-game activity asked students to practise the assigned target sentences, which not only provided opportunities for TL production but also facilitated students' intensive interaction, echoing interactionist perspectives on second language acquisition (SLA). The board-game activity also supports multiple combinations of CT problem-solving learning (Chen and Chi, 2020). The explicit facilitation of CT learning content embedded in the gamified mechanism can be joined with the promotion of language learning *via* a creative ER-integrated PP approach. Concisely, participants not only carry out the coding development using PP within groups, but also work on the target language and continually interact with each other between groups using the assigned sentence practice in the board-game activity.

When approaching language learners, it is essential to perceive the difference between learning a language in a second language environment (L2) and in a foreign language environment (FL). Past studies have shown that L2 settings are more effective for learning target languages than FL settings (Taguchi, 2008). L2 learners gain more context-rich target language access in natural communication situations compared to EFL learners who cannot access the target language in the immediate environment, although mass media may offer chances for target language practice (Longcope, 2009). While understanding that "contexts matter" significantly affects students' target-language learning (Sato and Storch, 2020), whether integrating interdisciplinary activities across different contexts shares similar results with a specific subject remains unknown. In addition, a pedagogically-informed instructional design of the learning approach (e.g., PP) and gamified activities (e.g., board-game) can be shared and adapted across contexts. It is pivotal to examine students' learning performance and to compare their learning behaviours in two distinct contexts; it would then be possible to identify potential challenges and offer insights into curriculum implantation and support for instructors.

While those complex elements are involved in interdisciplinary design to meet anticipated results, reducing potential TL learning anxiety for those who are L2 or FL learners must be taken into account. It was essential to offer a positive learning atmosphere to increase learners' participation in this cross-discipline study, as these two particular groups across contexts might demonstrate

unique learning behaviours, and so diverse and advanced methods were needed to cope with the given task when they first accessed CT concepts while engaging in TL acquisition. Concisely, exploring how these two particular groups of language learners coped with the interdisciplinary learning and potential language-learning anxiety in the proposed approach needed to be highlighted. In this paper, we investigate the impacts of the ER-integrated-PP approach on these two groups of CSL and EFL learners, and examined their learning anxieties and learning behaviours in the proposed approach and activity. Our study results can thus contribute to learning performance in cross-disciplinary development, and provide potential teaching implications and suggestions to support learners in their specific contexts and instructors in their curriculum design.

Using CT skills to deal with coding tasks needs to go through certain phases, while interacting with TL appears challenging for students. It may not be easy to unveil the results without examining students' behaviours regarding how they conduct ER activities to transform their ideas into problem-solving and solutions while engaging in TL practices. Chevalier et al.'s creative and computational problem solving (CCPS) model (2020) serves as a useful model for evaluating students' behaviours associated with CT development. It is a comprehensive model that illustrates these key phases: understanding the problem, generating ideas, formulating the robot's behaviour (the first loop), and programming and evaluating the solution (the second loop). They suggested that practitioners should be aware of ERs' immediate feedback and of students' rapid validation of strategies. As ER activities easily drive students into trial-and-error loops, students' learning without systematic examination of their strategies and justifying their reasons of decomposing problems has often been reported (Shute et al., 2017). Chevalier et al.'s CCPS model (2020) was adapted to help evaluate students' behaviours and examine the implementation of ER activities associated with CT and target-language development.

However, the above claims require further investigation because the effect of PP on ER-integrated interdisciplinary activities of CT and target-language learning, along with the board-game activity, for young learners is still not well documented. The aim of this study was therefore to design and assess an interdisciplinary activity catering for sixth grade children, and to investigate their learning behaviours when an instructional design took PP and board-game activities into account. An ER-integrated-PP approach with a pedagogy-informed-gamified design was deployed to enhance students' physical learning experience when compared to learning effects and differences of two groups. The ER-integrated-PP approach with gamified activities may act as a catalyst in promoting embodied learning experience for improving interdisciplinary integration, thereby fostering CT skills and language learning while lowering their anxiety about using the target language from the assigned sentences in this particular setting.

Therefore, the research questions are as follows:

1. Were there any differences in the interdisciplinary learning of the two groups (i.e., CSL and EFL) in the ER-integrated-PP context?
2. Were there any differences in the CT skills of the two groups in the ER-integrated-PP context?
3. Were there any differences in the FL anxiety of the two groups in the ER-integrated-PP context?
4. What were the differences in the learning behaviours of the two groups in the ER-integrated-PP context?

LITERATURE REVIEW

ER-Integrated-PP

To deal well with the challenges of the 21st century, attention has been paid to CT skill development. Programming has been a major tool to access CT skills (Hsu et al., 2018) because it allows students to access fundamental skills of abstraction, algorithmic thinking, critical thinking, debugging, and iteration (Shute et al., 2017).

Among the various strategies, PP presents a promising strategy to teach programming. The rationale of PP is that two people, a driver and a navigator, work side-by-side with one computer to generate coding collaboratively, while regularly switching roles. The driver mainly operates the computer to generate code, whereas the navigator contributes ideas and directions for solving problems (Williams and Kessler, 2002). Zhong et al. (2017) provided a clear summary of the benefits of PP, including cognitive development of CT and programming skills, increasing retention and learning satisfaction, and better communication, cooperation, and teamwork. Wei et al. (2021) further evidenced the effect of CT skills in PP while promoting the development of soft-skills (self-efficacy) for young learners. The studies of Zhong et al. (2017) and Wei et al. (2021) showed the similarities of the effectiveness of PP for learning CT skills *via* programming; however, ER activities integrated into PP to enhance CT skills and other disciplines have not been fully explored. As we were interested in exploring TL learning when students were situated in conversation practice using the target language, the experiment aimed to monitor the switching between roles of the driver and navigator within the pair of young students (within-group), and to ensure that they carried out their assigned conversation practice when they worked between groups.

Target-Language Production Enforcement

Although there has been no direct report on improving TL using ERs with PP, PP has been observed to improve soft skills such as communication skills (D'Angelo and Begel, 2017) and collaboration skills (Lewis, 2011) and to reduce learning frustration and anxiety (Zhong et al., 2017). However, TL acquisition in this cross-discipline study may not be easily attained without reinforcing intensive interaction during TL production, although production practice activities have been regular classwork along with well-designed comprehensive input in modern language classrooms (Sato and Storch, 2020). One gamified activity, the board-game activity, appears promising. Apart from sharing gamified characteristics such as immersion, flow, high motivation, and engagement (e.g., Chen and Chi, 2020; Cheng et al., 2020; Kuo and Hsu, 2020), it offers unique

advantages that facilitate learners' high interactivity opportunities and shared learning between participants (Chen and Chi, 2020). Board-game activities allow learners to employ open-ended, low-floored, shared, cooperative, strategic, and creative thinking to access CT (Chen and Chi, 2020). As participants in gamified activities must clearly comprehend all rules and information, and take turns playing to compete with their counterparts (Kuo and Hsu, 2020), such activities have gained popularity owing to the ease of classroom adaptation and implementation.

It is this specific turn-based nature involved in social interaction that allowed the attention on tailored TL production in the current study. It was not reasonable to expect students to communicate with each other using the target language, since both groups across contexts are FL and L2 learners. Rather, the question-and-response interaction underlying the turn-based nature of the board-game activity, along with the assigned sentences and some relevant vocabulary, was tailored to fortify students' TL production while they were working on their CT development. A series of turns in a session of the board-game activity could not only potentially accumulate TL practice; students could also develop their self-reflections by revising their previous errors in their upcoming turns (Chen and Chi, 2020). This meets the desired goals of the current study for CT development and TL production.

Although past studies presented empirical evidence of the effectiveness of ERs with board-game learning in terms of supporting learning (Hsu and Liang, 2021), such as self-directed and problem-solving abilities (Cheng et al., 2020), few have attempted to integrate board-games into cross-disciplinary and cross-contextual learning using ERs. The current study aimed to fill this gap by applying a pedagogically informed approach and activity that inculcated ER tools in an interdisciplinary learning scenario.

In addition, it must be clarified that there were two dimensions of strengthened collaborative learning involved in the current study: within-group and between-group design. The ER-integrated-PP approach stresses within-group interaction that involves students in meaning negotiation when they interact with partners to interpret or generate new understandings in the given task. The study of Cheng et al. (2020) showed that interactive skills using ERs improved FL learning (Cheng et al., 2020), while the research of Brennan and Resnick (2012) evidenced that ERs can facilitate students' abilities of problem solving, critical thinking, and cooperation to keep up with twenty-first century needs. Such strengthened collaborative learning on ERs with PP appear beneficial for fostering peer interaction and engagement in the within-group setting.

On the other hand, the question-and-response routine involved in board-game activities emphasises between-group interactions by strengthening target-language output production practice. While being involved in intense collaboration to complete a coding task *via* PP within groups, high interactivity between groups in target-language practices *via* the board-game activity was tailored to comply with interactionist perspectives in SLA, where explicit facilitation of target forms (e.g., sentences and vocabulary) was reinforced to facilitate oral development *via* communicative tasks (Blyth, 2018). Such reinforced interaction in both modes can also jointly create a positive learning environment, thus potentially reducing students' learning anxiety in this cross-discipline study. Briefly, ERs implemented with PP along with gamified activities in the within-group and between-group design is in line with the pedagogical support aims for achieving the desired goal.

Language Learning Anxiety

In language learning contexts, anxiety refers to the "worry and negative emotional reaction aroused when learning or using a second language" (MacIntyre, 1999, p. 27). Cumulative findings have shown that anxiety inhibits learners from participation in oral activities (e.g., Cakici, 2016), and negatively predicts outcomes in the second language (L2; e.g., Saranraj and Meenakshi, 2016). However, analysing the cause of anxiety in class with proper instructional design helps teachers understand students' problems, and they can then try to enhance students' learning performance (Hu and Wang, 2014). Assisting students in dealing with the conditional anxieties while making the learning environment less stressful are two key strategies to reduce students' anxiety (Horwitz et al., 1986).

Saranraj and Meenakshi (2016) stated that learning can be effective if anxiety can be appropriately handled or coped with in L2 or other language learning environments, because many learners experience a certain degree of anxiety during their learning, and use specific strategies to cope with it. Thus, instructors play a key role in reducing students' anxiety by properly designing activities as well as establishing a welcoming learning atmosphere for classroom activities. In the current study, ERs were a medium for the course activity that helped achieve effectiveness of the learning and the technology use, whereas PP strategies plus board-game activities could be a desirable design to offer a warm classroom setting while reducing students' learning anxiety.

Learning Behaviours

Analysing students' behaviours in ER activities is a springboard to understanding students' strategy use, and how they develop their cognitive processes associated with CT competences (Tsai et al., 2012). Chevalier et al. (2020) established a CCPS model that allows teachers to validate their instructional interventions, and to effectively facilitate students' CT development. For example, while instructional methods aim to cultivate students to have productive learning associated with reflection and planning of their strategies, they promote repetitive behaviours in the mechanical operation process, known as the trial-and-error approach (Wing, 2006) to improve skill development. Instructors can thus plan some proper instructional interventions that will assist students in building a well-settled strategy in the classroom practice of ERs (Chevalier et al., 2020).

This study adapted the CCPS model. To fit with the current context, two loops were taken from CCPS while one was expanded. The first loop (within-group) involves students in discussing and negotiating their proposed ideas, and reflection on their problem-solving strategies, including understanding the problem, generating ideas, and formulating the robot's behaviour (interpreted as negotiation loops). The second loop (within-group) is about programming and evaluating the solution (Chevalier et al., 2020),

which limits students' productive learning (identified as evaluation loops). The extended one is about the target-language interaction loop (between-group), where students are engaged in the assigned conversation practice with team-based learning. These three loops modified by CCPS were used to analyse the students' learning behaviours and find out their ways of dealing with the problem-solving task in the interdisciplinary learning.

RESEARCH METHOD

Participants

A total of 32 Grade 6 students participated in this study, 16 of whom were learning Chinese as a second language (CSL) in Singapore, while 16 were learning English as a Foreign Language (EFL) in Taiwan. None of the students had any previous experience of accessing interdisciplinary activities. Both groups participated in a language classroom with several weeks tailored for interdisciplinary activities. They were all volunteers to participate in the task. Both groups' language proficiency was considered to be at an elementary level. The research team cooperated with both the CSL teachers in Singapore and the EFL teachers in Taiwan to conduct the study in each of their specific contexts.

Instructional Design

The interdisciplinary activity was tailored to guide students to develop CT skills and target-language development using the interactive ER activity in the language learning classroom (**Figure 1**). The board-game activity was to collect required materials in the appointed place. While collecting materials, students were involved in developing CT skills by accessing logical sequences, executive conditions, and debugging. After ensuring all the students understood the rules, information, and the ways of controlling and programming the ERs, the CT task was conducted. Meanwhile, relevant vocabulary and sentence patterns for oral interaction were also organised for TL learning, where both groups had exactly the same content (conversation practice) but with different TL and programming interfaces (Chinese vs. English). Students worked in pairs to complete the coding task, and completed the set question-and-response interaction jointly.

Briefly, three constructs were established, namely ER tools with block-based programming (**Figures 2A,B**), CT learning (**Figures 2D–F**), and target-language materials (**Table 1**). The first are the ER tools. The study used physical motor-based ERs, where motors, sensors, and memory, and a map with infrared-reflection and identifiers were all included to control the ERs on the map (**Figure 2C**). This allowed students to simultaneously test their coding.

Second is CT learning. As the scenario of the board-game interactive activity asked students to construct city buildings by collecting needed materials (e.g., stones), students needed to identify their targets, decompose how the ERs could reach their intended destinations (e.g., wood), and come up with solutions (using algorithms like sequential logic or loop) or debugging (**Figure 2F**). They worked in pairs to control the robots by operating the block-based app programming to research the targeted place or to obtain the needed resources. Two teams worked at a table and competed with each other. CT learning occurred when the pairs of students controlled the ERs by moving them in the anticipated way.

The third construct was to arrange relevant vocabulary and simple sentence practice while students were working on their CT learning (see an example in **Table 1**). Students needed to apply these words and sentences in the CT activity. The vocabulary included words and phrases such as "move, step, forward and move, turn right or left, sand, construction, and stone." Two teams were involved in practising the conversations with each other using question-and-response exercises in the turn-based board-game activity.

Both groups accessed the same materials, including basic sequential rules and the algorithm of simplifying the steps, and vocabulary and sentence pattern teaching. They also received similar learning instruction and strategies (**Figure 3**), emphasising

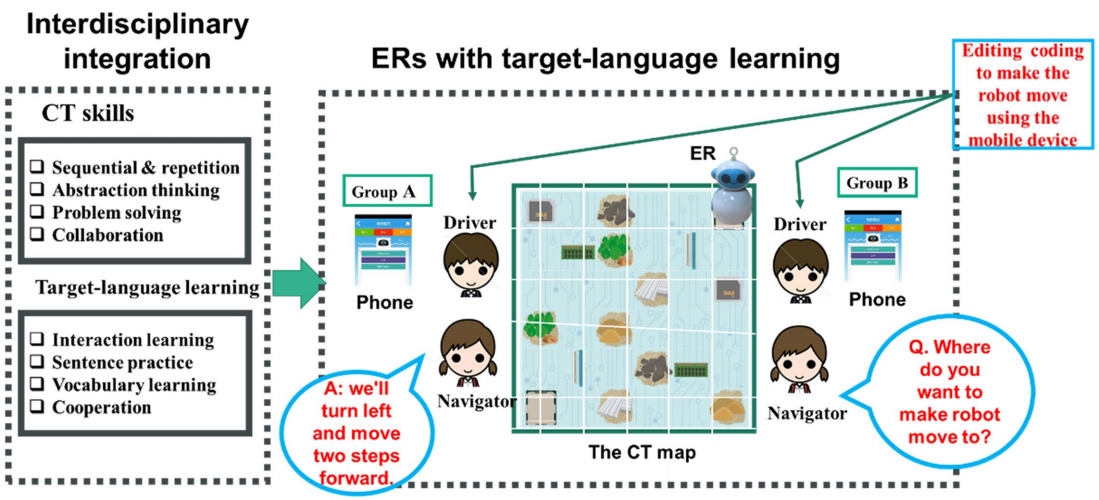


FIGURE 1 | Research framework.

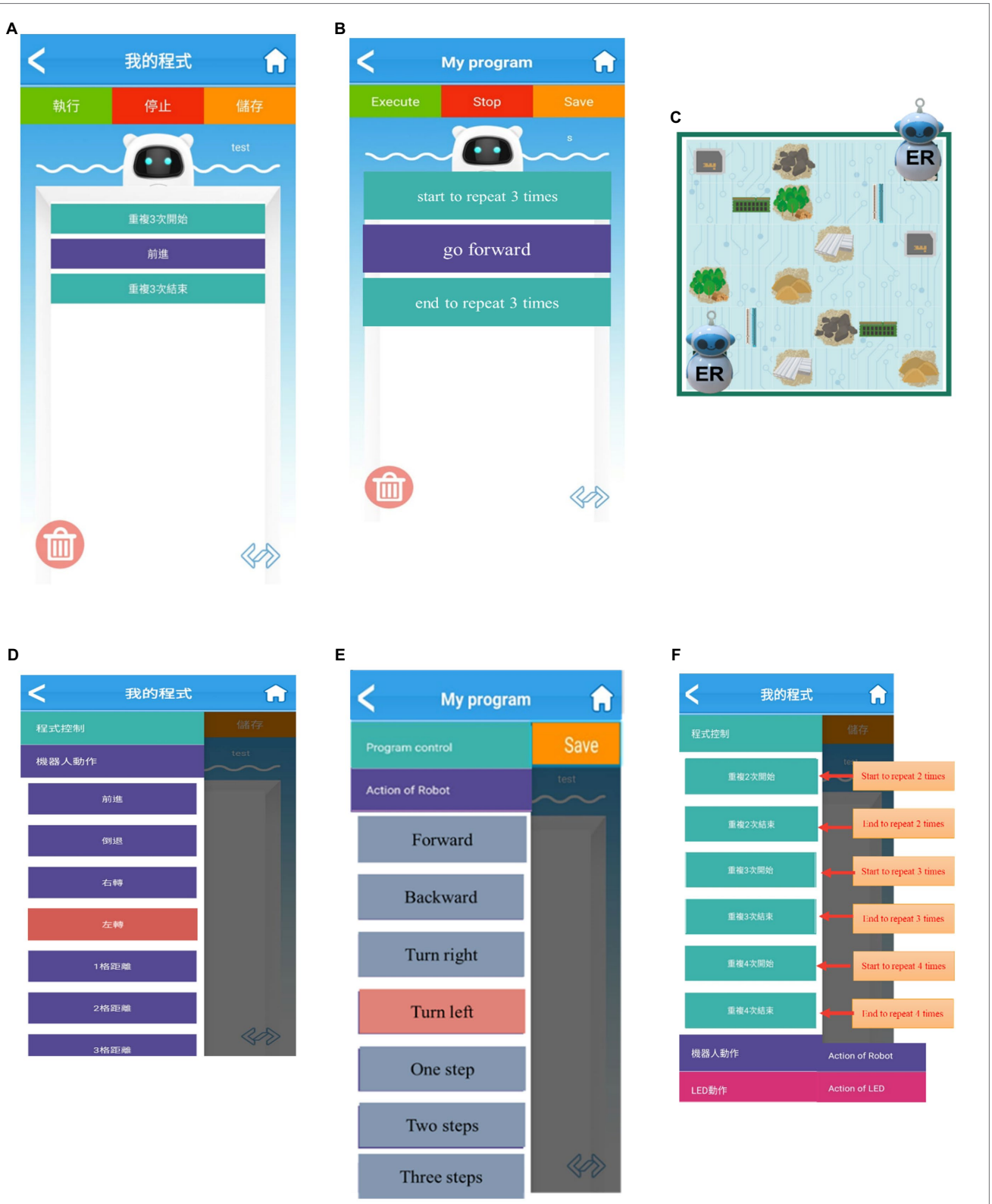


FIGURE 2 | (A) ER tools for EFL. **(B)** ER tools for CSL. **(C)** Interactive activity. **(D)** CT learning for EFL. **(E)** CT learning for CSL. **(F)** CT learning.

the roles of navigator and explorer, and TL conversation practice. After the explorer of one team operated the app to control the ERs, the navigator needed to ask the navigator of the opposite team to answer the questions in the target language (i.e., English or Mandarin). The navigator of the opposite team answered the question after completing coding. The two teams switched with their own partners to engage in TL practice. Shortly, the ER-integrated PP approach emphasised the two roles of explorer and navigator within teams (within-group interaction), while question-and-response interaction in the board-game activity involved language practice between groups (between-group interaction) in the interactive ER activity.

Research Process

A quasi-experimental method was adopted in this study, including a 3-week experiment period with one session of language learning for two weekly lectures of 1 h, and the other session of CT integration for one weekly activity of 2 h, giving a total of 4 h per week.

In the previous two periods, the students took the pre-test of language learning (either Mandarin or English) and completed the pre-questionnaire of the computational thinking scales (CTS) and the language classroom anxiety scale. They learnt the vocabulary and conversation practice in English for the EG and Mandarin, and received mini-guidance on the basic CT concepts. The last period was arranged to implement a hands-on activity in which students applied the learnt CT concepts and the relevant conversation practice involved in the interactive ER activities. Then, the post-test of English and Mandarin, and the post-questionnaires of CTS and language-learning anxiety were administered in the language classroom. Lastly, the learning process was video-recorded; their behaviours were further analysed for later discussion.

TABLE 1 | Example of target-language learning materials for both groups.

When the explorer of the opposite team asked the question, say:
Where do you want to make the robot move to? 你想要機器人往哪裡走?
The explorer of the team answered the question, saying:
1. *We want to move one step forward.* 我們想前進一步
2. *We want to move two steps forward and turn right/left.* 我們想前進二步，然後右/左轉
3. *We want to turn right/left and move three steps forward.* 我們想先右/左轉,然後前進三步

Instruments

The pre-test and post-test of CT competencies (50%) and language proficiency (50%) comprised: (1) sentence combination (five items worth 10 points) and multiple-choice questions for vocabulary items (10 items worth 40 points), with a full score of 50 for CT competences, and (2) the same arrangement of tests for language proficiency, with a full score of 50 for language proficiency. The test items were consistent with the goal, where

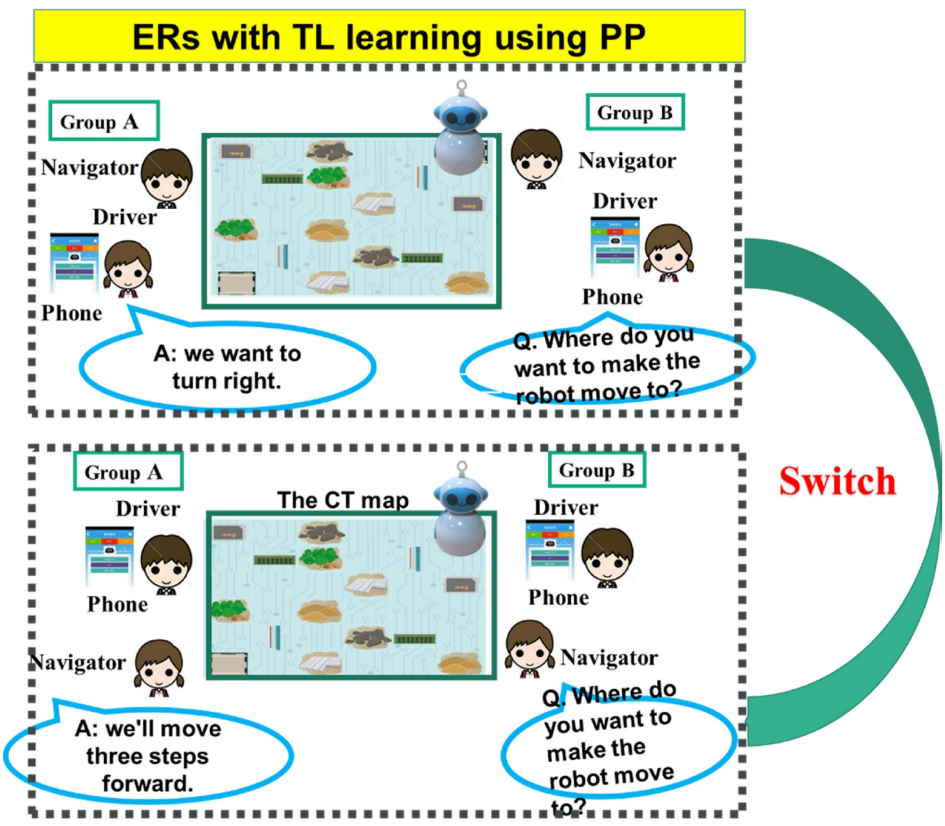


FIGURE 3 | The implementation of the ER-integrated-PP approach along with the question-and-response interaction for both groups.

students used TL in their participation in the question-and-response interaction. Both tests (Mandarin or English) had the same test content, but items were written in the different target languages (see **Appendix**). One experienced English teacher, one Mandarin teacher, and one technology education teacher were invited to validate both tests with the two different target languages. The researchers along with the two experts ensured the validity of the tests.

The questionnaire of CTS, adapted from the computational thinking scales by Korkmaz et al. (2017), with a 5-point Likert scale (1 = *strongly disagree*; 5 = *strongly agree*) was adopted to evaluate the students' knowledge, skills, and attitudes towards creativity, algorithmic thinking, cooperativity, critical thinking, and problem solving. This study adapted four dimensions of six items for the algorithmic-thinking dimension, four items for the operation dimension, five items for the critical-thinking dimension, and six items for the problem-solving dimension, with a Cronbach's alpha value of 0.82, showing acceptable reliability.

The language-learning anxiety scale was employed from Horwitz's FLCAS questionnaire (Horwitz, 1986), with a 5-point Likert scale (1 = *strongly disagree*; 5 = *strongly agree*). FLCAS was also modified for the CSL learners to examine their learning anxiety. This study used 10 items for speech anxiety, four for communication apprehension, three for negative evaluation, two for fear of making mistakes in target-language class, and nine for feeling uniquely unable to deal with the task of L2, with a Cronbach's alpha value of 0.78, showing acceptable reliability.

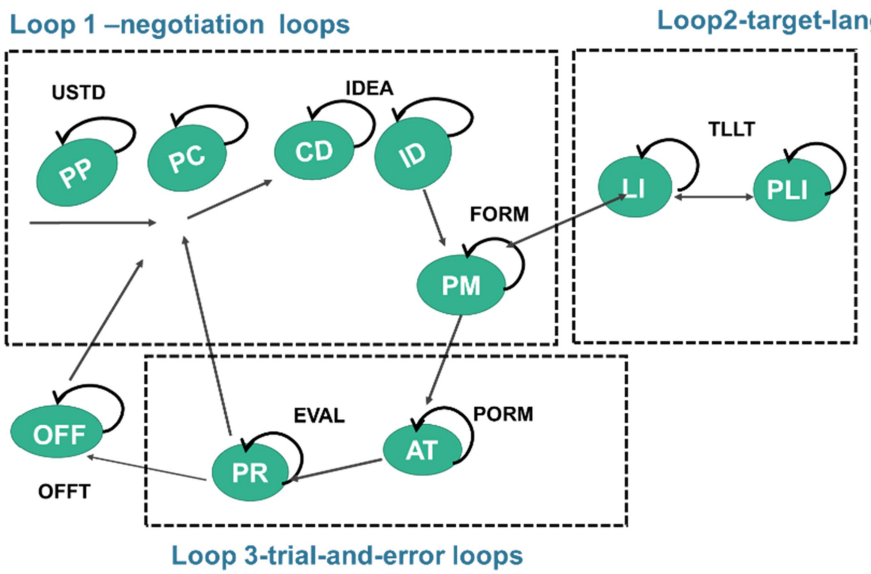
Coding Scheme

The study investigated students' learning behaviours regarding the CT and TL learning in the ER-integrated interdisciplinary task. Students' behaviours involved in tasks between and within teams were captured, and every 10s was a note to locate the main action in line with the video coding technique. To clearly identify students' behaviours during activities, two experts experienced in CT and target-language integration (FL and L2) were invited to confirm the coding schemes associated with the CCPS model to ensure the reliability of the behavioural analysis in this study. These experts together with the researchers confirmed the suitability of the codes and the corresponding CCPS definitions. Based on actual action on CT or FL interaction recorded in the video, the behaviours under the categories pertaining to CCPS were finally identified in the coding scheme listed in **Appendix 1** (**Figure 4**).

RESULTS

Learning Achievement

The purpose of this study was to examine if CSL and EFL had different learning outcomes when students were taking part in the interdisciplinary activities of language and CT integration. A significant difference was observed from the *t*-test results of the pre-test scores of the two groups ($t = -4.991$, $p > 0.05$), meaning that the homogeneous hypothesis of the two groups' achievements before the activity was violated. This implied that directly investigating the progress effects of dependent variables was reasonable. The result showed that no significant difference was found for language-learning progress in the independent sample *t* tests ($t = 0.23$; $p = 0.812 > 0.05$) between CSL ($M = 10.00$) and EFL ($M = 9.13$). However, a significant effect was observed for CT progress ($t = 3.02$; $p = 0.005 < 0.05$) and post-test progress ($t = 0.81$; $p = 0.009 < 0.05$). The CSL group had significantly higher progress performance in CT progress ($M = 19.75$) and post-test progress ($M = 29.75$) in comparison with the EFL group



Code	Phase content	CCPS model
PP	Within group discussion	Identifying the problems (USTD)
PC	Between group Communication	
ID	Individual Decision	Generating idea (IDEA)
CD	Cooperation Decision	
PM	Expression and explanation using physical language	Formulating the robot's behaviour(FORM)
AT	Coding using algorism	Programming (PROG)
PR	Examining ER's movement	Evaluating the solution (EVAL)
LI	Language Interaction	Target language-learning interaction (TLLT)
PLI	Teacher involving language interaction	
OFF	Off task or Irrelevant Message	Off the task (OFFT)

FIGURE 4 | The model for evaluating students' learning behaviours.

TABLE 2 | Progress scores of the independent sample *t*-test results between the two groups.

	CSL			EFL			*t*	*p*
	N	Mean	*SD*	*N*	Mean	*SD*		
Language progress	16	10.00	13.19	16	9.13	6 30	0.23	0.812
CT progress	16	19.75	17.71	16	5.63	5.88	3.02**	0.005
Total progress of learning achievement	16	29.75	20.20	16	14.75	7.52	2.81**	0.009

**$p < 0.01$.

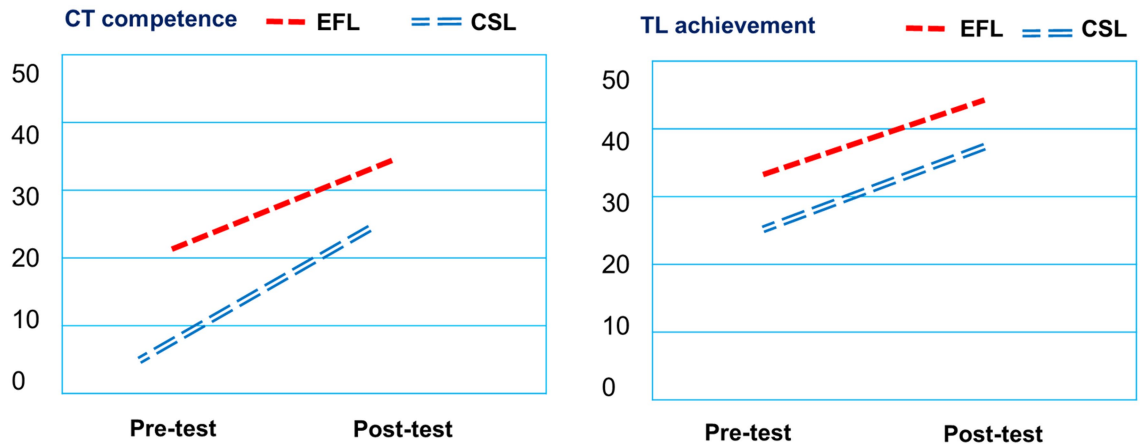


FIGURE 5 | Pre-test and post-test of CT and TL learning scores of the EFL and CSL students.

in CT progress ($M = 5.63$) and post-test progress ($M = 14.75$), when participating in this learning activity (**Table 2**).

Further, paired sample *t* tests were used to investigate the progress of both groups. Both groups significantly improved in their language learning ($t = -3.03$*; $p < 0.05$ for CSL, and $t = -5.79$***; $p < 0.05$ for EFL), CT capacity ($t = -4.46$***; $p < 0.05$ for CSL and $t = -3.83$**; $p < 0.05$ for EFL), and overall learning achievement ($t = -5.94$***; $p < 0.05$ for CSL and $t = -7.84$***; $p < 0.05$ for EFL). Both groups made significant improvement in their linguistic knowledge of target language process, CT process, and overall process of learning achievement, showing that the interdisciplinary activities were beneficial for integration acquisition (**Figure 5**).

Computational Thinking

The study aimed to examine the effects of the different learning groups on students' CT. One-way ANCOVA was first conducted using the pre-questionnaire scores of CT as a covariate. After verifying that the assumption of homogeneity of regression was not violated with $F = 1.24$ ($p > 0.05$), the post-questionnaire scores of the two groups were analysed. However, no significant effect was found between independent variables ($F = 0.247$, $p > 0.05$) on the students' CT skills (**Table 3**).

MANCOVA was further conducted using the pre-questionnaire scores of CT as a covariate to eliminate the learners' differences in their equivalent prior knowledge before the task, after verifying that the assumption of homogeneity of regression was not violated with $F = 0.913$ ($p > 0.05$) and that the Box's M test for homogeneity of covariance matrices was not violated with (Box's $M = 21.35$, $F = 1.82$, $p > 0.05$).

Table 4 shows that the four subscales of CT in the post-questionnaire differed significantly between the two groups (Wilks' lambda = 0.58, $F = 4.62$, $p = 0.006$, Eta = 0.42). The Bonferroni method was then used to analyse the confidence intervals. The results of the *post hoc* comparison indicated that the EFL group showed better cooperation than the CSL group, while the CSL group showed greater problem-solving capacities than the EFL group in these four dimensions of CT regarding algorithm, cooperation, critical thinking, and problem-solving skills.

Learning Anxiety

The study aimed to examine the effects of the different learning groups on students' anxiety. One-way MANCOVA was conducted using the pre-questionnaire scores of learning anxiety as a covariate, after verifying that the assumption of homogeneity of regression was not violated with $F = 1.939$ ($p > 0.05$) and that the Box's M test for homogeneity of covariance matrices was not violated with (Box's $M = 24.93$, $F = 0.158$, $p > 0.05$).

Table 5 presents that the five subscales of learning anxiety in the post-questionnaire differed significantly between the two groups (Wilks' lambda = 0.408, $F = 7.26$, $p = 0.000$, Eta = 0.59). The Bonferroni

TABLE 3 | Descriptive statistics of the two groups.

	CSL group ($N = 16$)					EFL group ($N = 16$)				
	Pre-test		Post-test			Pre-test		Post-test		
	M	*SD*	*M*	*SD*	*AdjM*	*M*	*SD*	*M*	*SD*	*AdjM*
Algorithm	2.98	1.02	3.27	0.82	3.27	3.28	1.13	3.30	1.02	3.30
Cooperation	3.57	0.90	3.30	0.68	3.30	4.05	0.98	4.03	0.92	4.03
Critical thinking	3.35	0.93	3.05	0.91	3.05	3.33	0.97	3.44	0.82	3.44
Problem-solving	2.96	0.72	3.38	0.76	3.38	2.17	1.13	2.47	1.03	2.47

TABLE 4 | *MANCOVA* analysis of CT for both groups.

SV	*Df*	*SSCP*				Wilks' lambda	*F*			
	1	Algor.	Coop.	Critical	Prob-solv		Algor.	Coop.	Critical	Prob-solv
Between Group	1	0.242	1.069	0.706	−1.226	0.584**	0.268^{NS}	6.176* (E>C)	2.677^{NS}	8.0444** (C>E)
		1.069	4.722	3.119	−5.418					
		0.706	3.119	2.060	−3.579					
		−1.226	−5.418	−3.579	6.217					
Pre-test	1	0.695	0.283	0.745	1.398	0.872	0.771	0.175	1.038	3.636
		0.283	0.115	0.304	0.569					
		0.745	0.304	0.799	1.498					
		1.398	0.569	1.498	2.810					
Within Group (error)	29	26.138	13.848	19.418	2.556					
		13.848	19.084	13.893	−2.097					
		19.418	13.893	22.316	6.692					
		2.556	−2.097	6.692	22.415					
Sum	32									

E, ESL; C, CSL; Algor., algorithm; Coop., cooperation; Critical, critical thinking; Prob-solv, problem-solving skills.
***p < 0.01; *p < 0.05; ^{NS}p > 0.05.*

method was then used to analyse the confidence intervals. The results of the *post hoc* comparison indicated that the EFL group showed lower learning anxiety than the CSL group for the dimensions of speech anxiety, communication apprehension, and fear of being negatively evaluated by other students.

Learning Behaviours

In answering the fourth research question, sequential behaviour analysis was executed to examine the differences between the learning behaviours of the two groups. The behavioural sequence reaches a significant level ($p < 0.05$) when the Z value is more than 1.96 ($Z > 1.96$; Bakeman and Gottman, 1997). **Figures 6** and **7** present the behavioural transition diagrams of the students involved in two learning groups; the *z*-scores are shown on the middle line and each line's direction represents its transfer direction. Three loops were analysed based on the analysis of Chevalier et al. (2020) of the CCPS model (Loop 1 and 3) and the loop for FL interaction (Loop 2).

The main differences between the two groups are that the CSL students presented two statistically significant behaviour sequences involved in loops 1 and 3, clarified as the negotiation loop and the trial-and-error loop, without loop 2, identified as target-language interaction. In comparison, the EFL students demonstrated three statistically significant behaviour sequences in these three loops (see **Figure 8**).

The CSL's two significant behaviour sequences are: PC → PP, and ID → PR → AT → ID. During the activities, the CSL students (freely switching roles in the task) demonstrated their behaviours of first identifying the question within or between groups (PC → PP). They then devoted themselves to working on the trial-and-error loop (loop 2), where they worked individually to decide the ER's routes and started checking the movement of robots to figure out the algorithm to make the ER reach the intended destination (ID → PR → AT → ID). Without focusing the negotiation on solutions and problem-solving strategies as was expected to be seen in Loop 1, the CSL students revealed no significant behaviours of target-language interaction (Loop 2). Students rarely participated in assigned target-language use and interaction.

Otherwise, the EFL's three significant behaviour sequences are: AT → CD, PM → ID, and LI → PLI. When aiming to reach the intended destination they demonstrated Loops 1 and 3. The EFL students collaboratively generated ideas by working on algorithms (AT → CD), and they physically expressed their ideas using gestures individually to justify their CT concepts to their partners (PM → ID). Such formulation fell into the essence of negotiation on problem-solving strategies, and thus

TABLE 5 | *MANCOVA* analysis of learning anxiety for both groups.

SV	*Df*	*SSCP*					Wilks' lambda	*F*				
	1	SA	*CA.*	FN	FM	UT		SA	*CA.*	FN	AM	FU
Between Group	1	14.92	11.502	11.063	9.650	5.377	0.408***	27.56*** (E<C)	14.03** (E<C)	12.36** (E<C)	5.15* (E<C)	3.26^{NS}
		11.50	8.867	8.528	7.439	4.145						
		11.06	8.528	8.203	7.155	3.987						
		9.65	7.439	7.155	6.241	3.478						
		5.37	4.145	3.987	3.478	1.938						
Pre-test	1	0.228	0.165	−0.195	0.165	0.331	0.862	0.421	0.161	0.252	0.099	0.806
		0.165	0.119	−0.141	0.119	0.239						
		−0.195	−0.141	0.167	−0.142	−0.283						
		0.165	0.119	−0.142	0.120	0.240						
		0.331	0.239	−0.283	0.240	0.480						
Within Group (error)	29	15.70	15.79	14.50	16.52	12.31						
		15.79	21.38	16.93	22.41	13.57						
		14.50	16.93	19.23	17.34	11.55						
		16.52	22.41	17.34	35.13	14.09						
		12.31	13.57	11.55	14.09	17.24						
	31											

E, ESL; C, CSL; SA, speech anxiety; CA, communication apprehension; FN, Fear of being negatively evaluated. AM: fear of making mistakes in class; FU, Feeling uniquely unable to deal with the task.

****p<0.001; **p<0.01; *p<0.05; ^{NS}p>0.05.*

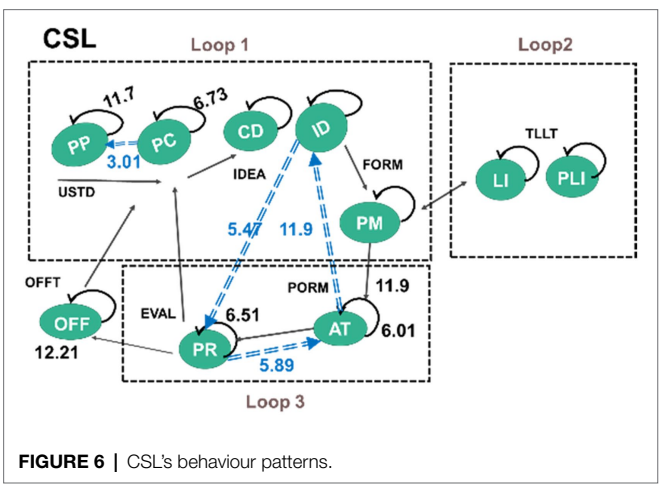


FIGURE 6 | CSL's behaviour patterns.

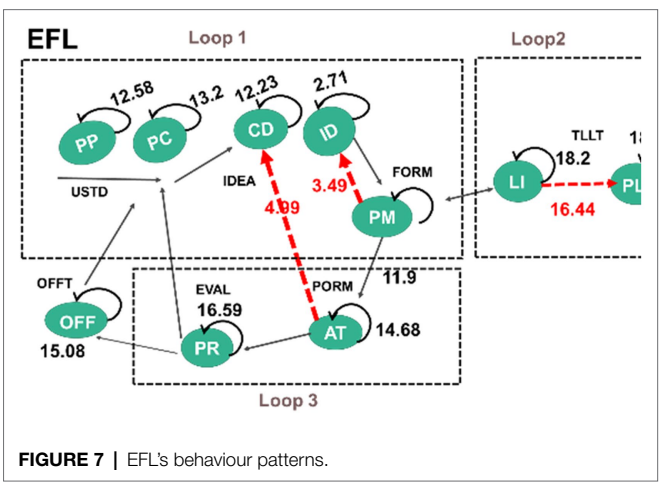
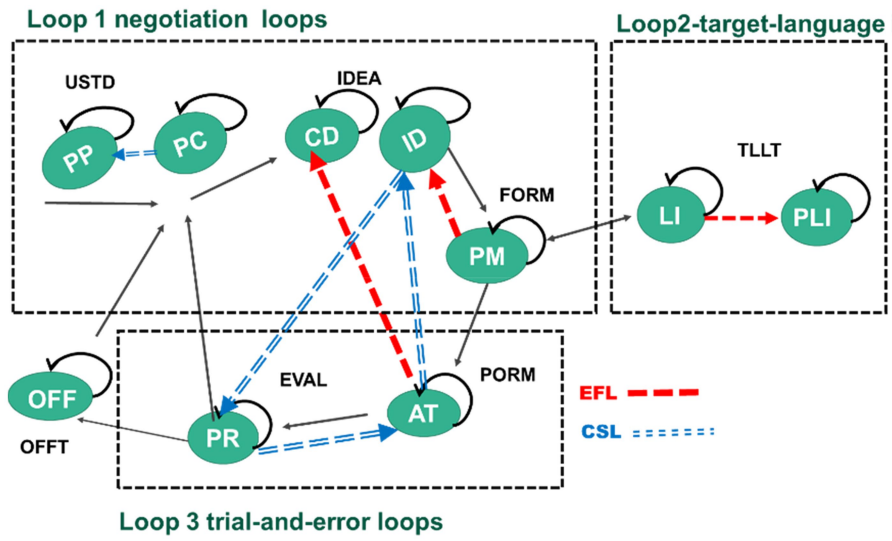


FIGURE 7 | EFL's behaviour patterns.

the EFL students revealed their behaviours of significantly engaging in target-language interaction in Loop 2 (LI → PLI). Following the PP task of coding and conversation practice, the EFL students frequently interacted with one another, kept concentrating on the robots' movements, and used assigned English sentences when it was their turn. If errors occurred, the teacher would come by and guide them to use the taught sentence in their interaction (LI → PLI).

DISCUSSION

Using ERs to acquire CT has led educators internationally to integrate CT into cross-content areas and to further go beyond STEM into other formal curricula (Cheng et al., 2020; Hsu and Liang, 2021). Understanding that this idea needs to be supported by meaningful task design and pedagogically-informed approaches, ER activities to enhance CT and TL must not solely focus on problem-solving skills. Rather, they must provide students with interactive learning settings to engage them in the cross-disciplinary learning, supporting their cognitive (CT) and TL development while reducing their TL anxiety. Coupled with existing studies that have successfully demonstrated learning outcomes across subjects using ERs (e.g., Hsu and Liang, 2021), critically designing activities with instructional strategies and examining their effects across contexts are scarce. Future research needs to consider whether any available tools



Code	Phase content	CCPS model
PP	Within group discussion	Identifying the problems (USTD)
PC	Between group Communication	
ID	Individual Decision	Generating idea (IDEA)
CD	Cooperation Decision	
PM	Expression and explanation using physical language	Formulating the robot's behaviour(FORM)
AT	Coding using algorism	Programming (PROG)
PR	Examining ER's movement	Evaluating the solution (EVAL)
LI	Language Interaction	Target language-learning interaction (TLLT)
PLI	Teacher involving language interaction	
OFF	Off task or Irrelevant Message	Off the task (OFFT)

FIGURE 8 | The differences in the two groups' behaviour patterns.

(e.g., ERs) with strategies and approaches are meaningful, suitable, and engaging for the particular context in which students are to engage in cross-disciplinary study, as simply putting relevant components together cannot guarantee the anticipated results.

Learning *via* a pedagogy-informed approach and tailored ER activities, both groups made significant improvement in their learning of CT and TL. We make no claims about which particular group is superior to the other when looking into differences in the learning behaviours of the two groups. However, the findings demonstrated in this study are worth considering as pioneering and challenging for cross-context instructors and CT practitioners, who need to be mindful of what requirements and expectations of interdisciplinary activities are deployed to children when they learn to code while accessing TL in coding tasks. Indeed, the results showed that the integration of essential CT skills and TL output production in within-group and between-group interactions could be jointly developed using PP strategies and board-game activities as proposed in this study. The ER-integrated activity, along with PP and board-game activities, can be developed using an interactive design offering an open environment for multiple problem-solving solutions but acceptable challenge for TL output practice. This design is regarded as important as studies investigating cross-disciplinary learning from educational settings argue that if children do not find coding tasks engaging, they will not be involved in discussion and negotiation with others regarding any potential CT strategies (Chevalier et al., 2020), nor will they interact with their counterparts while playing the board-game (Hsu and Liang, 2021).

While pedagogic approaches (e.g., PP) and gamified activities (board-game) are often shared across contexts, critically analysing how they affect students' learning is often limited to local contexts. Comparing students' learning behaviours in these two distinct contexts (L2 and FL) reveals potential challenges in interdisciplinary learning across contexts. The CSL students who had natural communication situations favoured individual working and decision making, whereas the EFL students who did not have natural communication situations preferred meaning negotiation and enjoyed collaboration with their partners. The CSL students showed more confidence in facing the challenge of the task by breaking the rules to work independently on solving the problems. However, analysis of their learning behaviours presented a typical trial-and-error loop, although there was a significant difference in their CT progress and problem-solving ability.

A trial-and-error loop is a universal strategy, particularly for novices, for building learning and action in enhancement learning (Mohr et al., 2018); it quickly supports progress (Sutton and Barto, 2018), but not skills development in the long run (Chevalier et al., 2020). In our study, the L2 students' (CSL) behaviours corresponded to the findings of Chevalier et al. (2020) about novices' ERs use for CT development. While ERs offer prompt feedback without intervention in the design, and students easily receive immediate evaluation of their strategy, they can fall into a trial-and-error loop. Although it could be possible that the CSL students were fully attracted by the ERs, such quick feedback from the ERs "reduces the potential of learning how to code to a problem, ignoring the expressiveness and communicative functions of programming" (Bers, 2020, p. 503).

While facing this challenge, Chevalier et al. (2020) successfully adopted an intervention of pause strategies in the ER interface to stop students from directly executing the code. Students showed better interaction and strategies as a result. Although their foci were neither interdisciplinary integration nor cross-context investigation, proper intervention to activate students' communication during programming should be reconsidered for CSL students in the design. If their interaction is activated, CSL students may possibly demonstrate less anxiety in their TL production, since individual work with little attention on between-group interaction often fails to develop social skills and interpersonal relationships, and negatively harms TL

acquisition. Although factors affecting students' language gains are rather complex, it is difficult for L2 students to acquire TL, even though the learners gain much exposure to the target language outside of class (e.g., Taguchi, 2008).

Rather than having a natural communication setting, the EFL students accessed the TL in the formal language classroom. During the task, they showed clear negotiation within groups and TL interaction loops between groups, favouring cooperative decisions. This finding echoes previous studies in EFL settings which found that the language development of learners in a formal FL classroom was facilitated, as they gained abundant input and language use opportunities over time with the help of instructors' explicit instruction and in-class collaborative peer activities (e.g., Taguchi, 2008; Sato and Storch, 2020). Such interaction can be attributed to the fact that they enjoyed the problem-solving task using a communicative task, since their social skills of cooperation in CTS and anxieties in TL anxiety were significantly better and lower, respectively, than those of CSL students.

PP along with the board-game activity, with an emphasis on switching roles in the coding task and the turn-based nature of playing, also contributed to anticipated loops for the EFL students, where they shifted their attention from using different strategies to negotiate with their partners, reinterpret their ideas, and practise conversational sentences within and between groups. PP interaction echoes the discussion of Zhong et al. (2017), and it was found that switching roles between drivers and navigators within groups enhanced negotiation. They suggested that negotiation is more crucial than switching because it is the essence of collaborative learning, although they neither used ERs nor included interdisciplinary activities in their study. The turn-based nature embedded in board-game activities is evidenced to afford creating a question-and-response interaction to TL output productions. Indeed, switching roles along with turn-based interaction can be considered optimal, as it also supports self-reflection and students learn from their mistakes (Chen and Chi, 2020; Wei et al., 2021).

While students enjoyed collaboration with their partners, they demonstrated lower anxiety, as shown by the EFL students. They had lower anxiety associated with CT development, including speech anxiety, communication apprehension, fear of being negatively evaluated, and fear of making mistakes in the class. Such lower anxiety is important for students' continued motivation and willingness to use the FL during the coding task. This confirms the finding of Dewaele and Pavelescu (2021) that less anxious learners are often associated with positive experience when trying to comprehend or speak the target language.

Lastly, several limitations should be clarified. One possible factor affecting the research findings is the novelty of the ERs, although both groups did improve their learning achievement and increased their CT skills. Exploring participants' learning effects in ER-integrated-PP strategies in the long run needs to be addressed. Second, the interactive coding task associated with the board-game mechanism was not described in detail in this paper due to the limited word count. Working on CT integration is not a privilege reserved only for the ER-integrated PP approach. The gamified mechanisms (e.g., level, credits, strategies, competition, and self-reflection) could have been influential variables that came into play to generate the outcomes. If the mechanisms are not appropriate and feasible with regard to the objectives, students may not find the learning task interesting, nor will they engage in it to gain any possible educational benefits (Hsu and Liang, 2021). Future studies addressing board game mechanisms integrated into ER using PP are recommended. Meanwhile, it is hard to generalise the result as the small sample size included in this study. However, the results from the current study serve as pioneering for cross-context instructors and CT practitioners. Lastly, although ER-integrated-PP methods support instructors in integrating interdisciplinary learning activities, it is necessary to confirm the findings from other settings or cross-subject integration.

CONCLUSION

A limited number of studies have focused on ER for interdisciplinary activities. Concerning the fact that PP along with board-game activities is one of the great instructional strategies, ER-integrated-PP activities for interdisciplinary activities which integrate CT and TL should not be ignored. This study investigated the feasibility of using ER-integrated-PP activities and critically validated its impact on interdisciplinary learning in the elementary setting. The result concluded that the ER-integrated-PP approach plus the board-game activity for the promotion of interdisciplinary learning was helpful for promoting the two groups of students' learning in terms of their CT competencies, TL learning, and CT skills, as well as lowering their FL learning anxiety. The results help expand the literature on the design of ERs in a PP way with the gamified activity for interdisciplinary activities.

Students' learning behaviours revealed that ER-integrated-PP education is adventurous in involving students in developing the CT process related to the essence of pair negotiation on problem identification and reinterpretation, and target-language interaction on the gamified mechanism. Drawing on the findings described above, ER activities for interdisciplinary integration can be feasible at the elementary education level. Researchers interested in ER-integrated-PP associated with board-game activities for interdisciplinary learning could consider if the same cross-context design can be reproduced in other cross-subject areas or institutions.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

T-CH: conceptualization, investigation, data curation, writing—original draft, and writing—review and editing. CC: formal

analysis, methodology, and writing—review and editing. L-KW and C-KL: formal analysis, data curation, and writing—review and editing. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We would like to thank the Industry-University Cooperation Project supported by the Institute for Information Industry (III), Taiwan in 2019 so as to connect the companies with researchers. The companies in Taiwan included both Reading & Rhythm Co., Ltd. and Ampus Technology Co., Ltd. We would like to thank III for providing the Robots to be controlled with App and the maps, to allow us to conduct the instructional experiments in Singapore and Taiwan.

FUNDING

This study is supported partly by the Ministry of Science and Technology of China under contract number MOST 108-2511-H-003-056-MY3.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.888215/full#supplementary-material

REFERENCES

Bakeman, R., and Gottman, J. M. (1997). *Observing Interaction: An Introduction to Sequential Analysis.* New York: Cambridge University Press.

Bers, M. U. (2020). *Coding as a Playground: Programming and Computational Thinking in the Early Childhood Classroom*. New York, United States: Routledge.

Blyth, C. (2018). Immersive technologies and language learning. *Foreign Lang. Ann.* 51, 225–232. doi: 10.1111/flan.12327

Brennan, K., and Resnick, M. (2012). "New Frameworks for Studying and Assessing the Development of Computational Thinking." in *Proceedings of the 2012 annual meeting of the American Educational Research Association*, Vancouver, Canada.

Burhans, D. T., and Dantu, K. (2017). "ARTY: fueling creativity through art, robotics and technology for youth." in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence.* 4765–4770.

Cakici, D. (2016). The correlation among EFL learners' test anxiety, foreign language anxiety and language achievement. *Engl. Lang. Teach.* 9, 190–203. doi: 10.5539/elt.v9n8p190

Chen, K. Z., and Chi, H. H. (2020). Novice young board-game players' experience about computational thinking. *Interact. Learn. Environ.* 28, 1–13. doi: 10.1080/10494820.2020.1722712

Cheng, Y. W., Wang, Y., Yang, Y. F., Yang, Z. K., and Chen, N. S. (2020). Designing an authoring system of robots and IoT-based toys for EFL teaching and learning. *Comput. Assist. Lang. Learn.* 34, 6–34. doi: 10.1080/09588221.2020.1799823

Chevalier, M., Giang, C., Piatti, A., and Mondada, F. (2020). Fostering computational thinking through educational robotics: A model for creative computational problem solving. *Int. J. STEM Educ.* 7, 1–18. doi: 10.1186/s40594-020-00238-z

Chung, C. C. J. (2014). "Integrated STEAM education through global robotics art festival (GRAF)." in *Proceedings of the IEEE Integrated STEM Education Conference.* 1–6.

D'Angelo, S., and Begel, A. (2017). "Improving communication between pair programmers using shared gaze awareness." in *Proceedings of the CHI Conference on Human Factors in Computing Systems.* ACM Press, 6245–6290.

Dewaele, J. M., and Pavelescu, L. M. (2021). The relationship between incommensurable emotions and willingness to communicate in English as a foreign language: a multiple case study. *Innov. Lang. Learn. Teach.* 15, 66–80. doi: 10.1080/17501229.2019.1675667

Horwitz, E. K. (1986). Preliminary evidence for the reliability and validity of a foreign language anxiety scale. *Tesol Quarterly* 20, 559–562.

Horwitz, E. K., Horwitz, M. B., and Cope, J. (1986). Foreign language classroom anxiety. *Mod. Lang. J.* 70, 125–132. doi: 10.1111/j.1540-4781.1986.tb05256.x

Hsu, T. C., Chang, S. C., and Hung, Y. T. (2018). How to learn and how to teach computational thinking: Suggestions based on a review of the literature. *Computers & Education* 126, 296–310.

Hsu, T. C., and Liang, Y. S. (2021). Simultaneously improving computational thinking and foreign language learning: interdisciplinary media with plugged and unplugged approaches. *J. Educ. Comput. Res.* 59, 1184–1207. doi: 10.1177/0735633121992480

Hu, L., and Wang, N. (2014). "Anxiety in foreign language learning." in *International Conference on Global Economy, Commerce and Service Science (GECSS-14)*, 122–124.

Korkmaz, Ö., Çakir, R., and Özden, M. Y. (2017). A validity and reliability study of the computational thinking scales (CTS). *Comput. Hum. Behav.* 72, 558–569. doi: 10.1016/j.chb.2017.01.005

Kuo, W. C., and Hsu, T. C. (2020). Learning computational thinking without a computer: how computational participation happens in a computational thinking board game. *Asia Pac. Educ. Res.* 29, 67–83. doi: 10.1007/s40299-019-00479-9

Lewis, C. M. (2011). Is pair programming more effective than other forms of collaboration for young students? *Comput. Sci. Educ.* 21, 105–134. doi: 10.1080/08993408.2011.579805

Longcope, P. (2009). Differences between the EFL and the ESL language learning contexts. *Studies in Language and Culture (SLC)* 30, 203–320. doi: 10.18999/STULC.30.2.303

MacIntyre, P. D. (1999). "Language anxiety: A review of the research for language teachers," in *Affect in Foreign Language and Second Language Teaching: A Practical Guide to Creating a Low-Anxiety Classroom Atmosphere*. ed. D. J. Young (Boston: Mc Graw-Hill), 24–45.

Mohr, H., Zwosta, K., Markovic, D., Bitzer, S., Wolfensteller, U., and Ruge, H. (2018). Deterministic response strategies in a trial-and-error learning task. *PLoS Comput. Biol.* 14:e1006621. doi: 10.1371/journal.pcbi.1006621

Papadakis, S. (2021). The impact of coding apps to support young children in computational thinking and computational fluency. a literature review. *Front. Educ.* 6:657895. doi: 10.3389/feduc.2021.657895

Saranraj, L., and Meenakshi, K. (2016). The influence of anxiety in second language learning: a case study with reference to engineering students in Tamil Nadu, India. *Indian J. Sci. Technol.* 9, 1–5. doi: 10.17485/ijst/2016/v9i42/102044

Sato, M., and Storch, N. (2020). Context matters: learner beliefs and interactional behaviors in an EFL vs. ESL context. *Lang. Teach. Res.* doi: 10.177/1362168820923582 [Epub ahead of print]

Shute, V. J., Sun, C., and Asbell-Clarke, J. (2017). Demystifying computational thinking. *Educ. Res. Rev.* 22, 142–158. doi: 10.1016/j.edurev.2017.09.003

Sutton, R. S., and Barto, A. G. (2018). *Reinforcement Learning: An Introduction. 2nd Edn*. Cambridge, MA: The MIT Press.

Taguchi, N. (2008). The role of learning environment in the development of pragmatic comprehension: a comparison of gains between EFL and ESL learners. *Stud. Second. Lang. Acquis.* 30, 423–452. doi: 10.1017/S0272263108080716

Tsai, M.-J., Hsu, C.-Y., and Tsai, C.-C. (2012). Investigation of high school students' online science information searching performance: the role of implicit and explicit strategies. *J. Sci. Educ. Technol.* 21, 246–254. doi: 10.1007/s10956-011-9307-2

Wei, X., Lin, L., Meng, N., Tan, W., and Kong, S. C. (2021). The effectiveness of partial pair programming on elementary school students' computational

thinking skills and self-efficacy. *Comput. Educ.* 160:104023. doi: 10.1016/j.compedu.2020.104023

Williams, L. A., and Kessler, R. R. (2002). *Pair Programming Illuminated*. Boston, MA: Addison-Wesley Longman Publishing Company.

Wing, J. M. (2006). Computational thinking. *Commun. ACM* 49, 33–35. doi: 10.1145/1118178.1118215

Zhong, B., Wang, Q., Chen, J., and Li, Y. (2017). Investigating the period of switching roles in pair programming in a primary school. *J. Educ. Technol. Soc.* 20, 220–233. doi: 10.1016/j.chb.2016.07.017

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Hsu, Chang, Wu and Looi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Psychology

CURRICULUM, INSTRUCTION, AND PEDAGOGY
published: 08 June 2022
doi: 10.3389/fpsyg.2022.875382

Combined Effects of Block-Based Programming and Physical Computing on Primary Students' Computational Thinking Skills

Oliver Kastner-Hauler[1]*, *Karin Tengler*[1], *Barbara Sabitzer*[2] and *Zsolt Lavicza*[2]

[1] Department of Media Education, University of Education Lower Austria, Baden bei Wien, Austria, [2] Linz School of Education – Department of STEM Didactics, Johannes Kepler University Linz, Linz, Austria

Basic Digital Education (BDE) is already planned to be integrated with the forthcoming curriculum for Austrian primary schools (6–10 years) as it was already implemented for lower secondary schools (10–14 years) in 2018. BDE includes the most essential and novel developments of Computational Thinking (CT), which are fundamentally responsible for nurturing students' problem-solving skills. Thus, evaluating teaching materials, scaffolding guidelines, and assessments is becoming increasingly important for the successful implementation of CT in Austrian classrooms. This study is a part of a longitudinal multi-cycle educational design research project aiming to explore how to foster CT and to raise the awareness, importance, and confidence of teachers and students in applying CT for everyday uses. Our paper focuses on a sub-study in which teaching units for grade 3 and 4 students (8–10 years) were designed by combining an Open Educational Resource (OER) textbook and Physical Computing with the micro:bit device. The designed learning environment consists of three units and was implemented in two classes over 3 weeks. The two classes were further split into two groups each, to ensure better support during implementation. The class teachers received upfront teacher training and conducted pre- and post-test assessments with the students. The resulting data was then analyzed to gain insights into the effects on CT skills of the young learners. Results showed that combining block-based programming and physical computing devices could become a promising approach to promote computational thinking skills in lower school grades. Furthermore, the observed direction of the designed units supports low-barrier access to increase the desired uses of CT in classrooms.

Keywords: microbit, physical computing, block-based programming, computational thinking, computer science, primary digital education, integrated learning environment, assessment

OPEN ACCESS

Edited by:
Stamatios Papadakis,
University of Crete, Greece

Reviewed by:
Hulya Kilic,
Yeditepe University, Turkey
Dayang Norhayati A. Jawawi,
Universiti Teknologi Malaysia, Malaysia
Tova Michalsky,
Bar-Ilan University, Israel
Jari Laru,
University of Oulu, Finland

***Correspondence:**
Oliver Kastner-Hauler
oliver.kastner@ph-noe.ac.at

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

Received: *14 February 2022*
Accepted: *20 April 2022*
Published: *08 June 2022*

Citation:
Kastner-Hauler O, Tengler K, Sabitzer B and Lavicza Z (2022) Combined Effects of Block-Based Programming and Physical Computing on Primary Students' Computational Thinking Skills. Front. Psychol. 13:875382. doi: 10.3389/fpsyg.2022.875382

1. INTRODUCTION

In our current knowledge and information-based society, Computational Thinking (CT) is becoming increasingly important due to the ongoing and widespread digitization. This digitization is also entering education, not least due to pandemic measures and required distance learning. As Wing (2006) amongst other leading computer scientists envisioned, Computational Thinking

should establish itself as the fourth cultural skill (Bollin and Micheuz, 2019) over the coming years—next to reading, writing, and arithmetic. This involves the development and application of problem-solving skills and thinking strategies *"... everyone, not just computer scientists, would be eager to learn and use."* (Wing, 2006). Broadening the set of skills, abilities, and attitudes that are key factors today for the successful implementation of digitization in education embraces the twenty-first Century Skills (P21–Battelle for Kids, 2019). These twenty-first century skills combined with the 5E instructional model for inquiry-based learning (Bybee, 2009) laid a substantial ground for the proposed learning design and make the haptic work with the physical computing device more effective for young learners.

CT has been implemented in a new curriculum (BMBWF, 2018) for lower secondary (10–14 years) in Austria as part of Basic Digital Education (BDE) and is already planned to be integrated into the forthcoming curriculum for primary (6–10 years). For broad introduction in lower secondary, an Open Educational Resource (OER) textbook for CT with the micro:bit (Bachinger and Teufel, 2018a) was developed, where the first author contributed example tasks in this resource (Bachinger and Teufel, 2018b). At the time of writing, there is still a lack of proven teaching material and scaffolding guidelines to conduct CT in class, particularly for primary teachers. Well-designed and working instructional material is urgently sought to provide support and get teachers started with the new tasks, especially for including less technically trained teaching staff (Papadakis, 2021). This paper aims to highlight how closing this gap can be achieved by establishing an integrated learning environment for CT with physical computing. Based on previous research cycles (Kastner-Hauler et al., 2021) and the feedback collected, selected examples from the OER textbook with the micro:bit were realigned and specially targeted for primary students. Students and teachers will be able to develop confidence by implementing the designed learning environment. Everyday CT use that builds on assessed material can emerge playfully. In addition, this way of administration supports the fulfillment of CT in an integrative way rather than seeing it as a separate subject.

In our work, we have been designing and developing a research project using the micro:bit physical computing device to support, promote, and evaluate the adoption of CT with BDE in schools. It is a longitudinal, multi-cycle research project to promote CT and increase awareness, importance, and confidence in its everyday applications. The sub-project and the focus of this article combines block-based coding with Makecode (Microsoft, 2016) and physical computing with the micro:bit (Sentance et al., 2017) in primary schools. Following this approach and building on the twenty-first century skills, the different aspects of CT can be demystified (Shute et al., 2017) and translated for classroom uses. In the following, we will demonstrate how the integrated learning environment guides the students gradually from computing at the desktop computer to physical computing with the micro:bit. We will show how the building blocks of the theoretical background work together to form a *"holistic approach of STEAM education through CT"* (Pears et al., 2019) and how the learning outcomes can be evaluated concerning CT (Zapata-Cáceres et al., 2021). STEAM refers to the fields of Science, Technology, Engineering, Arts, and Mathematics and is reportedly perceived as associated with micro:bit activities (Gibson and Bradley, 2017). Results give the comfortable impression, that inquiry-based learning enables the connection of all theory parts and provides playful discovery to foster the development of CT and problem-solving skills.

2. CONTEXT AND THEORETICAL BACKGROUND

2.1. Computational Thinking

The importance of CT in today's education is underlined by its increasing appearance in new curricula even at lower school levels. In Austria, CT was first implemented in 2018 as a mandatory element of Basic Digital Education (BDE) with the new curriculum for lower secondary education. There, CT is highlighted and defined as a separate area, alongside the user skills for common programs to manage everyday office life. The foundation for this trend in education was laid in an article by Wing (2006), which describes CT as a fundamental skill-set necessary for everyone to succeed in twenty-first century society. Wing further advances that CT skills can be applied to any problem, not just in computer science (CS), as a way of thinking within the problem solution process. The idea and mental concept behind CT originated from Papert (1980), who had also started the programming language Logo to promote and develop those skills. There, Papert also illustrates his practice-oriented constructionism that emphasizes on the learner's active role throughout the entire learning process while creating socially reflectible artifacts. The learning theory of constructionism connects the topics of this research, i.e., block-based programming, physical computing with the micro:bit, STEAM education, inquiry-based learning, and CT. For this purpose, a specially designed learning environment was rolled out and researched.

2.1.1. CT Frameworks

The development of a universally accepted CT definition is not yet complete. However, in recent years, a fairly consistent picture for defining CT has emerged, and CT is thus becoming increasingly demystified (Shute et al., 2017). Shute et al. (2017) found that the most common aspects of CT are abstraction, decomposition, algorithms, and debugging—adding iteration and generalization with their research. Slight variations of this definition including frameworks, practices, and related concepts are described in Barr and Stephenson (2011), Brennan and Resnick (2012), Grover and Pea (2013), and Selby and Woollard (2013). A thorough overview of what (Wing, 2006) started with her claim for CT skills development and how the different frameworks relate to each other can be found at Palts and Pedaste (2020, p. 118)—proposing an all encompassing model for CT development. Recent research by Li et al. (2020) examined CT definitions and conclude to see CT more as a mental model of thinking and approach for problem-solving than a sole computing skill—supporting our holistic approach and STEAM integration (Pears et al., 2019).

2.1.2. CT Assessment

When it comes to the assessment of CT, several researchers propose a system of assessments (Brennan and Resnick, 2012; Grover et al., 2015; Román-González et al., 2019) rather than one all-encompassing tool. For a comprehensive assessment of CT, Román-González et al. (2017a,b) developed, validated, and complemented a set of assessments to target all aspects of CT and the six levels of Bloom's (revised) taxonomy (Krathwohl, 2002). One test of the set of assessment tests developed is the Computational Thinking test (CTt). It was developed for 10–16 year-old students and is primarily used to test the levels 1 (Remember) and 2 (Understand) in Bloom's (revised) taxonomy. Due to the available time for the presented interventions of three units within 3 weeks, the higher levels 3 (Apply), 4 (Analyze), 5 (Evaluate), and 6 (Create) of the taxonomy cannot be expected to be achieved within this relatively short amount of time. Nevertheless, with the Beginners Computational Thinking test (BCTt) from Zapata-Cáceres et al. (2020, 2021) an adapted version of the CTt exists that specifically targets 5–10 year-old students. This perfectly fits the scope of this research, which targets 8–10 year-old students in primary school.

The BCTt follows the three-dimensional (3D) framework by Brennan and Resnick (2012) that groups CT aspects into three computational (ct) dimensions of ct-concepts, ct-practices and ct-perspectives. Brennan and Resnick (2012) identified seven ct-concepts that are highly useful for block-based coding projects and that also transfer to other non-programming contexts: *sequences, loops, parallelism, events, conditionals, operators, and data*. Applying the BCTt, 25 questions are assessed that can be answered independently of a programming environment either on screen or with paper and pencil. The assessment tests fully on ct-concepts, partially on ct-practices, and ignores ct-perspectives (Zapata-Cáceres et al., 2021). The BCTt is considered a scientifically sound and validated test for assessing computational thinking of primary school students (Román-González et al., 2017b; Zapata-Cáceres et al., 2020) and is recommended to be used within a *system of assessments* (Román-González, 2015) at the lower levels of Bloom's (revised) taxonomy.

2.2. Block-Based Programming

With the extended use of the computer in schools, the need to develop an understanding of its functioning and programming continues to increase. Especially in the school context, a distinction between text-based and block-based programming languages is essential for targeting the appropriate age group successfully. The text-based programming environments are preferably used for upper school levels and favor those who understand English well or are willing to learn it. Since the vocabulary of available commands is derived from the English language, one must have the commands memorized to be able to type them. Only the first few letters need to be typed from memory if auto-completion is offered by the programming environment. With block-based environments, there is no need to have prior experiences with the necessary vocabulary. The programmer can visually select from a given palette of command blocks, usually grouped by function, and then colored differently. Commands that do not show up in the programming environment simply do not exist. The block-based variants of programming languages are favored for lower school levels because of their lower entry barrier to start coding. A few common steps under teacher guidance are sufficient to then independently match the appropriate blocks (Weintrop and Wilensky, 2017) like a puzzle to construct a working program playfully. Additionally, in most block-based environments, the language of the blocks can be switched to other languages than English. Scratch from the MIT Media Lab (Resnick et al., 2009), which is considered as the mother of all block-based programming languages, is available in over 50 languages (Scratch-Wiki DACH, 2021). This compares to Makecode (Microsoft, 2016) for the micro:bit, which is used in the project and is also available in over 30 languages. Therefore, the integration of all learners in the event regardless of language level (Dasgupta and Hill, 2017) and vocabulary proficiency in English is assured. Block-based programming, even in early childhood ages (5–7 years), is reported to develop problem-solving, planning, and thinking strategies to gain social, language, and cognitive skills (Papadakis, 2022).

2.3. Physical Computing

Physical computing connects a computing device to the environment equipping it with the prerequisites for sensing (O'Sullivan and Igoe, 2004) and communicating (Igoe, 2017) and involves learners with the design and realization of tangible real world products from one's imagination (Przybylla and Romeike, 2014). A single-board computer such as the micro:bit is given access to the physical world utilizing sensors and controllers. These elements can sit directly on the single-board computer or be connected externally and augmented with electronic circuitry. The prevailing idea is that the machine can also handle some kinds of sensorimotor perceptions, albeit a bit more limited than humans. By engaging with physical matter and the environment, a way of human-machine interaction is obtained that is particularly conducive to grasping CS and CT concepts. Moreover, when a learner loads a self-created program onto the micro:bit, CS concepts take shape (Rock-Paper-Scissors Game—**Supplementary Figure 1**) and even get a face with the micro:bit LED display. We mainly focus on haptic tinkering, while playfully exploring the world of computing—functionalities and interrelationships can be developed more quickly in this constructionist learning setting. Furthermore, physical computing in the context of STEAM (Schulz and Pinkwart, 2015) promotes deeper understanding through the active creation of one's own learning experiences.

2.4. Inquiry-Based Learning and 5E Instructional Model

The essential features of classroom inquiry (National Research Council, 2000) were implemented by operationalizing the 5E instructional model (Bybee, 2009) and build an integral part of the proposed inquiry-based learning (IBL) design. Based on the model, the 5E learning cycle (BSCS.org, 2021) consists of five phases beginning with (1) engagement, (2) exploration, (3) explanation, (4) elaboration, and (5) evaluation. The framework of the 5E model provides enough elasticity to use

open-ended, guided, or direct questioning depending on the situation and the actual student's needs.

The learning materials offered were developed following a combined approach to stimulate interest in further investigation and playful tinkering during the thought process of problem-solving. The starting material contains only a subset of the available material in hard copy for a sample exercise using the micro:bit, but sufficient to understand the problem at hand and get started. Next, the Wiki website based on the textbook contains additional materials, but this must be uncovered by the learner during exploration as needed and is not presented at first sight. This behavior is accomplished through "spoiler" links to be clicked before new information becomes available and builds on previous research (Kastner-Hauler et al., 2021). In this way, an emphasis is placed on student-centered and self-directed learning pathways through the design of the material (Reitinger et al., 2016).

3. METHODOLOGY

The implementation of a newly designed integrated learning environment for CT with physical computing is investigated utilizing a pre- and post-test approach. Through further refining and adjusting of the material, a theory desirably emerges and will be investigated in follow-up research on a larger scale. For now, the following question is explored in this paper:

> RQ: To what extent do learning and teaching with the designed learning units for the micro:bit and Physical Computing effect students' Computational Thinking skills?

3.1. The Study

3.1.1. Participants

The participants in the study were selected from a primary school of a district capital. The sample consisted of 45 third and fourth-grade students, 19 female, and 26 male, aged 8–10. Students had previous experiences in using computers, and tablets with digital media. However, they had no experiences with the micro:bit and Makecode programming preceding the start of the study. At the end of the intervention, data from 41 students were available for analysis. All of these students completed both tests, the pre- and the post-test. Thus, 20 third-grade children, 7 girls (35%) and 13 boys (65%), as well as 21 fourth-grade children, 11 girls (52%) and 10 boys (48%), participated in both tests. To meet all legal and ethical research requirements, permission was obtained to conduct this study with the students. Confidentiality of the survey was also assured, and the identities of the participants were not recorded anywhere in the test. Pseudonymous unique codes were used to allow comparison of pre- and post-test at the stages of data processing.

3.1.2. Research Design

This pre- and post-test study is part of a long-term educational design research study that is conducted in iterative cycles (McKenney and Reeves, 2013, 2018). The research has been carried out since 2019 in selected primary and secondary schools with students in the age range of 8–14 years. Previously conducted studies on broadening awareness and application of computational thinking with physical computing included scaffolding material for learning and teaching CT with an Open Educational Resource (OER) textbook (Bachinger and Teufel, 2018a) and in combination with the Flipped Classroom method (Kastner-Hauler, 2020; Kastner-Hauler et al., 2021). Previous cycles have included studies of CT task difficulty recognition and design (Kastner-Hauler et al., 2020) and evaluation of the OER textbook usage to promote CT (Bachinger et al., 2021). For this paper, we focus on a sub-study in which instructional units were designed for third and fourth-grade students (ages 8–10) using the micro:bit. Two classroom teachers conducted three units each, and each class was divided into two groups for 3 weeks. Classroom teachers received in-service teacher training to avoid direct involvement of the authors and administered the pre- and post-tests (Zapata-Cáceres et al., 2021) for assessment. The resulting material was analyzed for observable effects on young learners' CT skills interacting with the learning environment (Basu et al., 2020, 2021).

3.1.3. Measurement Tools

The Beginners Computational Thinking test (BCTt) version 2 by Zapata-Cáceres et al. (2021) is used as the assessment tool for CT measurement (Zapata-Cáceres, 2021). Version 2 of the BCTt consists of 25 tasks and can be completed within a lesson unit. Each task provides four single-choice answers as possible solutions that must be selected. The assessment contains the following computational concepts (**Table 1**): sequences (No. 1–6), loops (No. 7–18), and conditionals (No. 19–25). A chick and its mother hen are the two main characters in the assessment tasks. The primary goal is to bring the chick to its mother (**Supplementary Figure 2**) by solving the task conditions of the maze (Zapata-Cáceres et al., 2020). The assessment test was conducted online at the beginning and end of the intervention in the computer lab with the students. Each student entered a unique identification code *via* the web browser to allow direct correlation of the tests before and after the intervention. Anonymity and confidentiality of the data were maintained by processing the data pseudonymously. During the assessment, before starting each new category of computational concepts (**Table 1**), the test was paused and an explanatory example (**Supplementary Figure 3**) was solved together with the students upfront, as recommended by the creators of the BCTt.

For the data analysis of the pre- and post-tests, all answers of students who took both tests were qualified. In the analysis, sociodemographic data on age and gender, as well as the resulting data of the items, were used. A *post-hoc* t-test for paired samples was utilized to determine a possible increase in CT skills. The t-test is an appropriate instrument for comparing the dependent samples of the pre- and post-tests with parametric data. The individual items of the BCTt were binary coded, e.g., with numbers of 1 for a correct response and of 0 for an incorrect response. Data analysis was performed using descriptive statistics and t-tests with the SPSS 27 software package.

TABLE 1 | Computational concepts in BCTt (Zapata-Cáceres et al., 2021).

Items	1. Sequences	Loops		Conditionals		
		2. Simple loops	3. Nested loops	4. If-then conditionals	5. If-then-else conditionals	6. While conditionals
1–6	■					
7–11		■				
12–18			■			
19–20				■		
21–22					■	
23–25						■

3.2. Resources

3.2.1. BBC Micro:Bit

The physical computing hardware used in the project is the BBC micro:bit. In 2013, the British Broadcasting Corporation (BBC) revived their past initiative "BBC Micro" from the 1980s to bring coding into every home, and school in the UK. The new initiative *Make it Digital* (BBC, 2015) aims to develop core skills in the STEAM fields (Tyrén et al., 2018) by inspiring creativity in the digital world. As of September 2016, the micro:bit project was transferred to the non-profit Micro:bit Educational Foundation (2016) independent of the BBC. Since then, the foundation has been working internationally to further disseminate and support the single-board computer—introducing coding even in primary school. Research on the use of the micro:bit in primary school shows support for collaboration, gamification, and individual work to develop problem-solving and programming skills (Kalogiannakis et al., 2021). The here described project works with both versions of the micro:bit, including the second version released in 2020.

The micro:bit combines all the features usually associated with a smartphone, such as Bluetooth connectivity, compass, acceleration, brightness, and temperature. With the listed features, the micro:bit qualifies as the ideal vehicle to promote physical STEAM activities with inquiry-based learning to bring CT to schools at an affordable price. The efficacy of this endeavor is confirmed by the first-year impact study (BBC and Discovery Research, 2017), which showed an increased interest in CS as a future subject option for students. After using the micro:bit in school, that was given for free to every grade 7 student in the UK, 70% more female and 25% more of all students considered CS (and CT) as subject focus after physical experimenting and live coding it.

3.2.2. Integrated Learning Environment

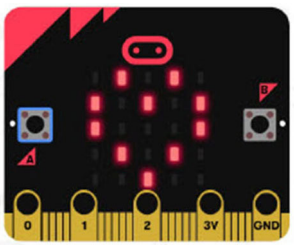
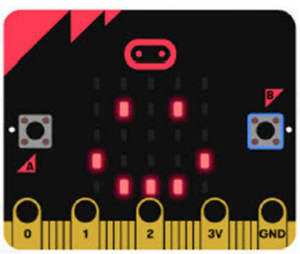
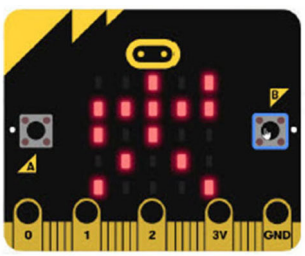
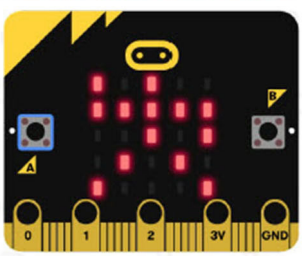
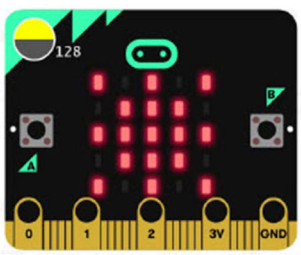
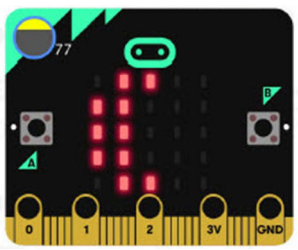
Three learning units (**Table 2**) that show immediate success and further motivate students were created for the intervention. The design of the units follows the 5E instructional model from Bybee (2009, 2014) for inquiry-based learning (IBL). Earlier research from the authors developed materials that were based on an OER textbook and an accompanying wiki website for lower secondary students (Kastner-Hauler et al., 2021). Building on previous cycles, materials for primary students were redesigned by reducing and rephrasing too verbatim explanations and shifting to pictorials more appropriate for younger students. Textbook, wiki, and embedded tutorials combine the approach to encourage interest in further investigation and playful exploration during the problem solution process. First, the printed book contains only a selected subset of the available material for a sample task using the micro:bit, but enough to understand the given problem and get started after initial tutorials. Second, the wiki website based on the textbook contains additional material for each step, but this initially hidden content needs to be actively revealed by the learner. Therefore, the wiki provides "spoiler" hyperlinks that must be clicked before new information is accessible to the learner. In complement, the haptic aspect of physical computing intertwined with the handling of the micro:bit device by using its sensors, buttons, and the display is shown for each unit in **Table 2**. This way of structuring and integrating the learning environment supports self-driven exploration, tinkering, and inquiry-based learning paths (Reitinger et al., 2016) whilst playful search for a solution to a problem task.

3.2.3. Lesson Plan Design Details

After an initial engagement and experimenting with how to connect the micro:bit and upload a program, unit 1 focused on display output with pre-set and custom 5 × 5-pixel graphics. Then, in unit 2, event-driven programming, loops, and animating the display output followed. In unit 3, advanced program constructs such as variables, and conditional branching were used and expanded to include physical computing with sensors. A period of 3 weeks was available for the overall intervention with three units (**Table 2**). The units are described in the following paragraphs in more detail.

Unit 1 consists of the tutorial "Flashing Heart." When running through the tutorial, the individual steps are presented with short videos and are additionally supported by hiding all unnecessary command blocks. The hiding makes the initial orientation and the focus on the essential parts for the entry immensely easier. After that, the Makecode programming environment, including the micro:bit simulator which is available online, is explained in more detail. By plugging the micro:bit into the computer and uploading the program, the code can then be tested directly on the physical device and the pixel-drawing can be checked for refinement. Finally, the learners were guided to make their own drawing appear on the micro:bit's display.

TABLE 2 | Lesson plan design—integrated learning environment.

Units	Unit 1	Unit 2	Unit 3
Objectives	- First program with makecode - Getting familiar with the micro:bit - Connect the micro:bit to USB/power - Transfer program to micro:bit	- Event-driven programming, Loops - Animate screen - Clear screen - Reset micro:bit - Display text - Repeat Loop	- Conditional programming, variables, sensors - If-then-else - Using accelerometer, light sensor - Using sound (computer only, micro:bit v1)
Activities	- Tutorial "flashing heart" - Hands-on micro:bit: computer connection with USB-cable - Drag and drop .hex program file onto micro:bit (USB) - Draw/display smiley - Draw/display stick figures	- When button A pressed → display built-in figure - When button B pressed → display other figure - When button A+B pressed → clear screen - When program reset/started → display single dot or letter - Animate Display → loop two figures	- Tutorial "rock, paper, scissors" → when "shaked" - Tutorial "sunlight sensor" → LEDs - When button pressed → play melody - Compose own melody
Physical computing	- Giving shape to algorithms - Use self-drawn stick figures	- Interact with device using buttons - Animate Screen	- Make the device feel (Shake, Light) - Make the device sound
Computational concepts	- Sequences - Loops (simple) - Events (on-start)	- Loops (advanced) - Events (buttons)	- Conditionals - Events (shake) - Variables (Data)
Computational practices*	- Being incremental and iterative	- Testing and debugging	- Reusing and remixing
Computational perspectives*	- Expressing	- Connecting	- Questioning
Learning outcome	1  2 	3  4 	5  6 

*"Computational practices and perspectives" are trained throughout the entire intervention and are not explicitly matched to specific units.
For example, "Testing and debugging" can occur early or later on depending on the first result.
The same applies to "Connecting" and asking for help from seat neighbors, et cetera.
These CT aspects are interwoven with the inquiry-based constructionist learning process and are present across all units.
Figures 1 and 2: Heart – Grin: Sample artifact for unit 1 with block-based Makecode programming environment. Source: Micro:bit Educational Foundation, 2022, https://makecode.microbit.org/_7rLK17EcuXko. Reproduced with permission. Figures 3 and 4: Maxi1 – Max2: Sample artifact for unit 2 with block-based Makecode programming environment. Source: Micro:bit Educational Foundation, 2022, https://makecode.microbit.org/_a4tfis7zuVsj. Reproduced with permission. Figures 5 and 6: Moon – Sun: Sample artifact for unit 3 with block-based Makecode programming environment. Source: Micro:bit Educational Foundation, 2022, https://makecode.microbit.org/_D6agprWL4iWv. Reproduced with permission.

Unit 2 introduces the concept of triggering an event by using buttons A and B, e.g., to start or clear the display. This is extended by using the reset button on the back of the micro:bit and by pressing both buttons A+B simultaneously. Finally, the display is animated, and two figures are shown alternately. The students are again encouraged to create their own figures after the joint exercise.

In unit 3, sensor functionality and more physical computing are used to start a program event-driven (through triggering of an event). For this purpose, the game "rock-paper-scissors" is translated to the micro:bit and the three symbols are displayed accordingly as soon as the micro:bit is shaken. A random number is generated, stored in a variable and the display is controlled with conditional branches. The next example uses the brightness sensor, which is hidden in some of the LEDs of the display and

measures the brightness of the environment. If the brightness value falls below a certain threshold, the micro:bit should display a symbol for darkness, e.g., the moon, otherwise, the sun should appear. Finally, the micro:bit is used to output sound and create one's own melody. The sound output is most easily done on the computer with the micro:bit simulation since version 1 of the micro:bit used here does not have a speaker. At the end, the students were able to give feedback and describe what they liked best about programming the micro:bit with Makecode.

4. RESULTS AND DISCUSSION

The BCTt was conducted twice in each group for both classes, before and after the three units of the intervention. The planned time for pre- and post-test was set equal to assure the same conditions for all students—although the post-test did not fully consume the planned time budget. This observable behavior is consistent with the findings of Zapata-Cáceres et al. (2020) who originally designed the assessment. Completing the intervention, pre- and post-tests were compared after initial data filtering for participation in both tests. The sample population appeared to be normally distributed with a sample size of $n = 20$ for third and $n = 21$ for fourth grade. Having met the prerequisites, it was possible to start the data analysis and examine the research question. Data processing was conducted from the first and second author under the four eyes principle. Findings from the elaborated data will be discussed subsequently.

4.1. Collected Data

The total score for each category of the BCTt shown in column *Avg. Sum* of the results in **Table 3** has a maximum of 25 correct answers. The individual maximum was reached two times in the pre-test, only for the fourth grade, and six times in the post-test for both grades. The individual maximum score of 14 and 13 in the pre-test, and of 17 and 18 in the post-test occurred for the third grade and the fourth grade, respectively. A better post-test mean as a control source demonstrates the positive effect of the intervention, as does a higher average sum of correct answers in each category for both grade levels—except in the category of *Simple loops*. A possible explanation will be discussed later in the next section. Moreover, a higher post-test mean and a higher post-test average sum of all participants in the last row *Total* show the amount of positive effect more clearly.

Further comparison of pre- and post-test with a *post-hoc* paired sample *t*-test (**Table 4**) showed the statistical significance of the intervention with $p < 0.050$. The intervention realized a total significance with a value of $p = 0.006$ for third grade and $p = 0.029$ for fourth grade. This clearly suggests that the designed units for Physical Computing with the micro:bit achieved a measurable positive effect on students' CT skills. Additional *t*-test comparison of the total number of correct answers shows a higher increase for third grade ($M = 1.950$, SD = 2.800) than for fourth grade ($M = 1.000$, SD = 1.949). Overall, third-graders were able to gain nearly double the amount of additional correct answers in the post-test than fourth-graders. On the other hand (**Table 3**), fourth-graders achieved a higher total number of correct answers upfront on the pre-test ($M = 21.429$, SD = 2.925) compared to the third-graders ($M = 20.000$, SD = 2.991).

TABLE 3 | Evaluation of computational concepts—Grade 3 and 4.

Computational concepts	Grade	*N*	BCTt	Mean	SD	Variance	Avg. sum
Sequences	3	20	Pre	0.883	0.196	0.038	17.667
			Post	0.933	0.100	0.010	18.667
	4	21	Pre	0.968	0.085	0.007	20.333
			Post	0.992	0.036	0.001	20.833
Simple loops	3	20	Pre	0.980	0.062	0.004	19.600
			Post	0.980	0.062	0.004	19.600
	4	21	Pre	0.933	0.115	0.013	19.600
			Post	0.933	0.097	0.009	19.600
Nested loops	3	20	Pre	0.764	0.238	0.056	15.286
			Post	0.836	0.219	0.048	16.714
	4	21	Pre	0.864	0.199	0.040	18.143
			Post	0.932	0.116	0.014	19.571
If-then conditionals	3	20	Pre	0.700	0.299	0.089	14.000
			Post	0.825	0.294	0.086	16.500
	4	21	Pre	0.810	0.249	0.062	17.000
			Post	0.881	0.218	0.048	18.500
If-then-else conditionals	3	20	Pre	0.400	0.348	0.121	8.000
			Post	0.725	0.413	0.170	14.500
	4	21	Pre	0.619	0.415	0.173	13.000
			Post	0.690	0.402	0.162	14.500
While conditionals	3	20	Pre	0.750	0.284	0.080	15.000
			Post	0.833	0.229	0.053	16.667
	4	21	Pre	0.635	0.315	0.099	13.333
			Post	0.762	0.261	0.068	16.000
Total	3	20	Pre	20.000	2.991	8.947	400
			Post	21.950	3.120	9.734	439
	4	21	Pre	21.429	2.925	8.557	450
			Post	22.429	2.039	4.157	471

To obtain a deeper understanding of the different aspects of the BCTt, a detailed group comparison of the six computational concepts (**Table 1**) was performed. The paired sample *t*-test values for mean, standard deviation, and significance grouped by the six categories of computational concepts are presented in **Table 4**. Although the comparison of the total of all categories showed statistical significance in the overall result, almost all the single categories lack a significant increase except one category. The category *If-then-else* provides a significance value of $p = 0.002$ for third grade.

To better illustrate the measured outcome of the interventions' effect on students' Computational Thinking skills, here characterized by the BCTt's categories of computational concepts, a combined chart of pre-test, post-test, and Cohen's d effect was elaborated in **Figure 1**. This graphical presentation depicts the highest gain of competencies in the category *If-then-else* for third grade (Change = 6.50, i.e., from $M = 8.00$ to $M = 14.5$) and in the category *While* for fourth grade (Change = 2.67, i.e., from $M = 13.33$ to $M = 14.00$). All other categories for both grades show an increase and effect (Cohen's d)—except

TABLE 4 | *Post-hoc* paired sample *t*-test (Grade 3 and 4).

Computational concepts	Grade	*N*	BCTt	Mean	SD	Sig. (2-tailed)	Cohen's d
Sequences	3	20	Pre + post	0.050	0.203	0.285	0.246
	4	21	Pre + post	0.024	0.096	0.267	0.249
Simple loops	3	20	Pre + post	0.000	0.092	1.000	0.000
	4	21	Pre + post	0.000	0.167	1.000	0.000
Nested loops	3	20	Pre + post	0.071	0.276	0.262	0.259
	4	21	Pre + post	0.068	0.195	0.125	0.349
If-then	3	20	Pre + post	0.125	0.393	0.171	0.318
conditionals	4	21	Pre + post	0.071	0.327	0.329	0.218
If-then-else	3	20	Pre + post	0.325	0.406	0.002	0.800
conditionals	4	21	Pre + post	0.071	0.427	0.452	0.167
While	3	20	Pre + post	0.083	0.322	0.262	0.259
conditionals	4	21	Pre + post	0.127	0.324	0.088	0.391
Total	3	20	Pre + post	1.950	2.800	0.006	0.696
	4	21	Pre + post	1.000	1.949	0.029	0.513

the category *Simple loops*. This category shows no explicit gain in competencies, but the connected sister-category of *Nested loops* shows an increase for both grades. Third-graders scored lower on pre- and post-test in almost all categories except for the category *While*, whereas fourth-graders scored lower than third-graders for both, pre- and post-test.

The overall effect of the intervention is measured with a Cohen's d value of 0.696 for third grade and 0.513 for fourth grade (**Table 4**). Having a Cohen's d value of >0.5 suggests a medium effect for both grades, achieving almost a large effect (Cohen's d > 0.8) for third grade. This outlines an overall increase in the categories of computational concepts related to the intervention, particularly the Computational Thinking skills that are effected by programming the micro:bit.

4.2. Discussion

This study aimed to evaluate newly designed teaching material for CT with the micro:bit and Physical Computing used in primary school. The BCTt was applied to assess the amount of the learning effect on students' CT skills with a 3 week intervention. The BCTt is a validated assessment (Zapata-Cáceres et al., 2020) for CT skills following the 3D framework focusing on different categories of computational concepts, partially on computational practices, and ignores computational perspectives (Zapata-Cáceres et al., 2021). Setup and implementation of the BCTt are easy to apply, making it widely available to researchers and teachers in the field. The above characteristics clearly signal the potential of the BCTt and its value for curriculum design and material development with CT assessment. In the following, the findings and limitations of the demonstrated study will be discussed.

4.2.1. Findings

The study indicates an increase in CT skills for both grades comparing the total sum of the BCTt score for pre- and post-test (**Table 3**). The figures suggest a higher effect for third-graders compared to fourth-graders, albeit both effect scores are within medium range. Both classes had no prior experiences with the micro:bit and Makecode programming before this intervention, but the primary school under study is equipped with Bee-Bots programmable floor robots. Teachers from the afternoon care regarding fourth grade were reportedly curious about this cute little robot and showed interest in trainings on this subject (Tengler et al., 2021) after collecting additional feedback. Therefore, we assume that fourth-graders in this study had first contact with the Bee-Bot during unofficial afternoon hours. On the other hand, the measured higher effect for third grade can be explained through a possible ceiling effect of fourth-graders in the assessment, which was also reported by the designers of the BCTt during the evaluation of their assessment instrument (Zapata-Cáceres et al., 2021). In addition, fourth-graders have more or less a year of development advantage given their age. The authors suggest further investigation to clarify these findings in the outlook section.

The category of *Simple loops* has the same value for pre- and post-test for both grades. Representing the highest achieved amount for third grade and second-highest for fourth grade suggests a possible explanation for the unmeasured difference. As a very high score on the pre-test diminishes the available room left for improvement on the post-test. Moreover, the trained computational concepts with Makecode and the micro:bit were mainly located in the sister-category *Nested loops*. Or simply, the high amount already known in this category makes much more effort necessary to improve measurably what the 3 week intervention was not able to provide.

Another interesting finding is depicted in **Figure 1** in the category *If-then-else*. The improvement for third grade in this category is extremely high yet it achieves overall the lowest score in pre-test. Having a large available range for improvement on the post-test is the nature of a relatively low pre-test score. In addition, the training deepens more on the computational concepts of the sister-categories *If-then* and *While* conditionals. The highest improvement for fourth grade can be seen in the category *While*, which suggests overall good effectiveness of the whole intervention. Since the difficulty level increases from one category to the next, the last category *While* can be seen as an indicative marker for the whole journey conveyed.

Problem-solving as a general cognitive ability is considerably linked to CT in the context of programming (Kalelioglu et al., 2016). It is important to train such problem-solving skills at an early stage of childhood to generate action-relevant knowledge and gain non-verbal intelligence and CT skills. Our presented study provides evidence that supports this point of view, which also compares to similar findings of other studies (e.g., Román-González et al., 2017b; Tsarava et al., 2019; Zapata-Cáceres et al., 2021).

4.2.2. Limitations

Some findings and observations of the study could be further explored for better understanding and overcoming limitations. The study showed that third-graders profited more than fourth-graders from the training. Therefore, it would be interesting to extend research into age group appropriate task design and tailor more specific for the participants' grades as the inventors

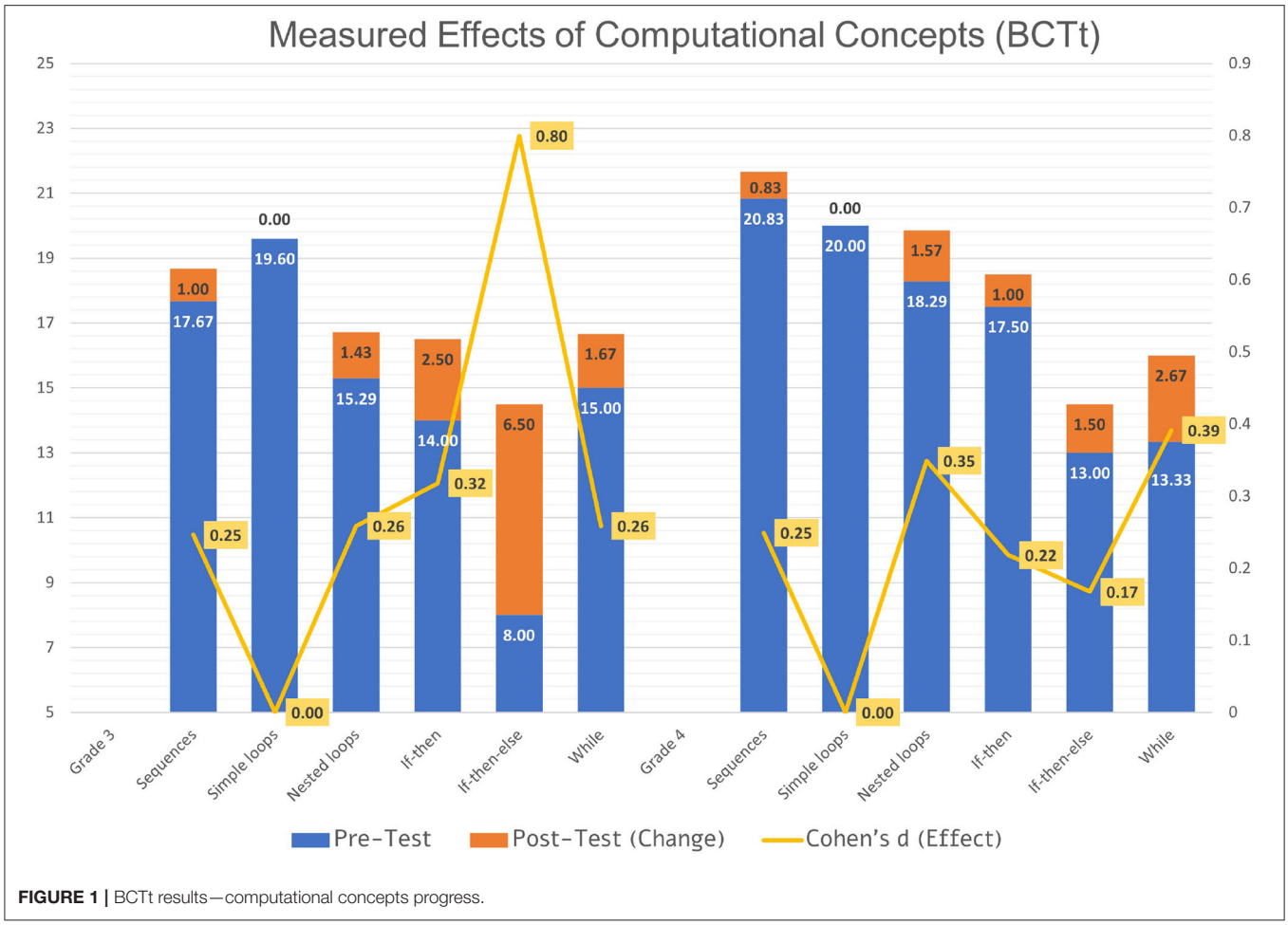


FIGURE 1 | BCTt results—computational concepts progress.

of the BCTt suggest to prevent ceiling effects. This applies to gender-specific tailoring as well, not touched in examination anywhere here. The available time for the intervention is another point to consider in more detail. The 3 weeks of training could be extended to six or even 3 months in further studies. And finally, the target group availability was limited due to pandemic constraints and the so resulting sample size of the participants should be enlarged in upcoming cycles of research.

The BCTt is only one part of the set of three complementary assessment tools as researched by Román-González et al. (2017a), but it is sufficient to trial the direction of newly created learning materials. Given the relatively short amount of available time for the intervention higher levels of Bloom's (revised) taxonomy cannot be expected to be achieved. Therefore, the BCTt is an appropriate instrument for assessing the reported intervention on the lower levels of Bloom's (revised) taxonomy linked to CT and problem-solving.

5. CONCLUSIONS AND OUTLOOK

The study presented in this article is part of a longitudinal multi-cycle educational design research (McKenney and Reeves, 2018) on teaching materials and scaffolding guidelines for CT implementation in the classroom. The overall research goal of the entire project aims to foster CT and to raise the awareness, importance, and confidence of teachers and students to apply CT as an everyday skill. The first step was to identify existing teaching materials and teacher readiness for CT uses in class. The need to vary the difficulty levels of existing example tasks from an identified source, the OER textbook and wiki for the micro:bit (Bachinger and Teufel, 2018b), was denoted and further investigated (Kastner-Hauler et al., 2020). The next step, during pandemic constraints, was to adapt an example task (Kastner-Hauler et al., 2021) for flipped classroom delivery (Lage et al., 2000; Bergmann and Sams, 2012; Buchner, 2018). In this paper, we focus on a sub-study in which three units for primary students (grades 3 and 4) were designed using an OER textbook for CT with the micro:bit—originally designed for lower secondary level (grades 5–8). The interventions laid out in **Table 2** were conducted by emphasizing the haptic and sensing elements of Physical Computing with the micro:bit device. The assessment of CT took place before and after the interventions with the BCTt (Zapata-Cáceres et al., 2021). The consecutive steps, not included in this article, will be developed and assessed during in-service teacher training for CT with design guidelines iteratively derived throughout the entire project. In the next phase of our

study, we will also investigate other CT assessments as mentioned and their operational applicability to further drive CT content development for classroom use.

With this study, we provide empirical evidence that teaching block-based coding combined with Physical Computing can foster CT skills of young learners of third and fourth grade (8–10 years) in primary school. Answering the research question, a significant value for the overall effect of the intervention was obtained and the increased values for the trained categories show a positive effect. A successful introduction to the concepts of CT and programming can be provided by the newly designed teaching material, even in upper primary school. The results of this study provide also the opportunity to further extend the research to lower secondary with a specifically targeted age group (10–14 years) assessment in upcoming cycles, i.e., the Computational Thinking test (CTt) (Román-González, 2015) or the abbreviated CTt (Tsarava et al., 2022). Further extending assessments to all six levels of Bloom's (revised) taxonomy could include the use of Bebras Tasks (Dagiene and Sentance, 2016; Bebras.org, 2022), Dr. Scratch (Moreno-León et al., 2015) or possibly Dr. Micro:bit—a version of Dr. Scratch to be developed in the future. Further cycles of materials development should incorporate self-authored tutorials within Makecode to eliminate clutter and focus on ct-concepts to learn. After that, *sidedoc* elements (collapsible, floating menus) within Makecode should enhance the use of the OER materials, eliminating the clutter and the need to switch between browser pages—reading instructions and programming can so be accomplished on one page.

Learning and teaching CT in the context of Physical Computing can be mastered in upper primary school levels (grades 3 and 4) with the here assessed materials. Results gave us positive feedback that our intervention had a substantial impact toward the desired direction. The combination of block-based programming with Makecode and the micro:bit are promising approaches to foster CT skills and introduce Basic Digital Education in primary schools with a playful, enjoyable, and tangible learning environment. Both students and teachers could be consequently empowered to experience a confident and good feeling in developing and applying their Computational Thinking skills whilst problem-solving.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of Education Lower Austria. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

OK-H: conceptualization, methodology, writing and re-writing, visualization, and project administration. OK-H and KT: investigation. OK-H and ZL: editing. BS and ZL: supervision. All authors have read and agreed to the published version of the manuscript.

FUNDING

Open access funding provided by Johannes Kepler University Linz, Austria.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.875382/full#supplementary-material

REFERENCES

Bachinger, A., Brandhofer, G., Kastner-Hauler, O., and Teufel, M. (2021). *Abschlussbericht zum Projekt "Denken lernen, Probleme lösen - Sekundarstufe I"*. Graz: Bundesministerium für Bildung, Wissenschaft und Forschung; Linz & Baden b. Wien.

Bachinger, A., and Teufel, M. (2018b). *Microbit – Das Schulbuch*. Available online at: https://microbit.eeducation.at/wiki/Hauptseite (accessed December 22, 2020).

Bachinger, A., and Teufel, M., editors (2018a). *Digitale Bildung in der Sekundarstufe - Computational Thinking MIT BBC Micro:Bit*. Grieskirchen: Austro.Tec. 107.

Barr, V., and Stephenson, C. (2011). Bringing computational thinking to k-12: what is involved and what is the role of the computer science education community? *ACM Inroads* 2, 48–54. doi: 10.1145/1929887.1929905

Basu, S., Rutstein, D., Xu, Y., and Shear, L. (2020). "A principled approach to designing a computational thinking practices assessment for early grades," in *Proceedings of the 51st ACM Technical Symposium on Computer Science Education* (Portland OR: ACM), 912–918.

Basu, S., Rutstein, D. W., Xu, Y., Wang, H., and Shear, L. (2021). A principled approach to designing computational thinking concepts and practices assessments for upper elementary grades. *Comput. Sci. Educ.* 31, 169–198. doi: 10.1080/08993408.2020.1866939

BBC (2015). *BBC - Make It Digital - Make It Digital: How the BBC and Its Partners Unleashed the UK's Digital Creativity*. Available online at: https://downloads.bbc.co.uk/learning/makeitdigital/Make_It_Digital_ebook.pdf (accessed March 22, 2022).

BBC and Discovery Research (2017). *Make It Digital - Microbit: Project*. Available online at: https://www.bbc.co.uk/mediacentre/latestnews/2017/bbc.com/mediacentre/latestnews/2017/microbit-first-year/ (accessed December 20, 2021).

Bebras.org (2022). *What is Bebras*? Available online at: https://www.bebras.org/about.html (accessed March 22, 2022).

Bergmann, J., and Sams, A. (2012). *Flip Your Classroom: Reach Every Student in Every Class Every Day*. International Society for Technology in Education.

BMBWF (2018). *Lehrplan Digitale Grundbildung - Bundesministerium für Bildung, Wissenschaft und* Forschung. Available online at: https://www.ris.bka.gv.at/Dokumente/BgblAuth/BGBLA_2018_II_71/BGBLA_2018_II_71.pdfsig (accessed January 1, 2020).

Bollin, A., and Micheuz, P. (2019). "Computational thinking on the way to a cultural technique: a debate on lords and servants," in *Empowering Learners*

for Life in the Digital Age, Vol. 524, eds D. Passey, R. Bottino, C. Lewin, and E. Sanchez (Cham: Springer International Publishing), 3–13.

Brennan, K., and Resnick, M. (2012). "New frameworks for studying and assessing the development of computational thinking," in *Proceedings of the 2012 Annual Meeting of the American Educational Research Association, Vol. 1* (Vancouver, BC), 25.

BSCS.org (2021). *Learn About BSCS's 5e Instructional Model*. Available online at: https://bscs.org/bscs-5e-instructional-model/ (accessed January 13, 2021).

Buchner, J. (2018). How to create educational videos: from watching passively to learning actively. *R&E-Source* 12, 1–10. Available online at: https://journal.ph-noe.ac.at/index.php/resource/article/view/584 (accessed April 29, 2022).

Bybee, R. W. (2009). *The BSCS 5e Instructional Model and 21st Century Skills*. Colorado Springs, CO: Biological Sciences Curriculum Study (BSCS), 24.

Bybee, R. W. (2014). The BSCS 5e instructional model: personal reflections and contemporary implications. *Sci. Child*. 51, 10–13. Available online at: https://pimser.org/wp-content/uploads/2022/01/BSCS_5E_Instructional_Model_Bybee-article.pdf (accessed April 29, 2022).

Dagiene, V., and Sentance, S. (2016). "It?s computational thinking! BEBRAs tasks in the curriculum," in *Informatics in Schools: Improvement of Informatics Knowledge and Perception, Vol. 9973*, eds A. Brodnik and F. Tort (Springer International Publishing) Cham:Switzerland, 28–39.

Dasgupta, S., and Hill, B. M. (2017). "Learning to code in localized programming languages," in *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale* (ACM) (Cambridge, MA), 33–39.

Gibson, S., and Bradley, P. (2017). A study of Northern Ireland key stage 2 pupils? perceptions of using the BBC micro: bit in STEM education. *STeP J.* 4, 15–41. Available online at: https://www.cumbria.ac.uk/media/university-of-cumbria-website/content-assets/public/education/images/documents/research/tean/step/Seanpaul-Gibson-and-Patrick-Bradley.pdf (accessed April 29, 2022).

Grover, S., and Pea, R. (2013). Computational thinking in k-12: A review of the state of the field. *Educ. Res.* 42, 38–43. doi: 10.3102/0013189X12463051

Grover, S., Pea, R., and Cooper, S. (2015). "Systems of assessments" for deeper learning of computational thinking in k-12," in *Proceedings of the 2015 Annual Meeting of the American Educational Research Association* (Stanford, CA), 15–20.

Igoe, T. (2017). *Making Things Talk: Practical Methods for Connecting Physical Objects, 3rd Edn*. San Francisco, CA: O'Reilly Media, Inc.

Kalelioglu, F., Gulbahar, Y., and Kukul, V. (2016). A framework for computational thinking based on a systematic research review. *Baltic J. Modern Comput*. 4, 583–596. Available online at: https://ojs.cumbria.ac.uk/index.php/step/article/view/374 (accessed April 29, 2022).

Kalogiannakis, M., Tzagkaraki, E., and Papadakis, S. (2021). "A systematic review of the use of BBC micro:bit in primary school," in *10th New Perspectives in Science Education - International Conference* Bologna, 486.

Kastner-Hauler, O. (2020). "Making und ICM mit BBC micro:bit - computational thinking als "flip" eines OER-schulbuchs," in *Tagungsband zur Tagung Inverted Classroom and Beyond 2020*, eds G. Brandhofer, J. Buchner, C. Freisleben-Teutscher, K. Tengler, V. F. fnma (Norderstedt: Books on Demand GmbH) Graz, 197–206.

Kastner-Hauler, O., Sabitzer, B., and Tengler, K. (2020). "BBC micro:bit experiments – spicing up computational thinking skills," in *INTED2020 Proceedings* Valencia, 5077–5083.

Kastner-Hauler, O., Tengler, K., Demarle-Meusel, H., and Sabitzer, B. (2021). "Adapting an OER textbook for the inverted classroom model - how to flip the classroom with BBC micro:bit example tasks," in *Proceedings* Lincoln, NE, 8.

Krathwohl, D. R. (2002). A revision of bloom's taxonomy: an overview. *Theory Pract*. 41, 212–218. doi: 10.1207/s15430421tip4104_2

Lage, M. J., Platt, G. J., and Treglia, M. (2000). Inverting the classroom: a gateway to creating an inclusive learning environment. *J. Econ. Educ*. 31, 30. doi: 10.2307/1183338

Li, Y., Schoenfeld, A. H., diSessa, A. A., Graesser, A. C., Benson, L. C., English, L. D., et al. (2020). Computational thinking is more about thinking than computing. *J. STEM Educ. Res*. 3, 1–18. doi: 10.1007/s41979-020-00030-2

McKenney, S., and Reeves, T. C. (2013). "Educational design research," in *Handbook of Research on Educational Communications and Technology*, eds J. M. Spector, M. D. Merrill, J. Elen, and M. J. Bishop (New York, NY: Springer New York), 131–140.

McKenney, S. E., and Reeves, T. C. (2018). *Conducting Educational Design Research, 2nd Edn*. New York, NY: Routledge.

Micro:bit Educational Foundation (2016). *The Micro:bit Educational Foundation*. Available online at: https://microbit.org/about/ (accessed December 20, 2021).

Microsoft (2016). *Microsoft MakeCode for Micro:bit*. Available online at: https://makecode.microbit.org/ (accessed November 11, 2021).

Moreno-León, J., Robles, G., and Román-González, M. (2015). Dr. scratch: automatic analysis of scratch projects to assess and foster computational thinking. RED. *Rev. Educ. Distanc*. 46, 23. Available online at: http://www.um.es/ead/red/46/moreno_robles.pdf (accessed April 29, 2022).

National Research Council (2000). *Inquiry and the National Science Education Standards: A Guide for Teaching and Learning*. National Academies Press, 9596.

O'Sullivan, D., and Igoe, T. (2004). *Physical Computing: Sensing and Controlling the Physical World With Computers*. Boston, MA: Thomson.

P21–Battelle for Kids (2019). *Framework for 21st Century Learning*. Available online at: http://static.battelleforkids.org/documents/p21/P21_Framework_DefinitionsBFK.pdf (accessed February 2, 2022).

Palts, T., and Pedaste, M. (2020). A model for developing computational thinking skills. *Inform. Educ*. 19, 113–128. doi: 10.15388/infedu.2020.06

Papadakis, S. (2021). The impact of coding apps to support young children in computational thinking and computational fluency. A literature review. *Front. Educ*. 6, 12. doi: 10.3389/feduc.2021.657895t

Papadakis, S. P. (2022). Can preschoolers learn computational thinking and coding skills with ScratchJr? A systematic literature review. *Int. J. Educ. Reform* 34:1–34. doi: 10.1177/10567879221076077

Papert, S. (1980). *Mindstorms; Children, Computers and Powerful Ideas*. New York, NY: Basic Books.

Pears, A., Barendsen, E., Dagiene, V., Dolgopolovas, V., and Jasute, E. (2019). "Holistic STEAM education through computational thinking: a perspective on training future teachers," in *Informatics in Schools. New Ideas in School Informatics, Vol. 11913*, Series Title: Lecture Notes in Computer Science (Cham:Springer International Publishing), 41–52.

Przybylla, M., and Romeike, R. (2014). Physical computing and its scope - towards a constructionist computer science curriculum with physical computing. *Inform. Educ*. 13, 225–240. doi: 10.15388/infedu.2014.14

Reitinger, J., Haberfellner, C., and Keplinger, G. (2016). "An overview of the theory of inquiry learning arrangements (TILA)," in *Theory of Inquiry Learning Arrangements: Research, Reflection, and Implementation*, eds J. Reitinger, C. Haberfellner, E. Brewster, and M. Kramer (University Press GmbH), 262. Available online at: http://dx.medra.org/10.19211/KUP9783737601450

Resnick, M., Maloney, J., Monroy-Hernández, A., Rusk, N., Eastmond, E., Brennan, K., et al. (2009). Scratch: programming for all. *Commun. ACM* 52, 60–67. doi: 10.1145/1592761.1592779

Román-González, M. (2015). "Computational thinking test: design guidelines and content validation," in *Proceedings of EDULEARN15 Conference* Barcelona, 2436–2444.

Román-González, M., Moreno-León, J., and Robles, G. (2017a). "Complementary tools for computational thinking assessment," in *Proceedings of the International Conference on Computational Thinking Education (CTE 2017)* (Hong Kong: The Education University of Hong Kong).

Román-González, M., Moreno-León, J., and Robles, G. (2019). "Combining assessment tools for a comprehensive evaluation of computational thinking interventions," in *Computational Thinking Education*, eds S. C. Kong and A. Harold (Singapore: Springer), 79–98.

Román-González, M., Pérez-González, J.-C., and Jiménez-Fernández, C. (2017b). Which cognitive abilities underlie computational thinking? Criterion validity of the Computational Thinking Test. *Comput. Hum. Behav*. 72, 678–691. doi: 10.1016/j.chb.2016.08.047

Schulz, S., and Pinkwart, N. (2015). "Physical computing in STEM education," in *Proceedings of the Workshop in Primary and Secondary Computing Education* (London: ACM), 134–135.

Scratch-Wiki DACH (2021). *Scratch in Many Languages - Scratch Wiki*. Available online at: https://en.scratch-wiki.info/wiki/Scratch_in_Many_Languages (accessed November 17, 2021).

Selby, C., and Woollard, J. (2013). *Computational Thinking: The Developing Definition*. Project Report. Available online at: http://eprints.soton.ac.uk/id/eprint/346937 (accessed January 1, 2020).

Sentance, S., Waite, J., Yeomans, L., and MacLeod, E. (2017). "Teaching with physical computing devices: the BBC micro:bit initiative," in *Proceedings of the 12th Workshop on Primary and Secondary Computing Education* (Nijmegen: ACM), 87–96.

Shute, V. J., Sun, C., and Asbell-Clarke, J. (2017). Demystifying computational thinking. *Educ. Res. Rev*. 22, 142–158. doi: 10.1016/j.edurev.2017.09.003

Tengler, K., Kastner-Hauler, O., and Sabitzer, B. (2021). "Tell, draw and code? teachers? intention to a narrative introduction of computational thinking," in *Informatics in Schools. Rethinking Computing Education, Vol. 13057*, Series Title: Lecture Notes in Computer Science, eds E. Barendsen and C. Chytas (Cham:Springer International Publishing), 29–42.

Tsarava, K., Leifheit, L., Ninaus, M., Román-González, M., Butz, M. V., Golle, J., et al. (2019). "Cognitive correlates of computational thinking: evaluation of a blended unplugged/plugged-in course," in *Proceedings of the 14th Workshop in Primary and Secondary Computing Education* (Glasgow, Scotland), 1–9.

Tsarava, K., Moeller, K., Román-González, M., Golle, J., Leifheit, L., Butz, M. V., et al. (2022). A cognitive definition of computational thinking in primary education. *Comput. Educ*. 179, 104425. doi: 10.1016/j.compedu.2021.104425

Tyrén, M., Carlborg, N., Heath, C., and Eriksson, E. (2018). "Considerations and technical pitfalls for teaching computational thinking with BBC micro:bit," in *Proceedings of the Conference on Creativity and Making in Education - FabLearn Europe'18* (Trondheim:ACM Press), 81–86.

Weintrop, D., and Wilensky, U. (2017). Comparing block-based and text-based programming in high school computer science classrooms. *ACM Trans. Comput. Educ*. 18, 1–25. doi: 10.1145/3089799

Wing, J. M. (2006). Computational thinking. *Commun. ACM* 49, 33–35. doi: 10.1145/1118178.1118215

Zapata-Cáceres, M. (2021). *BCTt - German Version*. Available online at: https://forms.office.com/r/FkXkgijqmj (accessed March 3, 2021).

Zapata-Cáceres, M., Martín-Barroso, E., and Román-González, M. (2020). "Computational thinking test for beginners: design and content validation," in *2020 IEEE Global Engineering Education Conference (EDUCON)* (Porto), 1905–1914.

Zapata-Cáceres, M., Martín-Barroso, E., and Román-González, M. (2021). "BCTt: beginners computational thinking test," in *Proceedings of the Raspberry Pi Foundation Research Seminar Series*, (Cambridge, UK: Raspberry Pi Foundation), 46–56.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Kastner-Hauler, Tengler, Sabitzer and Lavicza. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Psychology

ORIGINAL RESEARCH
published: 09 June 2022
doi: 10.3389/fpsyg.2022.872593

Integrating Computational Thinking and Empowering Metacognitive Awareness in Stem Education

Nagalaxmy Markandan, Kamisah Osman and Lilia Halim*

Faculty of Education, National University of Malaysia, Bangi, Malaysia

Education digitization highly enthuses learners for deeper learning and developing thought processes in formulating problems and their solutions effectively in their real-life circumstances. Implementing computational thinking skills through programming in Malaysian primary and secondary school STEM curriculum create huge challenges, especially among STEM educators. This study highlights the integration of four major theories in developing the Metacognitive Empowerment by Computational Thinking (ME-CoT) learning module by cultivating computational thinking through programming skills to promote metacognitive awareness in Biology students. Pilot research was conducted to investigate the reliability of the ME-CoT learning module. Since the study sample was less than 30 students then, the consistency of the measurements, Pearson's r was calculated to identify stability reliability. Findings revealed that the ME-CoT learning module has very strong stability reliability with a value of r = 0.974 and provides advantages such as assisting students to understand the content of the lesson more actively and in a fun way.

Keywords: computational thinking, text-based programming, Biology, constructionism, metacognitive awareness, STEM education

OPEN ACCESS

Edited by:
Stamatios Papadakis,
University of Crete, Greece

Reviewed by:
Insar Damopolii,
State University of Papua, Indonesia
Utiya Azizah,
Surabaya State University, Indonesia
Manne Kallio,
University of Helsinki, Finland
Arsad Bahri,
State University of Makassar, Indonesia

****Correspondence:***
Kamisah Osman
kamisah@ukm.edu.my

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

Received: *09 February 2022*
Accepted: *19 April 2022*
Published: *09 June 2022*

Citation:
Markandan N, Osman K and Halim L (2022) Integrating Computational Thinking and Empowering Metacognitive Awareness in Stem Education. Front. Psychol. 13:872593. doi: 10.3389/fpsyg.2022.872593

INTRODUCTION

Education digitalization is having a significant impact on 21st-century learning, proving the conceptual underpinning of integrating technology in education 4.0 (Maharani et al., 2019; Karimah et al., 2020; Samri et al., 2020) to solve problems effectively and efficiently with broad applicability by employing computational thinking (Tsarava et al., 2019). Computational thinking is a set of problem-solving abilities that today's learners must master and improve (Román-González et al., 2017), and it has progressively grown in importance as a means of thinking about addressing complicated or open-ended situations. Additionally, computational thinking equipped pupils with the ability to think critically, rationally, and systematically (Goyal et al., 2016; Susan and Nurfaradilla, 2019; Karimah et al., 2020, 2021; Samri et al., 2020) as well as to be a lifelong learner (Mohd Noor, 2012). Perhaps the most startling truth is that, because of its ability to debug or solve problems, computational thinking is always related to the metacognitive or cognitive domain. Facilitating problem-solving skills through programming skills will awaken students' thinking skills and boost their metacognitive awareness. Even though many academics point out that computational thinking is not the same as programming, several studies have demonstrated

that procedural and creative programming abilities can directly increase students' cognitive and metacognitive strategies (Román-González et al., 2017).

According to research, computational thinking is closely linked to cognitive concepts such as coding and programming (Kazimoglu et al., 2012; Arihasnida et al., 2017; Samri et al., 2020) both are cognitive processes. According to previous research, programming is an effective and strategic way to improve problem-solving skills in students (Tsarava et al., 2019). Furthermore, involving students in programming activities is synonymous with the development of computational thinking skills. Computational thinking skills are the foundation of a student's ability to think critically (Samri et al., 2020). These abilities are well-organized depending on the activities and tactics employed by the pupils to solve certain difficulties (Kamisah, 2022). When it came to computational thinking skills, the researcher introduced several of them, all of them based on computer programming or computing principles. Those skills were handled as a set of thinking skills in the way computer scientists will think, and it is a fundamental talent that everyone in the world should acquire (Wing, 2006, 2011; Kalelioglu et al., 2016; Zapata-Caceres et al., 2020).

The inclusion of computational thinking and programming skills in primary and secondary school curriculum exemplifies the Malaysian educational system's massive paradigm change. Furthermore, higher education has been identified as the location for implementing computational thinking. However, current paradigm shifts in STEM education, as well as the need for Education 4.0, have demonstrated that this fundamental skill is required from early childhood (Falloon, 2016; Papadakis and Kalogiannakis, 2019; Otterborn et al., 2020), through primary (Haslina et al., 2018; del Olmo et al., 2020), secondary (Gillott et al., 2020), and finally higher education (Kalelioglu et al., 2016; Román-González et al., 2017). In comparison to adults or teachers, pupils' ability to acquire computational thinking was crafted within Vygotsky's Zone Of Proximal Development (Kalelioglu et al., 2016; Kotsopoulos et al., 2017) and it is an advanced level (Puganesri and Puteh, 2019). As a result, Malaysia's Ministry of Education (MOE) has forged ahead and included computational thinking in the Standard Based Primary School Curriculum (Revised 2017) and Standard Based Secondary School Curriculum (Revised 2017), particularly in STEM education. Computational thinking is introduced in the classroom through an interdisciplinary approach in parallel to content knowledge. When traditional teaching methods are still used to deliver Biology content knowledge (Çimer, 2012; Lay Ah Nam and Kamisah, 2017; Bergan-Roller et al., 2018), a difficult scenario arises.

As we know most of the students who choose STEM Biology subjects are students who choose careers in medicine. Anatomy and physiology are core subjects in medical and health science programs which are often challenging programs compared to other disciplines (Periya and Moro, 2019). Moreover, student achievement in these anatomy and physiology subjects is closely related to student achievement in biology subjects starting from upper secondary school (Anderton et al., 2016). Teachers are responsible for contextual science knowledge, adapting it to the needs and demands of students and the curriculum to ensure significant learning occurs (Piaget, 1972a; Reinoso Tapia et al., 2019). In addition, topic content enriched with diagrams, processes, structures as well as biological literacy or facts describing biological processes (Reinoso Tapia et al., 2019) requires an interesting form of teaching and learning and relates to daily life as well as the mind-challenging questions (Fazilah et al., 2016) to attract students to create an active learning environment (Mohamad Masrizan, 2019; Reinoso Tapia et al., 2019). However, active learning cannot be practiced if rote learning (Fazilah et al., 2016) is practiced in schools. Symbiosis with that, abstract and complex facts are usually presented to students using lecture methods (Çimer, 2012; Lay Ah Nam and Kamisah, 2017; Bergan-Roller et al., 2018), as a contributing factor to passive learning. This method was chosen based on the simple factor of managing the class as well as simple techniques for completing the syllabus (Fazilah et al., 2016). In addition to raising issues of lack of motivation and interest (Lay Ah Nam and Kamisah, 2017), memorization learning methods increased the consequences of declining achievement in Biology subjects due to students lacking exposure to the problem -based learning leading to difficulty in answering HOTS questions. The memorization effect also led to a decrease in the achievement of the problem-solving test results of the Program for International Student Assessment (PISA) 2012 (Fazilah et al., 2016) and Trends in International Mathematics and Science Study (TIMSS) (KPM, 2016, 2017) in Malaysia. Thereby the number of students involved in Biology education decreases drastically in Malaysia. In addition, the participation of students in STEM is very worrying which can be seen through the issue of difficulty in achieving Policy 60:40 (Science/technical: literature).

Developing students' learning thinking processes has become a key challenge for teachers in the classroom, and it necessitates a methodical teaching strategy and instrument (Bergan-Roller et al., 2018). Due to a lack of appropriate guidance and modules to assist teachers and students in implementing computational thinking (Cheah, 2016; Lay Ah Nam and Kamisah, 2017; Puganesri and Puteh, 2019; Karimah et al., 2021), the researcher has the opportunity to develop a specific module to integrate computational thinking via programming skills. Computational thinking skills classification proposed by Kalelioglu et al. (2016) and Burbaite et al. (2018) has been applied in this research to develop a Metacognitive Empowerment by Computational Thinking Module (ME-CoT) in Biology Education.

In today's digital world, students should be well prepared with problem-solving abilities (Arihasnida et al., 2017; Samri et al., 2020), as noted in the Framework of Computational Thinking as Problem Solving (Kalelioglu et al., 2016). Burbaite et al. (2018) also blended Model Revised Bloom's Taxonomy. **Table 1** illustrates the computational thinking skills used to develop the ME-CoT module. As a result, the researcher took advantage of a former chance to combine the computational thinking skills from two well-developed frameworks to create an outstanding module with problem-solving computational thinking skills. Both frameworks stress essential computational problem-solving abilities such as abstraction, decomposition,

TABLE 1 | ME-CoT module's computational thinking skills.

The framework of computational thinking as problem solving (Kalelioglu et al., 2016)		**Model revised bloom's taxonomy and computational thinking** (Burbaite et al., 2018)		
		Computational thinking	**Cognitive domain**	**Knowledge domain**
Identify problem	**Abstraction** **Decomposition**	Abstraction Decomposition	Remember understand	Factual
Gathering/ representing and analyzing data	Data collection Data analysis **Pattern recognition** Conceptualizing Data representation	Pattern recognition Data representation	Apply Analyze	Conceptual
Generate, select and plan solution	Mathematic reasoning **Building algorithms and procedures**	Algorithms	Evaluate Creating	Procedural
Implement solution	Parallelization Automation **Modeling and stimulation**			Metacognitive
Assessing solution	Testing / **debugging**/ generalization			

pattern recognition, algorithms, modeling, simulation, and debugging. Those skills are well-organized to encourage students' metacognitive awareness whilst still accumulating problem-solving computational thinking. The table below illustrates computational thinking in the context of problem-solving. The ME-CoT module organizes each computational thinking skill according to the problem-solving sequence described in the Framework of Computational Thinking as a Problem-Solving Process (Kalelioglu et al., 2016). Despite focusing on computational thinking as a problem-solving skill, several studies have shown that programming skills can be used to teach computational thinking (Pedaste et al., 2015; Karimah et al., 2020). The interrelationship between programming and computational thinking demonstrates that programming and cognitive processes are inextricably linked (Samri et al., 2020; Karimah et al., 2021). According to the research, computational thinking is more closely linked to cognitive processes than to computing (Li et al., 2020).

However, because the thinking process is directly tied to programming abilities, which are more prominent in computing, it has been demonstrated that incorporating programming skills will help to enhance computational thinking (Karimah et al., 2021). Several software systems have been developed to assist students in efficiently learning Biology by stimulating and modeling. Visual programming, commonly known as scratch, is used by primary school students to build their computing skills (Oluk and Korkmaz, 2016; Haslina et al., 2018). Even though various programming tools such as scratch, Microsoft Small Basic, Alice, and Toontalk have been introduced in primary school and focus on visual aspects such as drag and drop the blocks (Karimah et al., 2021), it is more valuable if the teacher introduces more advanced programming skills using programming languages for secondary school students to create a higher level of interest in learning programming and Biology content.

Furthermore, Rubinstein and Chor (2014) stated that rather than focusing on visual block programming such as Scratch, secondary students should be introduced to computational concepts and practices related to real programming languages such as Java, C+, C#, and C++, particularly in biology education (Rubinstein and Chor, 2014). This module is designed for novices who want to do some basic programming using the C# programming languages. As a result, having adequate models of computation systems (Aho, 2012) is critical to ensuring that students use the ME-CoT Visual Studio to develop their presentation or activity product. This Visual Studio Community 2019 was chosen because of its ability to focus on visuals or images, which is relevant to Biology education. In form four Biology lessons, the ME-CoT Module was implemented using computer programming using Visual Studio Software on the topic of Respiratory Systems in Humans and Animals. Numerous researchers have identified the necessity to examine programming tasks utilizing block-based programming such as scratch in the context of primary and secondary education from various STEM education fields. However, there are just a few studies that use text-based programming to test students' computational thinking and metacognitive awareness skills. As a result, educational theories should be closely associated with the development and implementation of the Module using text-based programming. Thereby, the ME-CoT module was created based on the integration of three key learning theories: Cognitive Learning Theory, Social Learning Theory, and Constructivist Learning Theory, to measure students' computational thinking skills and metacognitive awareness.

RELATED WORK

Creating a module in the realm of education that is closely tied to educational theory (Higgs, 2013). The research proposed constructionism theory to enable students actively participate in learning processes and make a product as a sign of learning outcome. After digging deep into educational theory, constructionism is proven to strengthen students'

computational thinking capacity and trigger metacognitive awareness. Rather than focusing on "learning by doing," penetration of the constructionism theory in the ME-CoT Module was positively connected with "learning by making." Students achieve a sufficient level of mastery over the topic knowledge and the general computational thinking talent, as well as operation of the metacognitive awareness in themselves, by making such a product.

Constructivist Theory, according to Piaget, indicates that students play a significant role and actively participate in the construction of their knowledge. Teachers serve simply as facilitators, not as knowledge builders, for their students. As a result, this research is founded on the constructivism idea, which states that students are responsible for the construction of existing knowledge. Additionally, the materials or products that students create as a result of programming demonstrate that pupils are capable of more effective learning. Papert also demonstrated as Piaget suggested, that learning occurs through the construction and reconstruction of knowledge through experience. Papert's Constructionism, in particular, formulates learning in the context of a situation rather than looking at it from afar. Students' involvement in current events can help them gain a better awareness of the lack of interaction with the environment (Ackermann, 2001).

Additionally, Vygotsky's Social Theory of Constructivism is linked to both computational thinking and students' metacognitive awareness. John Flavell, the first scientist to investigate the phenomenon of memory, cleared the path for other researchers to investigate the topic. A social interaction-specific metacognitive-related theoretical proposal (Flavell, 1979). The development of concepts available in pupils' cognitive sets is aided by social contact. Cognitive development is linked to metacognitive development, which is the highest level of knowledge. Students' cognitive development is aided by social contact, collaborative learning, and cooperative learning. Vygotsky's Social Theory of Constructivism developed the notion of the ME-CoT Module, which provides a learning platform for students to connect with peers while completing tasks and implementing effective problem-solving skills in their respective Proximal Development Zones. Constructivism's Social Theory strives to establish an understanding that gives significance to what is taught. According to the Social Theory of Constructivism, students build concepts through interaction until a new concept emerges, resulting in knowledge transformation among students in their respective Proximal Development Zones (ZPD). The Proximal Developmental Zone is defined as the distance between a child's ability to perform a task under adult guidance and the child's ability to solve a problem on his or her own (Vygotsky, 1979a). The authentic problem assigned in ME-CoT Module engages novice learners or students in text-based programming skills within a discipline or field of studies. Thereby, the text-based programming is doable at the appropriate level for the students in upper secondary. The module is also designed perfectly under Proximal Development Zones (ZPD), which is situated in the social context and involves active students' participation as well as working as a community in a group.

The cognitive learning theory is essential since the implementation of each generated learning module takes place in the classroom, during teaching and learning activities to gain content knowledge in biology education. Cognitive learning theory explains that learning is a change - a change that occurs in the information available in a person's memory. Cognitive learning theory is a new view to replace behaviorism theory that focuses on external stimuli. Moreover, the curiosity instinct that drives the mental process to understand and know a concept is fundamental to the theory of cognitive learning. Since this learning module is more focused on metacognitive which involves sensory memory and long-term memory, hence Robert Gagne's Information Processing Theory is fundamental in the formation of students 'metacognitive awareness. Gagne (1970) has sought a variety of superior and perfect ways to ensure learning takes place. According to Gagne (1970), a person receiving, and processing information received through the senses is like the information processing of a computer. Plans are inputs processed by sensory memory and short-term memory. The information generated will be stored in long-term memory or used to act with the environment. The programming activities (input) especially text-based programming applied through this module support the construction of students' thinking through digital tools.

The metacognitive theory is closely related to Jean Piaget's Theory of Cognitive Learning (1849–1936) and the social theory of Vygotsky's Social Theory of Constructivism. Thinking about thinking is defined as metacognitive (Flavell, 1979; Sun, 2013; Mazli Sham and Saemah, 2014; Koc and Kuvac, 2016; Astuti et al., 2017). Thinking is a cognitive skill that entails mental activities that evolve by an individual's ability to adapt at each level with the organization of the structure of thinking, which includes schema, assimilation, accommodation, and adaptation (Piaget, 1972b). Metacognitive is a person's knowledge, control of thinking, and learning activities (Astuti et al., 2017; Chou, 2017). Effective learning will occur when students know what they know and what they need to know to fill the knowledge gap that exists where awareness exists within students. Metacognitive awareness refers to a person's awareness of what they know and doesn't know. Metacognitive strategies are methods that students employ to become more conscious of their thinking and learning processes. Students can use metacognitive awareness to help them govern their brains by planning, monitoring, and assessing what they've learned (Koc and Kuvac, 2016; Kyairaniah et al., 2017). Students are responsible for their knowledge. Metacognitive awareness is also very important in systematic problem solving as it is focused on computational thinking (Yağcı, 2019). Students' metacognitive awareness will be automatically awakened by exposing them to computational thinking and text-based programming skills. Students will be aware of the importance of programming abilities in ensuring the correctness and well-designed performance of the final product. The ME-CoT Module focused on demonstrating problem-solving that might occur with input programming tasks, as well as running the program to find any possible errors.

Since the module was created based on the four key theories, each activity was introduced in a staggered manner, beginning

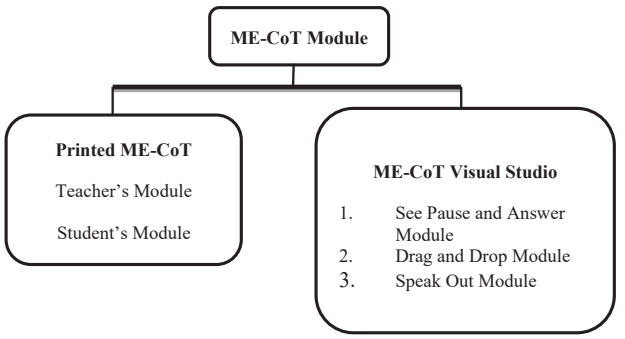


FIGURE 1 | Component of ME-CoT module.

with the ME-CoT Visual Studio and continuing with the Printed Module The students then participated in a hands-on session in ME-CoT Visual Studio which was followed by an authentic activity. **Figure 1** depicts the ME-CoT Module's two major components.

MODULE DEVELOPMENT

The theoretical integration is capable of developing the interdisciplinary approach to STEM Education and computer science which will be implemented in the teaching and learning process in the classroom. This theoretical integration cladding the principal interaction of three major learning theories namely Cognitive Learning Theory, Social Learning Theory, and Constructivist Learning Theory.

This learning theory can be specified into 4 main theories namely, (i) Theory of Constructivism, (ii) Theory of Cognitive Learning, i.e., Robert Gagne's Information Processing Theory, and (iii) Vygotsky's Social Theory of Constructivism, and (iv) Metacognitive Theory. **Figure 2** illustrates the learning theory combination that resulted in the creation of the ME-CoT Module Theoretical Framework. ME-CoT is a well-designed module based on the integration of these four main theories.

Theory of Constructionism

In the analysis and similarities that exist between the constructivist approach developed by Papert and the Constructivism created by Piaget, constructionism is created (Ackermann, 2001). Through programming activities, the ME-CoT module will create output in the form of visual products utilizing ME-CoT Visual Studio. Programming is frequently related to algorithms (Angeli and Jaipal-Jamani, 2018), which involve solving a problem step by step to get the desired result. As a result, programming is a set of activities that can help students develop computational thinking skills through making experiences. Programming is a computing activity that can produce an output with this clear as mentioned by Papert in Theory of Constructionism (Ackermann, 2001). In addition, Constructivist Learning Theory itself also has a very profound impact on this study.

Inquiry-based Learning (Bevins and Price, 2016) is founded on constructivist theory and directly relates to students' flexibility in choosing and conducting investigations based on their

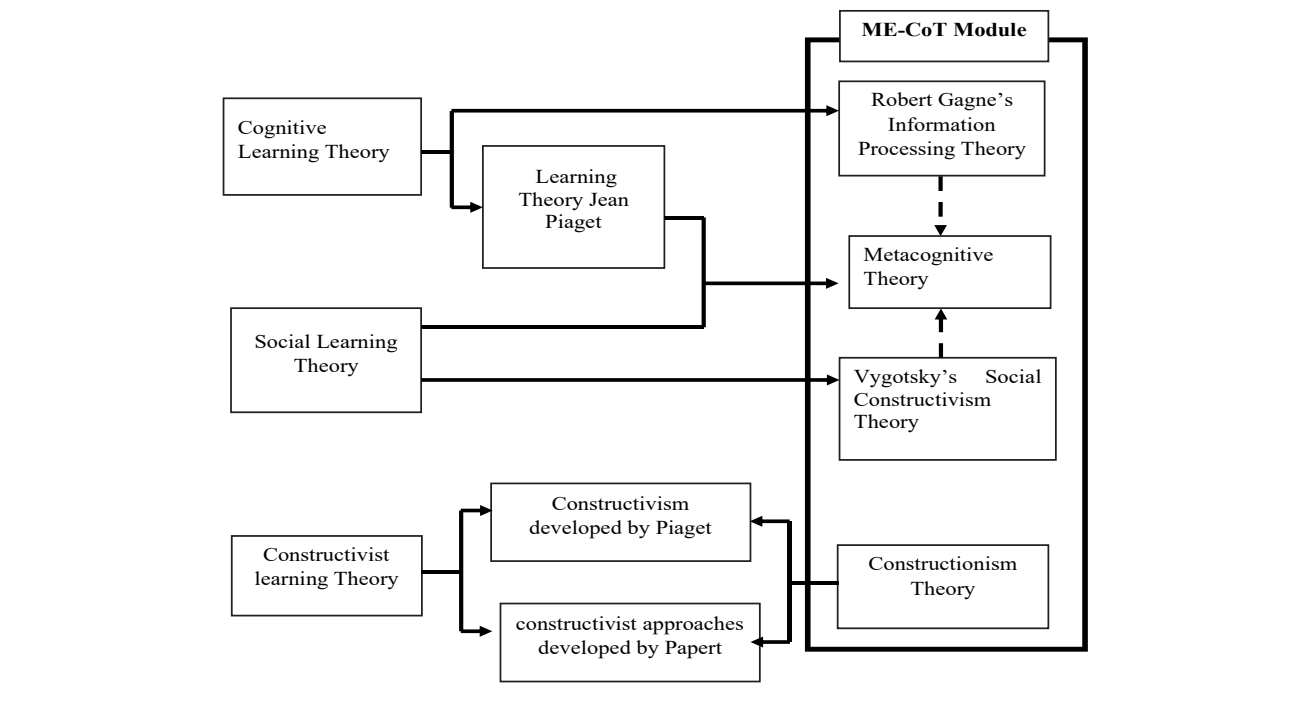


FIGURE 2 | Theoretical integration.

scientific knowledge. Furthermore, the inquiry-based activity approach utilized in teaching and learning is only focused on "learning by doing," but the theory of constructionism, as described by Papert (1996), prioritizes the development of products as a result of learning, also known as "learning by making."

Students' conceptual knowledge can be enhanced via inquiry learning based on the 5E model (Ping et al., 2020). Higher-Order Thinking Skills (HOTS) and Lower Order Thinking Skills (LOTS) are two types of conceptual knowledge. To develop metacognitive awareness, students must master both categories (Burbaite et al., 2018). The constructivist theory provides the foundation for guiding students through the process of mastering metacognitive awareness from a low to a high degree of difficulty. Although, as mentioned in the Theory of Behaviorism, Bloom's Taxonomy stresses the order of learning sequences according to the amount of difficulty (from low to high). The application of bloom's taxonomy in this study, however, is based on the guided inquiry learning approach recommended by Constructivist Learning Theory.

Cognitive Learning Theory (Robert Gagne's Information Processing Theory)

Cognitive learning theory explains that learning is a change – a change that occurs in the information available in a person's memory. Furthermore, since this research focuses on metacognitive awareness, which includes sensory memory and long-term memory, Robert Gagne's Information Processing Theory is essential in the development of students' metacognitive awareness. Gagne (1970) has sought a variety of superior and perfect ways to ensure learning takes place. A human absorbing and processing information obtained via the senses, according to Gagne (1970), is similar to a computer processing information. Plans are inputting that sensory memory and short-term memory process. The information generated will be stored in long-term memory or used to act with the environment showcasing the learning process. Furthermore, learning occurs when students attempt to comprehend the instructions, and it stimulates the learners' cognitive abilities through input from the learner's environment, resulting in change. This procedure is carried out with the assistance of media, which are used as vehicles to deliver important messages (Gagne, 1970). As a result, ME-CoT was purposefully designed around the Cognitive Learning Theory. Based on the ME-CoT activities, students will be assigned a group project to handle the ME-CoT Visual Studio, along with certain instructions that students must comprehend and apply to articulate the Visual Studio and generate the activity product or presentation product. This Module was designed in printed form to serve as a vehicle for delivering adequate knowledge to students for them to complete their assigned work.

Vygotsky's Social Constructivism Theory

The application of Vygotsky's Social Theory of Constructivism occurs due to the presence of interactions in groups or collaborative activities in solving problems. Additionally, metacognitive awareness is linked to Vygotsky's Social Theory of Constructivism because it provides a platform for students to collaborate with their peers in completing tasks and implementing successful problem-solving techniques in their Proximal Development Zones (SPDs). Students also manage to capture the computational thinking skills through coding activities easily. Students begin to employ cognitive strategies to arrange learning approaches to accomplish tasks that have been provided in the form of problems and inquiries once metacognitive awareness has been developed. By investigating to find answers, students will interact and work with classmates (in the Proximal Development Zone) or teachers. Furthermore, self-reflection displays the use of Vygotsky's Social Theory of Constructivism in building metacognitive awareness by evaluating the results of tasks.

Text-based programming had been categorized as a difficult programming task for students because it is using programming languages such as Java, C+, C++, and C#. ME-CoT Module is developed using a text-based programming tool (Visual Studio with the C# programming language) because kids are already familiar with block-based programming (Scratch) from primary school (Haslina et al., 2018; Kaufmann and Stenseth, 2021). ME-CoT Visual Studio has been crafted using Visual Studio which is user-friendly and well. Surprisingly, ME-CoT Visual Studio was created using very basic programming skills and the C# programming language. **Figure 3** shows the task the students did in ME-CoT Visual Studio. The students only need to complete the ME-CoT Visual Studio See Pause and Answer based on the algorithm they constructed in their printed ME-CoT Module during their discussion.

Metacognitive Theory

After performing a study on a group of preschool and primary school children on their capacity to learn and bind a set of things, Flavell proposed metacognitive theory. In metacognitive awareness, the age and cognitive growth of students are connected. Meanwhile, Flavell's research was tweaked to look at cognitive aspects in connection to social contact, revealing a link between social engagement and metacognitive awareness (Flavell, 1979; Pressley et al., 1985). Thereby, Metacognitive Theory was found to be strongly connected to Jean Piaget's Theory of Cognitive Learning (1849–1936) and Vygotsky's Social Theory of Constructivism (Flavell, 1979).

Metacognitive is referred to as thinking about thinking (Flavell, 1979; Sun, 2013; Mazli Sham and Saemah, 2014; Koc and Kuvac, 2016; Astuti et al., 2017). Thinking is a cognitive talent that incorporates mental activity that develops according to the individual's level and capacity to adjust at each level with the organization of thinking structure, such as schema, assimilation, accommodation, and adaptation (Piaget, 1972b; Ragbir Kaur Joginder Singh, 2011). The metacognitive theory is rooted in the cognitive dimension (Krathwohl, 2002), cognitive development begins in infancy, and Piaget (1972b) demonstrated that a child's thinking evolves through stages from infancy to adulthood. Gagne's information processing theory, on the other hand, helps with teaching-based learning and cognitive processes (Ragbir Kaur Joginder Singh, 2011).

```
private void Window_Loaded(object sender, RoutedEventArgs e)
{
    string ApplicationPath = System.IO.Path.GetDirectoryName(System.Diagnostics.Process.GetCurrentProce
    videopath = System.IO.Path.Combine(ApplicationPath, "SystemMedia", "sistemmanusia.mp4");
}

private void Label_MouseUp(object sender, MouseButtonEventArgs e)
{
    RootFrame = bodyFrame;

    Animation1 animationpage = new Animation1(videopath);
    animationpage.AskQuestion1("0:15", "1. Respirasi boleh dibahagikan kepada dua peringkat iaitu");
    animationpage.AskQuestion2("0:24", "2. Apakah itu respirasi luar");
    animationpage.AskQuestion3("0:30", "3. nyatakan gas yang terlibat dalam pertukaran gas");
    animationpage.AskQuestion4("0:57", "4. apakah fungsi rongga hidung");
    animationpage.AskQuestion5("1:03", "5. bagaimanakah keadaan rongga hidung");
    animationpage.StartSession(false); // if true video will mute; if false video sound will play
    RootFrame.Navigate(animationpage);

}
```

FIGURE 3 | Text-based programming in see, pause, and answer module.

The teaching or design of teaching approaches used by teachers plays an important role in enhancing metacognitive awareness (Çakiroğlu and Er, 2020). Social interaction is very important for effective learning to take place, so this importance shows the relationship between Metacognitive Theory and Social Learning Theory. The formation of Metacognitive Theory has a profound effect on the effectiveness of Metacognitive theory because it acts as an active monitoring and sequential control that occurs consciously over cognitive activity (Flavell, 1979). Thus, Metacognitive is a person's knowledge and control of thinking and learning activities (Astuti et al., 2017; Chou, 2017). Effective learning will occur when students know what they know and what they need to know to fill the knowledge gap that exists where awareness exists within students. A person's awareness of what is known and what is not known is also known as metacognitive. Metacognitive strategies are methods used by students to increase awareness of the process of thinking and learning that takes place in the students themselves. Metacognitive awareness can help students control their minds by planning, monitoring, and evaluating the information they have learned (Koc and Kuvac, 2016; Kyairaniah et al., 2017; Harrison and Vallin, 2018).

Students are responsible for their knowledge. Metacognitive awareness must be seen in terms of Planning, Monitoring, Evaluation, Information Management Strategies, Debugging Strategies, Declarative Knowledge, Procedural Knowledge, and Conditional Knowledge. Therefore, in this study metacognitive constructs consisting of aspects of Planning, Monitoring, Evaluation, Information Management Strategy, Debugging, Declarative Knowledge, Procedural Knowledge, and Conditional Knowledge (Schraw and Dennison, 1994; Harrison and Vallin, 2018) will be assessed using a questionnaire instrument the Metacognitive Awareness survey constructed by (Schraw and Dennison, 1994).

The study's conceptual framework is given in **Figure 4**, which is made up of a combination of Learning Theories. **Table 1** shows, the Framework of Computational Thinking as a Problem-Solving Process (Kalelioglu et al., 2016) and the revised Bloom's Taxonomy and Computational Thinking (Revised Bloom's Taxonomy and Computational Thinking Model) are used to explain how to apply computational thinking and foster metacognitive awareness. The use of Learning Theory, as well

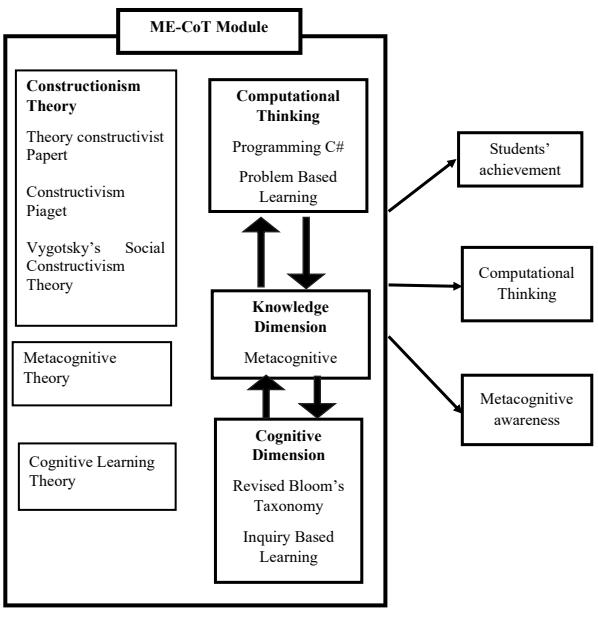


FIGURE 4 | Conceptual framework of the ME-CoT.

as the combination of the Computational Thinking Framework as a Problem-Solving Process (Kalelioglu et al., 2016) and the Revised Bloom's Taxonomy and Computational Thinking Model, to ensure effective learning in the classroom in terms of improving students' achievement, computational thinking, and metacognitive awareness among Form 4 Biology students.

Metacognitive Empowerment by Computational Thinking Visual Studio

Metacognitive Empowerment by Computational Thinking (ME-CoT) module is crafted using Morrison, Ross, Kalman, and Kemp (MRK) model. MRK model is a spiral model which is a very suitable and flexible model to be used in the classroom. ME-CoT is shaped based on the instructional design model which emphasizes nine main elements such as instructional problems, learner characteristics, task analysis, instructional objectives, content sequencing, instructional strategies, designing a message, development of instruction, and evaluation instrument. However, the content of the ME-CoT Module is aligned with the need of the educational Theory. ME-CoT Visual Studio is a text-based programming tool (Visual Studio with C# programming language) that eliminates the block-based programming (scratch) that the pupils were taught in primary school. As a result, the researcher was able to construct the ME-CoT Module, which critically reflects on content knowledge and has a positive impact during classroom implementation. Furthermore, ME-CoT Visual Studio was created with Visual Studio, which is user-friendly and has a design that allows for the inclusion of images and videos that are essential to Biology education.

See Pause and Answer Module

This "See, Pause, and Answer" module is a type of group project that instructs students on how to use video to generate questions and create a quiz-based activity product using the See, Pause, and Answer module. This module will be sent to other members of the group to assist them in answering the questions and to assist teachers in identifying students' misconceptions of the topic of study. This module was built using a combination of standard biology content knowledge and programming skills. Students will study and comprehend the content of the Respiratory System in Humans and Animals by viewing videos, and they will get better knowledge by preparing questions based on the video they have viewed.

Figure 5 shows the task that the students completed in ME-CoT Visual Studio. The students only need to complete the ME-CoT Visual Studio See Pause and Answer depending on the algorithm they generated in their printed ME-CoT Module. The algorithm consists of the time when the video should stop to pop out the question. In the brackets provided, students should write the time and questions. Only the words in red should be altered by the student.

The question instructions are presented during the product launch in the section labeled A, when the question comment is active. The portion labeled B, on the other hand, is a line of inactive questions. Students must use the symbol (//) to trigger a question. Students must also pay attention to the directions offered to them.

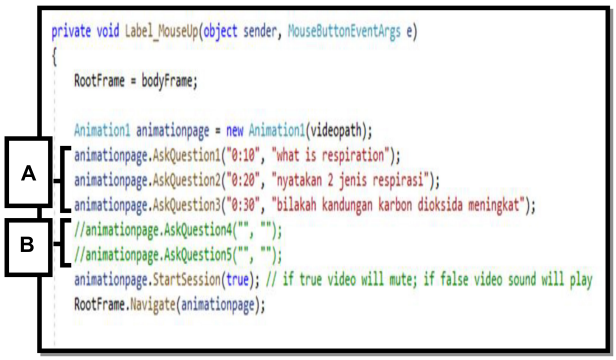


FIGURE 5 | C# Programming sheet in visual studio.

e.g., Animation.startsession (true); // if true video will mute; if false video sound will play

"True" indicates that the video included in the ME-CoT Visual Studio will play without sound or music, while "false" indicates that the video will work with audio.

Using Cognitive Learning Theory, students will begin to grasp computational thinking abilities through the use of the C# programming language throughout product development tasks. It has been demonstrated that students may make a product by creating the activity product, as indicated by the Theory of Constructionism. The product is assessed through debugging activities before being distributed to students in other groups to answer and learn more about the Respiratory System in Humans and Animals. This also allows students to detect faults in their understanding of the topic. This content employs metacognitive theory. Students will gain knowledge through making errors. Students can develop metacognitive awareness by identifying and improving deficiencies through the exercises given. The SPA Module "See, Pause, and Answer" is remarkable in that students must observe and comprehend the video that has been presented. The students should next create an activity product containing the questions to distribute to other groups of students. The creation of questions based on information from the video is an important tool for students to develop metacognitive awareness.

Drag and Drop Module

The Drag and Drop Module is a type of group assignment that instructs students on how to identify each respiratory component and its functions in the chapter on the Human and Animal Respiratory System. Students are expected to create activity items to assist other group members in answering the questions and to assist teachers in identifying students' misconceptions about the topic of study. This module was built using a combination of standard biology content knowledge and programming skills. The Respiratory System in Humans and Animals will be taught and understood by students.

Based on the topic and facts taught, students will create an activity product. Students begin to use Cognitive Learning Theory when they get a better understanding of content

standards. Students will next use images or graphics to create a "Drag and Drop" activity product.

In the ME-CoT Visual Studio Drag and Drop, the value of images on the topic of Biology is highlighted. Students will begin to grasp computational thinking abilities through the use of programming in the C# language during the activity product development.

Students learn how to recognize the ME-CoT Visual Studio Drag and Drop in greater detail, how to integrate photos and create responses in the Drag and Drop Software Module, and how to write a clue text in this part. The "Drag and Drop" activity demonstrates that students may construct a product, as indicated by the Theory of Constructionism. The product is assessed through debugging activities before being delivered to other groups to answer questions and learn about the Respiratory System in Humans and Animals. Additionally, utilizing ME-CoT Visual Studio Drag and Drop to create activity products allows students to pinpoint faults in conceptual knowledge. In this case, the metacognitive theory is used. Students can develop metacognitive awareness, or the ability to identify and fix deficiencies, by participating in tasks given by the ME-CoT Visual Studio Drag and Drop.

Speak Out Module

This "Speak Out" Software Module is a type of Visual Studio software module for expressing students' knowledge of the area being studied. This module teaches pupils how to spontaneously talk or articulate material they've studied. The "Speak Out" Software Module is unique because it can transform the audio produced along with the picture into video. Teachers evaluate the conclusions or summaries of the learning areas that have been covered using the module as a presenting output. This module was built using a combination of standard biology content knowledge and programming skills. Through classroom activities, students will study and grasp the topic of Respiratory systems in Humans and Animals.

The field of Biology is enriched with images that require a deeper understanding. Students should be able to utilize images to express concepts, processes, and information clearly and concisely. The concept of Cognitive Learning Theory plays a significant part in knowledge construction. Students will begin to acquire computational thinking abilities throughout the creation of the "Speak Out" output by using the C# programming language. Students can create a presenting product after completing the "Speak Out" Software Module, as described in Theory of Constructionism. The presentation product created by the "Speak Out" Software Module is a valuable tool for assessing students' knowledge of the Respiratory System in Humans and Animals topic. Students' inability to master the topic is indicated by errors in image interpretation. When a teacher or a coworker provides feedback, the pupils may be able to fix their errors. From the programming of the product through the assessment of the product derived from the Speak Out Module, Metacognitive Theory is used in this context. As a result, students begin to develop metacognitive awareness by planning, monitoring, and analyzing the provided activity. Furthermore, when students complete the assignment, they begin to enrich themselves with the subject knowledge they have acquired. Students will discover and correct problems at each level of the programming activity so that the presentation product launched can produce the desired results.

The module was developed by integrating Problem Solving Computational Skills, and these skills will arouse students' metacognitive awareness. The below table gives a clear explanation of the impact of using the ME-CoT Module on students. Each item found in the construct was able to assess students 'metacognitive awareness through the application of the ME-CoT Module. **Table 2** shows the application of the ME-CoT module to foster metacognitive awareness in students. Students who use the Me-CoT Module have the potential to make plans by speeding up the learning rate as well as setting specific goals before starting programming activities as set out in the ME-CoT Module. In addition to careful planning, students monitor themselves every time they make a programming activity. Students will consider several options for completing activity product assignments and presentation products. To further strengthen metacognitive awareness students should be able to make assessments. Self-evaluation and the success of an activity product or presentation product that has been built through programming activities. Students need time to understand something new information, especially information related to information management or information. Students will try to slow down the activities so that they can focus on important information. While using the ME-CoT Module, students can draw their examples and draw diagrams or pictures to ensure that they understand the new information they have learned.

EVALUATION AND RESULTS

Metacognitive Empowerment by Computational Thinking (ME-CoT) is a well-built tool that has been validated by seven professionals in the area. Following the validation, a preliminary investigation was carried out to determine its dependability. In this study, the general assessment was made using the Fuzzy Delphi measure where to obtain the expert agreement in determining the suitability of the ME-CoT module applied in the Biology form 4 classroom.

Validity

The content validity of the Visual Studio software module ME-CoT has been evaluated by seven experts. The Visual Studio software module ME-CoT is divided into 3 Modules. Each Module is carefully evaluated by experts. According to Mohd Sidek and Jamaludin (2005), if the content validity coefficient is equal to 70% or more, then the module built or produced has high validity. All items recorded a coefficient value exceeding 70%. According to Jamaludin (2008) and Kusuma et al. (2017), the value of the reliability coefficient of a measuring instrument or the activity of a module is the minimum that can be adopted is 0.50. Jamaludin (2012) has stated that the reliability coefficient value must be at least 0.60. While Jamaludin (2012) has stated that the value of the reliability coefficient of the measuring instrument or module activity is 0.90. Therefore, if the value of the reliability

TABLE 2 | Fostering metacognitive awareness by implementing the ME-CoT Module.

Metacognitive awareness	Fostering metacognitive awareness in the application of the ME-CoT Module
Planning	Students can plan in terms of time while preparing activity products. In addition, students should focus on every available information while building the activity product. Students should ascertain the type of product they are producing by providing an algorithm. Students read each step and understand each step that exists before starting the product production activity. Students should prepare the activity or presentation product within the time allocated for them.
Monitoring	Students should examine each step and ensure that each step is followed to achieve the goal. Revisions are very important to ensure that activity products and presentation products are successfully launched. Students should examine and focus on each of the options available in the formation of activity products and presentation products. For example, in the Speak Out module, students should provide algorithms in the image-based print module and provide image-related information. Next, the students should prepare a video presentation product. Students should use correct and accurate strategies to ensure that the video produced is accurate based on the images as well as the questions that have been posted. Students can directly test their level of achievement when students are exposed to a new technique in Visual Studio.
Evaluation	The evaluation is highly prioritized in the application of the ME-CoT Module. Students can assess their level of achievement after preparing an activity product. To further strengthen the understanding of students will use the results of the activity to assess their understanding of the students. Students will make sure they understand the content contained in the video that has been given before preparing the activity product. The See Pause and Answer module is a module that helps students to understand the whole information or content so that they can construct questions and answers based on the video before producing the activity product.
Information management strategy	Students must understand each step to produce a quality activity product or presentation product. In addition, to focus on image accuracy, students should focus on the image's size. Students will focus on each piece of information and each step to successfully launch an activity product or presentation product. Besides that, during the use of the module students have the potential to produce their algorithms based on the steps that have been given as well as students will produce their examples and their way of working to produce a presentation product.
Debugging	Students also get help and guidance from teachers or colleagues in understanding the activities that have been given. In addition, to examine the effectiveness of measures, students also have the potential to change the technique of each activity according to their suitability, for example when producing a product using programming learners will face various challenges, students who easily understand the concept will continue to produce products, while those who do not understand may need guidance and should make sketches or stop and review any new information that is less clear.
Declarative knowledge	While producing activity products, students can train themselves to know the important information that students should. Because, lesson content is the information available in textbooks and reference books students will read and understand as available, but students will focus more on computer science components such as programming. Students will try to understand the programming component's intellectual strengths and weaknesses. Students will also ascertain the objectives and information required by the teacher while producing the activity product or presentation product. Students are also able to control and assess their level of understanding of new information introduced to them. Students are also able to increase their interest in the topics studied when a new context is introduced to them.
Procedural knowledge	Students will provide two activity products and one product or presentation. Students will ensure each product is produced strategically and correctly. For example, students will follow each step in producing a product in their way. However, each strategy that the students use is based on their understanding and suitability to produce activity products and presentation products.
Conditional knowledge	Students need early exposure to the topics they are studying. Thus, the ME-CoT Module does not depart from the context emphasized in Vygotsky's Theory of Social Constructivism, where students have exposure to Biology content topics since primary school and during lower secondary, while for computer science, students are used to the size of the image and resizing images since lower secondary. Students are also good at using computers with basic information. However, programming using the C# programming language is new information introduced in a very simple and easy way through the ME-CoT Module. To successfully launch activity products and presentation products, students will always be highly motivated by using students intellectual strengths as well as balancing weaknesses. Students also know the appropriate and most effective strategies for producing activity products or presentation products by using the ME-CoT Module.

index of this module is between two minimum reliability values of 0.50 and a maximum of 0.90, then this module is acceptable and applicable. Meanwhile, the overall coefficient of the ME-CoT module also recorded a high value of 0.90. Although the content validity and coefficient have a good value, Biology instructors and students have provided several recommendations to help develop and improve the ME-CoT Module.

The Fuzzy Delphi sampling used is purposive sampling and criterion sampling, this is because each expert or sample is selected based on the purpose, based on the experience of experts in the field studied. Meanwhile, this method is categorized as judgment sampling because individual judgment is used to select the study sample based on the researcher's knowledge and study needs. There are seven main steps in implementing the Fuzzy Delphi method. The first step is the determination of expertise. Expert expertise is based on, qualifications, individual character, ability to compare differences, consistency and reliability, time, and experience. Meanwhile, tenure and teaching background recognition from certain experts and bodies, awards and certifications from certain institutions Educated student achievement and performance benchmarks are also classified as benchmarks for the expertise of an expert in a particular field of study (Ramlan and Ghazali, 2018). In the study, the selected experts are individuals who are skilled in the field of

programming, STEM field, Biology Education, and HOTS. The next step is the second step, namely, determining linguistic variables based on the Triangular Fuzzy Number (determining linguistic scale). The second step involves converting all linguistic variables into fuzzy triangle numbering. Likert scale data obtained in the first stage were analyzed using the Excel program for more neat scheduling. All Likert scale data is converted into a triangular Fuzzy number. The triangular Fuzzy Number represents the values of m1, m2, and m3 (Ramlan and Ghazali, 2018). The number will be written in the form (m1, m2, m3). The value of m1 means the minimum value, the value of m2 indicates a reasonable value and the value of m3 represents the maximum value. The three values of m1, m2, and m3 are used to produce a fuzzy scale. In this study, the 5 -point Fuzzy scale agreement level was used. Study data were then scheduled to obtain Fuzzy values (n1, n2, n3) as well as average Fuzzy values (m1, m2, m3) to obtain threshold values, expert percentage, defuzzification, and item ranking.

The third step is the distance determination step to identify the value of Threshold "d." The threshold value of "d" which is less than or equal to 0.2 indicates that the evaluated module is categorized as the expert agreement has been reached (Ramlan and Ghazali, 2018). The reading of the mean value of d in this study shows 0.153 where it is at a value below 0.2 then all the experts reach an agreement on the item referring to the general evaluation of the ME-CoT Module. Next, is the determination of the percentage of group consensus which is the fourth step, where the overall consensus (group consensus) should exceed 75% (>75%) for each item. If each item is equal to or exceeds 75% then each item has reached expert consensus. All items have passed 75% which indicates all items were accepted. General evaluation of the Me-CoT Module recorded the overall percentage of expert agreement is at 92% agreement value which is more than (>75%) means that the general evaluation items of the mE-CoT module have been accepted the conditions of expert agreement on general evaluation items of ME-CoT Module.

The next step is the fifth step which identifies the α-Cut value and the sixth step by identifying the α-Cut defuzzification value (average of fuzzy response). The α-Cut defuzzification value (average of fuzzy response) must exceed 0.5 (>0.5) and if the α-Cut value is less than 0.5 then the item should be dropped and the item does not qualify. The ranking of each item is sorted from the highest fuzzy rating value to the lowest value. Items 13, 11, and 9 show the highest α-Cut value of 0.767 with the first position. This shows that experts agree that, ME-CoT Module is a module containing activities that cultivate the cognitive domain and knowledge domain of students, built based on the new curriculum revision 2017 as well as suitable for use by students aged 16 to 17 years. Meanwhile, items 2,4,5,10, and 12 recorded an α-Cut value of 0.733 with the second position. This shows that the ME-CoT Module can cultivate computational thinking, help students master the field of learning, guide students to solve problems systematically, cultivate inquiry-based learning and the ME-CoT Module also focuses on "learning by making". Items 1, 3, and 7 also recorded an α-Cut value of 0.700 which is the item on the position.

This research included fourteen biology students from one school in one district in Malaysia. It will be examined how programming with Visual Studio helps students in gaining content knowledge of Biology Education. Each module has been assessed separately to have a better understanding of its usability and reliability.

The quasi-experimental study has a sample size of fewer than 30, with 15 students in the experimental group and 14 form 4 Biology students in the control group. According to the findings of the research, the pilot study sample should account for 10% of the total sample size for the real study (Lynne, 2008). With the number of pilot study samples containing 10 samples, the purpose of a pilot study that did not include the construction of new instruments but examined the performance of instrument items created by other researchers with new populations was appropriate (Isaac and Michael, 1995). However, in this study, the pilot study sample included 14 students, with more than 10% of the sample participating in the pilot study. A total of 14 Biology students participated in this pilot research. Furthermore, due to the COVID-19 pandemic, only 14 students were allowed to be used for the pilot study [permission from EraS (Ministry of Education, Malaysia) and the Secondary School Principal]. This study was conducted after receiving the permission letter from *Educational Research Application System* ERaS, the ethical approval was granted by the Ministry of Education Malaysia [Reference number: KPM.600-3/2/3-Eras (11306)].

Besides that, strict SOP in conducting the educational research in school restricted the number of students involved in the pilot study. This study is conducted 3 weeks after the reopening school announcement which had been made by the Ministry of Education, Malaysia. Due to that issue, a suitable statistical analysis had been carried out to analyze the pilot study finding in the reliability of the module.

Preliminary Study for Reliability

Reliability refers to the consistency of a measure while the Cronbach Alpha value is an estimate of the internal consistency of each item in a study instrument (Isaac and Michael, 1995; Hertzog, 2008; Lynne, 2008). The alpha coefficient depends on the variance of the items and the correlation between the items. To ensure the reliability of the measurement tool or study instrument is consistent, the alpha sedative value should exceed 0.75 for a pilot sample of 25 to 40 people. Since the study sample was less than 30 students then, the consistency of the measurements, Pearson's r was calculated to provide an estimate of reliability (Odom and Morrow, 2006) to further strengthen the findings of the study. There are three, reliability coefficients involving Pearson's r namely (i) stability coefficient, (ii) equivalence coefficient and, (iii) objectivity coefficient (Odom and Morrow, 2006). In this study, a stability coefficient is used where the consistency index is between two test times. The first evaluation was conducted immediately after the implementation of the pilot study while after 2 weeks the second evaluation was conducted. Estimates of the consistency index between the two tests were identified using Pearson's correlation coefficient r. Normally a value of Pearson's correlation coefficient r of at least 0.70 is acceptable stability but according to Hertzog (2008), the

TABLE 3 | Relationship strength according to the value of the correlation coefficient.

Size of correlation coefficient (r)	Correlation strength
± 0.81 to 1.00	Very strong
± 0.51 to 0.80	Strong
± 0.31 to 0.50	simple
± 0.21 to 0.30	weak
± 0.01 to 0.20	Very weak

value of Pearson's correlation coefficient r must be above 0.80 is highly recommended for samples with small pilot studies. The correlation Strength Scale by Cohen et al. (2002) was used in the study. **Table 3** shows the Correlation Strength Scale. The less quality measurement was used to identify the reliability due to the number of samples used for the pilot study is less than 15, which is only 14 students were used due to the restriction.

Reliability of Metacognitive Empowerment by Computational Thinking Visual Studio

The ME-CoT module is divided into two main parts, which are ME-CoT Visual Studio and the ME-CoT print module. Stability coefficients are used to determine the reliability of the Visual Studio ME-CoT Module. The first evaluation was conducted right after the pilot project was launched, and the second evaluation took place two weeks later. Estimates of the consistency index between the two tests were identified using Pearson's correlation coefficient r. Normally a value of Pearson's correlation coefficient r of at least 0.70 is acceptable stability but according to Hertzog (2008), the value of Pearson's correlation coefficient r must be above 0.80 is highly recommended for samples with small pilot studies.

Table 4 shows that there is a significant relationship between the first evaluation and the second evaluation of the See, Pause, and Answer Module with a value of $r = 1.000$ Sig = 0.000 ($p < 0.005$). The strong correlation coefficient index indicates the high usability and reliability of the ME-CoT Visual Studio module. The See, Pause, and Answer module, is the ME-CoT Visual Studio module that involves activities to produce activity products. The activity products produced by students involve various skills recommended in computational thinking which are highly emphasized in the ME-CoT module. Students must understand the content presented in the instructional video, then students must provide an algorithm that contains 3 questions (for a pilot study) along with the time for the questions displayed in the video. In addition, evaluation is emphasized in this study, where students prepare questions with answers. That answer is to be filled into answer.txt. in the ME-CoT Visual Studio module. This evaluation has a very important impact not only on computational thinking but on creating metacognitive awareness as well. Once the student prepares the activity product the student should launch the product, at this stage the student will identify their mistakes in building the product. Usually, this error occurs when students prepare the images (image size) and insert images into the ME-CoT Visual Studio. However, planning, monitoring, evaluating as well as strategically managing information applied by students can directly address the problem of a product launch by the students.

Secondly, it is the Drag and Drop Module, the relationship of the first evaluation with the second evaluation of the Drag and Drop Module with a value of $r = 0.967$ Sig = 0.000 ($p < 0.005$). The strong correlation coefficient index indicates the high usability and reliability of the ME-CoT Visual Studio module, especially the Drag and Drop module. This Drag and Drop module, focus on the images especially the characteristic of respiratory structure in human and animals. Pupils needed to draw the respiratory structure of the given task, then they need to search for the related picture on the internet. This activity took time. so, the students are advised to spend time wisely while searching and editing the picture. Then, the students were asked to crop the picture and name the images. Pupils then, list out the images and its characteristic which is known as an algorithm. After collecting the pictures needed, the students will produce the product activity using the Drag and Drop Module. Students will fill up the name of the image in Mainwindow.xaml (ME-CoT Visual Studio module) then the students will fill up the answer text and clue text. It shows that students need to follow the instruction given in the module and create the product activity. Students took time in collecting and downloading the images. Although students find difficulties in downloading the pictures, students were well equipped with the knowledge of editing and resizing images.

Lastly, Speak Out Module shows the relationship of the first evaluation with the second evaluation of the Speak Out Module with a value of $r = 0.974$ Sig = 0.000 ($p < 0.005$). The recorded findings have a very strong relationship based on the Relationship Strength Scale of Cohen et al. (2002). The strong correlation of Pearson's r the stability of the pilot study was above 0.95 which recorded a very encouraging value of above 0.80 for the small group of pilot studies (Hertzog, 2008). This finding revealed that students managed to create the presentation product. Students were given a question, then they need to read and understand the question. Then they must find the related images then talk about the question and relate it with the picture. Before beginning the speaking task, students need to list out the points and the number of images as an algorithm. Pupils will speak out the information and launch the product. The product will be presented as a video. Each group had created a video. Students were excited about using this Speak out module because they can hear their voices when the video was played in front of the classroom.

Meanwhile, to see the effectiveness of the ME-CoT module in improving student achievement in topics, fostering computational thinking and metacognitive awareness, a screening test was conducted during the pilot study. Assessments were made before and after the use of the ME-CoT Module.

The Statistical results of the Wilcoxon Signed Ranks Test in **Table 5** showed that there were differences in achievement test scores in the Respiratory System in Humans and Animals, before and after the use of the ME-CoT Module ($p = 0.001$, $p < 0.05$). The results of the analysis clearly showed that the mean value of the positive rank (mean rank = 7.50) was higher compared to the mean rank for the negative rank (mean

TABLE 4 | Pearson's correlation coefficient index r ME-CoT Visual Studio.

Relation	Second evaluation						Interpretation
	See Pause and Answer		Drag and Drop Module		Speak Out Module		
	r	Sig	r	Sig	r	Sig	
First evaluation See Pause and Answer Module	1.000	0.000					Very strong
First evaluation Drag and Drop Module			0.967	0.000			Very strong
First evaluation Speak Out Module					0.974	0.000	Very strong

TABLE 5 | Wilcoxon Signed Rank exams for pre and post exams for achievement scores.

Student's achievement	*N*	*M*	*SD*	Median	Z	Sig
Pre	14	25.86	6.225	24.5	−3.297	0.001
Post	14	46.00	10.735	50		

TABLE 6 | Pearson's correlation coefficient index r Computational Thinking as a problem solving.

Relation	Second evaluation computational thinking as a problem solving		Interpretation
	r	Sig	
First Evaluation Computational Thinking as a problem solving	0.995	0.000	Very strong

TABLE 7 | Pearson's correlation coefficient index r metacognitive awareness.

Relation	Second evaluation metacognitive awareness		Interpretation
	r	Sig	
First Evaluation Metacognitive Awareness	0.900	0.000	Very strong

rank = 0). Descriptive statistical data mean achievement test score (pre) for the pilot group was recorded as mean = 25.86 which is less than the mean of the post-test of the pilot group which is mean = 46.00. To determine the reliability of the computational thinking questionnaire instrument as a problem solution, Pearson's r correlation coefficient index was used with 2 assessments. The first assessment was on the day of the pilot study while the second assessment was conducted two weeks after answering the first assessment, **Table 6** shows the Pearson's r coefficient index of Computational Thinking as Problem Solving.

There was a significant relationship between computational thinking as a solution to the first assessment problem with the second assessment of students with a value of $r = 0.995$ sig = 0.000 ($p < 0.005$). Its strength is 0.995 which indicates a very strong relationship. Relationship strength is based on the Relationship Strength Scale Cohen et al. (2002). While **Table 7** shows a significant relationship between metacognitive awareness of the first assessment with the second assessment of students with a value of $r = 0.900$ sig = 0.000 ($p < 0.005$). Its strength is 0.900 which indicates a very strong relationship.

TABLE 8 | Pearson's correlation coefficient index ME-CoT Module.

Relation	Second Evaluation Printed ME-CoT		Interpretation
	r	Sig	
First Evaluation Printed ME-CoT	0.772	0.001	Strong

The findings of the pilot study showed that there was a noticeable effect of change before and after using the ME-CoT module. Despite very slight changes in achievement in test scores, the use of the ME-CoT Module can be seen to be effective. Meanwhile, a very strong relationship can be seen among students in terms of fostering computational thinking skills and metacognitive awareness.

In this study, the researcher used assessment test questions to analyze student achievement, while computational thinking as a problem-solving instrument (Yağcı, 2019) and Metacognitive Awareness Inventory (Schraw and Dennison, 1994; Harrison and Vallin, 2018) were used to analyze students' computational thinking and metacognitive awareness. The study's findings revealed that students who utilized the ME-CoT module performed better academically. Meanwhile, a substantial correlation demonstrates that computational thinking skills and metacognitive awareness are both very reliable. In the actual study, the development of computational thinking abilities and metacognitive awareness will be examined in depth.

Printed Metacognitive Empowerment by Computational Thinking Module

In addition to the Visual Studio Software Module, the printed material ME-CoT Module was also evaluated for its overall reliability index. **Table 8** shows the Pearson's Correlation Coefficient Index for the ME-CoT Module. The findings of the pilot study showed a significant relationship between the Printed ME-CoT module of the first assessment with the second assessment of students with a value of $r = 0.772$ Sig = 0.001 ($p < 0.005$). A value of Pearson's correlation coefficient r is less than 0.80 is not recommended. Nevertheless, a reading of 0.772 is reading above 0.70 acceptable stability (Cohen et al., 2002; Hertzog, 2008).

Meanwhile, students who used the ME-CoT module also displayed feelings of happiness and fun when using the ME-CoT Module in Lessons. Students also stated that they enjoyed listening to their voices as well as their explanations on the topics

TABLE 9 | Teacher's views and suggestions.

Criteria	Teacher's views and suggestions
Advantages	• The Me-COT module is appealing and simple to operate. Contains colorful images that draw pupils' attention. • Students can master the material of the course because they construct their questions, answers, and explanations. • Group activities are a lot of fun for students. Students may create their activity items and are delighted to see their percentage marks.
Weakness	• Students take a long time to grasp the concept of programming in the early stages of introduction; nevertheless, a quick video presentation on each step of programming helps students master the processes of programming.
Suggestions for improvement	• Please include a topic header on each page for easy reference. • Students can be provided a video presentation of the visual studio programming module's essential material, which can also be attached in softcopy form. • In the module, provide a student information page. Students can save time by using the module if appropriate photos and sizes are provided in one folder.

given which automatically increased students' self-confidence in learning Biology. Meanwhile, not only does it help self-learning, but the ME-CoT Module also helps peers to prepare activity products to assess peers' mastery of the units studied also indirectly attracts the attention of students. The findings of the pilot study showed that there was a noticeable effect of change before and after using the ME-CoT module. despite very slight changes in achievement in test scores, the use of the ME-CoT Module can be seen to be effective. Meanwhile, a very strong relationship can be seen among students in terms of fostering computational thinking skills and metacognitive awareness.

Views and Suggestions of Biology Teachers

The ME-CoT Module was tested at a school under the auspices of PPD Jempol and Jelebu on a Form 4 Biology teacher and 14 students. The Me-CoT module was implemented in the classroom by the Biology teacher. The ME-CoT Module is implemented by Biology teachers under the supervision of researchers. The opinions and ideas of teachers are solicited and studied to develop the ME-CoT module. **Table 9** illustrates the opinions and ideas of Biology teachers who helped with the ME-CoT Module usability test.

Adding a heading to each page sheet is one of the enhancement ideas that help students when using the ME-CoT module. Opening and searching pages from a list of contents take a lengthy time for students. Meanwhile, students would be able to spot problems more readily if module usage videos are distributed to them in softcopy form. Aside from that, the teacher had pointed out a page for students to record their names in the module, which should be added to the students' module. In addition, an image folder and a video folder were proposed to be added to the module to assist students in making the most use of the ME-CoT Module. This module, according to the Biology teacher, is very interesting not only because it encourages students to master the content of Biology, but also because it helps students pay attention to both the subject of Biology and the field of programming, which has the potential to produce high-quality and interesting learning products.

The ME-CoT module is developed based on the principles interlinked between the three main Learning Theories, namely, Cognitive Learning Theory, Social Learning Theory, and Constructivist learning Theory. If traced, the ME-CoT module was developed by combining four learning theories namely; Robert Gagne's Information Processor Theory, Metacognitive Theory, Vygotsky's Social Theory of Constructivism, and Theory of Constructivism. As proposed in the Theory of constructionism, the result of the ME-CoT module is three products consisting of two activity products and one presentation product. Meanwhile, learning by making mentioned by Papert in Theory of constructionism can be seen clearly in this study. This study produces an output (activity product or presentation process) in the form of an assignment in technology media through programming activity (C#programming language). This can show the effectiveness of the Me-CoT module in integrating computational thinking into learning. Students 'thinking can be constructed with systematically arranged problem-solving activities. In addition, students' thinking can be guided through the production of new products using technological tools (Kafai and Burke, 2014). Besides that, this study emphasizes the need for coding is not for making unnecessary products, but the product created should have value such as creating games, stories, and animations that can be shared with others (Kafai and Burke, 2014). With this, the Theory of constructionism which focuses on the formation of products through programming activities that develop computational thinking is accurately classified as training students to think, as well as fostering metacognitive awareness in students.

Furthermore, the metacognitive theory is closely related to thinking about thinking (Flavell, 1979; Sun, 2013; Mazli Sham and Saemah, 2014; Astuti et al., 2017). Thinking involves a cognitive activity which is a change in mental activity that develops according to the level and ability of students. Metacognitive awareness guides students to think from a low level to a high level. This thinking also guides students in improving achievement (Mazli Sham and Saemah, 2014; Mohamad Masrizan, 2019). In this study, students plan each activity by setting a time and finding information. Next, students begin to monitor the activities they have done or review their assignments. After monitoring, students will make an assessment. This process is continued with the application of information management that has been obtained. According to Metacognitive Theory, students begin to think of identifying weaknesses and finding solutions then students have begun to seek help from teachers or peers. As students are exposed to

new information students begin to analyze their intellectual strengths and weaknesses, especially in this study the application of programming can guide students to focus on the programming component. Meanwhile, appropriate strategies are highly focused by students on ensuring that learning takes place in students. At the same time, students can master student content more efficiently as well leads to improvement in achievement. Thus the ME-CoT Module does not deviate from the context emphasized in Vygotsky's Theory of Social Constructivism, where students have exposure to Biology content topics since primary school and during lower secondary, while for computer science, students are sensitive about image size and resizing image since lower secondary. Students are also good at using computers with basic information. However, programming using the C# programming language is new information introduced in a very simple and easy way through the ME-CoT Module. Vygotsky's Social Theory of Constructivism Learning experiences can be built-in in students.

Meanwhile, the third theory namely Vygotsky's Theory of Social Constructivism is one of the most innovative learning theories in the 21st century (Kozulin, 2015) at the point of emphasis in the production of the ME-CoT Module. The interaction of historical, social, and personal experiences forming a psychological consciousness is an important foundation in Vygotsky's Social Theory of Constructivism (Vygotsky, 1979b). Social interaction is very important in the ME-CoT module. Where students interact with each other in groups and carry out group activities to produce activity products or presentation products. In addition, guidance from teachers or peers is also preferred, whereas students who have difficulty in solving a problem need guidance from teachers or peers (Pressley et al., 1985). Furthermore, the selection of the programming language also played a role in this study. To ensure that students do not experience cognitive load, the ME-CoT module using a programming language is in line with Vygotsky's Social Theory of Constructivism where new concepts are developed in the Proximal Development Zone. Based on the study by Stripeikaitė (2017) showed that students who use C-Syntax master programming in more depth. In addition, students who are exposed to C-Syntax programming (such as C ++, Java, and C#) can master the programming language more easily compared to students who learn block-based programming such as Scratch (Stripeikaitė, 2017). Furthermore, according to Krpan et al. (2017), the use of C# and Python languages has been widely introduced among primary and secondary school students. Both programming languages are very popular and have the advantage of being transferred or expanded to other languages (Krpan et al., 2017). Therefore, the C# programming language used in this study is flexible and has been introduced since primary school abroad (Lin and Weintrop, 2021), so it is undeniable that the C# programming language is very appropriate and is in the Proximal Development Zone proposed in Vygotsky's Social Theory of Constructivism.

In addition, Robert Gagne's Information Processor Theory is one of the theories that formed the basis of the development of this ME-CoT module because the principles and laws of the theory that emphasize the cognitive development of a student. Learning is closely related to the changes that occur in the information available in the memory of students. Students 'memory develops in line with the development of students' metacognitive awareness because this learning theory involves sensory memory and long-term memory. Meanwhile, technological tools such as laptops and Visual Studio software modules used in enhancing long-term memory are closely related to activity product production activities and presentation products produced by students through each level of problem-solving suggested in computational thinking skills. Meanwhile, each skill trains students to think of ways to solve problems systematically while fostering metacognitive awareness in the students themselves. When students try to understand a problem to find a solution then students use long-term memory that exhibits the occurrence of the learning process. The use of technological tools (C#programming) guides computational thinking and ensures that the learning process occurs through metacognitive awareness and in turn plays a role in improving student achievement.

LIMITATIONS, FUTURE DIRECTION, AND RECOMMENDATION

This study, like any other, has limitations. To begin with, this research is only focused on problem-solving computational thinking skills. As a result, the ME-CoT Module was built around six main problem-solving computational thinking skills. Those problem-solving computational skills have a close correlation to the construct of metacognitive awareness, and they help students develop metacognitive awareness. The ME-CoT is based on the respiratory topic, which has been identified as one of the most crucial topics for students to master to perform well on the general Biology Exam. Abstraction, decomposition, algorithmics, pattern recognition, modeling and simulation, and debugging are some of the skills. Second, this research is limited to a particular topic that was chosen via a need analysis, and it covers 34 biology students and 10 biology teachers. Finally, because of the COVID-19 pandemic's impact and the SOP that must be followed while doing research in schools, especially when it includes children, this study was done with a limited number of students.

There are multiple areas for further research that stem from this study. Increasing the number of problem-solving computational skills will be helpful to see the impact of ME-CoT implementation on students' problem-solving skills through computational skills. Besides that, not only metacognitive awareness but students' interest in learning the specific topic should be focused on to identify whether this ME-CoT module arouses pupils' interest in learning Biology. Furthermore, the goal of this study is to see the students' achievement in Biology education, thereby, other topics related to human and animal physiology can be focused on. Finally, the restriction on conducting research in school during the COVID-19 pandemic limited the number of students for this study.

Multiple challenges arose from implementing a curriculum plan to merge STEM education (Biology education) with computer science (computational thinking) to build an interdisciplinary approach. A well-organized learning module

with relevant learning goals should be constructed for long-term implementation. Based on the four primary educational theories, this study describes the design and implementation approach for incorporating text-based programming ME-CoT learning Module into current Biology Syllabuses. This study will add some valuable information regarding implementing coding using text-based learning which remains unclear (Papadakis, 2022). We aimed to further experiment to determine the effectiveness of the ME-CoT Learning Module associated with student achievement in Biology, computational thinking, and metacognitive awareness.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ministry of Education, Malaysia. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

NM contributed in the module development, experimentation, and analysis of findings. KO and LH contributed significantly in the theoretical aspect of module development as well discussion of findings and future direction. All authors contributed to the article and approved the submitted version.

FUNDING

The open access publication of this article was supported by UKM (TAP-K006677).

ACKNOWLEDGMENTS

We would like to thank the Ministry of Education, Malaysia for giving us this golden opportunity to conduct research in schools, although there were various constraints from the impact of the COVID-19 pandemic that occurred in Malaysia. Besides that, we would like to thank all participating Negeri Sembilan State Education Department, Biology Teachers, and students for their time in completing this research.

REFERENCES

Ackermann, E. (2001). Piaget's constructivism, Papert's constructionism: what's the difference? *Future Learn. Group Publ.* 5:438. doi: 10.1111/j.1526-4610.2005.t01-1-05013.x

Aho, A. V. (2012). Computation and computational thinking. *Comput. J.* 55, 832–835. doi: 10.1093/comjnl/bxs074

Anderton, R. S., Evans, T., and Chivers, P. T. (2016). Predicting academic success of health science students for first year anatomy and physiology. *Int. J. High. Educ.* 5, 250–260. doi: 10.5430/ijhe.v5n1p250

Angeli, C., and Jaipal-Jamani, K. (2018). "Preparing pre-service teachers to promote computational thinking in school classrooms," in *Computational Thinking in the Stem Disciplines* (Cham: Springer), 127–150.

Arihasnida, A., Nurazmiera, M. R., Norhasyimah, H., Tamil Selvan, S., and Siti Nur Kamariah, R. (2017). Tahap kemahiran visualisasi bagi mata pelajaran lukisan kejuruteraan di uthm. *J. TVET Practit.* 2, 1–8.

Astuti, D., Khasanah, U., and Romadon, S. (2017). *The Effect of Thinking Empowerment by Questioning (TEQ) through Team-Assisted Individualization (TAI) Learning Model to Metacognitive and Learning Achievement*. Cham: Springer.

Bergan-Roller, H. E., Galt, N. J., Chizinski, C. J., Helikar, T., and Dauer, J. T. (2018). Simulated computational model lesson improves foundational systems thinking skills and conceptual knowledge in biology students. *BioScience* 68, 612–621. doi: 10.1093/biosci/biy054

Bevins, S., and Price, G. (2016). Reconceptualising inquiry in science education. *Int. J. Sci. Educ.* 38, 17–29. doi: 10.1080/09500693.2015.1124300

Burbaite, R., Drasute, V., and Stuikys, V. V. (2018). Integration of computational thinking skills in stem-driven computer science education. *Glob. Eng. Educ. Conf.* 2, 1824–1832. doi: 10.1109/EDUCON.2018.8363456

Çakiroğlu, Ü, and Er, B. (2020). Effect of using metacognitive strategies to enhance programming performances. *Inform. Educ.* 19, 181–200. doi: 10.15388/INFEDU.2020.09

Cheah, H. M. (2016). Enhancing creative teaching using computational thinking. *Int. Conf. Teach. Learn.* 2016, 26–42.

Chou, M. H. (2017). A task-based language teaching approach to developing metacognitive strategies for listening comprehension. *Int. J. Listen.* 31, 51–70. doi: 10.1080/10904018.2015.1098542

Çimer, A. (2012). What makes biology learning difficult and effective: students' views. *Educ. Res. Rev.* 7, 61–71. doi: 10.5897/ERR11.205

Cohen, L., Manion, L., and Morrison, K. (2002). *Research Methods in Education*. Abingdon: Routledge.

del Olmo, J., Cózar-Gutiérrez, R., González-Calero, J. A., del Olmo-Muñoz, J., Cózar-Gutiérrez, R., and González-Calero, J. A. (2020). Computational thinking through unplugged activities in early years of Primary Education. *Comput. Educ.* 150:103832. doi: 10.1016/j.compedu.2020.103832

Falloon, G. (2016). An analysis of young students' thinking when completing basic coding tasks using Scratch Jnr. On the iPad. *J. Comput. Assist. Learn.* 32, 576–593. doi: 10.1111/jcal.12155

Fazilah, R., Othman, T., and Azraai, O. (2016). Aplikasi kemahiran proses sains dalam pembelajaran berasaskan masalah untuk matapelajaran biologi. *J. Kurikulum Pengajaran Asia Pasifik* 4, 38–46.

Flavell, J. (1979). Metacognition and cognitive monitoring: a new area of cognitive - development inquiry. *Notes Queries* 34, 906–911. doi: 10.1093/nq/CLVII.dec14.424-a

Gagne, R. M. (1970). Learning theory, educational media, and individualized instruction. *Learn. Theory* 22.

Gillott, L., Joyce-Gibbons, A., and Hidson, E. (2020). Exploring and comparing computational thinking skills in students who take GCSE Computer Science and those who do not. *Int. J. Comput. Sci. Educ. Sch.* 3, 3–22. doi: 10.21585/ijcses.v3i4.77

Goyal, S., Vijay, R. S., Monga, C., and Kalita, P. (2016). "Code bits: an inexpensive tangible computational thinking toolkit for K-12 curriculum," in *TEI 2016 - Proceedings of the 10th Anniversary Conference on Tangible Embedded and Embodied Interaction*, (New York, NY: ACM), 441–447.

Harrison, G. M., and Vallin, L. M. (2018). Evaluating the metacognitive awareness inventory using empirical factor-structure evidence. *Metacogn. Learn.* 13, 15–38. doi: 10.1007/s11409-017-9176-z

Haslina, H., Aliza, A., Lye Guet, P., and Mohd Saferi, M. H. (2018). *Using Scratch Programming To Engage Primary School Penggunaan Scratch Programming*

Untuk Meningkatkan Penglibatan Murid Sekolah Rendah. Bangi: Universiti Kebangsaan Malaysia, 49–57.

Hertzog, M. A. (2008). Considerations in determining sample size for pilot studies. *Res. Nurs. Health* 31, 180–191. doi: 10.1002/nur

Higgs, L. (2013). Theory in educational research and practice in teacher education. *Bulg. Comp. Educ. Stud.* 11, 105–111.

Isaac, S., and Michael, W. B. (1995). *Handbook in Research and Evaluation: A Collection of Principles, Methods, and Strategies Useful in the Planning, Design, and Evaluation of Studies in Education and the Behavioral Sciences*. Princeton, NJ: EdITS publishers.

Jamaludin, A. (2008). *Modul dan Pengendalian bimbingan Kelompok*. Serdang: Universiti Putra Malaysia.

Jamaludin, A. (2012). *Modul Motivasi Diri*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Kafai, Y. B., and Burke, Q. (2014). *Connected Code: Why Children Need to Learn Programming*. Cambridge, MA: MIT Press.

Kalelioglu, F., Gulbahar, Y., and Kukul, V. (2016). A framework for computational thinking based on a systematic research review. *Baltic J. Modern Comput.* 4, 583–596.

Kamisah, O. (2022). *Contextualizing Computational Thinking Disposition Framework From an Affective Perspective*. Bangi: Universiti Kebangsaan Malaysia, 390–412.

Karimah, M., Noraidah, S. A., Tengku Siti Meriam, T., and Noorazean, M. A. (2021). Validation of the components and elements of computational thinking for teaching and learning programming using the fuzzy delphi method. *Int. J. Adv. Comput. Sci. Appl.* 12, 80–88. doi: 10.14569/IJACSA.2021.0120111

Karimah, M., Noraidah Sahari, A., Tengku Siti Meriam, W. T., and Noorazean, M. (2020). Analysis of the requirements of computational thinking skills to overcome the difficulties in learning programming. *Int. J. Adv. Comput. Sci. Appl.* 11, 244–253. doi: 10.14569/ijacsa.2020.0110329

Kaufmann, O. T., and Stenseth, B. (2021). Programming in mathematics education. *Int. J. Math. Educ. Sci. Technol.* 52, 1029–1048. doi: 10.1080/0020739X.2020.1736349

Kazimoglu, C., Kiernan, M., Bacon, L., and MacKinnon, L. (2012). Learning programming at the computational thinking level via digital game-play. *Proc. Comput. Sci.* 9, 522–531. doi: 10.1016/j.procs.2012.04.056

Koc, I., and Kuvac, M. (2016). Preservice science teachers' metacognitive awareness levels. *Eur. J. Educ. Stud.* 0, 43–63.

Kotsopoulos, D., Floyd, L., Khan, S., Namukasa, I. K., Somanath, S., Weber, J., et al. (2017). A pedagogical framework for computational thinking. *Digital Exp. Math. Educ.* 3, 154–171. doi: 10.1007/s40751-017-0031-2

Kozulin, A. (2015). Vygotsky's theory of cognitive development. *Int. Encyclopedia Soc. Behav. Sci.* 25, 322–328. doi: 10.1016/j.nepr.2019.03.010

KPM (2016). *Buku Penerangan Kurikulum Standard Sekolah Menengah (KSSM)*. Putrajaya: Bahagian Pembangunan Kurikulum.

KPM (2017). *Laporan Kebangsaan TIMSS 2015*. Putrajaya: KPM, 1–112.

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy an overview. *Am. J. Psychol.* 41, 219–225. doi: 10.1207/s15430421tip4104

Krpan, D., Mladenovic, S., and Zaharija, G. (2017). "Mediated transfer from visual to high-level programming language," in *Proceedings of the 2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2017*, (Rijeka: MIPRO), 800–805.

Kusuma, M. D., Rosidin, U., Abdurrahman, A., and Suyatna, A. (2017). The development of higher order thinking skill (Hots) instrument assessment in physics Study. *IOSR J. Res. Method Educ.* 07, 26–32. doi: 10.9790/7388-0701052632

Kyairaniah, A. H., Mohd Isa, H., and Lubis, M. A. (2017). Tahap kemahiran metakognitif murid sekolah menengah di kawasan felda dalam pembelajaran pendidikan Islam. *ASEAN Comp. Educ. Res. J. Islam Civil.* 1, 94–106.

Lay Ah Nam, and Kamisah, O. (2017). Developing 21st century skills through a constructivist-constructionist learning environment. *STEM Educ.* 3, 205–216.

Li, Y., Schoenfeld, A. H., diSessa, A. A., Graesser, A. C., Benson, L. C., English, L. D., et al. (2020). Computational thinking is more about thinking than computing. *J. STEM Educ. Res.* 3, 1–18. doi: 10.1007/s41979-020-00030-2

Lin, Y., and Weintrop, D. (2021). The landscape of Block-based programming: characteristics of block-based environments and how they support the transition to text-based programming. *J. Comput. Lang.* 67:101075. doi: 10.1016/j.cola.2021.101075

Lynne, C. (2008). PS1 pilot studies. *Medsurg Nurs.* 17:411.

Maharani, S., Kholid, M. N., Pradana, L. N., and Nusantara, T. (2019). Problem solving in the context of computational thinking. *Infinity J.* 8:109. doi: 10.22460/infinity.v8i2.p109-116

Mazli Sham, A., and Saemah, R. (2014). Gaya pembelajaran dan kesedaran metakognitif dalam kalangan pelajar aliran sains. *Proc. Sosial Sci. Res.* 12, 712–719.

Mohamad Masrizan, A. P. (2019). Exploring metacognition regulation among biology student at labuan matriculation college mohamad. *J. Kurikulum* 53, 1–16.

Mohd Noor, A. (2012). *Redesigning Classroom Pedagogy*. Singapore: National Institute of Education Singapore.

Mohd Sidek, N., and Jamaludin, A. (2005). *Pembinaan Modul: Bagaimana Membina Modul Latihan Dan Modul Akademik*. Serdang: Universiti Putra Malaysia.

Odom, L. R., and Morrow, J. R. (2006). What's this r? A correlational approach to explaining validity, reliability, and objectivity coefficients. *Meas. Phys. Educ. Exerc. Sci.* 10, 137–145. doi: 10.1207/s15327841mpee1002_5

Oluk, A., and Korkmaz, Ö (2016). Comparing students' scratch skills with their computational thinking skills in terms of different variables. *Int. J. Modern Educ. Comput. Sci.* 8, 1–7. doi: 10.5815/ijmecs.2016.11.01

Otterborn, A., Schönborn, K. J., and Hultén, M. (2020). Investigating preschool educators' implementation of computer programming in their teaching practice. *Early Childhood Educ. J.* 48, 253–262. doi: 10.1007/s10643-019-00976-y

Papadakis, S. (2022). Can preschoolers learn computational thinking and coding skills with scratchjr? A systematic literature review. *Int. J. Educ. Reform* 2022:10567879221076077.

Papadakis, S., and Kalogiannakis, M. (2019). "Evaluating a course for teaching advanced programming concepts with scratch to preservice kindergarten teachers: a case study in Greece," in *Early Childhood Education*, eds D. Farland-Smith (Hoboken, NJ: Wiley).

Papert, S. (1996). An exploration in the space of mathematics educations. *Int. J. Comput. Math. Learn.* 1, 95–123. doi: 10.1007/BF00191473

Pedaste, M., Mäeots, M., Siiman, L. A., de Jong, T., van Riesen, S. A. N., Kamp, E. T., et al. (2015). Phases of inquiry-based learning: definitions and the inquiry cycle. *Educ. Res. Rev.* 14, 47–61. doi: 10.1016/j.edurev.2015.02.003

Periya, S. N., and Moro, C. (2019). Applied learning of anatomy and physiology: virtual dissectiontables within medical and health sciences education. *Bangkok Med. J.* 15, 121–127. doi: 10.31524/bkkmedj.2019.02.021

Piaget, J. (1972a). Intellectual evolution from adolescence to adulthood. *Hum. Dev.* 15, 1–12. doi: 10.1159/000271225

Piaget, J. (1972b). The role of imitation in the development of representational thought. *Int. J. Ment. Health* 1, 67–74. doi: 10.1080/00207411.1972.11448598

Ping, I. L. L., Halim, L., and Osman, K. (2020). Explicit teaching of scientific argumentation as an approach to developing argumentation skills, science process skills, and biology understanding. *J. Baltic Sci. Educ.* 19, 276–288. doi: 10.33225/jbse/20.19.276

Pressley, F. D., MacKinnon, G., and Waller, T. G. (1985). *Metacognition, Cognition, and Human Performance*. Cambridge, MA: Academic Press.

Puganesri, K., and Puteh, S. (2019). Computer science education in Malaysia schools: the challenges of enhancing computational thinking skills. *Int. J. Eng. Adv. Technol.* 8, 441–444. doi: 10.35940/ijeat.F1080.0986S319

Ragbir Kaur Joginder Singh (2011). *Panduan Ilmu Pendidikan Komprehensif unutk KPLI(Sekolah Rendah)*. Selangor: KUMPULAN BUDIMAN SDN.

Ramlan, M., and Ghazali, D. (2018). *Aplikasi Kaedah Fuzzy Delphi Penyelidikan Sains Sosial*. Kuala Lumpur: Penerbit Universiti Malaya.

Reinoso Tapia, R., Delgado-Iglesias, J., and Fernández, I. (2019). Learning difficulties, alternative conceptions, and misconceptions of student teachers about respiratory physiology. *Int. J. Sci. Educ.* 41, 2602–2625. doi: 10.1080/09500693.2019.1690177

Román-González, M., Moreno-León, J., and Robles, G. (2017). Complementary tools for computational thinking assessment. *Proc. Int. Conf. Comput. Think. Educ.* 17, 154–159.

Rubinstein, A., and Chor, B. (2014). Computational thinking in life science education. *PLoS Comput. Biol.* 10:e1003897. doi: 10.1371/journal.pcbi.1003897

Samri, C., Osman, K., and Nayan, N. A. (2020). Level of computational thinking skills among secondary. *Sci. Educ. Int.* 31, 159–163. doi: 10.33828/sei.v31.i2.4

Schraw, G., and Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemp. Educ. Psychol.* 19, 460–475. doi: 10.1006/ceps.1994.1033

Stripeikaitė, I. (2017). "Skipping the baby steps": the importance of teaching practical programming before programming theory. *Lecture Notes Comput. Sci.* 22, 319–330. doi: 10.1007/978-3-319-70111-0_30

Sun, L. (2013). The effect of metacognitive learning strategies on English learning. *Theory Pract. Lang. Stud.* 3, 2004–2009. doi: 10.4304/tpls.3.11.2004-2009

Susan, S., and Nurfaradilla, M. N. (2019). Teachers' concern towards applying computational thinking skills in teaching and learning. *Int. J. Acad. Res. Bus. Soc. Sci.* 9, 296–310. doi: 10.6007/ijarbss/v9-i1/5398

Tsarava, K., Leifheit, L., Ninaus, M., Román-González, M., Butz, M. V., Golle, J., et al. (2019). "Cognitive correlates of computational thinking," in *Proceedings of the 14th Workshop in Primary and Secondary Computing Education*, Glasgow, 1–9.

Vygotsky, L. S. (1979a). Consciousness as a problem in the psychology of behavior. *Sov. Psychol.* 17, 3–35. doi: 10.2753/rpo1061-040517043

Vygotsky, L. S. (1979b). The development of higher forms of attention in childhood. *Sov. Psychol.* 18, 67–115. doi: 10.2753/rpo1061-0405180167

Wing, J. (2011). *Research Notebook: Computational Thinking—What and Why? The Link Magazine*. Available online at: http://www.cs.cmu.edu/link/research-notebook-computational-thinking-what-and-why (accessed June 23, 2015).

Wing, J. M. (2006). Computational thinking. *Concurrences* 49, 33–35.

Yağcı, M. (2019). A valid and reliable tool for examining computational thinking skills. *Educ. Inf. Technol.* 24, 929–951. doi: 10.1007/s10639-018-9801-8

Zapata-Caceres, M., Martin-Barroso, E., and Roman-Gonzalez, M. (2020). "Computational thinking test for beginners: design and content validation," in *Proceedings of the IEEE Global Engineering Education Conference, EDUCON, 2020-April*, (Piscataway, NJ: IEEE), 1905–1914.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Markandan, Osman and Halim. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Psychology

ORIGINAL RESEARCH
published: 10 June 2022
doi: 10.3389/fpsyg.2022.870090

Developing Teaching Practice in Computational Thinking in Palestine

*Abdel Ghani[1], David Griffiths[2], Soheil Salha[1], Saida Affouneh[1], Fakher Khalili[1], Zuheir N. Khlaif[1] and Daniel Burgos[12]**

[1]Faculty of Educational Sciences, An-Najah National University, Nablus, Palestine, [2]Research Institute for Innovation & Technology in Education (UNIR iTED), Universidad Internacional De La Rioja, Logroño, Spain

Sporadic efforts have been made to introduce computational thinking methods into K-12 education in Palestine, but these have been held back by the challenging educational environment. However, a recent in-service training initiative, funded and organized by the Ministry or Education of Palestine, constitutes a significant effort to embed computational thinking in K-12 practice. The middle school teachers who participated in the training course were invited to participate in the present study, and 38 did so. A qualitative approach involving both interviews with teachers and classroom observations was used in data collection. All the teachers agreed to be observed in their classrooms, while 20 of the 38 also agreed to participate in the interviews. The findings showed that teachers of a range of topics, including social sciences and languages, employed computational thinking skills in teaching their students, but they were confronted by a number of challenges, including technical infrastructure and support, and a lack of time to prepare CT classes and space in the curriculum to deliver them. The results indicate that the most appropriate action to support teachers' delivery of CT would be to provide peer exchanges and expert coaching in the integration of CT in the curriculum.

Keywords: in-service training, computational thinking, pedagogy, thinking steps, challenges to adoption

OPEN ACCESS

Edited by:
Stamatios Papadakis, University of Crete, Greece

Reviewed by:
Elizabeth A. Boyle, University of the West of Scotland, United Kingdom
Sigrid Schefer-Wenzl, University of Applied Sciences Wien, Austria
Fan Ouyang, Zhejiang University, China

***Correspondence:**
Daniel Burgos
daniel.burgos@unir.net

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

Received: *05 February 2022*
Accepted: *16 May 2022*
Published: *10 June 2022*

Citation:
Ghani A, Griffiths D, Salha S, Affouneh S, Khalili F, Khlaif ZN and Burgos D (2022) Developing Teaching Practice in Computational Thinking in Palestine. Front. Psychol. 13:870090. doi: 10.3389/fpsyg.2022.870090

INTRODUCTION

This paper reports on the challenges experienced in Palestine in the adoption of computational thinking (CT) methods in K-12 schools, and on an intervention made by the Ministry of Education of Palestine (MoE) to address these challenges. In doing this, it provides insight into a previously invisible body of CT practice in a country living with conflict and economic insecurity, and an exploration of how the challenges of implementing CT can be addressed in such a context.

In alignment with the worldwide trend to the inclusion of CT in the curriculum, many schools and teachers in Palestine have introduced elements of computational thinking concepts and skills. This has frequently been in the absence of any official program, often through participation in national and international programming competitions. Indeed, the teachers who were responsible for this practice were often unaware that the activities that they were carrying out could be classified as CT.

Teachers, schools, and the MoE in Palestine are all keen to contribute to the development of CT skills in the classroom but are faced with two problems. Firstly, there is a lack of practical guidance on how CT can best be implemented with K-12 learners in different curricular

areas. Secondly, schools, teachers, and learners vary greatly around the world, and there is a lack of studies in diverse K-12 educational contexts, making it challenging to develop appropriately focused policies, to provide effective support for teachers, or to share practice. As a result, teachers lack guidance on how to approach the teaching of CT in their own context, and this difficulty is often compounded by their misconceptions of what CT involves.

In response to these challenges, in 2019 the MoE established a training program to support and extend emerging CT practice in K-12 schools, with a focus primarily on STEM subjects and language classes, with a much smaller participation from social sciences teachers. The research reported here was carried out within the context of this intervention, taking advantage of the opportunity to collect valuable data about CT practice. It aims to provide insight into the challenges faced by teachers of CT in K-12 schools in Palestine, the support which is provided for them in designing CT strategies in their teaching, and the strategies that they have adopted in using CT approaches in the classroom.

LITERATURE REVIEW

The practice of the authors in developing support for teachers in the adoption of CT approaches has been informed by ongoing engagement with the literature. In literature review, we summarize the publications which have guided our practice and research. We outline the concepts and approaches to CT which have been merged in our research framework and identify the gaps which we address in this paper.

Approaches to Computational Thinking

The origins of Computational Thinking (CT) in primary and secondary schools have been discussed in many publications and can only be addressed briefly here. The term was introduced by Papert (1980) but became prominent following influential interventions by Wing (2006, 2011). She proposed that in higher education it would benefit everyone to learn how to think like a computer scientist and made a recommendation to "expose pre-college students to computational methods and models" (Wing, 2006, p. 35). Accordingly, she suggested that "to reading, writing, and arithmetic, we should add computational thinking to every child's analytical ability" (Wing, 2006, p. 33). Wing (2011) defined computational thinking a "the thought processes involved in formulating problems and their solutions so that the solutions are represented in a form that can be effectively carried out by an information-processing agent."

Wing's proposals became increasingly prominent and were supported by the National Science Foundation in the United States, which provided substantial funding to implement the approach in schools. In Europe, similar priorities were emerging at the time of Wing's initial paper, and the Key Competences for Lifelong Learning set out that.

Individuals should have skills to use tools to produce, present and understand complex information and the ability to access, search and use internet-based services. Individuals should also be able use IST (Information Society Technologies) to support critical thinking, creativity, and innovation (European Council, 2006, p. 1).

This concern with digital skills, and within that category computational thinking, was in large part driven by the concern that countries would be unable to compete as the digital economy unfolded, as acknowledged by the European Commission: "skills will to a great extent determine competitiveness … Many sectors are undergoing rapid technological change and digital skills are needed for all jobs" (European Commission, 2016). In support of this strategy, studies have indicated that CT activities in the classroom can increase students' self-efficacy in both CT and STEM (Feldhausen et al., 2018; Kwon et al., 2019).

Over a number of years, work has been underway to provide evidence of the benefits of the use of CT in K-12 classrooms. For example, Israel et al. (2015, p.273) found that "struggling learners generally thrived in the computing environments," while Hava and Ünlü (2021) have identified a correlation between the middle school students' computational thinking skills and their STEM career interest and attitudes toward inquiry. The documentation of the benefits of CT in such studies has led to widespread initiatives to include CT in the educational process and curriculum, with a concomitant need to provide appropriate teacher training. Early in the emergence of the CT approach, Barr and Stephenson (2011) identified a lack of the knowledge and skills required to integrate CT into school curricula. More recently, the implications of CT for teacher education were reviewed by Yadav et al. (2017). Work has also been carried out to identify appropriate pedagogical approaches (Yadav et al., 2016), and studies of teacher education in CT have been carried out, for example by Hirsh and Baronak (2020) in the context of pre-service training, and by Rich et al. (2021) in relation to continuous professional development.

As described above, Wing's original formulation of CT involves "thinking like a computer scientist," and so, it is perhaps unsurprising that the terms CT and Computer Science (CS) are often used interchangeably in both theory and practice. However, it is necessary to distinguish them in order to provide guidance for teachers. This distinction has proved hard to finalize, and this has given rise to the many definitions in the literature. An overview of some of the definitions and approaches that have been proposed are available and Bocconi et al. (2016). An influential recent formulation is that of Denning and Tedre (2019):

Computational thinking is the mental skills and practices for:

- Designing computations that get computers to do jobs for us.
- Explaining and interpreting the world as a complex of information processes (Denning and Tedre, 2019, p. 4).

To support teachers in the task of supporting children's learning, a many specialized computer applications has been developed, including mobile and tablet-based apps. A long-standing example is Scratch and, more recently, Scratch Jr., and the impact of the latter is reviewed by Papadakis (2021). Papadakis (2022) also comments that "Many apps offering various programming lessons, puzzles, and challenges to teach core

coding concepts to children have increased in recent years," but adds that it is not yet clear what children learn from them. Many of these apps take a ludic approach to CT, as championed in Bers' book "Coding as Play" (Bers, 2020). An important class of applications is block-based programming environments enabling children to manipulate code through touch screens, Lin and Weintrop (2021) review work taking this approach and discuss its relationship with textual programming environments.

Theoretical Framework

The authors have worked over a number of years to encourage school students to apply to study STEM subjects at university level, with a particular emphasis on the participation of women. In the course of this practice, conceptions of CT have been identified which are valuable in addressing the challenges which are experienced in the Palestinian context, and these have come together to constitute our research framework. We make the broad distinction that CS is a discipline that focuses on the study and development of computers and algorithmic processes, whereas CT is a broader set of strategies that can help to solve problems in ways that computers can understand. More specifically, we follow the definition provided by the ISTE (2011) which operationalizes CT as a problem-solving process that is based on:

a. Formulating problems in a solvable way.
b. Organizing and analyzing data logically.
c. Use abstractions to represent data.
d. Automating solutions through algorithmic thinking.
e. Identifying, analyzing, and implementing possible solutions with the goal of achieving the most efficient and effective combination of steps and resources.
f. Generalizing and transferring this problem-solving process to a wide variety of problems.

From this perspective, the integration of CT into the curriculum is differentiated from teaching the discipline of computer science, in which learners are asked to follow the detailed processes of computation, and, in effect, to think like a computer when dealing with problems in any field (Grover and Pea, 2013). Nevertheless, many computer science principles can be included among CT concepts or processes, and, as a result, it is not a simple matter to tease out the implications for teaching practice of a decision to adopt a CT approach. In our research, we identify the following thinking steps as the components or processes of CT:

a. Algorithmic thinking: clearly specifying and arranging a sequence of steps to solve a problem, using mathematical, symbolic, logical, and textual expressions (Weese, 2017; Denning and Tedre, 2019, p.84).
b. Abstraction: a mental activity that isolates a specific characteristic of something on a conceptual level, while discarding information (Istikomah and Budiyanto, 2018; Park, 2019). Abstraction can also involve creation of a generalized representation or model of a complex problem (Weese, 2017).
c. Problem decomposition: the process of dividing a complex problem into small elements or subproblems, so that each element can be dealt with separately (Istikomah and Budiyanto, 2018; Rijke et al., 2018; Palts and Pedaste, 2020).
d. Data management: tasks include collection, representation, and analysis of the quantitative or qualitative values of variables related to a thing or phenomenon. Data management are generally facts or information that are collected for reference or analysis.
e. Parallelization: the "simultaneous processing of smaller tasks from a larger task to reach a common goal" (Sampson et al., 2018, p.196), which may also improve efficiency (Palts and Pedaste, 2020).
f. Control flow; the process determining the completion of the steps of an algorithm (Istikomah and Budiyanto, 2018), which may repeat specific steps many times, complete steps under certain conditions, skip steps, or stop a process before all steps are finished.
g. Visualization: the use of different representations, such as maps, photos, and drawings, to create and improve conceptual understanding among learners (Cetin and Andrews-Larson, 2016)

It is these thinking steps which we have sought to develop in our practice, and which we use as an analytical tool in examining the MoE summer school which is the focus of this study. Rather than seeing these thinking steps as discrete curricular items, we follow the approach recommended by Pears et al. (2019), who argue for seeing CT as an integrative element across the curriculum, rather than as a separate study. Given the practical challenges of defining the scope of CT in the classroom across the curriculum, we seek to bring simplicity and clarity to our aims and activities by providing support to teachers in developing their pedagogic skills in the three areas defined the framework set out by Brennan and Resnick (2012). "computational concepts (the concepts designers engage with as they program, such as iteration and parallelism), computational practices (the practices designers develop as they engage with the concepts, such as debugging projects or remixing others' work), and computational perspectives (the perspectives designers form about the world around them and about themselves)." In their framework, Brennan and Resnick were building on their work with the Scratch programming environment, and we have worked extensively with this environment in our training activities.

Gaps in Knowledge of Teacher Education and Computational Thinking

Since the emergence into prominence of CT, a great deal of research has been carried out to examine its impact on the way that students solve problems, specifically in mathematics and science subjects, for example, Yuen and Robbins (2014), Chen et al. (2017), Park (2019), and Hava and Ünlü (2021). However, Voogt et al. (2015) identified that relatively little attention had been paid to teachers' thinking, and to teacher education in CT. In recent years work, some work has been

carried out to fill this gap, with researchers starting to explore in-service and pre-service teachers' knowledge and their practices in including CT in classrooms.

In their review of 2019, Mason and Rich found that few studies had been carried out into the education of teachers at elementary level, or their impact:

(a) few studies have been published about training elementary school teachers to teach computing, coding, and programming, with slightly more studies on preservice teacher training than in-service PD; (b) interventions have focused more on developing elementary CS teachers' content knowledge than their pedagogical knowledge; (c) studies overwhelmingly showed that training can improve teachers' self-efficacy, attitudes, and knowledge, even over relatively short interventions; and (d) the literature has said little about whether or to what extent changes in self-efficacy, attitudes, and knowledge lead to changes in actual practice or improved student learning (Mason and Rich, 2019, p. 809).

The need for practical studies with teachers which can shed light on how they can be best supported is echoed by Mäkitalo et al. (2019). Writing with Matti Tedre (co-author of the influential book *Computational Thinking* (Denning and Tedre, 2019) whose definition of CT we discuss above), they put matters rather more strongly: "CT literature has pointed out a dire need for more research on understanding teacher learning and how to support their learning in the best way in pre- and in-service teacher education." They identify three problems that CT is facing, the third of which is a lack of understanding on "How do teachers learn to synthesize CT with existing content and pedagogical strategies?" Little substantial work has been carried out in this area, and in particular concerning in-service training, and the valuable studies that have been carried out, for example Kong et al. (2020), have only started to fill the gap.

The present research is a contribution toward the effort to fill these gaps in the knowledge of how to address teachers' needs for training and support in introducing CT, how they integrate CT into their practice. We also address another gap, which is largely unidentified in the research literature: to understand the challenges of adopting CT approaches in contexts which are quite different from the wealthy and powerful countries in which the approach was first conceived. In the following section, we document and describe this context.

The Educational Context

A Challenging Environment

Historically, the Palestinian education system has faced many challenges and difficulties, and for several decades the formal education has been administered and controlled by external authorities as a consequence of political instability. In 1967, Israel occupied the West Bank, including East Jerusalem, and the Gaza Strip. Israeli military authorities targeted the education system especially during the first intifada (1987–1992; Abo Hommos, 2013), including measures, such as closing down many schools and universities for more than 2 years, arresting or killing teachers or students, and imposing arbitrary measures to create poor learning environments (Ramahi, 2015). Israeli policies have had an enormous impact on Palestinian education, leading to a long-lasting decline in academic standards at all levels of education (Nicolai, 2007), and arbitrary Israeli policies have resulted in the weakening of the Palestinian education system in general. However, these Israeli policies and arbitrary measures did not entirely prevent Palestinians from preserving their cultural and historical identity. Palestinian teachers have taken it upon themselves to confront these policies; by adopting hidden curriculum, enhancing teaching methods, engaging in qualifying courses to improve their competencies, and employing informal education to create effective learning environments (Wahbeh, 2003).

After the Oslo Accords in 1993, the Palestinians became responsible for the Palestinian education system. Consequently, the MoE was established in 1994, and has sought to prepare Palestinian children for citizenship to build their state. Currently, the MoE administers about 75% of students in the West Bank, East Jerusalem, and the Gaza Strip. The United Nations Relief and Works Agency (UNRWA) administers 15%, while private institutions administer10%. The MoE has faced great difficulties and challenges in its work to rebuild and reform the Palestinian education system, including in providing schools with qualified teachers, improving the infrastructure, and designing modern educational curricula to meet the aspirations of the Palestinian people (Ramahi, 2015).

Teacher Education and Professional Development

In 1998, the MoE tried to standardize teacher qualifications, in order to enhance the professional capacity of the workforce. A bachelor's degree was designate as the minimum qualification for a teacher entering the profession (Nicolai, 2007), but this minimum requirement did not help in improving the education system and the learning environment (Bivins, 2015). The demand for certain types of teachers, specifically those proficient in STEM, was higher than the supply during that time (Bivins, 2015). In response, teachers with math and science experience were often hired without any teaching background at all (Nicolai, 2007). Ministry of Education of Palestine (2008) established a developed a Teacher Education Strategy, and this plan remains the current strategic plan and goals. In accordance with this plan, the MoE has made strenuous efforts to requalify teachers, and, 2010 it allocated 4.5% of its annual budget to develop teachers' competencies (UNESCO, 2010). It is therefore unsurprising that, again according to UNESCO (2010), traditional teaching strategies are still common and prominent in most Palestinian schools. These strategies include "teacher-centered" approaches, consisting of rote-learning, lecturing, and dictation. Moreover, Palestine has participated three times in the (TIMSS) study of international trends in mathematics and science for three times (2003, 2007, and 2011), and the results indicated that Palestine is still in the ranks of the lowest achievement countries (Afana, 2021), and these results are consistent with the rudimentary teaching strategies still used in these disciplines. Recent statistics are not available, but there is no reason to suppose that this negative situation has changed in recent years.

Recently Afana (2021) conducted a study of mathematical learning in Palestinian schools in challenging circumstances. According to this study, teaching loads are high, teachers'

teaching preparation is low, pre-service training does not exist, and in-service training is not effective, and universities' teaching programs are not practical but theoretical. The significant decrease in students' mathematics scores for students taught by teachers who studied mathematics education rather than pure mathematics was remarkable. At the same time the "Palestinian universities admission students for education with low final secondary exam scores, unlike medicine and engineering. Only few of them decide to study education with high scores because they like to become teachers" (Afana, 2021, p. 129).

Use of CT Strategies in K-12 Schools in Palestine

The historical and professional circumstances described in the previous sections have created an extremely challenging environment for pedagogic innovation, and as a result the widespread adoption CT by Palestinian teachers in their classrooms is still out of reach. It is therefore unsurprising that when we carried out a literature review of the integration of computational thinking by Palestinian teachers in their classrooms, we found no direct evidence. Nevertheless, there are indications that some progress has been made. The framework of the National curriculum asked teachers to develop thinking skills among students as observation, interpretation, and using various strategies to solve problems, and in 2019, the MoE initiated a training program in six of the 16 districts under its jurisdiction, to coordinate and extend this work. This initiative was the context for the research which is reported on in this paper.

RESEARCH DESIGN

Research Questions

The present researchers, some enthusiastic teachers, and the MoE of Palestine are convinced of the importance of integrating CT in schools as a means of improving educational achievement and as a contribution to the development of the economy. Indeed, the workshops which provided the opportunity for data gathering for this study are evidence of the commitment of the MoE. However, if CT is to be more widely adopted in Palestine, it is necessary to create a picture of the current situation which is both more detailed and more accurate. At present it is not clear to what extent CT has been established in existing teaching practice, what challenges teachers face in introducing CT in their lessons or what support would help them in overcoming those challenges. This knowledge is needed in order to provide guidance to schools in introducing CT approaches and to design appropriate policies and supports. In addressing this need, we provide information from the Palestinian context which helps to fill the gap identified by Mason and Rich (2019, p. 809) cited above, who note the lack of studies of the training of elementary school teachers in CT methods, and of the results of that training in terms of teachers' self-efficacy, attitudes, and knowledge.

In light of these needs, the following research questions were formulated:

1. How is computational thinking integrated into STEM and language classroom activities across different subject areas?
2. What types of support do teachers require to integrate computational thinking into STEM and language classroom instruction?
3. What pedagogical strategies do teachers report work well for using computational thinking approaches in teaching and learning?
4. What challenges do teachers face during the adoption of computational thinking in classrooms?

Population and Data Collection

The study took advantage of the opportunity to collect data that was presented by six summer workshops that provided CT training for teachers from K-12 public schools, organized by the MoE in cooperation with the AlNayzek non-profit organization. One workshop was organized in each of the districts of Ramallah, Nablus, Jenin, Tulkarem, Qalqilia, and Salfit. Each workshop had two facilitators: an educational supervisor, whose duties include assessing the performance of teachers in their classrooms, and an expert in computer science with experience in education. In each workshop, teachers met for 2 days per week for 4 hours for 2 months. The workshops included an introduction to computational thinking and modeling of computing tools and gave teachers the opportunity to practice designing activities to integrate CT into classroom activities. During the training sessions, the teachers talked about their experience with CT in their classrooms, and training was carried out through a learning by design approach which was designing activities to be used in the class. Permission was obtained from the MoE to visit these sessions and talk with teachers about their experience.

The teachers who attended in the summer workshops were contacted through the documentation and communication channels of the workshop organizers and invited to participate in the study. Most of the teachers taught a STEM subject, but the participants were extended to include Arabic, English, and social sciences. All the teachers had made a commitment to integrate CT into their instruction, and all were willing to participate in the workshops. One of the teachers had participated in national competitions for technological initiatives. It was therefore expected that the teachers would have their own ideas about CT skills prior to the workshops, even their fields were not related to computer science. The population of the study was constituted by the 235 teachers attending the workshops, who were all invited to participate in the study, and data collection was carried out through the activities of the workshops, and in interviews with participants carried out in parallel with the workshops.

Methods

The research approach taken in the research was selected to make the most of the unique opportunity for data collection offered by the training course. In a context where little is known of how teachers address CT in their practice, the highest priority was to obtain insight into teaching activities

directly from teachers, in the real-life contexts in which they occurred (Yin, 2003). Quantitative methods were not seen as being appropriate to the study, because the self-selecting nature of the population precluded extrapolation to teaching practice across Palestine. Rather, the focus was on understanding in greater depth the practice which was being conducted by teachers who were committed and enthusiastic practitioners of CT teaching. The principal method used in the research was therefore one-to-one interviews, which ensured that teachers were not intimidated by the presence of colleagues or supervisors in discussing their practice. In view of the fact that the researchers were seeking information about an unknown body of practice, a semi-structured method was adopted for the interviews, to ensure that teachers had plenty of opportunity to raise the issues which they considered to be important. Thematic analysis was used to analyze, categorize, and report on themes within the data. Teachers' introspection in interviews provides reliable evidence of the way in which teachers think about their practice and its purpose, but it may not be as reliable in understanding what actually occurs in classrooms during lessons, and the problems and successes which emerge. To address this, the interviews were supplemented with classroom observations carried out by the research team.

Semi-structured Interviews

The interview questions were developed to collect the data required to answer the research questions. The interviewer asked questions to explore teachers' experience of integrating CT approaches into their teaching different topics in K-12 settings. The identification of teachers was based on the following criteria: attendance at all the summer workshops sessions; the willingness to integrate CT into their instruction; participation in national competitions in technological initiatives; and previous use of CT in their classroom activities. The selection process also took into consideration the need to cover a range of subjects, with different backgrounds, in order to provide data on CT integration in different contexts. The researchers scheduled the interviews at times and locations convenient to the participants.

In the interviews teachers were asked to talk about their journey using CT in teaching their topic, using the following questions (translated from the Arabic).

- Could you please give us your opinion of the workshops that you have attended?
- What did you learn from the workshops regarding CT?
- How would you describe your experience of enhancing CT strategies in your teaching?
- Could you please describe the classroom activities related to CT that you have implemented in your teaching?
- What were the main challenges that you faced while implementing CT in your class?
- You have been trained on different strategies for CT, how will you use them in your class?
- Could you please give some examples of how you implemented CT in your class and in your subjects?

After transcription of the audio files, the researchers sent the text files to the participants to make sure of they truly represented the intention of their answers in the interviews. The participants were invited to edit/add/delete the text as they saw fit. After receiving their transcripts, none of the teachers changed their answers, but few of them did add more information about the training sessions and their experiences.

Classroom Observations

Classroom observations enabled the researchers to examine the strategies used by teachers from different backgrounds to integrate CT in real classes, and the procedures they used to implement CT classroom activities. In order to protect privacy, no photographs were taken that showed the faces of teachers or students. During the classroom observations, field notes were taken describing the instruction of the teachers and the activities of the students. The researchers observed and documented the strategies and procedures used to integrate CT, identified the challenges encountered, and noted how teachers overcame those challenges. All the field notes were shared with participants and other members of the research team to conduct data analysis.

Ethical Considerations

As noted above, the workshops were conducted by team of two, one of whom was an educational supervisor from the MoE. The supervisors' role is monitoring, advising, and supporting teachers in improving their performance. Moreover, the supervisors have considerable influence on the teachers' career development, and the supervisors visit all teachers twice each year to write reports about their performance. A positive report on teaching practice and outcomes can lead to promotion for teachers who use new technologies in teaching, and to nomination for professional development in educational strategies and the use of technology. It was therefore essential that none of the supervisors knew which teachers participated in the study. The research was discussed with the teachers at breaks in the workshop session, and the consent process was also completed at that time. The teachers were able to choose a preferred date and time to conduct the interview.

Data Analysis Procedures

Analysis started during the process of data collection through writing field notes describing what was observed by the researcher. These were compiled and contrasted with the results of the interviews in identifying the findings. The procedures set out in Marshall and Rossman (2011) were applied in the data analysis. A researcher transcribed the audio recorded files as soon as an interview was completed. The names of the interviewees were anonymized to protect the identity of the participants. The transcribed file was cleaned, and unnecessary information deleted before sending it to the interviewees for review. After finishing all the interviews and cleaning, two researchers read the transcripts

files in depth to identify the main ideas, concepts, and topics related to the research questions. A coding book was used as to guide the process of data analysis.

Two researchers conducted an exhaustive manual coding of the interview transcripts in order to identify CT concepts and skills. In this process, each researcher read through the transcripts and highlighted significant sections of text. In the analysis, we were only interested in studying CT concepts and skills used by practitioners in their work, and hence we focused on the portion of each interview where the interviewees described their current practice. The two researchers developed the themes and subthemes related to the research questions individually, then they met to discuss the themes and subthemes to achieve agreement on the final themes. Any disagreement appeared was solved by negotiation among the researchers.

RESULTS

The 235 teachers who attended in the CT workshops were invited to participate in the study, and of these, 38 gave their agreement. The teachers were all from different schools, from all six districts involved in the workshops. **Table 1** shows the number of male and female teachers participating in the experiment according to the type of subject they are studying. It shows that the number of teachers who were observed during the experiment was (38), of which (22) were female, and (16) were male. And the number of teachers interviewed was (20) male and female, males (10) and females (10).

The observed classes covered a range of topics taught in fifth to ninth grade classes in different schools. Over half of the participating teachers were female.

Selection of the interviewees was made through the facilitators of the workshops, who were asked to suggest four teachers at each of the six training locations, from the 38 who had declared their willingness to participate in the project. The individual interviews ranged between 20 and 30 min and were carried out in the training centers. Of the 20 teachers who were interviewed, half were female. The interview recordings were transcribed manually in Arabic, and sections were only translated into English if this was required for publication.

We now discuss the analysis of the data as it relates to our four research questions.

TABLE 1 | Teaching area and gender of the participants in the study.

Teaching topic	Observations		Interviews	
	Male	Female	Male	Female
Science	2	3	1	2
Math	3	4	2	1
Technology	4	6	3	4
English language	2	4	1	2
Arabic language	3	2	2	0
Social sciences	2	3	1	1
Total	**16**	**22**	**10**	**10**

Research Question 1: How Was Computational Thinking Integrated Into Classroom Activities Across Different Contexts?

Group Work and Classroom Seating

Most of the activities we observed were group work activities (30 out 45 activities) and 15 were individual activities. We did not find a specific pedagogical approach which was associated with CT. Rather, the way that the teachers in this study used CT was conditioned by their approach to organizing instruction in a class, for example employing whole-class instruction or a student-centered approach. Consequently, the levels of explicit instruction that they used while implementing CT activities varied. Several teachers from different schools included CT as a student-centered activity to teach math concepts using group work in their classrooms. Their instruction began with giving directions and demonstrations to the whole class, followed by individual and collaborative work. Teachers explained the various CT tasks to the students, who then worked on those tasks collaboratively and individually.

Most of the teachers we observed (30 out of 38) changed student seating arrangements to facilitate classroom activities with CT. Some teachers changed the seats into a U shape, others ensured that students worked in a group and the rest used whole-class (traditional classroom) instruction.

Flexible Lesson Plans

In our observations we noticed that half of the teachers adjusted the levels of instruction to meet the needs of subgroups within the classroom and enhance students' engagement in learning. They provided different activities for *beginners* (focused on collecting data), *novices* (who are more advanced, carrying out data analysis), and *multi-level*, all of which were carried out in the same classroom at the same time. This required flexibility and open-mindedness on the part of teachers, and some teachers modified the activities many times to meet students' needs in the classroom. For example, teacher 5 said "Why not, I am flexible, it is easy for me to change to tasks based on students' needs… The purpose is to integrate CT in my activities how and when… it depends on students' needs" [Teacher 5]. On the other hand, some teachers did not depart from their existing plans, with eight teachers using a maintaining a pre-established linear approach, while six used branching structures, with different possible routes to resolve the problem.

Open-Ended Tasks

Seven teachers used open-ended tasks, rather than following the curriculum, especially in teaching programming. This strategy was not without risk for the teachers, as there was a danger that if they did not complete the curriculum plan by the end of the semester, there would be a negative impact on their prospects for promotion. The approach adopted by these teachers was to allow students to explore different computational thinking skills and apply them to their classroom activities. In the interviews, seven teachers reported that open tasks could be beneficial for students. For example, teacher 3 said that

TABLE 2 | Thinking steps used by teachers in their instruction.

Thinking step	Definition	Topic
Data gathering	Collecting data from different resources (Internet, local community)	Mathematics, Science
Data analysis	Finding the patterns among data and understanding the characteristics of samples within teams	Mathematics, Science
Data representation	Organizing data in a suitable way using tables charts, and infographics	Mathematics, Science, Technology
Abstraction	Identifying and extracting relevant information from the activities to define main ideas	Mathematics, Languages, Social Sciences
Algorithm design	Using ordered consequences instructional steps for solving similar problems or for implementing an activity	Mathematics, Science, Technology
Decomposition	Breaking the problem and procedures into smaller parts to manage and solve it.	Mathematics, Science, Technology
Visualization	Using visual content as an easier way to understand and find patterns	Mathematics, Science, Languages, Social Sciences
Debug and correct errors	Finding mistakes in the steps to solve a problem/programming a task and fixing it	Mathematics, Technology

"open-ended tasks can allow students to explore new computational thinking skills and programming skills." These types of activities commonly involved programming tools, such as Scratch, and languages, such as Kotlin in the eighth and ninth grades.

Examples of Activities and Computational Thinking Integration

Teachers used a range of activities to integrate CT into STEM and language instruction, including hands-on activities (ten activities), focusing on the processes and practices of CT, experimentation (12 activities), inquiry (nine activities), and Scratch and robots (14 activities). In math and science classes, nine teachers asked students to get data from Google Dataset about the weather, or statistical data about Middle East countries, and to visualize it to find patterns. Teachers started by showing how data could be found through a Google Dataset search, and how to access the Google data center. They then provided step-by-step instructions for students by introducing CT and instructions about how to download the data and transfer it into an Excel sheet to visualize it.

All the teachers in observed classrooms used CT steps defined by Yadav et al. (2016) in their instruction, as shown in **Table 2**.

As may be seen, all of the thinking steps identified in our framework were applied in one or more of the seven mathematics classes which were observed. In the five science classes, all thinking steps were applied with the exception of *abstraction* and *debug and correct*. While it is perhaps unsurprising that debut and correct was not used beyond mathematics and technology, the absence of *abstraction* is interesting, given the importance that it has in scientific thinking. In the ten technology classes, the observed thinking steps aligned with some of the basic skills of programming: *data representation, algorithm design, decomposition, debug, and correct.* This raises a concern that CT activities in technology may be too tightly focused on programming, and it indicates that it would be valuable to focus training on other aspects of CT which are important aspects of technology in general and computing in particular: *data gathering, data analysis, data representation and visualization.* The 16 Languages and social sciences classes observed made use only of *abstraction* and *visualization* and thus had the narrowest range of activities despite being the largest number of classes observed. This indicates that teachers outside of the STEM subjects had difficulty in designing learning activities which could apply the concepts of CT in their classes and that it will be important to focus on this in future training if the aim of applying CT across the curriculum is maintained.

Research Question 2: What Types of Support Did Teachers Require to Integrate Computational Thinking Into Classroom Instruction?

Thirty teachers in the study reported that they were encouraged by their school administration and supervisors to learn about ways of using CT in their instruction. Twenty-eight participants said that technical and instructional support was provided for teachers in their schools by the school administration and their colleagues. Interestingly, the nine teachers who expressed a need for training on designing CT activities emphasized the instructional design aspects rather than a need for technical support. Thirty-three said they had attended only two workshops on designing activities to use in instruction and that they felt a need for more.

Professional Development of Teachers

Thirty teachers said that they had the basic skills necessary to teach CT, but they did not know how to integrate these skills into their classroom activities. "If I had designed activities related to my field, I would use them… I know myself; I do not know how to integrate these skills into my classroom activities… If we had access to open-source activities, it would be beneficial for me" [Teacher 4]. It therefore appears that teachers need training in finding and using activities related to their teaching topics that use open-source applications.

During the period of this study, the teachers who participated received training to equip them with the knowledge and skills essential for CT integration in classrooms. Thirty-two teachers of thirty-eight of the participants in study were satisfied with the workshops and valued the opportunity to mix with others and learn from each other. Teacher 10 said "I am happy to be in this workshop. It provides me with skills and knowledge to implement activities." Ten teachers in the interviews believed that these workshops would not be sufficient to gain information

TABLE 3 | Learning strategies used in the classroom.

Learning strategy	Description
Problem-based learning	Six teachers of math and science divided the students into teams and then provided them with a task. The students explored the learning solution individually and reported their own learning conclusions and feedback to the team. Each student described their problem-solving procedures to the team
Group learning	Many teachers in different areas (science, technology, and English language) used this strategy. Following the instructions of the teacher, the students completed the tasks into two ways. In one class, five teams divided the work into subtasks and solved them individually, and then gathered the partial results into a final output. In another class three teams, the members were required to complete the task together, negotiate, and share their ideas to solve it
Project-based learning	Twenty-two teachers from different fields used this strategy. Teachers organized the integration of CT around projects. They introduced the tasks, which were in the form of questions, and asked students to collect, analyze and present data from Google. The students were engaged in design, problem-solving, and in making decisions based on the pattern in the data
Scaffolding	Technology, science, and English teachers used various forms of scaffolding, such as using teaching aids, chunking the activities, and modeling in their instruction to help students to carry out the CT activities
Teacher-centered lecture	Some teachers used this strategy at specific points to introduce CT concepts and to provide examples to demonstrate CT skills
Gamification	Many science and math teachers used the features of gamification (points, rewards, badges…) to introduce CT skills and concepts into instruction to enhance extrinsic motivation

to integrate CT into instruction. This indicates that it would be valuable to have coaching available on site in school, so teachers can derive more benefits than can be achieved in centralized workshops.

Availability of Expertise

Twenty-seven out of thirty-eight teachers who participated in the study mentioned their need to communicate with experts in their subject area while integrating CT into classroom activities. Five teachers expressed their concerns about computational thinking in STEM classrooms, since they did not have a clear idea about how they would integrate it and had had bad experiences with it. In particular, the English language teachers complained about the lack of guidance on the integration of CT in classroom activities in their field. The summer professional development program went some way to meeting this need, and twenty-five teachers out of thirty-eight teachers recognized the importance of exchanging their skills and knowledge about CT during the workshops, and in the interviews, teachers said that they were more confident about integrating CT into their classroom instruction after the workshop.

Nevertheless, it is clear that there remains an unmet need for ongoing access to support from experts and peers. Thirteen teachers reported that they wanted to hear their colleagues' stories of integrating CT into their teaching, and many teachers were influenced by those colleagues, who inspired them to implement CT skills. In the absence of such expert input and contact with peers, eleven teachers said that they used social media to ask for help, and sixteen teachers from different schools shared open resources to teach concepts in science on their pages on Facebook. Teacher 11 attested that "These activities were helpful for me."

This shows that the opportunity to exchange experiences and expertise with their peers was highly valued by teachers.

Lack of Time

Time is a crucial factor that influences teachers' use of CT approaches in their instruction. All the teachers mentioned that they needed to learn about CT, as well as about how it could be used in classrooms. However, they did not have enough time to prepare activities to integrate CT into their instruction, because of the time taken up by administrative tasks.

Support From School Principals Was Important for Teachers

An invitation to participate in the workshops constituted a recognition of teachers' work, and an opportunity for teachers to develop their computational thinking, knowledge and skills, as well as helping teachers develop lesson plans and activities. Moreover, trainers could encourage teachers to develop activities and lessons, and enhance their confidence in integrating CT in classrooms.

School principals developed a clear vision and school statement about integrating CT into instruction and put it on their calendars during the academic year. Twenty teachers in the interviews mentioned that their school had clear visions about CT integration into instruction. School principals in five schools as reported by five teachers tried to minimize the risk of implementing CT into instruction. Overall, school administrations (principals and managerial staff) and teachers appreciated the administrative support, as well as colleagues' support for computational thinking integration in classrooms.

Research Question 3: What Pedagogical Strategies Do Teachers Report Work Well for Using Computational Thinking in Teaching?

The participating teachers used a number of different learning strategies to integrate CT concepts and practices into their instruction. **Table 3** sets out the learning strategies that were used in classroom instruction, as observed and reported by teachers. These included gamification, teacher-centered lectures, problem-based learning, project-based learning, scaffolding, and collaborative learning.

Classroom observations showed that the teachers used a range of pedagogical strategies to teach the students the skills

TABLE 4 | Challenges and strategies in adoption of CT.

Challenges identified		Strategies implemented to overcome challenges
Challenges to CT integration in instruction	Lack of suitable resources	Design activities/collaborate with other teachers to design/use existed activities.
	Lack of technical support	Ask colleagues for support/use social media to ask for help.
	Access to technology	Rotate students to the computer lab/library to use interactive projects/Internet.
	Large classes	Dividing students into groups
Challenges relating to teachers	Limited time	Break down/use video.
	Assessment	Use reflection papers and rubrics.
	Differentiation	Use peer mentoring/ scaffolding/collaboration.
	Lack of training	Use social media to get support.
Challenges related to students	Student not engaged	Use a variety of activities.
	Solving problems	Practice a lot/use many activities

and concepts of CT. These strategies appeared to vary according to the subject in which the CT was embedded. For example, the science teachers focused on using collaborative activities to teach students CT concepts and skills. **Table 3** shows the type of teaching strategy used by teachers in the classroom and its interpretation, as it was divided into problem-based learning, group learning, project-based learning, scaffolding, teacher-centered lecture, and gamification.

The richness of the pedagogical strategies contrasts with the need for pedagogical support expressed by teachers. In combination with the appreciation shown by teachers for the opportunity to exchange practice with peers, this indicates that the required pedagogic expertise is present in the community of teachers, but that the structures and time that would enable the sharing of this practice is not available to teachers. This indicates that the challenge is not to develop of appropriate pedagogical strategies for use by teachers, but rather to support them through peer exchanges and coaching.

Research Question 4: What Challenges Do Teachers Face During the Adoption of Computational Thinking in Classrooms

In the responses to the questions in the interviews, all the teachers described challenges in integrating CT into instruction. We also noticed these challenges during classroom observations. Interestingly, when the teachers discussed these barriers, they described the challenges within the context of their instruction. Eight teachers describe the source of challenges from the infrastructure of the schools. Ten teachers described the source of the challenge from their individual characteristics, such as social commitments and lack of time. **Table 4** summarizes the major challenges reported by teachers and observed during integration of CT activities, as well as the strategies that teachers used in overcoming those challenges. The barriers to integrating computational thinking reported in teachers' interviews, together with those noted during classroom observations, were gathered into three categories, relating to students, to the learning environment, and to the teachers, respectively. We now discuss these.

Challenges Related to the School Environment

Thirty teachers described many factors that prevented them from integrating CT into STEM instruction including the lack of time, differentiation of strategies, assessment and training. Four challenges indicated by all the teachers and in the observations related to the learning environment including weaknesses in the technical infrastructure, large classes, lack of technical support, and lack of resources.

All the teachers mentioned that the lack of instructional time was a major challenge, although they were able to compensate for this factor by breaking down the CT activities into small chunks and carrying them out in different classes at different times. In addition, six teachers used freely available videos to explain the concepts, and designed activities to ensure that students focused on the video while viewing them. All the teachers mentioned that they were committed to finishing the mandated textbooks, while at the same time integrating CT into their instruction. "My supervisor came and asked about my progress in the curriculum. She did not ask about activities, her concern is whether I am following the plan to finish the curriculum on time or not, not that I am doing my best to use CT in class activities" [Teacher 7].

Challenges Related to Student Engagement

Most of the teachers mentioned that students did not fully engage in the activities because they were new to them. Teachers mitigated this challenge by using group work to encourage students to learn from each other, and to create mutual encouragement to complete the activities. Some teachers mentioned the lack of student engagement was due to the class size and the quality of the activities which were suggested to them for use in the classroom. "I believe students do not engage in the activities because it does not attract their attention, or they were feeling bored because of it… With time, maybe the quality of CT activities will be better and meet students' needs" [Teacher 15]. **Table 4** shows that teachers of computational thinking faced a challenge in the process of integrating various topics. They also experienced a lack of resources and of technical support, difficulties in access to technology and making it available to students, and overcrowded classes. At a professional level, teachers also felt that CT was being held back by their lack of time to enjoy the educational process, a lack of appreciation of teachers' work by others, and the need to trade-off between the subjects required for the student to successfully complete the curriculum. Teachers also identified that their lack of experience and training was a

challenge. At a pedagogic level, a challenge was the presence in classes of students who are not engaged in the educational process and are unable to follow the CT activities which they are given.

CONCLUSION, LIMITATIONS, AND IMPLICATIONS FOR FUTURE RESEARCH

The present study set out to provide information on the challenges faced by teachers of CT in K-12 schools in Palestine, the support which is provided for them in designing CT strategies in their teaching, and the strategies that they have adopted in using CT approaches in the classroom. We now summarize the insight we have gained into our research questions.

How Is Computational Thinking Integrated Into STEM Classroom and Language Activities Across Different Subject Areas?

The majority of the activities observed (30 of 45) involved group work and reorganization of the classroom space, but no particular pedagogic approach was used by teachers of CT, who tended to apply their default method of instruction. Half of the teachers demonstrated flexibility in providing different activities for groups of students with different levels of prior learning, while seven teachers exposed themselves to some risk by giving their students open-ended tasks rather than following the curriculum. The spread of activities across the different subject areas provided interesting results. All of the thinking steps identified in our framework were applied in one or more of the mathematics classes which were observed, and all but two steps were observed in science. In technology classes, however, the thinking steps were closely aligned with the procedures of computer programming, while languages and social sciences made use only of *abstraction* and *visualization,* despite being the largest group of classes observed. Thus, an important result is to highlight the need to focus in future training on supporting teachers outside the disciplines of mathematics and science in applying the whole range of CT thinking steps in their teaching.

What Types of Support Do Teachers Require to Integrate Computational Thinking Into STEM Classroom Instruction?

It is interesting to note that the requirement for support was largely pedagogic rather than in development of technical skills. Most teachers said support was provided in the schools, but those who requested training emphasized a need for instructional design rather than technical support. They valued the opportunity to share practice with their peers in the workshops, but the need was expressed for ongoing coaching on site in school. The teachers unanimously identified a lack of time to learn about and integrate new methods as a significant barrier to adoption of CT methods.

What Pedagogical Strategies Do Teachers Report Work Well for Using Computational Thinking Approaches in Teaching and Learning?

The data did not enable us to track and assess the outcomes of different pedagogic strategies. However, a rich variety of strategies was deployed by teachers, who also valued the opportunity to exchange practice with peers. This indicates that the pedagogic challenge is not to develop pedagogical strategies or resources for use by teachers, but rather to support them through peer exchanges and coaching.

What Challenges Do Teachers Face During the Adoption of Computational Thinking in Classrooms?

The teachers unanimously identified challenges in the weaknesses in the technical infrastructure, large classes, lack of technical support and lack of resources. In the findings for research question 2, we saw that teachers did not need support in their technical skills, but they do need suitable equipment that is adequately maintained. All teachers also identified lack of instructional time as a major challenge. Mitigating strategies included the use of videos and adaptation of existing activities, support from colleagues and social media, and rotating groups of students from library activities to CT activities, and peer support and coaching would be valuable in helping teachers to meet the challenges they face.

Limitations

The limitations of this study are related to the population, which was drawn on a self-selecting sample of teachers who were involved in the MoE training course within which the research was carried out. This approach offered great advantages in access to teachers, but inevitably other teachers, who might have had valuable insights to offer, or who could have contrasted the results, were not included in the study. In addition, a smaller number of teachers were observed in their classrooms compared with the number of interviews, because not all interviewees volunteered to be observed. As a consequence of this self-selection of respondents, the research design adopted a qualitative approach.

Implications and Future Work

The study represents a first step toward understanding teachers' experiences and students' engagement in CT in Palestine. More research is necessary to evaluate the support that teachers need to understand the integration process, and the instructional strategies that could motivate different students to engage in CT. The current study offers valuable information for decision-makers in Palestine, and evidence which provides encouragement for further efforts to include more CT in K-12 settings in the Palestinian context. A repeated theme in these conclusions is the importance of exchanges with peers and coaching from both peers and experts. This is a strong indication of where

resources would be best deployed in support of teachers in implementing CT methods in their teaching. However, it should be recognized that many of the challenges identified, particularly in our analysis of research question 4, are systemic in the educational environment in Palestine, and not limited to CT. These conditions inevitably act as a constraint on what can be achieved through peer exchanges and coaching.

In future research, it is planned to make use of a mixed methods approach, to involve more teachers in a larger sample, and to include students. The researchers do not differentiate in the data analysis between the results for different genders and teaching topics, but recognize that this is an important aspect to be investigated in future work. Finally, the fact that the researchers did not focus on a specific level of school, with the teachers coming from the three levels of K-12 education in Palestine (elementary, middle school, and high school). Further research should investigate integrating computational thinking into instruction at different levels. The study took place over 7 months of integrating CT into instruction, a process which changed over the duration of the study. In the future, it will be important to conduct research to understand how to create a sustainable culture of CT among schools and teachers and to identify the impact of using different learning strategies to engage students in CT concepts and skills in different curriculum areas.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

DB: data analysis, discussion, and editing. DG: lead writer, methodology, and data validation. SA: design, data analysis, and conclusion. SS: writing and literature review. AG: data validation and discussion. FK: design, and methodology. ZK: writing and conclusion. All authors contributed to the article and approved the submitted version.

REFERENCES

Abo Hommos, N. (2013). PA the day after: Collapse and its effect on education. palestinian center for policy and research, Ramallah.

Afana, Y. (2021). Moving beyond TIMSS An Explanatory Sequential Mixed Methods case Study of Mathematics Education in the Palestinian context. Doctoral dissertation, University of Leicester.

Barr, V., and Stephenson, C. (2011). Bringing computational thinking to K-12: what is involved and what is the role of the computer science education community? *Inroads* 2, 48–54. doi: 10.1145/1929887.1929905

Bers, M. U. (2020). *Coding as a Playground: Programming and Computational Thinking in the Early Childhood Classroom*. New York, United States: Routledge.

Bivins, A. L. (2015). *The State (And non-state) of Education Development in the West Bank*. BA thesis, University of Chicago.

Bocconi, S., Chioccariello, A., Dettori, G., Ferrari, A., Engelhardt, K., Kampylis, P., et al. (2016). Developing computational thinking in compulsory education - implications for policy and practice. Joint Research Centre (Seville site) Working Paper.

Brennan, K., and Resnick, M. (2012). New frameworks for studying and assessing the development of computational thinking. In *Proceedings of the 2012 Annual Meeting of the American Educational Research Association*; April, 12–18, 2012; Vancouver, Canada. 25.

Cetin, I., and Andrews-Larson, C. (2016). Learning sorting algorithms through visualization construction. *Comput. Sci. Educ.* 26, 27–43. doi: 10.1080/08993408.2016.1160664

Chen, G., Shen, J., Barth-Cohen, L., Jiang, S., Huang, X., and Eltoukhy, M. (2017). Assessing elementary students' computational thinking in everyday reasoning and robotics programming. *Comput. Educ.* 109, 162–175. doi: 10.1016/j.compedu.2017.03.001

Denning, P. J., and Tedre, M. (2019). *Computational Thinking*. Cambridge, MA: MIT Press.

European Commission (2016). Communication on new skills agenda, human capital, employability and competitiveness: "A new Skills Agenda for Europe." Available at: http://www.ipex.eu/IPEXL-WEB/dossier/document/COM20160381.do [Accessed May 8, 2022].

European Council (2006). Recommendation of the European Parliament and the council on key competencies for lifelong learning. Official Journal of the European Union, 10–18. Available at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:394:0010:0018:en:PDF [Accessed May 8, 2022].

Feldhausen, R., Weese, J. L., and Bean, N. H. (2018). "Increasing student self-efficacy in computational thinking via STEM outreach programs." in *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*; February 21–24, 2018; (302–307).

Grover, S., and Pea, R. (2013). Computational thinking in K–12: A review of the state of the field. *Educ. Res.* 42, 38–43. doi: 10.3102/0013189X12463051

Hava, K., and Ünlü, Z. K. (2021). Investigation of the relationship between middle school students' computational thinking skills and their STEM career interest and attitudes toward inquiry. *J. Sci. Educ. Technol.* 30, 484–495. doi: 10.1007/s10956-020-09892-y

Hirsh, R. A., and Baronak, K. (2020). Empowering early childhood pre-service teachers with tech fluency. *Creat. Educ.* 11, 2730–2748. doi: 10.4236/ce.2020.1112200

Israel, M., Pearson, J. N., Tapia, T., Wherfel, Q. M., and Reese, G. (2015). Supporting all learners in school-wide computational thinking: A cross-case qualitative analysis. *Comput. Educ.* 82, 263–279. doi: 10.1016/j.compedu.2014.11.022

ISTE (2011). Teacher Resources. Available at: https://www.iste.org/explore/articledetail?articleid=152 [Accessed May 8, 2022].

Istikomah, I., and Budiyanto, C. (2018). The contribution of educational robotics and constructivist approach to computational thinking in the 21st century. In *Proceedings of the 1st International Conference on Computer Science and Engineering Technology*. Universitas Muria Kudus, Kudus, Indonesia; November 27–28, 2018 (610–616).

Kong, S. C., Lai, M., and Sun, D. (2020). Teacher development in computational thinking: design and learning outcomes of programming concepts, practices and pedagogy. *Comput. Educ.* 151:103872. doi: 10.1016/j.compedu.2020.103872

Kwon, K., Ottenbreit-Leftwich, A. T., Sari, A. R., Khlaif, Z., Zhu, M., Nadir, H., et al. (2019). Teachers' self-efficacy matters: exploring the integration of mobile computing device in middle schools. *TechTrends* 63, 682–692. doi: 10.1007/s11528-019-00402-5

Lin, Y., and Weintrop, D. (2021). The landscape of block-based programming: characteristics of block-based environments and how they support the transition to text-based programming. *J. Comp. Lang.* 67:101075. doi: 10.1016/j.cola.2021.101075

Mäkitalo, K. H., Tedre, M., Laru, J., and Valtonen, T. (2019). "Computational thinking in Finnish pre-service teacher education," In *Proceedings of International Conference on Computational Thinking Education*; June 13–15, 2019; 105–108.

Marshall, C., and Rossman, G. B. (2011). *Designing Qualitative Research*. New York: Sage publications.

Mason, S. L., and Rich, P. J. (2019). Preparing elementary school teachers to teach computing, coding, and computational thinking. *Contem. Issues Technol. Teach. Educ.* 19, 790–824.

Ministry of Education of Palestine (2008). Teacher Education Strategy in Palestine. Available at: http://wbgfiles.worldbank.org/documents/hdn/ed/saber/supporting_

doc/MNA/West_Bank_Gaza/Section%203_Teacher%20Education%20Strategy-%20English,%20WBG.pdf [Accessed May 8, 2022].

Nicolai, S. (2007). *Fragmented Foundations: Education and Chronic Crisis in the Occupied Palestinian Territory*. Paris, France: International Institute for Educational Planning (IIEP).

Palts, T., and Pedaste, M. (2020). A model for developing computational thinking skills. *Informat. Educ.* 19, 113–128. doi: 10.15388/infedu.2020.06

Papadakis, S. (2021). The impact of coding apps to support young children in computational thinking and computational fluency. *Front. Educ.* 6, 1–3. doi: 10.3389/feduc.2021.657895h

Papadakis, S. (2022). Can preschoolers learn computational thinking and coding skills with ScratchJr? A systematic literature review. *Int. J. Educ. Reform.* 3–5. doi: 10.1177/10567879221076077

Papert, S. (1980). *Mindstorms: Children, Computers, and Powerful Ideas*. Basic Books, New York.

Park, J. (2019). Evaluation of computational thinking through code analysis of elementary school Students' scratch projects. *J. Korean Assoc. Inform. Educ.* 23, 207–217. doi: 10.14352/jkaie.2019.23.3.207

Pears, A., Barendsen, E., Dagiene, V., Dolgopolovas, V., and Jasute, E. (2019). "Holistic STEAM education through computational thinking: a perspective on training future teachers." in *Informatics in Schools: Situation, Evolution, and Perspectives. Lecture notes in computer science (Springer)*, 41–52.

Ramahi, H. (2015). Education in Palestine: current challenges and emancipatory alternatives. *Rosa Luxemburg Stiftung Regional Office Palestine*, 1–51.

Rich, P. J., Mason, S. L., and O'Leary, J. (2021). Measuring the effect of continuous professional development on elementary teachers' self-efficacy to teach coding and computational thinking. *Comp. Educ.* 168:104196. doi: 10.1016/j.compedu.2021.104196

Rijke, W. J., Bollen, L., Eysink, T. H., and Tolboom, J. L. (2018). Computational thinking in primary school: An examination of abstraction and decomposition in different age groups. *Inform. Educ.* 17, 77–92. doi: 10.15388/infedu.2018.05

Sampson, D., Ifenthaler, D., Spector, J. M., and Isaías, P. (eds.) (2018). *Digital Technologies: Sustainable Innovations for Improving Teaching and Learning*. United States: Springer International Publishing.

UNESCO (2010). World data on education. Principles and general objectives of education. Available at: http://www.ibe.unesco.org/fileadmin/user_upload/Publications/WDE/2010/pdf-versions/Palestine.pdf [Accessed May 8, 2022].

Voogt, J., Fisser, P., Good, J., Mishra, P., and Yadav, A. (2015). Computational thinking in compulsory education: towards an agenda for research and practice. *Educ. Inf. Technol.* 20, 715–728. doi: 10.1007/s10639-015-9412-6

Wahbeh, N. A. (2003). Teaching and learning science in Palestine: dealing with the new Palestinian science curriculum. *Mediterranean J. Educ. Stud.* 8, 135–159.

Weese, J. L. (2017). Bringing computational thinking to K-12 and higher education. Doctoral dissertation, Kansas State University.

Wing, J. (2006). Computational thinking. *Commun. ACM* 49, 33–35. doi: 10.1145/1118178.1118215

Wing, J. (2011). Research notebook: computational thinking-what and why. The link magazine, 6.

Yadav, A., Gretter, S., Good, J., and McLean, T. (2017). "Computational thinking in teacher education," in *Emerging Research, Practice, and Policy on Computational Thinking* (United States: Springer), 205–220.

Yadav, A., Hong, H., and Stephenson, C. (2016). Computational thinking for all: pedagogical approaches to embedding a 21st century problem solving in K-12 classrooms. *Tech. Trends.* 60, 565–568. doi: 10.1007/s11528-016-0087-7

Yin, R. K. (2003). *Case Study Research: Design and Methods* (*3rd Edn*). New York: Sage.

Yuen, T. T., and Robbins, K. A. (2014). A qualitative study of students' computational thinking skills in a data-driven computing class. *ACM Trans. Comp. Educ.* 14, 1–19. doi: 10.1145/2676660

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Ghani, Griffiths, Salha, Affouneh, Khalili, Khlaif and Burgos. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Education

ORIGINAL RESEARCH
published: 17 June 2022
doi: 10.3389/feduc.2022.872479

Evaluating Technology-Enhanced, STEAM-Based Remote Teaching With Parental Support in Luxembourgish Early Childhood Education

*Ben Haas[1,2], Zsolt Lavicza[2], Tony Houghton[2] and Yves Kreis[1]**

[1] Department of Education and Social Work, Institute for Teaching and Learning, University of Luxembourg, Esch-sur-Alzette, Luxembourg, [2] Linz School of Education, Johannes Kepler University Linz, Linz, Austria

OPEN ACCESS

Edited by:
Stamatios Papadakis, University of Crete, Greece

Reviewed by:
Seyum Getenet, University of Southern Queensland, Australia
Michail Kalogiannakis, University of Crete, Greece

***Correspondence:**
Yves Kreis
yves.kreis@uni.lu

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Education

Received: *09 February 2022*
Accepted: *29 April 2022*
Published: *17 June 2022*

Citation:
Haas B, Lavicza Z, Houghton T and Kreis Y (2022) Evaluating Technology-Enhanced, STEAM-Based Remote Teaching With Parental Support in Luxembourgish Early Childhood Education. Front. Educ. 7:872479. doi: 10.3389/feduc.2022.872479

During COVID-19 confinement, we observed numerous challenges in using educational technology in early childhood Science–Technology–Engineering–Arts–Mathematics (STEAM) education in Luxembourg. Thus, we designed a conceptual framework on parent-assisted remote teaching with active uses of educational technology supported by cycles of design-based research. After a previous study utilizing computer-aided design (CAD) software and three-dimensional (3D) printing in primary education, we used our initial findings to work with 12 early childhood students (ages 4–6), together with their teachers and parents in the second remote teaching period in Luxembourg. We created a STEAM modeling task with CAD software on robots and collected data through chat responses, messageboards, and online communication channels during a 3-week period. Here, we observed new roles in the parent–child relationship while learning STEAM in remote teaching with technology, and new opportunities in using educational technology overall in early childhood education. In this article, we have described findings that are likely to influence students' learning and parent-assisted teaching, in particular parents and students' perceptions and motivations, together with the way in which parents provide technical knowledge and support in remote early childhood STEAM education.

Keywords: early childhood, STEAM, remote teaching, educational technology, parents

INTRODUCTION

Remote teaching due to COVID-19 restrictions in Luxembourg (Kreis et al., 2020) was relatively short compared to the worldwide average, only 15 weeks from 2020 to 2021 according to the United Nations Educational, Scientific and Cultural Organization (UNESCO) (UNESCO, 2021). In Luxembourg, elementary school pupils from ages 4 to 12 received during this time access to various

online learning resources. Among these were the automated tutoring system in mathematics education, "MathemaTIC" (Haas et al., 2020) and an online platform with documents, videos, and challenges on different languages and Science–Technology–Engineering–Arts–Mathematics (STEAM) disciplines.[1] These educational resources, used over the past years (e.g., MathemaTIC from 2016 on) by most school classes (ages 7–12), and thus known by pupils and teachers, were quickly integrated into the "schooling at home" teachings in primary education. However, these resources were not designed for early childhood education students, where teaching with educational technology is not usual in Luxembourgish schools, similar to international trends reported by the Organization for Economic Co-operation and Development (OECD) in "Using Digital Technologies for Early Education during COVID-19" (OECD, 2021). The imposed physical and social distancing, where both teachers and pupils were not acquainted with working with technology-supported lessons in early childhood education in Luxembourg, reduced in various classes the learning to a less student-centered approach, and facilitated less active but rather repetitive tasks in learning (e.g., filling out paper and pencil exercises). Throughout previous observations on the use of educational technology and remote teaching in schools in Luxemburg and results from a survey among parents by the Ministry of Education, Children, and Youth (Ministère de l'Éducation Nationale, de l'Enfance et de la Jeunesse—MENJE) (MENJE, 2020a), we observed a high level of frustration including many confrontations between parents and children, and difficulties in understanding tasks or connecting to teachers or pupils vice versa for parents, pupils, and teachers (Haas, 2021). Hence, new educational needs became perceptible during remote teachings, such as understanding the possibilities of using educational technology, connecting through social media platforms, and supporting parents in the teaching process at home. Parents of pupils in early childhood education request more support (MENJE, 2020a), such as how to engage with the teachers' tasks, regulate stimuli, motivation, or monitor media choices, and media consumption (Hirschland, 2008; Neumann, 2018).

Over the past years, several researchers experimented with digital modeling of shapes, forms, or mathematical functions with computer-aided design (CAD) software and augmented reality (AR) technologies in elementary schools to demonstrate motivational factors, new learning opportunities, and a different learning approach (Steen et al., 2006; Liu et al., 2019; Ng and Chan, 2019). These technologies reported higher motivation among children (Bacca Acosta et al., 2019). Moreover, these technologies were recently used among pupils with mathematical learning disabilities (MLDs) to increase visual-spatial memory in cycle 2 to cycle 4 (Haas et al., 2021a).

However, most studies did not directly test these technologies in remote teaching in early childhood education, especially in Luxembourg, without immediate physical or digital teacher assistance and with parents' assistance. A first attempt to use these technologies at home was made in 2020 during the first distance learning in Luxembourg. We conducted a remote teaching study (e.g., egg cup creation) with CAD software and three-dimensional (3D) printing to experiment with their use in remote teaching and to identify the parent's role in the teaching process. In this first study (Haas et al., Under Review) of design-based research (Wang and Hannafin, 2005), we established a framework for learning at home with parent–assisted teaching. In this framework, which is presented in "Methodology and Methods" section, we identified four elements (perception of STEAM and teaching, motivation of students, technology-knowledge of parents, and parent-assisted teaching and scaffolding) influencing students' remote learning experience.

Based on this framework, we designed a second study of remote early childhood STEAM education with CAD software and 3D printing, which we is reported in this article. We proposed these tasks to 12 pupils aged 4–6 years studying with parental support. Tasks were integrated in a thematic school week on robots and consisted of designing, showing in AR, and printing an own-designed robot with geometric shapes. Since this happened in remote teaching and due to the young age of the children, parents were highly involved. We sought to answer the following research questions to confirm and question our framework from the first study:

- How do parents and children perceive STEAM education and teaching with technology in remote early childhood education?
- How do students' motivation evolve while using technology in remote early childhood STEAM education?
- How do the parents' technological knowledge influence students' learning while using technology in remote early childhood STEAM education?
- How do parents assist their children and what scaffoldings do they use in remote early childhood STEAM education?

To investigate these questions, we collected data from 12 pupils and each parent, from a heterogeneous socioeconomic background per child over 3 weeks. This article presents theoretical references, task design, the used qualitative methods, and the discussion of our findings. In particular, the opportunity offered by the COVID-imposed restrictions provided a unique opportunity to analyze parents' and children's perceptions and motivation, together with the way in which parents provide technical knowledge and support in remote early childhood STEAM education.

LITERATURE REVIEW

In the following literature review, we outlined the importance of a STEAM transdisciplinary approach, the use of technology (especially CAD and AR) for 3D modeling in class, and the 3D printing of the designed model. Furthermore, parents' role in "schooling at home" and their view on technology use at home are examined. In each subsection, we identified gaps in previous practices or research that to be covered in the present research.

[1] https://www.schouldoheem.lu

STEAM Transdisciplinary Approach

Although STEAM disciplines are still taught separately, the current curriculum in early childhood education in Luxembourg (MENJE, 2020b) requires teachers and educators to work in class with a transdisciplinary approach (Rausch et al., 2021) for STEAM disciplines (Liao, 2016), where arts are integrated in STEM (Lavicza et al., 2018). The integrated approach of STEAM disciplines (Kelley and Knowles, 2016; Haas et al., 2021b) has been shown to support students in applying discipline skills and understanding content more easily than taught separately (Burnard et al., 2018; Lavicza et al., 2020; El Bedewy et al., 2022). Moreover, this approach offers more creativity to the tasks and, thus, is likely to positively affect problem-solving skills of learning (Dana-Picard et al., 2021).

In early childhood education, this is mostly done in Luxembourg through thematic project weeks (e.g., colors in nature or, in our case, robots) where daily activities are interconnected to a specific theme or project (MENJE, 2018). The positive effects of such an approach, close to project-based learning (PBL) (Cesarone, 2007), are in socializing students and connecting content and process skills to real-world situations. Such an approach is further recommended for integrated science teaching (Haatainen and Aksela, 2021) and, thus, enabling STEAM-integrated teaching in early childhood education. Nevertheless, this practice can be conceived as a multi- or interdisciplinary approach, whereas our research will focus especially on transdisciplinarity.

Use of Technology for 3D Modeling in Class

In recent years, technology-enhanced learning (e.g., tablets in class, cameras, or digital measuring tools) has featured within STEAM in primary education and early childhood (Chaudron et al., 2015; Jablonski and Ludwig, 2020; Guntur and Setyaningrum, 2021). Although tutoring systems (Steenbergen-Hu and Cooper, 2013) prevailed in classes in primary education in past years, other technologies, which connected directly to the environment of students (e.g., CAD software, AR, or 3D printing), became more frequent in teaching. Thus, mixed-realities, such as the use of AR, where digital information is applied to the real-world information, were used to support students in visualizing mathematical concepts or explaining cultural or architectural phenomena (Ng, 2017; Liu et al., 2019; Lavicza et al., 2020). A representation of a shape in three dimensions in a real-world environment instead of a 2D representation gives the students a better understanding. Furthermore, the use of AR in an educational context was found to increase students' motivation (Bacca Acosta et al., 2019; Sarkar et al., 2020). In addition to AR, CAD software allows students to modulate polygons, shapes, lines, or other geometry concepts in three dimensions (Stone et al., 2020). As a result, students can reproduce objects from the real world or conceptualize mathematical objects. According to Ng and Chan (2019), 3D modeling with CAD software is likely to support students in learning informally and less in a procedural and formulae-driven learning setting. This further connects STEAM disciplines to real-world objects or situations (Haas, 2021). However, most of the time technology was used in class and not in distance learning.

3D Printing of Designed Model

In extension to AR and CAD software, 3D printing showed in several research stances to further support students in transferring learned STEAM skills to real-world problems or objects (Ng et al., 2020; Pearson and Dubé, 2021). The 3D printing process is an additive manufacturing process based on a designed CAD model, where a digital object is printed as a physical, real-world object. According to Lieban (2019), 3D printing engages and motivates students in a creative process where they can modulate and combine digital and physical objects to understand mathematical concepts better.

Moreover, 3D printing in the process itself (e.g., adjusting temperature or preparing the print) makes students experience an engineering task while applying different STEAM skills. In combination, the three mentioned technologies (AR, CAD, and 3D printing) offer the possibility to combine STEAM disciplines, apply skills transdisciplinary (Takeuchi et al., 2020), and modulate digital and physical objects. This task is quite complex and cannot be left to young children alone. Preparations can be done remotely, but the printing task requires physical presence at 3D printers.

Parents' Role in "Schooling at Home"

Since technologies are not necessarily constrained to physical presence courses, a pedagogical use in remote teaching is possible (Chilton and Mccracken, 2017). According to Kreis et al. (2020), there are different remote learning forms. The situation in Luxembourg during the COVID-19 lockdown could be best described as "schooling at home," where parents received tasks, documents, and instructions from teachers. Parents took the role of an assistant-teacher and instructed their children. Like teachers in class, parents are likely to use different forms of support (e.g., affective) in teaching (Wood et al., 2016; Neumann, 2018) and use their own school experiences and educational knowledge (Livingstone et al., 2015; Elliott and Bachman, 2018). The first framework on parents' "schooling at home" was developed in our first study on an eggcup creation (Haas et al., Under Review). In this study, we obtained valuable data on parent-assisted STEAM remote learning and with the presented second study in this article, we were able to confirm some of the data from the first experience and add new insights and findings.

Parents' View on Technology Use at Home

Our experiment is designed to explore technology-enhanced remote teaching with parental support in Luxembourgish early childhood education. Thus, the parents' view on technology use at home is crucial to our research. Touchscreen devices have been used by children for a number of years even before the age of 2 (Burns and Gottschalk, 2019; Dardanou et al., 2020; Rizk and Hillier, 2020). In contrast, parents' knowledge about the quality of apps is rather limited, and they are mostly concerned about

the use of technology could harm their children (Chaudron et al., 2015; Papadakis et al., 2019, 2022). In our research, we aimed to address how the technological knowledge of parents influence their children in remote learning.

TASK DESIGN AND EXPERIENCE IN EARLY CHILDHOOD EDUCATION

The remote teaching in STEAM education in early childhood education with AR, CAD software, and 3D printing was planned with a pedagogical team of three class teachers in Luxembourg. For the thematic project weeks on "Robots and Geometry," the team agreed to propose a remote teaching activity to pupils aged 4–6 years. During the experience, schools were closed, and all the teaching happened remotely. The duration of the experience was limited to 3 weeks, the length of the project weeks. Furthermore, participation was opened to parents who were willing to join and participate actively. The 3D printing happened after 2 weeks remotely, in the presence of the pupils and teachers. We used the following project timeline (**Figure 1**).

Parents who enrolled with their children received two sessions of introduction to the modeling task, a design of a robot, and the basic explanation of the used CAD software and the AR function, with worked examples (Sweller et al., 2011). During these sessions, parents and their children asked questions and received details on the software's manipulations. In our first experience with remote teaching and technologies like CAD or AR, we observed the importance of such tutorial sessions before beginning the modeling task. After the tutorial week, parents worked with their children for 2 weeks on the modeling task (see **Figure 2**). The tasks were designed according to the four principles of Dienes (Dienes, 1960; Lieban, 2019): Construction, Multiple embodiments, Dynamic, and Perceptual variability. These principles supported different requirements to manage the heterogeneous group of students for the modeling process (e.g., dynamic principle with reproducing, constructivity principle with free design). Therefore, we offered children and their parents' several entries and goal levels. The assignments ranged from reproducing a robot with given shapes to designing a completely new robot with a free choice of elements of constructions. Parents exchanged through online chats, video conference tools, e-mails, and messageboards in the modeling task process with the researchers and teachers. Furthermore, they exchanged as peers on the messageboard.

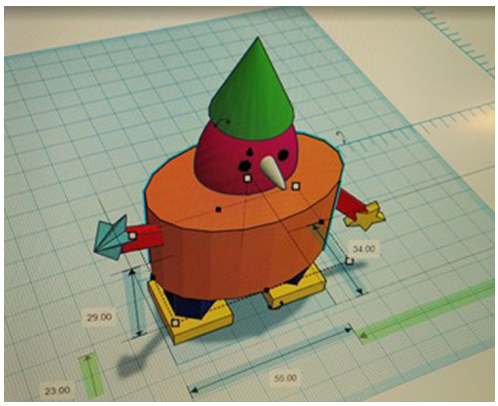


FIGURE 2 | A designed robot with the CAD software.

At the end of the modeling task, teachers, and researchers performed the 3D printing process of the robots in class with the pupils since it seemed important to show pupils the engineering part of the 3D printing in real time (e.g., functions of the printer, time to print, heat, and transformation of the filament).

METHODOLOGY AND METHODS

The iterative design-based process started with the first study in primary education, focusing on the parent's perspective within the learning process while designing a cultural artifact (Haas et al., Under Review). That study was design-based research on technology-enhanced environments (Wang and Hannafin, 2005) in remote teaching using AR, CAD, and 3D printing technologies. The second study with early childhood students, presented in this article, chose a qualitative approach (similar to the grounded theory approach) in a quasi-experimental design with participant's selections and no control group (Campbell and Stanley, 1963).

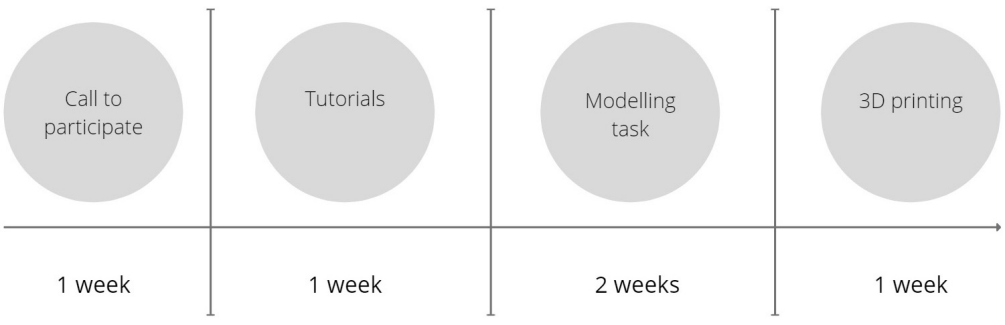


FIGURE 1 | Experience timeline in remote STEAM teaching in early childhood education.

TABLE 1 | Communication channels and their uses.

Communication channel	No. of parents using it
Online chats	$n = 8$
Online video conference tools	$n = 6$
Online email	$n = 5$
Online messageboard	$n = 12$

Participants

The participants of this study volunteered from a class of 36 pupils aged 4–6 in an early education class in Luxembourg. Parents were encouraged to enroll their children in the class communication channel. Of the 36 possible participants, 12 pupils with one parent each chose to join the experiment. Conditions to participate were to assist children in solving the tasks with technical and pedagogical support, responding to questions, and collecting data. Every participant finished the study and participated in the data collection. The group of participants was from a heterogeneous socioeconomic background. Since access to the Internet and remote learning content was assured by the school community in procuring hardware and software to families, there was no accessibility restriction observable.

Data Collection

We offered the parents with different communication opportunities: online chats, video conference tools, email, and messageboard. Parents chose to use different channels, and we collected all communication from each channel. As indicated in **Table 1**, parents used the communication channels differently. Some parents used only one channel, others multiple channels. During the study, we collected overall 129 messages from parents *via* the different communication channels, with a clear dominance from the messageboard. Communication from emails and videoconferences was used mainly in the week on tutorials and less often than other communication sources.

Data Analysis

The collected quotes were coded using a grounded theory approach (Glaser and Strauss, 2005; Corbin and Strauss, 2014) involving the application of inductive reasoning. Details of the coding are explained in the last paragraph of this subsection. Grounded theory differs from traditional scientific models of research that follows the order (1) theoretical framework, (2) hypothesis, and (3) data collection (to assess the validity of the hypothesis). When using a grounded theory approach, hypotheses and theories emerge from the analysis of the collected data. Thus, this methodology starts with a wide range of data collection and then subsequent detailed analysis.

In this second study on remote teaching with technology in early childhood education, we compared the coded wordings to our developed framework (**Figure 3**) in remote teaching with CAD and 3D printing (Haas et al., Under Review). This framework identified four major elements influencing remote teaching with parent assistance and digital modeling tasks. First, the perception of STEAM courses and teaching and students' motivation influence the learning in remote teaching and technical knowledge of parents, and scaffolding of parent-assisted teaching influences the teaching by parents. Based on these four elements and focusing on our mentioned research questions in early childhood education, we coded the collected data.

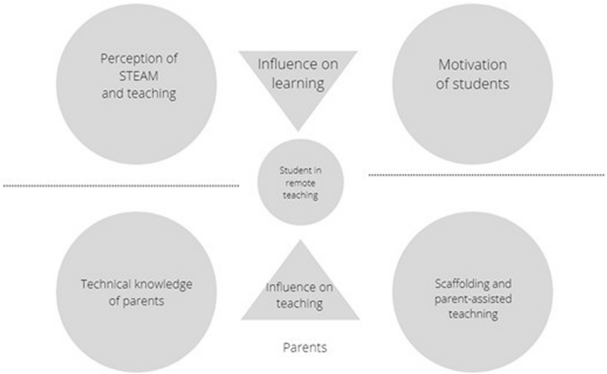


FIGURE 3 | Framework on learning in remote teaching with parent-assisted teaching.

Quotes from parents using different communication channels (e.g., video conference tool, email/messageboard, and online chats) were thus coded and connected to the elements in **Figure 3**. Our iterative coding process identified findings for every element we analyzed, described, and discussed in more detail in this section, with a clear focus on early childhood education. Different codes were compared to the first study in remote teaching from the design-based research and discussed.

We coded the messages in three rounds, starting with open coding (Glaser and Strauss, 2005). In the second round, we tried to connect the coded message to the framework, established from our first study, into those influencing the learning of pupils and those influencing teaching of children by parents. In the last coding, we coded the messages to the four categories on the perception of STEAM course and teaching, motivation of students, parents' technical knowledge, and parent-assisted teaching and scaffolding. Six quotes were not possible to categorize, as they were not relevant for the topic (e.g., messages on other teaching activities, school openings, and private appointments).

In **Table 2**, we regrouped recurrent message examples and coded categories. This section highlighted our Data Analysis, and in the following "Discussion" section, we explained in more detail the collected and categorized data.

RESULTS AND DISCUSSION

In this study, we observed if these tasks were likely to positively affect the STEAM learning behavior of pupils and how the parents were involved in their learning. Recalling our research questions relating to parent and student perception, students' motivation changes, parents' technological knowledge, and parental assistance, we presented two elements of perception

TABLE 2 | Examples of parents' messages and codes.

Categories	Examples of parents' messages
Perception on STEAM course and teaching	X used Geometry to build the robot What are the names of the objects you can move? I asked him the names of the shapes Will the robot stand on these cubic feet? Is the weight in the robot balanced? I showed X how to transform a shape into another
Students' motivation	This is better than doing it on paper X is happy to print her own robot We can't wait to see if the robot could be printed It is easier to do this than to complete the worksheets My son and I spent a good time trying out the different shapes on the robot
Technical knowledge of parents	How do you change the view? I can't find all the shapes Are there videos or explanations on how to drill into a shape? I find it hard to work on the Ipad, it is rotating too fast Will it save automatically?
Scaffolding and parent-assisted teaching	I ask questions like yes or no to see if he understands X has done so well, I am telling her all the time Should we share the great work of students? X if you show your son first how to do it, he can do it himself

STEAM, Science–Technology–Engineering–Arts–Mathematics.

of STEAM courses and teaching and motivation of students' learning behavior. Then, with the elements of technical knowledge of parents and scaffolding and parent-assisted teaching, we described the parents' perception in a similar role as a teacher. Overall, we can observe the high participation of the 12 parents and children, a positive acceptance of such tasks and strong interconnectivity between parents during the remote teaching. Furthermore, the acceptance of such tasks by the teachers evolved during the experience was subject to integration into the upcoming regular courses.

Parent and Child Perception of STEAM Course and Teaching

Parents and children from the study perceived the teaching of STEAM skills in early childhood in general interconnected but mentioned at the beginning of the modeling tasks with the CAD software, the importance of learning mathematics more in-depth separately. This might be related to the fact that currently only mathematics and no STEAM education is included in the curriculum (MENJE, 2018). According to early childhood research in mathematics education, parents' behavior toward learning (Elliott and Bachman, 2018) and the parents' socioeconomic backgrounds (DeFlorio and Beliakoff, 2015) impacts the learning behavior and perception of the students. Hence, the more parents invest in playful and continuous support, the more likely the pupils succeed in early childhood education (Cheung and Kwan, 2021).

Nonetheless, according to Vandermaas-Peeler et al. (2009) parents with low socioeconomic backgrounds tend to use less specific knowledge and vocabulary in supporting their children. Thus, beliefs and perceptions in education play a crucial role for early childhood students in learning STEAM skills. As mentioned earlier, parents supported strong interdisciplinarity with physical, face-to-face teaching in early childhood. However, parents did not know that the remote modeling tasks supported their children in STEAM skills (e.g., recognition and modulation of shapes). Beliefs and perceptions were similar to our previous research findings (Haas, 2021) from the first remote teaching experience on eggcup creation, where we observed a change of perception of STEAM courses and teaching among children and their parents. In addition, during the modeling tasks in early childhood, we observed a change in perception by all participating parents (Haas et al., Under Review). Moreover, with this second experience in early childhood, we confirmed a similar perception change as for the first experience.

Parents reported, mainly in the messageboard, that their children used "geometry," had to "think about the names of the shapes," or "combined cubes and transform shapes to obtain a robot." Thus, pupils used math words to describe their creations in the CAD software and observed the diversity of a given shape. Compared to non-technology-enhanced tasks, such as building a robot with wooden construction blocks, shapes can be transformed and combined with CAD software. The manipulation with the software supported parents and children to use math vocabulary to construct the robot together.

Furthermore, parents reported on the added value by modeling shapes in every direction, size, and color. The usefulness of such tasks for parents was reported several times during the online chats, and both parents and children appreciated the playful approach. In addition to mathematical vocabulary, parents reported on physical rules they discussed with their children. A question like "Will the robot stand on these cubic feet? Is the weight in the robot balanced?" Critics on technology use in early childhood were often criticized for not connecting educational tasks to a real-life situation, which was reported essential in the early childhood development (Rushton, 2011). Regarding the parents' report, they perceived the modeling task, once engaged, as a learning activity that playfully connects to real life.

Figure 4 shows 4 printed CAD software designs of robots. Each design presented differences in stabilization and positions. Parents worked with their children on shapes and a more general STEAM combined approach. This scientifically supported approach in the parent–child relationship was thus observable throughout the modeling task and supports a positive learning development with explorations as requested by research in learning in early childhood (Hu et al., 2021).

In the upcoming section, we described results related to the motivation of students, based on coded data from parents.

Student Motivation Changes

Several studies researched and discussed motivational factors using technologies like AR or CAD software (Ng, 2017; Bacca Acosta et al., 2019; Haas et al., 2021a). Thus, technology-enhanced tasks with digital modeling are more likely to support students in understanding 3D geometry, support the

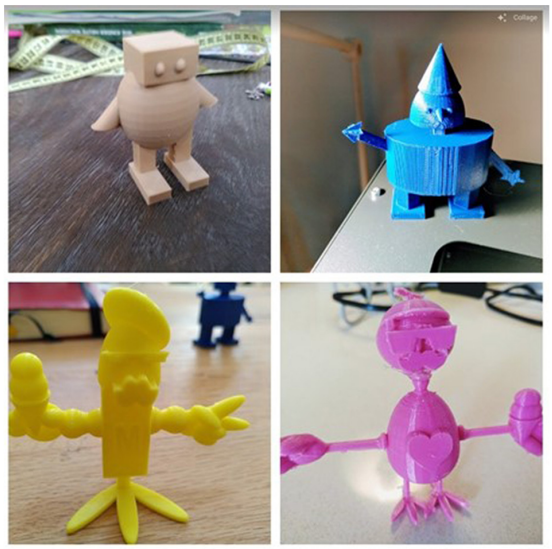


FIGURE 4 | Printed robots of the modeling task.

development of additional solving strategies, and delve into new learning behaviors (e.g., manipulations of geometric objects, experiment with objects, or transform properties). In our experience in early childhood, both parents, and children reported a higher motivation to solve the tasks than the standard tasks received during the confinement (e.g., paper–pencil tasks with a closed setting). Moreover, as for the first experience, they confirmed that working together with such a task was positive. Modeling the robot in the CAD software was perceived as a "playful construction game" with a "really unique result" at the end of the process. Parents reported that their children were motivated to experiment with the software and "see if the robot could be printed" and "how it could vary by adding or changing shapes and objects." Consequently, it appears that the experimental approach in verifying the feasibility of the design (e.g., Can it stand detached? Is it printable?), the modulation of shapes (e.g., What happens if we raise the number of sides of a pyramid? What varies?) or the creative combination of shapes, forms, and given objects, renders the tasks highly motivational in terms of learning. Pupils seemed to be in a flow (Nakamura and Csikszentmihalyi, 2014), similar to our first experience, where the task supports a positive learning motivation.

Moreover, Carlton and Winsler (1998) described tasks that support the free exploration of a student's environment and allowed self-regulation of goals to foster intrinsic motivation in early childhood. In addition, parents reported a positive effect on learning since the modeling task required a standard action between parents and children. This is a general trend in using technology in education with parents, called co-using (Chaudron et al., 2015). According to Dias and Brito (2021), co-using could lead to a negative motivation for children since parents decide on the educational content and manipulation. However, since the teachers gave the tasks, tools, and the criteria for the result, we observed equity of choosing in the designs between parents and children instead. Thus, pupils needed guidance and support from their parents and were motivated to create a design with their parents. Parents reported that pupils were "eager to present a robot" or "my son and I spent a good time trying out the different shapes on the robot." However, what we observed as well was that a high amount of time invested was only possible for parents who did not need to work outside the house (e.g., who were not in the home office or permanently at home). Another vital element within the modeling task was the technical knowledge of parents, which is described in the next section.

Parents Technical Knowledge

Over the last few years, parents have become more comfortable in letting their children in the early childhood age use technology for educational purposes and acknowledge their value-added for learning (Livingstone et al., 2015). In their role as digital mediators, they decide on the validity of educational technology or tasks and if it is suitable for their children. Parents who are more positive about using technology are thus more supportive of using it than parents, not in favor of educational technology (Nikken and Jansz, 2006; Dias and Brito, 2021). In the modeling tasks, parents were better prepared than in our first experience. This could be attributed to the proposed 1-week tutorial support from the teachers' and researchers' side. In worked-examples, messageboards, or videos, parents learned the content and manipulations of the used CAD software. Based on their replies, parents were less intimidated and insecure than in our previous experiences.

Furthermore, parents linked the modeling task to the everyday activities of their children (e.g., building with wooden bricks) and discussed the possibilities the software could offer. On the other hand, we observed a similarity to our first experience, the difference for parents in using the software. Parents working with similar software in their work-life (e.g., interior architect, road planer, or construction site manager) knew about more functions and manipulation strategies than those with more minor technology-enhanced jobs. The simple experience of using the smartphone or tablet for social media, games, or video streaming did not prepare parents, as reported in the online chats, for the manipulations in CAD software. Yet again, there was a high discrepancy between socioeconomic backgrounds. Manipulations with software like CAD are not generalized in public education and are thus reserved for those visiting higher education or specialized educational training (e.g., technology-enhanced works).

The tutorial week, however, aided parents to discover and manipulate the different functions, which helped the children in the manipulation of the task. Some parents reported that for them it seems easier to work on a computer with a mouse and for their children, the tablet (**Figure 5**) was "much easier" to manipulate for the digital design of the robot. Furthermore, the AR function on the application on the tablet allowed the transfer of the digital design into the real-world environment (**Figure 6**).

This was not done by every parent and child, since this manipulation was "too difficult" or "not understandable to



FIGURE 5 | Pupils working on a tablet.

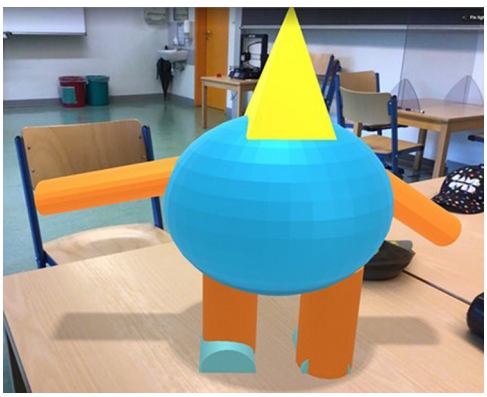


FIGURE 6 | AR visualization of a robot done by a child.

produce" by some of the parents. AR can develop new strategies but need to be used in a well-prepared and proficient-learning setting (Billinghurst and Duenser, 2012). Half of the parents noted that they would need additional training to use AR and to work with their children in this function.

Independently of the technological knowledge, the mediator role or parents' performance in manipulating CAD software, parents offered similar scaffoldings and parent-assisted teaching, which we described in the upcoming section.

Parental Assistance

During the modeling task with the CAD software in early childhood, the interactions between parents and their children were high and similar to interactions teachers would experience in class with their pupils. Similar to the research findings of Aram (2008) and Kucirkova et al. (2013), the digital task can support new, more intense interactions between the learner and the teacher (e.g., parent). Within the experience, parents used different scaffoldings to assist their children, similar to what a teacher would do (e.g., by dichotomous questions, rewards, subtargets in the construction process, or technical help). According to Neumann (2018), and this was similar to our first experience, parents use cognitive, affective, and technical scaffoldings when working with their children on a task. In **Table 3**, we grouped recurrent examples of the scaffoldings parents reported in one of the communication channels.

Based on recurrent messages in the communication channels, it appears that every parent used cognitive, affective, and technical scaffoldings to certain degrees of intensity. Parents who were described by teachers in general as more supportive showed in the data higher-affective support messages and showcased their children's work. We assume that the intensity of support in the modeling tasks with parent-assisted teaching is strongly related to their usual child–parent relationship, which was confirmed by the teachers in later discussions. However, similar to the findings of Wood et al. (2016), all the participating parents intended to positively support their children's STEAM learning experience during the task. The exchange on the messageboard, between parents, revealed moreover a desire to support the whole group of participating children. Thus, we observed how parents assisted each other in finding solutions to function manipulations or on how to best use the software in general. At the end of the tasks, parents acknowledge the importance of having a communication channel with support from teachers and exchange with peers. Peer learning among parents (Shilling et al., 2013), as well highlighted in our previous remote research experience, seems to have a supportive and likely positive effect on the accomplishment of remote tasks with parent-assisted teaching. According to Guldberg (2008), peer learning among adults can support the community of practice, which could lead to an improvement in support for the entire community. Compared to regular school teaching activities, the participation of parents and the exchange on school subjects and pedagogical techniques were much higher in this remote modeling task. A continuation of using exchange platforms during regular STEAM education could eventually improve the overall participation of all parents in the learning process.

At the end of the modeling task, the designs of the pupils were printed with the school 3D printer in their presence and tested on functionality (**Figure 7**). Each child received his/her robot and took it back home. Parents posted the designs and the prints of their children on the messageboard, and pupils received positive feedback from the community of participants. While we initially wanted to observe how parents and children react to the remote teaching with CAD software in STEAM education, we realized at the end of the experience, how important the community of parents and children is in the learning process overall. There are often barriers (Hornby and Blackwell, 2018) for parents to be involved in the educational community (e.g., bad own experience in school, low verbal communication skills, and social fears) and be supportive of their children. This is leading to exclusions, which do have strong effects on students' scholarly and social development (Sime and Sheridan, 2014). The community within the modeling task, however, worked like an enabler for parents' engagement in school activities and gave guidance in supporting their children. It could be a supportive place where parents could work together, similar to a real-world community, but mixed in its socioeconomic factors, backgrounds, and skills. As stated by Ainsworth (2002) that it takes a community (or village) to raise a child and support it, the proposed remote practice with support could eventually reduce the gap between parents' communal possibilities to support their children in STEAM education.

TABLE 3 | Observed scaffoldings done by parents.

Type of scaffolding	Description of the scaffolding	Recurrent examples
Cognitive	As a cognitive scaffolding, we considered supports on conceptual and procedural understanding of the modeling task.	Parents used dichotomous questions to explain differences in shapes or to support students in combining several shapes. Parents showed examples and asked students to recreate shapes or combinations of shapes. Parents used examples to explain the functionalities of a design.
Affective	As an affective scaffolding, we considered supports that were positively encouraged to realize and stay on task within the modeling task.	Parents encouraged students by saying positive words about the work or the designs of the children. Parents complimented their children on their capabilities to design with the software and to create their own robot.
Technical	As a technical scaffolding, we considered supports that procured help with the features and manipulations within the CAD software.	Parents explained and showed functions within the software, as the drag and drop function or to change the view of the design. Parents showcased with examples how to combine shapes, change lengths and heights, place shapes on the platform, or change colors of the designs. Parents showed videos from the Internet to explain functions to their children.

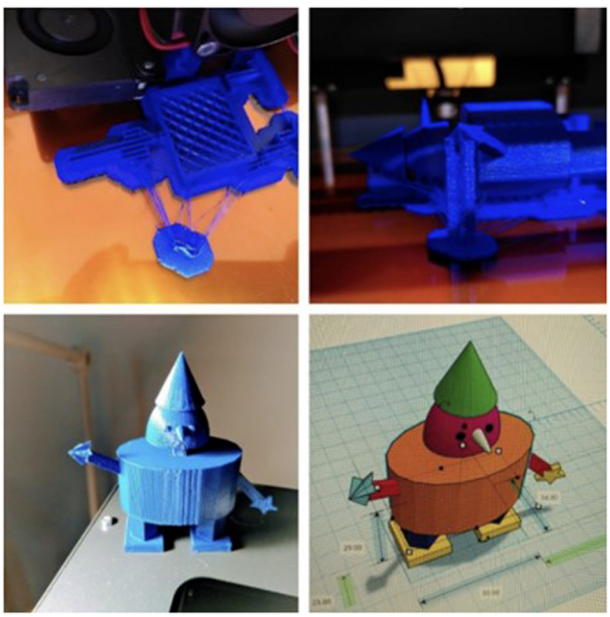


FIGURE 7 | Printing result and process.

CONCLUSION

The present study was designed to evaluate technology-enhanced, STEAM-based remote teaching with parental support in Luxembourgish early childhood education. We formulated four research questions to confirm and supporting the framework from our first study (Haas et al., Under Review). We observed several similarities and thus confirmations on children and parents in remote teaching with CAD software. Our framework with the four elements fitted the collected and coded data during the modeling task. In relation to the research questions we found that (1) STEAM was perceived as interconnected which is characterized by the use of mathematical vocabulary and the testing of physical rules, (2) tasks were highly motivational in terms of learning, (3) manipulations of CAD software were not common and differed based on socioeconomic background, whereas the use of AR was too difficult to use for half of the parents, and (4) interactions between parents and their children were comparable to the ones between teachers and students, and strongly related in intensity of the usual child–parent relationship. An interesting fact is that parents preferred to use computers with a mouse, whereas children preferred touch-based tablets for the modeling tasks.

Accordingly, there were changes in the perception that could have effects beyond the modeling task for parents, pupils, and in the retrospective for the teachers. What started as an experience of new technology used in remote teaching revealed opportunities not only on the content and skill level in STEAM but also, overall, in the educational process. The experience led parents to participate actively with their children, question STEAM contents, experiment with scaffoldings, and connect to each other. We observe scientific approaches and high motivation among early childhood students, who actively designed and created real-world objects with STEAM skills.

Of course, this experience was reduced to 12 participants and one parent each, and findings should again be confirmed and discussed in future research. Questions on community learning for parents and children in remote teaching for STEAM education became clearer and need more data, which should be collected in the next research stances. In particular, socioeconomic backgrounds need detailed consideration. Probably, an eco-learning system, where communication channels, digital technologies, like CAD software, AR, and 3D printing, and professional and peer resources are made available, would provide a promising setting. Finally, we need to consider the long-term effects of remote learning in STEAM education in early childhood. Children and parents benefit from active learning technology and social interactions through proposed communication channels with professional support. Training and workshops in teacher education should prepare schools to propose similar experiences in future.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because the dataset cannot be shared, as agreed with participants' legal representatives. Requests to access the datasets should be directed to BH, ben.haas@outlook.com.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

BH produced the first draft of the manuscript and edited the research design. ZL consulted on the methodology and methods and edited and reviewed the manuscript. TH reviewed and edited the manuscript. YK consulted on the research design, reviewed, edited, and formatted the manuscript. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We thank the Millermoaler School in the Echternach community for their participation.

REFERENCES

Ainsworth, J. W. (2002). Why does it take a village? The mediation of neighborhood effects on educational achievement. *Soc. Forces* 81, 117–152. doi: 10.1353/sof.2002.0038

Aram, D. (2008). Parent–child interaction and early literacy development. *Early Educ. Dev.* 19, 1–6. doi: 10.1080/10409280701838421

Bacca Acosta, J. L., Baldiris Navarro, S. M., Fabregat Gesa, R., and Kinshuk, K. (2019). Framework for designing motivational augmented reality applications in vocational education and training. *Australas. J. Educ. Technol.* 35, 102–117. doi: 10.14742/ajet.4182

Billinghurst, M., and Duenser, A. (2012). Augmented reality in the classroom. *Computer* 45, 56–63. doi: 10.1109/MC.2012.111

Burnard, P., Dragovic, T., Jasilek, S., Biddulph, J., Rolls, L., Durning, A., et al. (2018). "The art of co-creating arts-based possibility spaces for fostering STE(A)M practices in primary education," in *Arts-Based Methods in Education Around the World River Publishers Series in Innovation and Change in Education - Cross-Cultural Perspective*, eds T. Chemi and X. Du (Gistrup: River Publisher), 245–279. doi: 10.13052/rp-9788793609372

Burns, T., and Gottschalk, F. (2019). *Educating 21st Century Children: Emotional Well-Being in the Digital Age*. Paris: OECD Publishing. doi: 10.1787/b7f33425-en

Campbell, D. T., and Stanley, J. C. (1963). "Experimental and quasi-experimental designs for research on teaching," in *Handbook of Research on Teaching*, ed. N. L. Gage (Chicago, IL: Rand McNally).

Carlton, M. P., and Winsler, A. (1998). Fostering intrinsic motivation in early childhood classrooms. *Early Child. Educ. J.* 25, 159–166. doi: 10.1023/A:1025601110383

Cesarone, B. (2007). ECAP Report: early childhood learning standards: effects on teaching practices. *Child. Educ.* 84, 60–61. doi: 10.1080/00094056.2007.10522974

Chaudron, S., Beutel, M. E., Černikova, M., Donoso Navarrete, V., Dreier, M., Fletcher-Watson, B., et al. (2015). *Young Children (0-8) and Digital Technology: A qualitative exploratory Study Across Seven Countries*. Luxenbourg: Publications Office of the European Union.

Cheung, S. K., and Kwan, J. L. Y. (2021). Parents' perceived goals for early mathematics learning and their relations with children's motivation to learn mathematics. *Early Child. Res. Q.* 56, 90–102. doi: 10.1016/j.ecresq.2021.03.003

Chilton, H., and Mccracken, W. (2017). New technology, changing pedagogies? Exploring the concept of remote teaching placement supervision. *High. Educ. Pedagog.* 2, 116–130. doi: 10.1080/23752696.2017.1366276

Corbin, J., and Strauss, A. (2014). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*, 4th Edn. Thousand Oaks, CA: SAGE Publications, Inc.

Dana-Picard, T., Hershkovitz, S., Lavicza, Z., and Fenyvesi, K. (2021). Introducing golden section in the mathematics class to develop critical thinking from the STEAM perspective. *Southeast Asian J. STEM Educ.* 2, 151–169.

Dardanou, M., Unstad, T., Brito, R., Dias, P., Fotakopoulou, O., Sakata, Y., et al. (2020). Use of touchscreen technology by 0–3-year-old children: parents' practices and perspectives in Norway, Portugal and Japan. *J. Early Child. Lit.* 20, 551–573. doi: 10.1177/1468798420938445

DeFlorio, L., and Beliakoff, A. (2015). Socioeconomic status and preschoolers' mathematical knowledge: the contribution of home activities and parent beliefs. *Early Educ. Dev.* 26, 319–341. doi: 10.1080/10409289.2015.968239

Dias, P., and Brito, R. (2021). Criteria for selecting apps: debating the perceptions of young children, parents and industry stakeholders. *Comput. Educ.* 165, 1–13. doi: 10.1016/j.compedu.2021.104134

Dienes, Z. P. (1960). *Building up Mathematics*. London: Hutchinson Educational.

El Bedewy, S., Lavicza, Z., Haas, B., and Lieban, D. (2022). A STEAM practice approach to integrate architecture, culture and history to facilitate mathematical problem-solving. *Educ. Sci.* 12:9. doi: 10.3390/educsci12010009

Elliott, L., and Bachman, H. J. (2018). SES disparities in early math abilities: the contributions of parents' math cognitions, practices to support math, and math talk. *Dev. Rev.* 49, 1–15. doi: 10.1016/j.dr.2018.08.001

Glaser, B. G., and Strauss, A. L. (2005). *Grounded theory : Strategien qualitativer Forschung*. Bern: Huber.

Guldberg, K. (2008). Adult learners and professional development: peer-to-peer learning in a networked community. *Int. J. Lifelong Educ.* 27, 35–49. doi: 10.1080/02601370701803591

Guntur, M. I. S., and Setyaningrum, W. (2021). The effectiveness of augmented reality in learning vector to improve students'. spatial and problem-solving skills. *Int. J. Interact. Mob. Technol.* 15, 159–173. doi: 10.3991/ijim.v15i05.19037

Haas, B. (2021). *Transition from in-Class to Outdoor Learning with Real-world Mathematical Modelling*. Linz: Johannes Kepler University Linz.

Haas, B., Kreis, Y., and Lavicza, Z. (2020). "Fostering process skills with the educational technology software MathemaTIC in elementary schools," in *Proceedings of the 10th ERME TOPIC CONFERENCE (ETC10) on Mathematics Education in the Digital Age (MEDA)*, eds A. Donevska-Todorova, E. Faggiano, J. Trgalova, Z. Lavicza, R. Weinhandl, A. Clark-Wilson, et al. (Linz, Austria: Johannes Kepler University), 199–206.

Haas, B., Kreis, Y., and Lavicza, Z. (Under Review). Parent's experience in remote learning with digital and physical mathematical modelling. *Res. Pract. Technol. Learn.*

Haas, B., Kreis, Y., and Lavicza, Z. (2021a). Case study on augmented reality, digital and physical modelling with mathematical learning disabilities students in an elementary school in Luxemburg. *Int. J. Technol. Math. Educ.* 28, 125–132. doi: 10.1564/tme_v28.3.02

Haas, B., Kreis, Y., and Lavicza, Z. (2021b). Integrated STEAM approach in outdoor trails with elementary school pre-service teachers. *J. Educ. Technol. Soc.* 24, 205–219.

Haatainen, O. M., and Aksela, M. (2021). Project-based learning in integrated science education: active teachers' perceptions and practices. *LUMAT Int. J. Math. Sci. Technol. Educ.* 9, 149–173. doi: 10.31129/LUMAT.9.1.1392

Hirschland, D. (2008). *Collaborative Intervention in Early Childhood: Consulting with Parents and Teachers of 3- to 7-Year-Olds*. Oxford: Oxford University Press.

Hornby, G., and Blackwell, I. (2018). Barriers to parental involvement in education: an update. *Educ. Rev.* 70, 109–119. doi: 10.1080/00131911.2018.1388612

Hu, X., Chiu, M. M., Leung, W. M. V., and Yelland, N. (2021). Technology integration for young children during COVID-19: towards future online teaching. *Br. J. Educ. Technol.* 52, 1513–1537. doi: 10.1111/bjet.13106

Jablonski, S., and Ludwig, M. (2020). "Development of an Intensive Study Programme on Outdoor Mathematics Teaching with Digital Tools," in *Research on Outdoor STEM Education in the Digital Age: Proceedings of the ROSETA Online Conference in June 2020 Conference Proceedings in Mathematics Education*, eds M. Ludwig, S. Jablonski, A. Caldeira, and A. Moura (Münster: WTM), 111–118.

Kelley, T. R., and Knowles, J. G. (2016). A conceptual framework for integrated STEM education. *Int. J. STEM Educ.* 3, 1–11. doi: 10.1186/s40594-016-0046-z

Kreis, Y., Haas, B., Reuter, R., Meyers, C., and Busana, G. (2020). "Reflections on our teaching activities in the initial teacher training during the COVID-19 crisis: from 'onsite classes' to 'schooling at home'," in *Self and Society in the Corona Crisis: Perspectives from the Humanities and Social Sciences The Ends of the Humanities*, eds G. Mein and J. Pause (Esch-sur-Alzette: Melusina Press). doi: 10.26298/c4j7-5x48

Kucirkova, N., Messer, D., Sheehy, K., and Flewitt, R. (2013). Sharing personalised stories on iPads: a close look at one parent–child interaction. *Literacy* 47, 115–122. doi: 10.1111/lit.12003

Lavicza, Z., Fenyvesi, K., Lieban, D., Park, H., Hohenwarter, M., Mantecon, J. D., et al. (2018). "Mathematics learning through arts, technology and robotics: multi- and transdiscpilinary STEAM approaches," in *Proceedings of the 8th ICMI-East Asia Regional Conference on Mathematics Education* (Taipei: National Taiwan Normal University), 110–122.

Lavicza, Z., Haas, B., and Kreis, Y. (2020). "Discovering everyday mathematical situations outside the classroom with MathCityMap and GeoGebra 3D," in *Research on Outdoor STEM Education in the digiTal Age: Proceedings of the ROSETA Online Conference in June 2020 Conference Proceedings in Mathematics Education*, eds M. Ludwig, S. Jablonski, A. Caldeira, and A. Moura (Münster: WCTM), 23–30. doi: 10.37626/GA9783959871440.0.03

Liao, C. (2016). From interdisciplinary to transdisciplinary: an arts-integrated approach to STEAM Education. *Art Educ.* 69, 44–49. doi: 10.1080/00043125.2016.1224873

Lieban, D. (2019). *Exploring Opportunities for Connecting Physical and Digital Resources for Mathematics Teaching and Learning*. Linz: Johannes Kepler University.

Liu, E., Li, Y., Cai, S., and Li, X. (2019). "The effect of augmented reality in solid geometry class on students' learning performance and attitudes," in Smart Industry & Smart Education Lecture Notes in Networks and Systems (LNNS), eds M. E. Auer and R. Langmann (Cham: Springer), 549–558. doi: 10.1007/978-3-319-95678-7_61

Livingstone, S., Mascheroni, G., Dreier, M., Chaudron, S., and Lagae, K. (2015). *How Parents of Young Children Manage Digital Devices at Home: The Role of Income, Education and Parental Style*. London: EU Kids Online, LSE.

MENJE (2018). *Plan-cadre Pour L'éducation Précoce au Luxembourg*. Luxembourg: Ministére de l'Éducation Nationale, de l'Enfance et de la Jeunesse.

MENJE (2020a). *L'apprentissage á Distance, bien Vécu par La Majorité des Enseignants et des Parents*. Available online at: https://gouvernement.lu/fr/actualites/toutes_actualites/communiques/2020/04-avril/01-apprentissage-script.html (accessed February 9, 2022).

MENJE (2020b). *Plan D'études - Contenus Prioritaires : C1 - C4*. Luxembourg: Ministère de l'Éducation Nationale, de l'Enfance et de la Jeunesse.

Nakamura, J., and Csikszentmihalyi, M. (2014). "The concept of flow," in *Flow and the Foundations of Positive Psychology: The Collected Works of Mihaly Csikszentmihalyi, ed. M. Csikszentmihalyi* (Dordrecht: Springer Netherlands), 239–263. doi: 10.1007/978-94-017-9088-8

Neumann, M. M. (2018). Parent scaffolding of young children's use of touch screen tablets. *Early Child Dev. Care* 188, 1654–1664. doi: 10.1080/03004430.2016.1278215

Ng, O., and Chan, T. (2019). Learning as making: using 3D computer-aided design to enhance the learning of shape and space in STEM-integrated ways. *Br. J. Educ. Technol.* 50, 294–308. doi: 10.1111/bjet.12643

Ng, O.-L. (2017). Exploring the use of 3D computer-aided design and 3D printing for STEAM learning in mathematics. *Digit. Exp. Math. Educ.* 3, 257–263. doi: 10.1007/s40751-017-0036-x

Ng, O.-L., Shi, L., and Ting, F. (2020). Exploring differences in primary students' geometry learning outcomes in two technology-enhanced environments: dynamic geometry and 3D printing. *Int. J. STEM Educ.* 7:50. doi: 10.1186/s40594-020-00244-1

Nikken, P., and Jansz, J. (2006). Parental mediation of children's videogame playing: a comparison of the reports by parents and children. *Learn. Media Technol.* 31, 181–202. doi: 10.1177/0956797610362670

OECD (2021). *Using Digital Technologies for Early Education during COVID-19: OECD Report for the G20 2020 Education Working Group*. Paris: OECD Publishing. doi: 10.1787/fe8d68ad-en

Papadakis, S., Alexandraki, F., and Zaranis, N. (2022). Mobile device use among preschool-aged children in Greece. *Educ. Inf. Technol.* 27, 2717–2750. doi: 10.1007/s10639-021-10718-6

Papadakis, S., Zaranis, N., and Kalogiannakis, M. (2019). Parental involvement and attitudes towards young Greek children's mobile usage. *Int. J. Child Comput. Interact.* 22:100144. doi: 10.1016/j.ijcci.2019.100144

Pearson, H. A., and Dubé, A. K. (2021). 3D printing as an educational technology: theoretical perspectives, learning outcomes, and recommendations for practice. *Educ. Inf. Technol.* 27, 3037–3064. doi: 10.1007/s10639-021-10733-7

Rausch, A., Bold, E., and Strain, P. (2021). The more the merrier: using collaborative transdisciplinary services to maximize inclusion and child outcomes. *Young Except. Child.* 24, 59–69. doi: 10.1177/1096250620922206

Rizk, J., and Hillier, C. (2020). "Everything's technology now": the role of technology in home- and school-based summer learning activities in Canada. *J. Child. Media* 15, 272–290. doi: 10.1080/17482798.2020.1778498

Rushton, S. (2011). Neuroscience, early childhood education and play: we are doing it right! *Early Child. Educ. J.* 39, 89–94. doi: 10.1007/s10643-011-0447-z

Sarkar, P., Kadam, K., and Pillai, J. S. (2020). Learners' approaches, motivation and patterns of problem-solving on lines and angles in geometry using augmented reality. *Smart Learn. Environ.* 7:17. doi: 10.1186/s40561-020-00124-9

Shilling, V., Morris, C., Thompson-Coon, J., Ukoumunne, O., Rogers, M., and Logan, S. (2013). Peer support for parents of children with chronic disabling conditions: a systematic review of quantitative and qualitative studies. *Dev. Med. Child Neurol.* 55, 602–609. doi: 10.1111/dmcn.12091

Sime, D., and Sheridan, M. (2014). 'You want the best for your kids': improving educational outcomes for children living in poverty through parental engagement. *Educ. Res.* 56, 327–342. doi: 10.1080/00131881.2014.934556

Steen, K., Brooks, D., and Lyon, T. (2006). The impact of virtual manipulatives on first grade geometry instruction and learning. *J. Comput. Math. Sci. Teach.* 25, 373–391.

Steenbergen-Hu, S., and Cooper, H. (2013). A meta-analysis of the effectiveness of intelligent tutoring systems on K–12 students mathematical learning. *J. Educ. Psychol.* 105, 970–987. doi: 10.1037/a0032447

Stone, B., Kay, D., Reynolds, A., and Brown, D. (2020). 3D Printing and service learning: accessible open educational resources for students with visual impairment. *Int. J. Teach. Learn. High. Educ.* 32, 336–346.

Sweller, J., Ayres, P., and Kalyuga, S. (2011). *Cognitive Load Theory*. New York, NY: Springer. doi: 10.1007/978-1-4419-8126-4

Takeuchi, M. A., Sengupta, P., Shanahan, M.-C., Adams, J. D., and Hachem, M. (2020). Transdisciplinarity in STEM education: a critical review. *Stud. Sci. Educ.* 56, 213–253. doi: 10.1080/03057267.2020.1755802

UNESCO (2021). *Education: From Disruption to Recovery*. Available online at: https://en.unesco.org/covid19/educationresponse (accessed February 9, 2022).

Vandermaas-Peeler, M., Nelson, J., Bumpass, C., and Sassine, B. (2009). Numeracy-related exchanges in joint storybook reading and play. *Int. J. Early Years Educ.* 17, 67–84. doi: 10.1080/09669760802699910

Wang, F., and Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educ. Technol. Res. Dev.* 53, 5–23. doi: 10.1007/BF02504682

Wood, E., Petkovski, M., De Pasquale, D., Gottardo, A., Evans, M. A., and Savage, R. S. (2016). Parent scaffolding of young children when engaged with mobile technology. *Front. Psychol.* 7:690. doi: 10.3389/fpsyg.2016.00690

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Haas, Lavicza, Houghton and Kreis. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Psychology

ORIGINAL RESEARCH
published: 21 June 2022
doi: 10.3389/fpsyg.2022.904761

Educational Robotics Intervention to Foster Computational Thinking in Preschoolers: Effects of Children's Task Engagement

Anaclara Gerosa[1], Víctor Koleszar[1], Gonzalo Tejera[1,2], Leonel Gómez-Sena[1,3] and Alejandra Carboni[1,4]*

[1]Centro Interdisciplinario en Cognición para la Enseñanza y el Aprendizaje, Espacio Interdisciplinario, Universidad de la República, Montevideo, Uruguay, [2]Instituto de Computación, Facultad de Ingeniería, Universidad de la República, Montevideo, Uruguay, [3]Laboratorio de Neurociencias, Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay, [4]Instituto de Fundamentos y Métodos, Facultad de Psicología, Universidad de la República, Montevideo, Uruguay

Computational thinking (CT) is a broadly used term in education to refer to the cognitive processes underlying the application of computer science concepts and strategies of problem-solving. Recent literature has pointed out the value of children acquiring computational thinking skills (i.e., understanding and applying concepts, such as conditionals, iteration, or generalization), especially while learning STEM subjects. Robotics has been used as a tool to introduce computational thinking and STEM knowledge to children. As physical objects, robots have been proposed as developmentally appropriate for the early childhood setting, promoting motivation and allowing young learners to represent abstract ideas in a concrete setting. This study presents a novel educational robotics (ER) intervention using RoboTito, a robot programmable through tangible elements in its environment designed for kindergarteners. We used a quasi-experimental design with an active control group. In addition, we conducted a structured observation of the filmed material of the sessions to gather data on children's attention and motivation throughout the activities. Fifty-one children (male = 33; mean age = 66 months, SD = 5.49 months) attending level 5 (kindergarten) at a Uruguayan public school participated in the study. Children in our experimental condition participated in an intervention programming RoboTito using tangible elements, while children in our control condition played with the robot through sensory-motor activities using a remote control and did not engage in programming. Motivational and attentional factors were assessed through video-recorded sessions of the ER activities. Four trained observers blind to the experimental conditions participated in the coding. Children's interactions were assessed in four categories: task engagement, distractibility, oral participation, and objective fulfillment. Our results suggest children's task engagement mediated their gains in CT after the intervention; *post-hoc* Tukey contrasts revealed non-significant pre-test to post-test gains for the control and low engagement groups, and significant for the high engagement group. Overall, we conclude task engagement played a central role in children's learning gains and our robotics intervention was successful in promoting CT

OPEN ACCESS

Edited by:
Stamatios Papadakis, University of Crete, Greece

Reviewed by:
Spyros Vosinakis, University of the Aegean, Greece
Kan Kan Chan, City University of Macau, Macao SAR, China

***Correspondence:**
Anaclara Gerosa
agerosa@psico.edu.uy

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

Received: *25 March 2022*
Accepted: *02 May 2022*
Published: *21 June 2022*

Citation:
Gerosa A, Koleszar V, Tejera G, Gómez-Sena L and Carboni A (2022) Educational Robotics Intervention to Foster Computational Thinking in Preschoolers: Effects of Children's Task Engagement. Front. Psychol. 13:904761. doi: 10.3389/fpsyg.2022.904761

for engaged children. We discuss the practical implications of our results for early childhood education and developmentally appropriate ER targeted for young learners.

Keywords: computational thinking, robotics, task engagement, cognitive development, early childhood, preschool

INTRODUCTION

Several efforts in the last decades have been done to introduce computational thinking (CT) to educational practice and curriculums in several countries throughout the world (Bocconi et al., 2016). The term CT became popular in education after Jeanette Wing's 2006 on computational thinking (Wing, 2006), which has, since then, been cited multiple times as a highly relevant contribution toward this field (Grover and Pea, 2013; Voogt et al., 2015). Wing (2006) defined CT as solving problems and designing to understand human behavior by drawing from computer science and later as a thought process involving representing problems and their solutions algorithmically so that they can be solved by a computer (Wing, 2011). Since then, CT has been embraced in educational settings to describe the thought processes behind computer science and programming as well as the socio-emotional predispositions which make these possible. For example, the International Society for Technology in Education (ISTE) and the Computer Science Teacher Association (CSTA) proposed an operational definition that describes CT as a problem-solving process that spans characteristics, such as formulating problems algorithmically, logically organizing data, achieving representation through abstraction, automatization, procuring time and resource efficiency, and generalization, but also highlight confidence, persistence, tolerance to ambiguity and communication as supporting factors (ISTE, 2015).

CT has been included in various educational settings, ranging from early childhood and preschool education to university levels (Grover and Pea, 2013; Lyon and Magana, 2020; Fagerlund et al., 2021). The inclusion of CT notions in formal education has taken many forms, which include its integration in both computer science courses and its embedding in different disciplines, such as math (Weintrop et al., 2016; Hickmott et al., 2018), science (Sneider et al., 2014; Swanson et al., 2019), and art (Bell and Bell, 2018). Moreover, CT has reached classrooms during the instruction of programming (Zhang and Nouri, 2019; Papadakis, 2021, 2022) through robotics (Ioannou and Makridou, 2018) and unplugged activities, such as board games or storybooks (Huang and Looi, 2021).

Particularly in early childhood settings, CT has often been included through the use of educational robotics and unplugged activities. Programmable robots have been proposed as a developmentally appropriate tool to introduce young children to CT under the rationale that as physical objects, robots could allow preschool children to learn in a non-restrictive embodied way, supporting gross motor development (Bers, 2021). Moreover, robots are tangible elements similar to the toys children manipulate daily, thus providing intuitive interfaces for early development (Horn and Bers, 2019).

Despite several tools being available to enhance children's CT (Yu and Roque, 2019) their assimilation for learning purposes in compulsory education has not been as straightforward (Repenning et al., 2010). Several challenges, from lack of teacher training and professional development opportunities (Caeli and Bundsgaard, 2020), to cost to classroom management (Bers, 2008) have been previously mentioned in the literature. Moreover, academic reporting of successful small-scale studies should consider possibilities for scalability of their results and discuss the adaptability of their findings into real-world classrooms (Bakala et al., 2021).

An underreported aspect of the inclusion of educational robotics to promote CT into classrooms is the effect of motivational and attentional factors, such as task engagement in children's learning outcomes. Engagement has been shown to be an essential part of learning. Fredricks et al. (2004) defined engagement as a meta-construct that includes behavioral (time performing a task), emotional (i.e., interest and motivation), and cognitive engagement (i.e., self-regulation). Specifically, behavioral engagement has been defined as the correspondence between the child's behavior and the situation's demands (Ponitz et al., 2009), and has been positively associated with academic achievement (Lei et al., 2018). Recent evidence from educational robotics (ER) has shown learning motivation to be associated with performance in problem-solving and computational thinking during primary school (Stewart et al., 2021). Thus, understanding children's engagement during ER tasks is highly relevant to their later performance.

The present study aimed to test the effectiveness of a set of ER activities for preschoolers in their CT performance. In order to accomplish this, we used a quasi-experimental design with an active control group. Additionally, we conducted structured observation on filmed material from the ER sessions to account for children's task engagement, distractibility, participation, and task objective fulfillment. Thus, our study contributes to the field of ER interventions aimed at promoting CT in the early childhood setting by providing empirical data on the effectiveness of this approach. Moreover, it is to our knowledge the first study in which observable behavioral aspects, such as children's engagement during tasks, is assessed and analyzed as a factor of children's performance in CT and ER tasks.

BACKGROUND

CT's Relation to Cognitive Development

Interventions which target children during early childhood have shown to have long-term effects on their academic achievement. Executive function (EF) skills refer to several top-down neurocognitive processes needed to regulate thoughts, emotions, and goal-oriented behavior (Zelazo et al., 2005; Diamond, 2013; Blair et al., 2016). During this stage in their development, children experience an exponential improvement in these skills, supported structurally by their prefrontal cortex (Perone et al., 2018). Thus, during this stage in development, children increase

their autonomy and are able to engage in goal-oriented behaviors (Doebel, 2020). Executive functions include abilities, such as attention shifting (flexibility), planning, and working memory (maintaining and manipulating information in mind; Diamond, 2013), and have previously been associated with CT (Robertson et al., 2020), in particular to programming and debugging.

Recently, researchers in the intersection of computing education and developmental psychology have studied the association between CT and young children's cognitive development (Gerosa et al., 2021; Tsarava et al., 2022) in an attempt to define CT empirically through studying its relation with well-established cognitive abilities at different points in development. These findings suggest there is an association between CT skills and early math skills at an early age, specifically during preschool education. Findings from older children attending middle school (10–14 years of age) found no correlation with math skills but found strong associations with language abilities and problem-solving (Román-González et al., 2017). Taken together, these findings suggest early numerical, and math skills are relevant for CT early on in development but later become less relevant as children acquire written language, and CT becomes increasingly intertwined with programming.

Using Educational Robotics to Promote CT in Young Learners

Educational robotics (ER) have been used to promote CT in young learners. Previous studies have tested the effectiveness of ER curriculums and activities toward teaching computational thinking. Kazakoff et al. (2013) showed a 1-week robotics intervention could improve kindergarten children's sequencing scores, while Bers et al. (2019) concluded that children as young as 3 years old could grasp CT concepts *via* robotics. Studies with slightly older children (Papadakis et al., 2016; González and Muñoz-Repiso, 2018; Jung and Won, 2018) have reached similar conclusions. However, despite a wide variety of commercial and non-commercial robots and kits being available (Sapounidis and Demetriadis, 2016; Yu and Roque, 2019), only a handful of them have been used for research purposes in an applied setting. BeeBot (Stoeckelmayr et al., 2011) has been used in educational robotics interventions in preschool education by several researchers. For example, Angeli and Valanides (2020) used Beebot with different scaffolding strategies, namely, narratives and cards to promote children's CT. Through a randomized control trial design, Di Lieto et al. (2017) found that an ER intervention using bee-bot improved preschoolers' executive function after 13 sessions. Similarly, Muñoz-Repiso and González (2019) found statistically significant post-test differences between its control and ER group on sequencing, action-instruction correspondence, and debugging which indicates a better performance of the group of 3- to 6-year-olds which took part in an ER intervention. Different versions of LEGO robotics (Kazakoff et al., 2013; Sullivan et al., 2013; Bers et al., 2014; Cho and Lee, 2017) and KIBO (Sullivan et al., 2017; Bers et al., 2019) have also been used in interventions targeted at preschool children and were successful in promoting CT, positive technological development and sequencing skills.

Bringing ER for CT Into Classrooms

In a recent review of the characteristics of educational robotics interventions to promote CT in preschoolers, we found several reporting gaps which could hinder their reproducibility and scalability (Bakala et al., 2021). Among the underreported details in interventions, we found several elements which are highly relevant for educators and practitioners willing to implement these findings, such as ER session's duration and frequency, children's group size, and adult's role in scaffolding the activities. Moreover, sutil differences in these factors such as classroom organization or overly large groups could affect whether children benefit from interventions. These findings also highlight that some of our current knowledge gaps regarding ER interventions to promote CT in young children rely on the contextual variables that appear when transferring contained interventions into everyday preschool settings. Moreover, ER activities have historically been a part of extracurriculars, camps, and competitions (Eguchi, 2007), which self-select for children who are more likely to be intrinsically motivated with these tasks. Thus, if CT through ER aims to be incorporated into classrooms, it is of special relevance to study factors, such as task engagement and participation as proxies to attentional control and motivation to tailor interventions that impact the most extensive possible set of children.

METHODOLOGY

Design and Procedure

We used a quasi-experimental design with pre-test and post-test assessments. Children were randomly assigned to either experimental condition or control conditions. Our experimental condition consisted of an ER intervention with RoboTito (Bakała et al., 2019; Gerosa et al., 2019), a robot designed to be programmable through the arrangement of objects in its environment and has been successfully used by young children. Children in the experimental condition took part in an 11-session educational robotics intervention designed to promote CT. We implemented an active control group, meaning children assigned to the control condition got to play with the same robot but did not program its movements through manipulating its environment. Instead, children in the control condition played sensory-motor games with the robot controlling it remotely with a tablet, thus excluding the programming requirements present in our experimental condition. Groups were matched in gender, mean age and their pre-test scores in the fluid intelligence task.

All pre-test and post-tests assessments were conducted in three sessions of up to 25 min each. All children were assessed at school by trained researchers. Evaluations took place in the morning between 9 and 11 am. Paper-based assessments were applied individually in a 1:1 child–adult ratio, while computerized tablet-based measurements were applied concurrently in groups, following a 4:1 ratio between children and adults. Further information on the assessed variables during pre-test and post-test assessments is provided in section "Instruments."

Research Context and Sample

Fifty-one children (male = 33, female = 18; overall mean age = 66 months old, SD = 5.49 months) attending level 5 (kindergarten) at a public school in Montevideo, Uruguay participated in the study. Convenience sampling was implemented. Sociocultural levels for our sample were characterized as middle–high according to Uruguay's national administration of public education. Inclusion criteria consisted of children attending preschool level 5 (aged 4–6 years) with typical development. One child was excluded from our sample due to having a diagnosed developmental disorder. Informed consent was obtained from parents/caregivers, and the study was approved by the Research Ethics Committee of the School of Psychology, University of the Republic. All methods were performed in accordance with the Declaration of Helsinki.

About RoboTito

Preschool children recognize and name colors and basic shapes, make plans about playing, building, or drawing, and understand broad concepts of time. In this sense, the robot must allow the development of these skills, taking into account the appropriate cognitive abilities for that age.

The main design characteristics are ease of assembly in limited production runs, use of standardized and widely available components, robust enough to be used by children, flexibility for modifications, and an open hardware and software specification.

Robotito is a robot that defines its behavior according to the physical disposition of the elements found in its environment, presenting behaviors, such as following or dodging the elements, performing the trajectories, or looking for evading the elements. Robotito dimensions are 16.5 cm in diameter and 7.2 cm in height. The robot has no explicit front, it can freely move in any direction while simultaneously rotating around a vertical axis. The mobility base has three omnidirectional wheels. These wheels are composed of rotating sections that allow wheels to slip sideways freely (**Figure 1**).

As we intend to experiment with several robot–environment interaction modalities, and those depend on the sensing abilities of the robot, we equipped the robot with a basic sensor set. The sensor set mounted in the robot includes two types of sensors. First, it has installed six laser rangefinder sensors distributed equidistantly around the perimeter. The second type of sensor

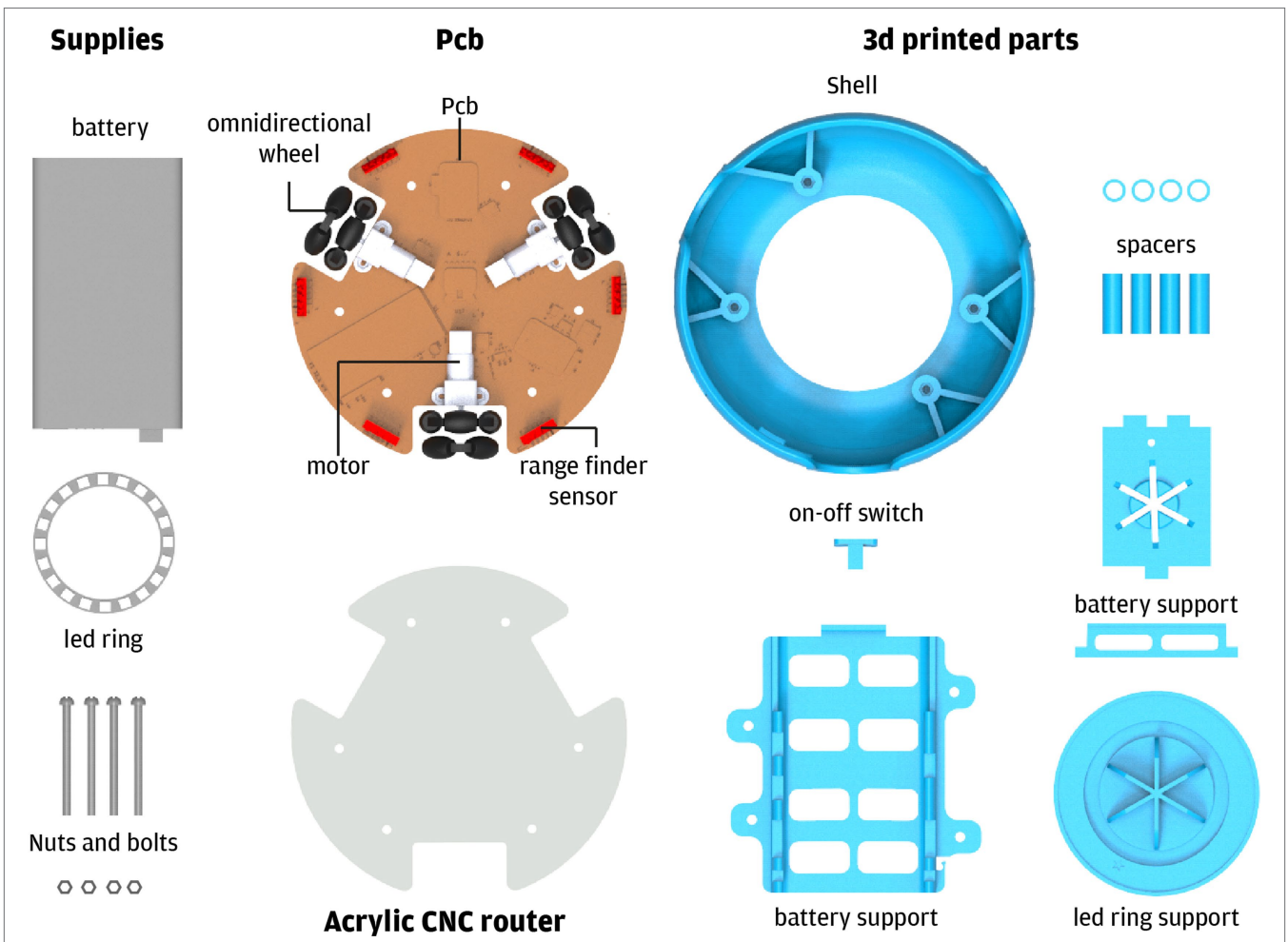


FIGURE 1 | Main components of RoboTito.

is a single combined color, distance, and gesture unit placed under the robot, in the geometric center, and pointing downwards.

The laser sensors measure distances to obstacles from a few millimeters out to about a meter. They allow the robot to react to objects placed in the vicinity, for example, feel an attractive or repulsive force from objects. The color sensor allows the robot to change wheels velocity based on color patches placed on the floor. For example, a specific color could cause the robot to move in a particular direction or rotate in place. Additionally, this sensor detects when the robot is picked up and disables the motors.

The robot logic follows a reactive paradigm. This organization means that the robot control is composed of simple behaviors, each one a simple rule that associates the input from sensors to an action. We developed two behaviors that allow the robot to interact with the environment. One uses the color sensor to indicate a direction to move. The other uses the distance sensor to allow the robot to be attracted by objects. Educators can create new interaction modes using the robot integrated development environment.

The Robotito user interface allows students to understand what the robot is sensing, and inform possible internal states that justify the robot's actions. The interface is provided by an array of 24 RGB LEDs placed in a circle. LEDs can be lighted on, control their intensity and color, and provide the user feedback on what the robot is doing and sensing. For example, LEDs could turn red on the side where an obstacle is detected or blink when the robot senses some color on the floor to represent that it reached its home. Also, the robot has installed a buzzer capable of emitting musical notes. Behaviors can use the buzzer to emit auditive cues or play a joyful tune on mission accomplishment.

Instruments

In the following sections, we briefly describe each of the assessments used during our study. Sections "CT Assessment" and "Fluid Intelligence" describe assessments implemented during baseline and post-test, while "Educational Robotics Task Analysis" describes the observation method applied to the video recordings of the ER sessions throughout the intervention.

CT Assessment

Adapted questionnaire based on Yune Tran's CT questionnaire (Tran, 2019). This questionnaire explores five CT constructs, namely, the ability to create algorithms, loops, debugging, inferring from a conditional statement, and sequencing. Children's answers for each task were dummy coded for scoring (scoring range: 1–12). For example, one item in the questionnaire required children to create an algorithm using arrows which depicted four directions (right, left, backward, and forward) to reach an objective in the plane. Scale reliability was acceptable (Cronbach's alpha: 0.72).

Fluid Intelligence

Tablet-based version of Raven's colored progressive matrices (Raven and Court, 1986). This instrument has seen widespread use in research contexts, has undergone validation in a Latin American population (Pasquali et al., 2002), and has shown stability across time and cultures (Raven, 2000). The task requires children to identify the correct missing pattern from the stimuli in a six-option multiple-choice format. Fluid intelligence was used as a control variable in the context of our regression analyses.

Educational Robotics Task Analysis

Video-recorded sessions of ER activities were analyzed for the experimental condition ($N = 27$, male = 18, female = 9, mean age = 5.4 years, SD = 5.8 months). Five-minute intervals of each session (starting point set to the time point in which each task's objective was first instructed to children) were used for data analysis. Four trained observers participated in the coding of each session. Inter-observer reliability was high, ranging from 83 to 100%. Four variables were explored in the experimental condition:

- ER task engagement: Defined as the total amount of seconds the child engages in either manipulating the robot or the intervention materials, answering the coordinator's inquiries, or pointing or directing his or her attention in a task-relevant way.
- Number of switches: Defined as the number of times the child transitions between engaged and disengaged states throughout the observation.
- Number of relevant oral participations: Total number of times the child participates orally during the task in ways that are relevant to solving it (whether his or her proposals lead to the correct solution or not).
- ER task objective fulfillment: Children's performance during the task was coded in regards to their accomplishment of objectives. A task score of 2 = totally accomplished objectives, 1 = partially accomplished objectives, or 0 = did not accomplish objectives. An "insufficient information" score was used if behavioral cues were deemed insufficient for observers to make a judgment of accomplishment of objectives. Objective fulfillment scores were added up in order to create a final score.

ER Intervention Structure and Description

ER sessions were implemented in groups of 5–7 children and lasted 25–30 min each (Gerosa et al., 2019). A spare classroom within the school was used to carry out the activities, and their frequency averaged at 1.5 sessions per week. Each group of children had its own mat and a robot to play with. A maximum of two groups were able to participate in the activities simultaneously in the space. A total of 11 sessions were carried out and led by two members of the research team who worked as group coordinators (one in each group working simultaneously or alternating if there was only one group). **Figure 2** depicts a typical setting during our ER activities.

Group coordinators would propose and explain each activity to the children, answer their questions and lead the session. A maximum of two undergraduate students per group participated in assisting the session and helping children who required extra help.

Sessions 1–7 consisted of children programming the robot using cards to engage its color sensor. Session 1 consisted of an introductory activity. In this session, the group coordinator would talk to children about the general rules for the workshop and introduce them to the idea of playing with a robot to solve

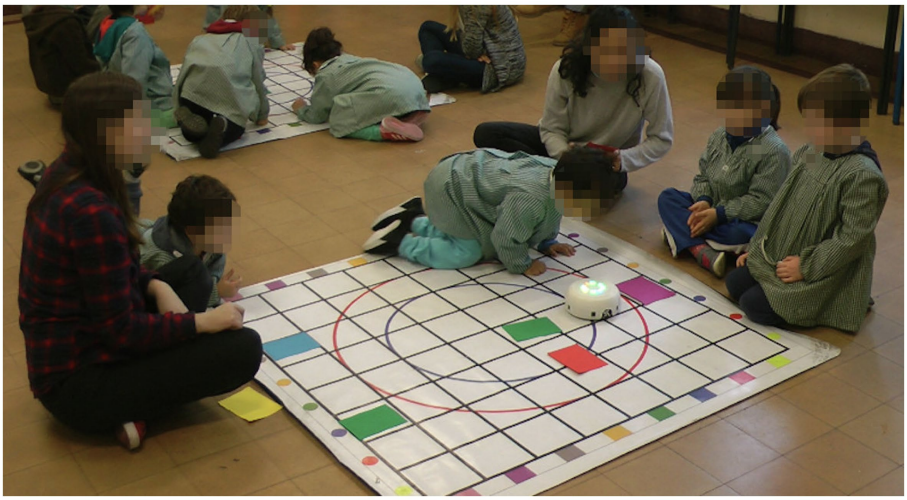


FIGURE 2 | Example of a group of children trying to solve one of the proposed challenges during the intervention with RoboTito.

different situations. We talked about their pre-existing notions about what a robot is, how it looks and its purpose. Lastly, we introduced the robot RoboTito and worked with them to identify its sensors, how it moves and how to turn it on and off. Session 2's main objective was to establish a simple goal and work on spatial concepts, such as backward, forward, left, and right. We asked children to create programs for short trajectories and reviewed how each card influenced the robot's direction and movement. Lastly, we asked children to arrange the cards in a way that the robot would never stop moving. Given the functioning of the robot using the color sensors, this task required children to arrange the cards in the correct order in the shape of a square. A detailed explanation of this task is shown in **Figure 3**.

Session 3 involved children using the previously learnt notions about the robot's lights and its associations with color and direction to complete sequences to reach a predetermined objective (**Figure 4**). We introduced the purple card as a target which served as feedback and showed to be a motivator for children to reach their objective. In session 4, we asked children to look at a given configuration of cards in the mat and try to predict the robot's behavior. Children would point and signal to their hypothesized trajectory given the predefined setting of the cards. Then, we asked them to modify the robot's trajectory using the cards based on their observations. Session 5 consisted of planning and creating sequential movements while being prompted to focus on resource efficiency. Explicitly, children were asked to use the least number of cards possible and think about the shortest trajectory when creating their sequence. In session 6, we incorporated distracting objects to promote children inhibiting irrelevant elements in the robot's environment setting. This increased the task's difficulty slightly, as children were required to avoid the unnecessary objects while planning their sequence. Session 7 consisted of children looking at a pre-set erroneous configuration and taking the necessary steps to reach the objective, thus debugging the given program.

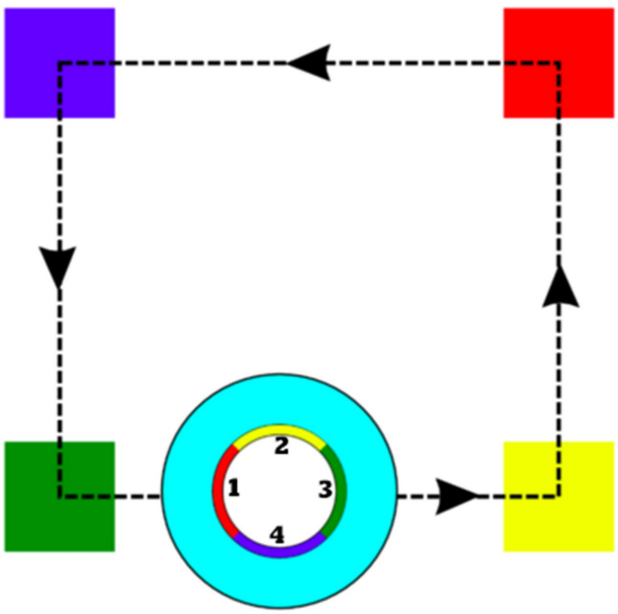


FIGURE 3 | Card configuration for programming the robot to do a continuous trajectory in a square shape. Numbered areas signal the LED lights in the robot, which signal the direction the robot will take after sensing each color card (In this case: if green to the right, if yellow forward, if red to the left, if blue backward).

Sessions 8–11 were implemented using the robot with its distance sensors. Session 8 introduces and familiarizes the children with this new way of functioning of the robot and identifying its sensors. Children were tasked to try to move the robot placing their hands in front of the sensors and to pay attention to what happens when their hands are closer or further away from them. This allowed us to introduce the notion of range when dealing with the distance sensors. Session

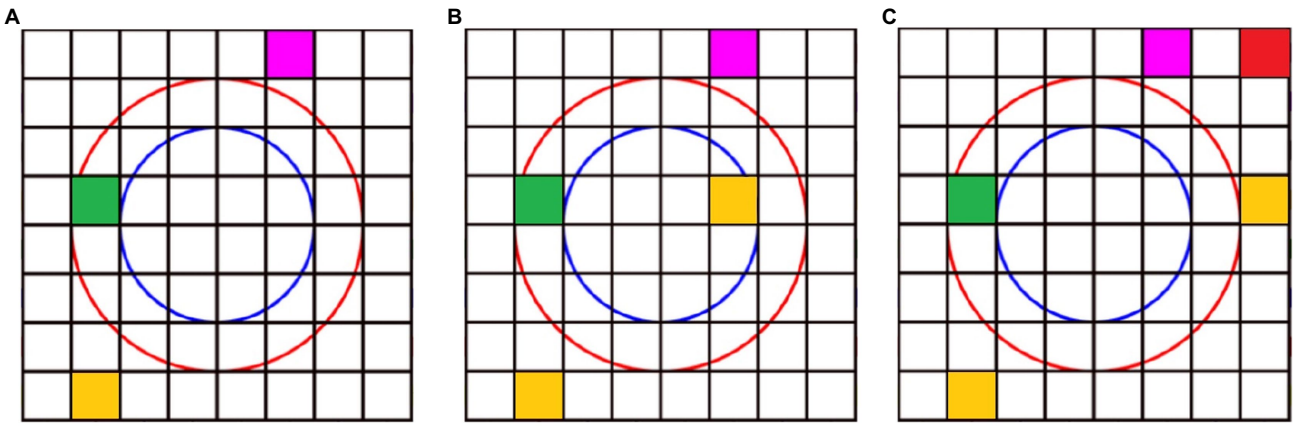


FIGURE 4 | Example of a task using the robot's color sensors. This task is analogous to those presented in sessions 3 and 4. Taking into account the robot would be positioned in the way shown in **Figure 3**, the representation in **(A)** shows the setting arranged before asking children the following: "If we have these cards set and cannot move them, which card would we need, and where would we put it to reach our objective of the purple card?" **(B)** Shows the solution to this question using just one card in the correct position. **(C)** Shows a correct answer to reach its objective, albeit using more cards.

TABLE 1 | Structure of the intervention plan.

Session	Active sensor	Phase
1	Color	Introductory
2	Color	Simple goal-setting
3	Color	Simple goal-setting
4	Color	Predicting behavior
5	Color	Planning and resource efficiency
6	Color	Incorporating distractors
7	Color	Debugging
8	Distance	Introductory
9	Distance	Trying to guess how these sensors work
10	Distance	Embodied experience pretending to be robot
11	Distance	Simple goal-setting and debugging

9 consisted of children trying to infer the rules of functioning of the robot under the distance sensor modality. Specifically, we allowed for free exploration time with the robot and tasked them to guess if they could find a pattern in its actions through testing different conditions. We explicitly introduce the robot's underlying rules in session 10, explaining to children that the robot in this modality had two main rules that guided its behavior: firstly, it cannot sense an object out of range for the sensors. Secondly, once it senses its surrounding objects within range, it will approach the object that is furthest away from it. We used an embodied approach to facilitate their learning of these rules through asking them to imagine they were the robot themselves. Thus, children were tasked to perform the correct movements using these rules and considering their current settings. Finally, session 11 involved predicting the robot's behavior, integrating the knowledge from sessions 8–10, and later implementing it using the robot. If the target was not met, we asked them to create hypotheses on what happened and to try to alter the setting to debug their configurations to obtain the desired result. **Table 1** presents a summary of our intervention.

Data Analysis

Statistical analysis was performed using R and R Studio software (Team R, 2019). Mixed-effects linear models (MLM) were implemented to test for the effects of the intervention. We included principal and interaction effects of time (pre and post-test measures) and group (experimental groups and control), fluid intelligence scores were used as a control variable, while random effects were composed of individuals nested within classrooms. In order to test whether our task observation variables (task engagement, objective fulfillment, participation, and switching) were factors capable of modulating intervention effects, we divided children in our experimental group into high and low engagement groups. Each variable was thus discretized into two separate factor levels using the median. Thus, this allowed us to divide children into three groups for comparison (control, low engagement, and high engagement groups). Fluid intelligence was used as a control variable in order to prevent a confounding effect. *Post-hoc* Tukey tests were performed in order to test the existence of within-group effects of performance gains before and after assessments.

RESULTS

Descriptive Statistics

Table 2 displays general descriptive statistics for each group and our overall sample, including their age, baseline scores in fluid intelligence, and gender.

Intervention Effects on Children's CT

Table 3 presents each group of children's average performance on the CT evaluation both before and after the intervention. Children in our experimental condition (i.e., programming the robot through rearranging objects in its environment) obtained overall higher scores post-test; however, these differences in results did not show statistical significance to the control group.

TABLE 2 | Descriptive statistics.

	Control group (*N* = 24)	Experimental group (*N* = 27)	Overall (*N* = 51)
Age (months)			
Mean (SD)	66.8 (5.03)	65.3 (5.87)	66.0 (5.49)
Median [Min, Max]	68.0 [54.0, 75.0]	66.0 [55.0, 76.0]	67.0 [54.0, 76.0]
Missing	2	2	4
Fluid intelligence			
Mean (SD)	10.6 (5.55)	10.3 (5.72)	10.4 (5.58)
Median [Min, Max]	11.0 [0, 20]	9.0 [2, 21]	11.0 [0, 21]
Gender (N)			
Boys	15	18	33
Girls	9	9	18

Age (in months), baseline fluid intelligence scores, and gender for each group.

TABLE 3 | Pre-test and post-test mean scores in CT for each our control and experimental groups.

	Control group (*N* = 24)	Experimental group (*N* = 27)
CT score (mean, SD)		
Pre-test	3.58 (2.18)	3.76 (2.58)
Post-test	4.99 (2.57)	5.79 (3.11)

Effect of Observational Outcomes on Children's CT Scores Per Group

Figure 5 shows children's CT scores before and after our ER intervention according to their grouping factor (control group, children who presented low levels of task engagement, children who presented high levels of engagement). Our results show children's task engagement mediated their gains in CT after the intervention [$F(2) = 4{,}25$; $p < 0.05$]. We performed *post-hoc* Tukey contrasts which revealed significant pre-test to post-test gains for the high engagement group ($p < 0.01$) yet non-significant for the control ($p = 0.92$) and low engagement ($p = 0.99$) groups.

Figure 6 shows children's CT scores before and after our ER intervention according to their grouping factor for objective fulfillment (A), oral participation (B), and switching (C). While these variables did not show statistical significance, they present a similar pattern to that of **Figure 6** which shows a tendency to favor the experimental condition.

Spearman Correlation Between Pre-test and Post-test Gains and Observational Measures

Table 4 shows Spearman correlations between CT gains (post-test minus pre-test) and each of the observed variables. We found positive, significant correlations between CT gains and children's

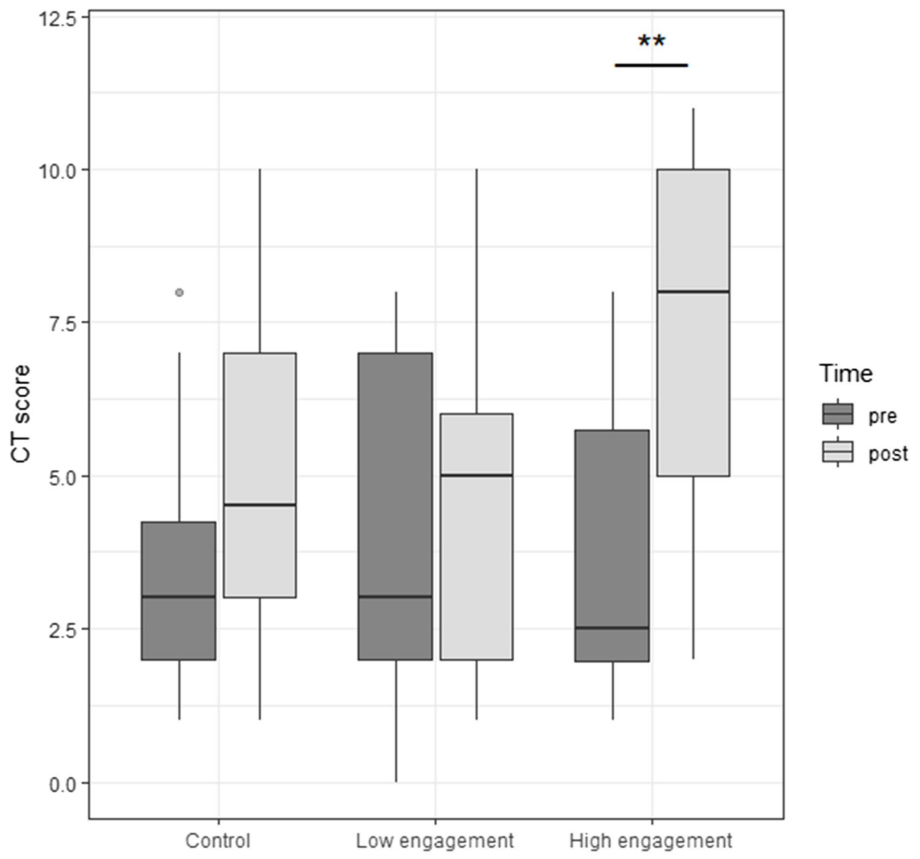


FIGURE 5 | Children's pre-test and post-test CT score for control and different levels of engagement in the experimental condition.

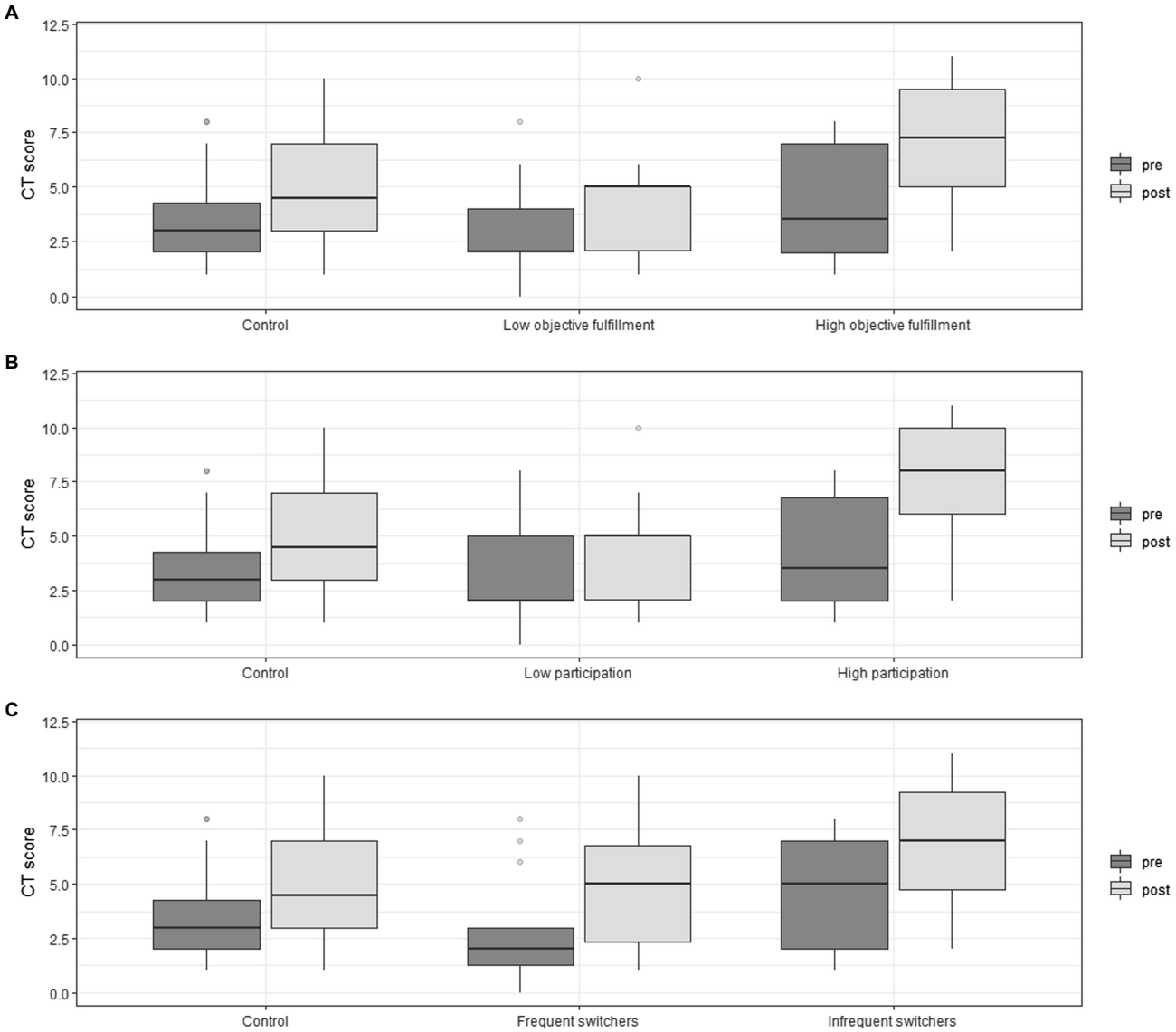


FIGURE 6 | Children's pre-test and post-test CT score for control and different levels of **(A)** objective fulfillment, **(B)** Oral participation during tasks, and **(C)** number of switches between ON-task state and OFF-task state in the experimental condition.

TABLE 4 | Spearman correlation between CT gains (Δ) and children's task engagement, switching, oral participation, and objective fulfillment scores.

	CT (Δ)
Observable variables	
Engagement	0.51**
Switching	−0.3
Oral participation	0.46*
Objective fulfillment	0.39

(*) significance level below threshold of 0.05.
(**) significance level below threshold of 0.01.

time on-task (task engagement), mean number of oral participations during task, and average score in objective fulfillment.

DISCUSSION

This study investigated whether a controlled ER intervention using a robot programmable through its environment had positive effects on young children's CT. Our research involved a quasi-experimental design with an active control group and explored motivational and attentional variables throughout the intervention. These variables were recorded through structured observation of our filmed material and included children's task engagement, number of switches between engaged and disengaged states, number of oral participations throughout the task, and ER objective fulfillment.

Our intervention activities in ER were designed taking into account previous literature on educational robotics which was

targeted specifically at preschool children (Kazakoff et al., 2013; Sullivan et al., 2013, 2017; Bers et al., 2014; Ioannou and Makridou, 2018) and involved children solving goal-oriented problems through programming the robot's environment accordingly to ensure it navigated the space correctly.

Our results showed 5 year old children in the experimental condition who presented high engagement in the activities significantly increased their overall CT skills, while children with low engagement and children in the control condition did not. Overall, our results highlight that attentional and motivational factors, such as children's engagement, are relevant and could modulate the benefits of ER on children's CT.

Generally, our results are aligned with previous evidence on the possibility of improving CT through ER at an early age and contributed to understanding how context-related factors might impact controlled interventions. While studies have shown evidence that ER is an effective way to introduce young children to CT (Bakala et al., 2021), the effects of environmental factors, such as task engagement, were not reported. Previous evidence from older children (Sharma et al., 2019) has shown engagement as a relevant factor in children's attitudes toward robot programming (i.e., their self-confidence in the task); however, to our knowledge there has not been a quantitative study in which task engagement is specifically related to CT and ER outcomes in young children, despite engagement being pointed out as a relevant variable in these stages (Critten et al., 2021).

One of the strengths of the present study is that CT was assessed through a questionnaire independent of the intervention tools, while most previous studies opted to rely on ER performance as a proxy to CT (Sullivan et al., 2017; González and Muñoz-Repiso, 2018; Bers et al., 2019; Saxena et al., 2020), thus we were able to infer that any benefits would be indeed related to a cognitive skill rather than resulting from training in a specific task.

Moreover, much of the current evidence on ER interventions is often limited by the lack of control groups and quantitative assessments. In a recent review of empirical studies on CT through robotics for preschoolers Bakala et al. (2021) found that only 26% of the reviewed studies reported the use of control groups and experimental or quasi-experimental designs.

Only a few of the previous empirical studies in ER to promote CT did include assessments that were independent from the intervention tools. Such is the case of work by Nam et al. (2019), who used picture sequencing and mathematical problem-solving tasks as proxies to CT and Cho and Lee (2017) in which children were asked to self-report on their efficacy and interest in the subject. However, the independent assessments are dissimilar between studies, involving a wide range of abilities that include problem-solving but also socio-emotional skills, such as self-confidence to perform the tasks. In the last year, diagnostic CT assessments which could be independently applied to young children have been developed and validated. Thus future studies should gradually incorporate these types of assessments (Relkin and Bers, 2019; Zapata Cáceres et al., 2020).

Examining context-related variables through structured observation of the experimental condition allowed us to shed light into some of the factors that could enhance or prevent the success of these types of interventions. Thus, our results highlight the importance of maintaining children's engagement and fostering their interest throughout the process. Further studies should examine how individual factors, such as children's interest in robotics, as well as previous exposure to similar activities, could enhance their ability to succeed in these tasks. Furthermore, aspects, such as scaffolding techniques, group size, child:adult ratio, and other variables that could potentially impact proper engagement, should be further examined in order to identify the best practices for maximizing positive results. So far, most of the existing data consist of case studies or small-scale research (Jung and Won, 2018). For example, a case study by Janka (2008) indicated that introducing storytelling to their activities was an integral part of promoting meaningful learning instances using educational robots. Additionally, the authors recommend small groups with up to five children per teacher as an adequate way to organize classrooms for effective learning, which was approximately the amount of children per group used in the present study.

Recent studies, such as those performed by Angeli and Valanides (2020) and Zhong and Si (2021), pose interesting questions and provide budding evidence on the way different scaffolding techniques impact children and teenagers' performance during robotics' tasks. Moreover, a recent review by Atmatzidou et al. (2018) confirms that studies with strong levels of guidance generally obtain better results, while their own experimental data from 11- to 16-year-olds showed groups that received more questions and prompts to help understand the problems, design and evaluate solutions throughout the tasks were more successful than those who were allowed to explore freely. All of the aforementioned variables are determinant to the feasibility and scalability of the ER interventions proposed. Thus, further evidence is required to identify best practices and extract useful guidelines for teachers interested in introducing ER and CT as classroom activities.

The intervention designed for the present study generally meets these recommendations and our results showed our intervention was successful in promoting computational thinking skills for highly engaged preschoolers. Our results regarding the different effects for children with low and high engagement call for the need to control these variables during interventions and design interventions which maintain children with high levels of engagement.

LIMITATIONS

The confounding nature of attention and motivation in a natural educational setting does not allow us to infer which of these processes is causing this effect. Children who are highly motivated by ER are probably more likely to pay more attention to the tasks and tools, while children with better executive functioning might have better cognitive resources to engage in the tasks and thus be more attentive throughout the activities. Further research should be conducted to control these variables: for example, this could be achieved by including questionnaires to account for children´s intrinsic motivation toward ER before the intervention.

Another limiting factor might be that our ER assessment scores were extracted from structured observation of the natural

ER learning setting. Filmed material often lacks the flexibility of in-person assessment and the control provided by individual evaluation. Further studies might consider adding a brief individual ER assessment through a structured task before and after the intervention in order to have an independent measure. For example, temperamental factors or personality traits at play during group dynamics might have skewed the external observer's ability to determine children's skills. Arguably, more extroverted children might have had more chances to showcase their skills than introverted children.

Despite this, our observational approach to the intervention could also be considered a strength, as it allowed us to gather data that was highly ecological in nature and depicts the group setting and dynamics similarly to those of real-life classrooms.

CONCLUSION

This study assessed the effects of an 11-session educational robotics intervention in preschoolers' CT skills. The intervention consisted of a set of activities using a robot programmable through tangible objects in its environment. We used a quasi-experimental design and active control group to test the effects of the intervention. Our results show evidence for positive effects of this particular intervention in children who were highly engaged throughout the activities. These findings have implications for educational practitioners and researchers, as it sheds light into the importance of designing engaging interventions and assessing children's attentional and motivational factors throughout the activities to assure engagement is maintained.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics committee from the School of Psychology, University of the Republic, Uruguay (UdelaR). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

AG: writing—original draft, formal analysis, and investigation. VK: formal analysis and investigation. GT: conceptualization, supervision, writing—original draft, and writing—review and editing. LG-S: conceptualization and supervision. AC: conceptualization, supervision, and writing—review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

The present study was funded by Uruguay's National Agency for Research and Innovation (ANII) funds FSED_2_2017_1_138793 as well as the Interdisciplinary Cognition Center for Research and Learning (CICEA-UDELAR).

ACKNOWLEDGMENTS

The authors wish to thank all children, parents, teachers, and administrators who participated in the present project, as well as Uruguay's national agency for research and innovation (ANII) for funding this research. We would also like to thank Nicolás Capricho who collaborated on the project and created the images used to depict RoboTito in the present paper.

REFERENCES

Angeli, C., and Valanides, N. (2020). Developing young children's computational thinking with educational robotics: An interaction effect between gender and scaffolding strategy. *Comput. Hum. Behav.* 105:105954. doi: 10.1016/j.chb.2019.03.018

Atmatzidou, S., Demetriadis, S., and Nika, P. (2018). How does the degree of guidance support students' metacognitive and problem solving skills in educational robotics? *J. Sci. Educ. Technol.* 27, 70–85. doi: 10.1007/s10956-017-9709-x

Bakala, E., Gerosa, A., Hourcade, J. P., and Tejera, G. (2021). Preschool children, robots, and computational thinking: a systematic review. *Int. J. Child Comput. Interact.* 29:100337. doi: 10.1016/j.ijcci.2021.100337

Bakała, E., Visca, J., Tejera, G., Seré, A., Amorin, G., and Gómez-Sena, L. (2019). "Designing child-robot interaction with Robotito." in *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)* (IEEE), 1–6.

Bell, J., and Bell, T. (2018). Integrating computational thinking with a music education context. *Inf. Educ.* 17, 151–166. doi: 10.15388/infedu.2018.09

Bers, M. U. (2008). *Blocks to Robots Learning with Technology in the Early Childhood Classroom*. New York, NY: Teachers College Press.

Bers, M. U. (2021). *Coding as a Playground: Programming and Computational Thinking in the Early Childhood Classroom. 2nd Edn*. New York, NY: Routledge.

Bers, M. U., Flannery, L., Kazakoff, E. R., and Sullivan, A. (2014). Computational thinking and tinkering: exploration of an early childhood robotics curriculum. *Comput. Educ.* 72, 145–157. doi: 10.1016/j.compedu.2013.10.020

Bers, M. U., González-González, C., and Armas-Torres, M. B. (2019). Coding as a playground: promoting positive learning experiences in childhood classrooms. *Comput. Educ.* 138, 130–145. doi: 10.1016/j.compedu.2019.04.013

Blair, C., and McKinnon, R. D. Family Life Project Investigators (2016). Moderating effects of executive functions and the teacher–child relationship on the development of mathematics ability in kindergarten. *Learn. Instr.* 41, 85–93. doi: 10.1016/j.learninstruc.2015.10.001

Bocconi, S., Chioccariello, A., Dettori, G., Ferrari, A., Engelhardt, K., Kampylis, P., et al. (2016). Developing computational thinking in compulsory education. European Commission, JRC Science for Policy Report, 68.

Caeli, E. N., and Bundsgaard, J. (2020). Computational thinking in compulsory education: a survey study on initiatives and conceptions. *Educ. Technol. Res. Dev.* 68, 551–573. doi: 10.1007/s11423-019-09694-z

Cho, Y., and Lee, Y. (2017). Possibility of improving computational thinking through activity based learning strategy for young children. *J. Theor. Appl. Inf. Technol.* 95, 1–9.

Critten, V., Hagon, H., and Messer, D. (2021). Can pre-school children learn programming and coding through guided play activities? A case study in computational thinking. *Early Childhood Educ. J.* 1–13. doi: 10.1007/s10643-021-01236-8

Di Lieto, M. C., Inguaggiato, E., Castro, E., Cecchi, F., Cioni, G., Dell'Omo, M., et al. (2017). Educational robotics intervention on executive functions in preschool children: a pilot study. *Comput. Hum. Behav.* 71, 16–23. doi: 10.1016/j.chb.2017.01.018

Diamond, A. (2013). Executive functions. *Annu. Rev. Psychol.* 64, 135–168. doi: 10.1146/annurev-psych-113011-143750

Doebel, S. (2020). Rethinking executive function and its development. *Perspect. Psychol. Sci.* 15, 942–956. doi: 10.1177/1745691620904771

Eguchi, A. (2007). "Educational robotics for elementary school classroom." in *Society for Information Technology and Teacher Education International Conference* (Association for the Advancement of Computing in Education (AACE)), 2542–2549.

Fagerlund, J., Häkkinen, P., Vesisenaho, M., and Viiri, J. (2021). Computational thinking in programming with scratch in primary schools: a systematic review. *Comput. Appl. Eng. Educ.* 29, 12–28. doi: 10.1002/cae.22255

Fredricks, J., Blumenfeld, P., and Paris, A. (2004). School engagement: potential of the concept, state of the evidence. *Rev. Educ. Res.* 74, 59–109. doi: 10.3102/00346543074001059

Gerosa, A., Koleszar, V., Gómez-Sena, L., Tejera, G., and Carboni, A. (2019). "Educational robotics and computational thinking development in preschool." in *2019 XIV Latin American Conference on Learning Technologies (LACLO)* (IEEE), 226–230.

Gerosa, A., Koleszar, V., Tejera, G., Gómez-Sena, L., and Carboni, A. (2021). Cognitive abilities and computational thinking at age 5: evidence for associations to sequencing and symbolic number comparison. *Comput. Educ. Open* 2:100043. doi: 10.1016/j.caeo.2021.100043

González, Y. A. C., and Muñoz-Repiso, A. G. V. (2018). "A robotics-based approach to foster programming skills and computational thinking: pilot experience in the classroom of early childhood education." in *Proceedings of the Sixth International Conference on Technological Ecosystems for Enhancing Multiculturality*, 41–45.

Grover, S., and Pea, R. (2013). Computational thinking in K–12: a review of the state of the field. *Educ. Res.* 42, 38–43. doi: 10.3102/0013189X12463051

Hickmott, D., Prieto-Rodriguez, E., and Holmes, K. (2018). A scoping review of studies on computational thinking in K–12 mathematics classrooms. *Digit. Exp. Math. Educ.* 4, 48–69. doi: 10.1007/s40751-017-0038-8

Horn, M., and Bers, M. (2019). Tangible computing. *Cambridge Handb. Comput. Educ. Res.* 1, 663–678. doi: 10.1017/9781108654555.023

Huang, W., and Looi, C. K. (2021). A critical review of literature on "unplugged" pedagogies in K-12 computer science and computational thinking education. *Comput. Sci. Educ.* 31, 83–111. doi: 10.1080/08993408.2020.1789411

Ioannou, A., and Makridou, E. (2018). Exploring the potentials of educational robotics in the development of computational thinking: a summary of current research and practical proposal for future work. *Educ. Inf. Technol.* 23, 2531–2544. doi: 10.1007/s10639-018-9729-z

ISTE (2015). Computational thinking operational definition.

Janka, P. (2008). "Using a programmable toy at preschool age: why and how." in *Teaching With Robotics: Didactic Approaches and Experiences. Workshop of International Conference on Simulation, Modeling and Programming Autonomous Robots*, 112–121.

Jung, S. E., and Won, E. S. (2018). Systematic review of research trends in robotics education for young children. *Sustainability* 10:905. doi: 10.3390/su10040905

Kazakoff, E. R., Sullivan, A., and Bers, M. U. (2013). The effect of a classroom-based intensive robotics and programming workshop on sequencing ability in early childhood. *Early Childhood Educ. J.* 41, 245–255. doi: 10.1007/s10643-012-0554-5

Lei, H., Cui, Y., and Zhou, W. (2018). Relationships between student engagement and academic achievement: a meta-analysis. *Soc. Behav. Personal. Int. J.* 46, 517–528. doi: 10.2224/sbp.7054

Lyon, J. A., and Magana, J. A. (2020). Computational thinking in higher education: a review of the literature. *Comput. Appl. Eng. Educ.* 28, 1174–1189. doi: 10.1002/cae.22295

Muñoz-Repiso, A. G. V., and González, Y. A. C. (2019). Robótica para desarrollar el pensamiento computacional en Educación Infantil. *Comunicar* 59, 63–72. doi: 10.3916/C59-2019-06

Nam, K. W., Kim, H. J., and Lee, S. (2019). Connecting plans to action: the effects of a card-coded robotics curriculum and activities on Korean kindergartners. *Asia Pac. Educ. Res.* 28, 387–397. doi: 10.1007/s40299-019-00438-4

Papadakis, S. (2021). The impact of coding apps on young children computational thinking and coding skills. A literature review. *Front. Educ.* 6:657895. doi: 10.3389/feduc.2021.657895

Papadakis, S. (2022). Can preschoolers learn computational thinking and coding skills with ScratchJr? A systematic literature review. *Int. J. Educ. Reform.* 1–34. doi: 10.1177/10567879221076077

Papadakis, S., Kalogiannakis, M., and Zaranis, N. (2016). Developing fundamental programming concepts and computational thinking with ScratchJr in preschool education: a case study. *Int. J. Mob. Learn. Org.* 10, 187–202. doi: 10.1504/IJMLO.2016.077867

Pasquali, L., Wechsler, S., and Bensusan, E. (2002). Matrizes progressivas do Raven infantil: Um estudo de validação para o Brasil. *Aval. Psicol.* 2, 95–110.

Perone, S., Almy, B., and Zelazo, P. D. (2018). "Toward an understanding of the neural basis of executive function development," in *The Neurobiology of Brain and Behavioral Development.* (Elsevier: Academic Press), 291–314.

Ponitz, C. C., Rimm-Kaufman, S. E., Grimm, K. J., and Curby, T. W. (2009). Kindergarten classroom quality, behavioral engagement, and reading achievement. *Sch. Psychol. Rev.* 38, 102–120. doi: 10.1080/02796015.2009.12087852

Raven, J. (2000). The Raven's progressive matrices: change and stability over culture and time. *Cogn. Psychol.* 41, 1–48. doi: 10.1006/cogp.1999.0735

Raven, J. C., and Court, J. H. (1986). *Raven's Progressive Matrices and Raven's Coloured Matrices*. London: HK Lewis.

Relkin, E., and Bers, M. U. (2019). "Designing an assessment of computational thinking abilities for young children," in *STEM in Early Childhood Learners: How Science, Technology, Engineering and Mathematics Strengthen Learning.* eds. L. E. Cohen and S. Waite-Stupiansky (New York, NY: Routledge), 83–98.

Repenning, A., Webb, D., and Ioannidou, A. (2010). "Scalable game design and the development of a checklist for getting computational thinking into public schools." in *Proceedings of the 41st ACM Technical Symposium on Computer Science Education*, 265–269.

Robertson, J., Gray, S., Toye, M., and Booth, J. (2020). The relationship between executive functions and computational thinking. *Int. J. Comput. Sci. Educ. Sch.* 3, 35–49. doi: 10.21585/ijcses.v3i4.76

Román-González, M., Pérez-González, J. C., and Jiménez-Fernández, C. (2017). Which cognitive abilities underlie computational thinking? Criterion validity of the computational thinking test. *Comput. Hum. Behav.* 72, 678–691. doi: 10.1016/j.chb.2016.08.047

Sapounidis, T., and Demetriadis, S. (2016). "Educational robots driven by tangible programming languages: a review on the field." in *International Conference EduRobotics 2016* (Cham: Springer), 205–214.

Saxena, A., Lo, C. K., Hew, K. F., and Wong, G. K. W. (2020). Designing unplugged and plugged activities to cultivate computational thinking: an exploratory study in early childhood education. *Asia Pac. Educ. Res.* 29, 55–66. doi: 10.1007/s40299-019-00478-w

Sharma, K., Papavlasopoulou, S., and Giannakos, M. (2019). Coding games and robots to enhance computational thinking: how collaboration and engagement moderate children's attitudes? *Int. J. Child Comput. Interact.* 21, 65–76. doi: 10.1016/j.ijcci.2019.04.004

Sneider, C., Stephenson, C., Schafer, B., and Flick, L. (2014). Exploring the science framework and NGSS: computational thinking in the science classroom. *Sci. Scope* 38:10. doi: 10.2505/4/ss14_038_03_10

Stewart, W. H., Baek, Y., Kwid, G., and Taylor, K. (2021). Exploring factors that influence computational thinking skills in elementary students' collaborative robotics. *J. Educ. Comput. Res.* 59, 1208–1239. doi: 10.1177/0735633121992479

Stoeckelmayr, K., Tesar, M., and Hofmann, A. (2011). "Kindergarten children programming robots: a first attempt." in *Proceedings of 2nd International Conference on Robotics in Education (RIE)*.

Sullivan, A. A., Bers, M. U., and Mihm, C. (2017). *Imagining, Playing, and Coding with KIBO: Using Robotics to Foster Computational Thinking in Young Children*. Hong Kong: Siu-cheung KONG The Education University of Hong Kong, 110.

Sullivan, A., Kazakoff, E. R., and Bers, M. U. (2013). The wheels on the bot go round and round: robotics curriculum in pre-kindergarten. *J. Inf. Technol. Educ.* 12, 203–219. doi: 10.28945/1887

Swanson, H., Anton, G., Bain, C., Horn, M., and Wilensky, U. (2019). "Introducing and assessing computational thinking in the secondary science classroom," in *Computational Thinking Education.* eds. S. C. Kong and H. Abelson (Singapore: Springer), 99–117.

Team R (2019). *R: A Language and Environment for Statistical Computing (Version 3.5. 2)*. R Foundation for Statistical Computing 2018.

Tran, Y. (2019). Computational thinking equity in elementary classrooms: what third-grade students know and can do. *J. Educ. Comput. Res.* 57, 3–31. doi: 10.1177/0735633117743918

Tsarava, K., Moeller, K., Román-González, M., Golle, J., Leifheit, L., Butz, M. V., et al. (2022). A cognitive definition of computational thinking in primary education. *Comput. Educ.* 179:104425. doi: 10.1016/j.compedu.2021.104425

Voogt, J., Fisser, P., Good, J., Mishra, P., and Yadav, A. (2015). Computational thinking in compulsory education: towards an agenda for research and practice. *Educ. Inf. Technol.* 20, 715–728. doi: 10.1007/s10639-015-9412-6

Weintrop, D., Beheshti, E., Horn, M., Orton, K., Jona, K., Trouille, L., et al. (2016). Defining computational thinking for mathematics and science classrooms. *J. Sci. Educ. Technol.* 25, 127–147. doi: 10.1007/s10956-015-9581-5

Wing, J. M. (2006). Computational thinking. *Commun. ACM* 49, 33–35. doi: 10.1145/1118178.1118215

Wing, J. (2011). Research notebook: computational thinking—what and why. *Link Mag.* 6, 20–23.

Yu, J., and Roque, R. (2019). A review of computational toys and kits for young children. *Int. J. Child Comput. Interact.* 21, 17–36. doi: 10.1016/j.ijcci.2019.04.001

Zapata Cáceres, M., Martín-Barroso, E., and Román-González, M. B. C. T. (2020). "Beginners computational thinking test." in *Understanding Computing Education, Proceedings of the Raspberry Pi Foundation Research Seminar Series*, online, 1.

Zelazo, P. D., Qu, L., and Müller, U. (2005). Hot and cool aspects of executive function: relations in early development.

Zhang, L., and Nouri, J. (2019). A systematic review of learning computational thinking through scratch in K-9. *Comput. Educ.* 141:103607. doi: 10.1016/j.compedu.2019.103607

Zhong, B., and Si, Q. (2021). Troubleshooting to learn via scaffolds: effect on students' ability and cognitive load in a robotics course. *J. Educ. Comput. Res.* 59, 95–118. doi: 10.1177/0735633120951871

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Gerosa, Koleszar, Tejera, Gómez-Sena and Carboni. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Education

ORIGINAL RESEARCH
published: 21 June 2022
doi: 10.3389/feduc.2022.872637

How Might We Raise Interest in Robotics, Coding, Artificial Intelligence, STEAM and Sustainable Development in University and On-the-Job Teacher Training?

Jannik Henze, Carina Schatz, Shalina Malik and André Bresges*

Institute for Physics Education, University of Cologne, Cologne, Germany

OPEN ACCESS

Edited by:
Stamatios Papadakis, University of Crete, Greece

Reviewed by:
Ángel Freddy Rodríguez Torres, Central University of Ecuador, Ecuador
Sila Kaya-Capocci, Ağrı İbrahim Çeçen University, Turkey

***Correspondence:**
Jannik Henze
jannik.henze@uni-koeln.de

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Education

Received: *09 February 2022*
Accepted: *20 April 2022*
Published: *21 June 2022*

Citation:
Henze J, Schatz C, Malik S and Bresges A (2022) How Might We Raise Interest in Robotics, Coding, Artificial Intelligence, STEAM and Sustainable Development in University and On-the-Job Teacher Training? Front. Educ. 7:872637. doi: 10.3389/feduc.2022.872637

Schools are searching for strategies to foster 4C competencies (Creativity, Cooperation, Communication and Critical Thinking) in children. Scientific Reasoning, Critical Thinking, and the ability to debunk myths are already important competencies that can be fostered with science education. How can we approach the majority of seventh grade students in a given school to create innovative approaches for the future, and leverage their skills in science, art and (digital) technology along the path? And are the teachers ready to guide them on this path? This article focuses on the questions: how did the teachers adopt both the STEAM approach, and the use of digital tools while being supervised by researchers and student teachers and how did this change their beliefs about technology in education. As a pathway, we aimed to connect Robotics, Coding, Artificial Intelligence (AI) with the Sustainable Development Goals (SDGs) of the United Nations. To end poverty, protect the environment, and ensure that all people enjoy peace and prosperity by 2030, the SDGs are incorporated into national policies and school curricula. With this, citizens, teachers, and governments alike struggle with strategies on how these goals can be reached by 2030, facing the growing challenges in an ever increasingly complex and insecure world. It is clear that technology will take a dominant role in this development. Based on the STEAM paradigm and the 5E approach of the Biological Sciences Curriculum Study (BSCS), we have developed a pedagogical concept that encompasses both the technological aspects, AI and the SDGs. We tested this concept as part of an on-the-job teacher training project with 60 education science student teachers and 8 teachers in their classrooms, together with their 116 7th grade

students and found out that STEAM-based projects with a sixth phase in addition to the 5E approach can be carried out promisingly with the help of digital creativity tools. We found that the 5E model with an additional sixth phase is well suited for bringing STEAM into the classroom.

Keywords: STEAM, robotics, computational thinking (CT), Sustainable Development Goals (SDGs), teacher—education, Piaget, Vygotski

INTRODUCTION

Scientific reasoning, critical thinking, and the ability to debunk myths are important competencies that can be fostered with science education. But how can a majority of students in a given school leverage their skills in science, art and technology to create innovative paths that will lead them to a positive future, and how can teachers guide them on their journey?

The 5E model, which dates to the 1980s (Bybee et al., 2006), serves as the basis for this study. Since then, many digital innovations have found their way into the lives of students. Likewise, their everyday living has changed. Due to becoming an internal part of the modern school system, it became necessary to investigate whether sustainability and digitalization are compatible with a 40 year old teaching model. Furthermore, modern, and digital education is lacking in the German school system.

In a large STEAM (Science, **T**echnology, **E**ngineering, **A**rts, and **M**athematics)-based project we aimed to connect Robotics, Coding, Artificial Intelligence (AI) with the Sustainable Development Goals (SDGs) of the United Nations. These Goals are implemented worldwide into curricula "as a universal call to action to end poverty, protect the planet, and ensure that by 2030 all people enjoy peace and prosperity" (United Nations Development Programme, 2021). This means that governments and education must strive to develop and implement strategies on how these goals can be communicated within their classrooms and how it is even possible to reach them before 2030. Can advancing the "smart" use of technology be a possible solution to achieve these goals?

In the following, the theoretical framework of the research will first be outlined. This includes the presentation of developmental psychological aspects, the 5E model, based on the works of Bybee et al. (2006) and Bybee (2009), the explanation of what digital creativity tools is as well as the connection between STEAM education and the SDGs. We then describe the research questions, our approach, and the materials and methods we used. Finally, we present and discuss the results, draw a conclusion and give a brief outlook on possible future research.

THEORETICAL FRAMEWORK

When we look at digital creativity tools, at the first glance they remind us of the toys that students are playing games with. This is our motivation to start by briefly examining the developmental psychological perspective on the process of playing as a concept of learning. Next, the 5E-model, on which the field study is based, the aspect of STEAM digital creativity tools and STEAM education will be presented. Lastly, a brief description of the school where the test was conducted is given.

Developmental Psychological Aspects of Playing as a Concept of Learning

Playing can mean several different things that children can engage in. According to psychologist Lev Vygotski, *playing*, be it with toys or a game, is triggered by situations that might be relevant for children's lives and engages them to transform these certain situations into a game. For example: when a child observes a stagecoach driving by, it might react by playing "stagecoach driver." Within this game-situation, the child prepares himself to engage in a situation where it might become, eventually, an actual stagecoach driver.

Further on, according to psychologist Jean Piaget, playing—as a concept of how Vygotsky described it—can be divided into two different developmental stages that describe how, and to what extent, a child benefits from playing. The first of these stages is practice. Here, the physical development with respect to play takes place by imitating known basic principles and understanding the uses of objects, thus satisfying the intrinsic urge to explore (Leong and Bodrova, 1996), which can be applied to this study by letting the children explore the given tools and partaking in construction games. The next stage, according to Piaget, is symbolic play, in which mental models are created, where every object can be a placeholder for something else, which are then applied in play (Leong and Bodrova, 1996). The advantage of play is that it gives the learners a sense of self-control, which serves as a base to take on new challenges more self-efficiently (Leong and Bodrova, 1996). Both Lev Vygotsky and Jean Piaget assume an interiorization process in their theories, in which learners develop their conceptions, ideas and models with the help of concrete actions (Aebli, 1985). This means that, through playful situations, complex interrelationships can be modeled in an understandable way (Kircher et al., 2014). One of the big ideas of STEAM Education and the Maker Movement is linking basic knowledge in physics with everyday technology using construction games. Within these games, students can explore complex socio-technical issues in a playful situation. This enables learners to be creative during the construction process and thus to realize many ideas (Kircher et al., 2014).

The 5E-Model

To facilitate the learning of physical concepts, learners must be engaged in appropriate learning activities. These activities

should be designed in three parts to be as effective as possible. In the first part, goals should be identified. Following this, the current learning status should be discussed. The last part should determine the means by which the learners can reach the identified goal from their current position (Etkina et al., 2006). This tripartition can be expanded into more parts to allow learners to delve deeper into the given subject matter. The 5E-model was developed based on constructivist learning theory and cognition psychology as well as proven methods in science education (Duran and Duran, 2004) to create lessons in a student instead of teacher centered way (Turan, 2021). The model can be used within single or few hours as well as for longer units. Teachers that participated in studies in which the 5E-Model was applied, said to have more confidence and are more comfortable in teaching sciences (Duran and Duran, 2004). Nevertheless, studies also showed that it is hard for teachers to find suitable activities and materials for different phases of the 5E-Model (Namdar and Kucuk, 2018). Furthermore, several studies have shown different barriers, like classroom management and time issues, that hinder teachers from implementing the 5E into their own lesson planning (Turan, 2021).

The 5E-Model consists of the following five phases: Engagement, Exploration, Explanation, Elaboration, and Evaluation. In the first phase, learners are confronted with the learning content, which activates their existing knowledge and their curiosity (Bybee et al., 2006). It is also possible to determine what students might already know about the topic or what (mis)conceptions they have (Duran and Duran, 2004). Accordingly, the learners are confronted with a problem to solve. The phase is successful when the pupils are engaged in the problem and are intrinsically motivated to solve it (Bybee, 2009). In Vygotski's approach, the motivation and the need for action are to be located here.

In the Exploration phase, learners are given the opportunity to freely explore and become familiar with the essential skills and concepts that are made necessary by the problem posed in the engagement phase (Duran and Duran, 2004). This phase should be designed so that all learners have the same experience to build knowledge and skills. The role of the teacher in this phase can be seen as merely accompanying to allow students to explore as freely as possible (Bybee, 2009) and explicitly not giving away any kind of explanation, which is reserved for the following phase (Duran and Duran, 2004).

The Explanation phase allows learners to demonstrate their understanding of the concepts by explaining certain aspects or the entire concept itself (Bybee et al., 2006). In this way, the Explanation phase helps to ensure that learners develop a consistent vocabulary related to the problem, and present the concepts, information, and skills they have grasped in an understandable way (Bybee, 2009). Furthermore, a teacher should only fill in with explanations, if the Student's way of explaining is not sufficient or contains misconceptions (Duran and Duran, 2004; Namdar and Kucuk, 2018).

In the Elaboration phase, learners can consolidate their abilities and understanding regarding the topic, thereby leading them to a deeper understanding and adapted skills (Bybee et al., 2006). In this phase, learners can build on the concepts and skills they have already understood by, for example, applying them to new concepts within the problem. For this purpose, the interaction between learners in groups can be seen as a major factor for the success of the phase. The group discussions and collaborations provide opportunities to receive feedback from other learners on the one hand and to enter an exchange about their knowledge on the other hand. The goal of the elaboration phase is the transfer of knowledge from previous phases to new problems (Bybee, 2009). Here, as in the Exploration phase, the playful approach emphasized by Vygotsky is followed.

In the Evaluation phase, the learners are given the opportunity to reflect on their learning journey (Bybee et al., 2006). In this final phase they also receive feedback on their learning progress, skills, and insights (Bybee, 2009). This should give the teachers proof of the Student's learning success and can be conducted in a formal or informal way (Duran and Duran, 2004).

In the Exchange phase, a sixth phase we added to the 5E in the last week of the field study, we provided an opportunity for all participants to reflect and exchange on what and how they learned. This phase was added to emphasize the exchange between learners as well as between learners and teachers. We found this to be a very profitable addition to the 5E-Model, to get insight into the students as well as the teachers' experience of the whole project to enrich the Evaluation phase. Accordingly, this phase focuses more on meta-cognitive skills than the other phases. In Vygotskian thinking, the Engage phase would stimulate the children to open their Zone of Proximal Development, while the Explore and Elaborate phase provide the necessary playground for the learners to simulate the situation they engage in, test and improve their competencies, and simulate possible outcomes. The Explain and Exchange phase with their focus on inter-group communication provide the students with the necessary opportunity to negotiate the rules of their game in Vygotskian theory. From a social-constructivist perspective, these phases provide the opportunity to exchange insights, models and world-views and assess the relevance for life in the view of their peers. While the "Explore" phase is generally open and playful, the "Elaborate" Phase targets the development of a testable prototype that might be evaluated in the subsequent "Evaluate" Phase. This connects to the learning theory of Piaget, where children test their hypotheses by play.

STEAM Education

The core idea behind STEM (Science, Technology, Engineering and Math) Education is to connect the sciences, rather than teaching them in isolation (Krakower, 2018). But even though the relationship between different disciplines was recognized, the creative aspects of them were missing. Due to becoming more influential and significant in this digital and global world, such aspects were incorporated into the STEM framework (Yakman, 2008), resulting in the existence of the STEAM approach. The natural science disciplines are not only complemented by the arts, but also by methods to encourage creativity and innovation. These methods, like visual thinking, were derived from artistic fields (Thomas and Huffman, 2020). If Art would be used in a narrow sense, e.g., just in the form of painting, learners would not see where this is connected to and relevant for STEAM problems.

Art can only be integrated into the learning process if it is used in a broader sense. Here learners progress by integrating the arts in the area of problem solving (Quigley et al., 2020). By integrating the Arts aspect, more individuals can be reached, who have little interest in traditional STEM contexts (Thomas and Huffman, 2020). In the context of STEAM education, collaboration, and mutual feedback among learners worked very well, as has been observed by Cassie Quigley from the University of Pittsburgh. This was due in part to the use of technology and assignments that encourage collaborative work. Each learner in a group was assigned a task according to their abilities to solve a problem cooperatively as a group (Quigley et al., 2020). This cooperative and problem-solving approach of learning is at the forefront of STEAM education (Jackson et al., 2020).

Sustainable Development Goals

The United Nations formulated 17 goals to improve human life on earth in the near future. They are known as the Sustainable Development Goals or SDGs. Each of these goals aims for different aspects of life and contains different targets and possible actions to reach it. Some of the SDGs are already covered by the STEAM education definition. For example, SDG 09 promotes to "Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation" (United Nations Department of Economic and Social Affairs, 2021). SDG 07 is to "Ensure access to affordable, reliable, sustainable, and modern energy for all" (United Nations Department of Economic and Social Affairs, 2021). The Art aspect includes considerations of societal developments, and the aspect of Engineering relates to the SDGs via the creative as well as logical use of technical tools to solve global problems (Yakman, 2008).

STEAM Digital Creativity Tools

STEAM tools aim to support the training of creative thinking as well as the competence of problem solving and critical thinking (Makeblock, 2019). The intrinsic drive for discovery postulated by Piaget can also be nurtured with the help of STEAM tools by encouraging learners to find creative solutions to specific problems. In this regard, digital creativity tools, such as those used in the field study we conducted, are suitable for this purpose, as they are child-friendly and contain many technical resources which are relevant for the teaching of physics. Digital creativity tools refer to various devices that, among other things, can be used to stimulate the learners' creativity in order to find solutions to problems. A digital creativity tool can be used to integrate STEAM education in schools and develop problem solving skills, creativity, and boost the students motivation (Kalogiannakis et al., 2021). A study commissioned by LEGO Education, conducted in 2019 by Harris Insights and Analytics, examined Students' confidence in the context of STEAM education and digital creativity tools. In this study, only 14% of German students reported being very confident in learning STEAM content (Harris Insights and Analytics, 2019). Furthermore, a study on the physical area of light and optics showed a significant increase in both learning success and creativity among students who learnt these topics using STEAM methods (Wandari et al., 2018). Accordingly, compared to traditional instruction, there is a significant positive difference in the use of digital technologies in STEAM-based instruction (Tamim, 2011). A study that examined the use of another digital creativity tool (BBC micro: bit) found teachers being more open about using such a tool if this has a connection to the everyday life of a student and is generally useful (Kalogiannakis et al., 2021). They also show to be more positive about using digital tools if they have multiple uses in school contexts and allows students and teachers to learn from it (Papadakis, 2022).

Recent past works like the ones from Kalogiannakis et al. (2021) and Papadakis (2022) mainly focused on the usage of digital tools like apps and programming languages in a school context and lacked physical tools. Therefore we wanted to gain more insight on several educational tools presented in the following.

In our study, the results of which are presented in this paper, we used four different tools (Makeblock mTiny, Makeblock Cody Rocky, Makeblock Neuron, and DJI Tello Edu Drone) each of which has different characteristics.

The Makeblock mTiny was chosen for the project because it offers screen-free programming and thus reduces screen time, on the other hand it enables inexperienced students to experience and understand complex programming in a playful and uncomplicated way, while at the same time teaches the basics of computational thinking. A meta study regarding ScratchJr which is similar to the way the mTiny is programmed, shows it to be useful in introducing young students to STEAM education (Papadakis, 2022).

The Makeblock Codey Rocky was selected because this tool is a further development of the mTiny. It contains many sensors with which learners can program various commands and then see if Codey Rocky reacts to them. It can also be controlled directly using an app or be programmed using a block-based programming language. It was chosen as an addition to the mTiny, because this robot cannot be controlled via Joystick and has to be programmed or controlled with the help of a tablet device.

The Makeblock Neuron set was chosen for the project because its properties allow it to be easily used as a versatile construction kit. This is based on a number of sensors that allow various measurements and, on the other hand, a large number of actuators that can be attached and controlled, even remotely to simulate an Internet of Things (IoT) environment. In addition, it is possible to connect the Neuron set with the Codey Rocky and thus exploit a potential for mobile or robotics applications. The set can be programmed without a screen by connecting individual blocks in a certain order to build simple measuring devices. In addition, it can be programmed via app in a block-based coding environment, which, according to Kalogiannakis et al. (2021) seems to help students understand the general concept of a programming language.

We selected to use a DJI Tello Edu drone for the project because drones can enable students to study and control an object that can freely move in a 3-dimensional, Cartesian space. In mathematics and physics education, this option was only accessible in simulations or thought experiments before the introduction of drones. In addition, drones are becoming

increasingly present in today's world and students should therefore learn how to handle them in a safe manner that obeys rules of privacy. We chose the Tello Edu drone as its small size and weight allows it to be used in the classroom, making it very suitable for this project. An app makes it possible to program this drone using a block-based language. For classroom use, it is important that students can test their code within the app in a simulated flight environment before they are provided with the actual drone. This enhances safety, reduces the need to load batteries and reduces the overall cost for the school.

MATERIALS AND METHODS

Digital STEAM Creativity Tools were used for teaching and learning Robotics, Sensors, Artificial Intelligence and Computational Thinking together with a Vygotskian teaching approach in a large scale, school-spanning field study. For our research, a mixed method design was conducted with different focus areas.

The field study project was structured using a modified 5E model (Bybee et al., 2006; see **Figure 1**), which was used in the context of the university with the English terms but translated for the students with appropriate German terms. The individual phases and activities are briefly described below. Across all the phases of the project, each unit was transparently accompanied by appropriate presentations by the staff. At the beginning of each lesson, the learners were thus offered a classification of the respective day in the overall project as well as an overview of the daily schedule.

The *Engagement* phase took place in the first week of the project. The thematic introduction was done by means of two videos on different SDGs, of which each learning landscape watched one video. The first video focused on SDGs 2 *Zero Hunger* and 6 *Clean Water and Sanitation* (see **Supplementary Figure 1**), whereas the second video focused on SDGs 3 *Good Health and Wellbeing* and 11 *Sustainable Cities and Communities* (see **Supplementary Figure 2**).

Following these videos, the students reflected on what they had seen with the help of in-depth questions on a padlet and conducted their own research on the SDGs. In this phase, the staff of the University of Cologne continued to explain the entire project process to the 7th grade students in detail. In the following second project week, the *Exploration* phase, the students, guided by university students, got to know the digital creativity tools. The teachers only played a minor role in supporting the 7th graders as well as the university students in terms of classroom management. Again, a distinction was made between the learning landscapes, so that in learning landscape A the devices mTiny, Codey Rocky and Neuron were used, and in learning landscape B the Codey Rocky, Neuron and the Tello (Edu) drone were used to test how the provided tools influence the designed solutions. During the *Explanation* phase, which was carried out in the third week, learners had to explain the possibilities of one of the devices they tested on a digital worksheet in Google Classroom. This was then evaluated as part of the study. In this way, the positive and negative characteristics, functions, and programming possibilities of the digital creativity tools in the perspective of the student could be studied. In addition, core groups had to display and explain the STEAM tools they had researched to other core groups so that an exchange could take place about all the devices and each student saw a short presentation about each tool. The *Elaboration* phase was extended to the fourth to sixth week. The phase started with the learners working in small groups to choose a topic related to the SDGs and one of the creativity tools, and then working on either their own or pre-determined research questions. Their aim was to find a solution to a problem that could be modeled using the tools.

This led to the fifth phase, the *Evaluation* phase. In this phase the students mainly prepared the presentation of their projects to a public audience. Since the learners were free to work on their project, prepare a presentation or do both at the same time, this phase blurs with the preceding Elaboration until the day of presentation: The Barcamp. This event was designed to resemble a design pitch to raise venture capital, or to raise public awareness for a project. The learners presented their solutions to the public in the form of a video conference. Access to the video conference was possible for everyone after prior registration. After each presentation, the audience had the opportunity to ask the presenters questions and give feedback on their prototypes. In the *Exchange* phase, the sixth phase we added to the 5E, in the last week of the field study, there was an opportunity for both learners and teachers to share and reflect on the project. An intervention on Artificial Intelligence was also conducted during this week to give the students a perspective on what modern technology could enable their own projects to do. Furthermore, the students filled out various surveys regarding final university student theses. Because of the burden of the surveys on the students we chose not to collect measurable data.

Due to the COVID-19 pandemic, the project had to be launched exclusively digitally. Accordingly, the students took part in a video conference led by University of Cologne staff. In the following week, due to the pandemic situation, it was possible to switch to a hybrid state that lasted for another 2 weeks. In this state, the groups were separated into subgroups, which alternated in daily visits to the school. While one group was able to go to school, the second group was connected to the lessons with the help of tablet PCs. From the fourth week of the project onward, the restrictions were eased, and the core groups were then present until the end of the project. Students who were not employed at the University of Cologne and who conducted some parts of the project were connected via tablet PC video conferencing during the project in order to minimize the risk of contagion for all involved.

The project was carried out at the Helios School—Inclusive University School of the City of Cologne. This school is designed by University of Cologne education scientist Kersten Reich in the tradition of John Dewey's laboratory school at Chicago University, but under today's conditions (Reich, 2018). Dewey anticipated already 100 years ago the needs of education that we consider crucial today, namely the multiperspectivity and broad access to learning. His vision of a school included the participation of students in social processes where they would

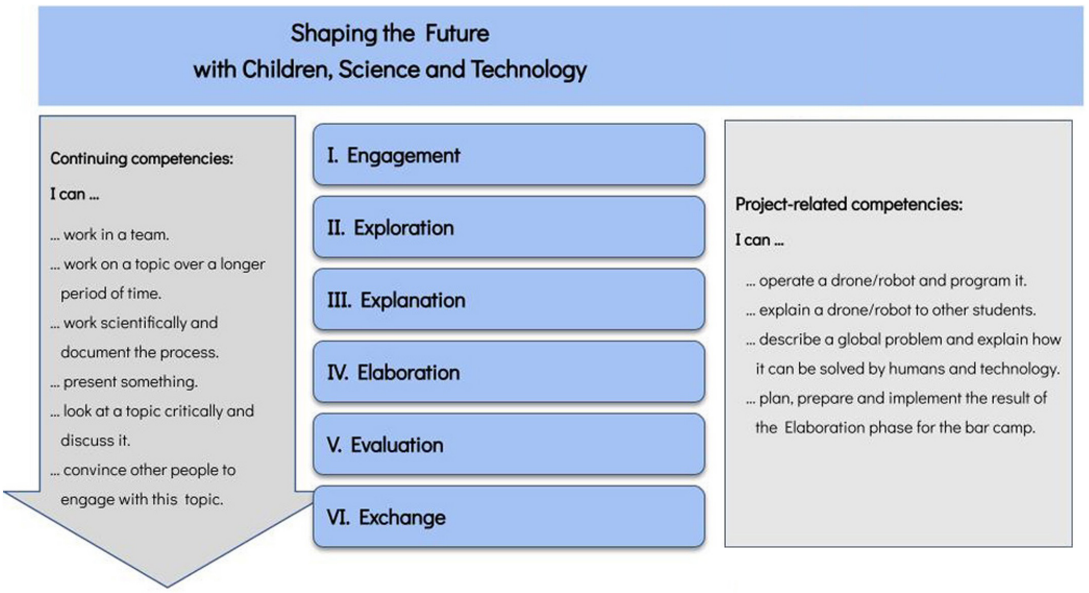


FIGURE 1 | Advanced organizer used in the project to show the students the competencies they can acquire.

build on their skills in communication as well as problem solving. One of the schools main foundations is, according to Deweys as well as Reichs research, the principle of learning and teaching through *learning by doing* (Reich, 2018). The Helios School was founded under a constructivist perspective toward education but had to face two major problems. The first being the heavy focus of the German educational system on the attainment of a degree rather than social equity. The second problem lies in the German teacher training system, which is split into theoretical and practical units (Reich, 2018).

The participants comprised about 116 7th grade students of the Helios Inclusive University School of Cologne, Germany together with their 14 teachers. The age range of the teachers was between 28 and 46 years. The teachers had been in teaching for between less than one and more than 16 years at the time of the study. The teachers' subjects ranged from social studies over languages to STEM and physical education, also one of the teachers was a special education teacher who did not specify further subjects. Furthermore over 200 students of the Bachelor and Master programs of University of Colognes STEM Teacher Training Department took part in this study, of which 40 were actively involved in the implementation of this field study, while the rest supported them with templates and feedback. All 40 actively involved university students and 116 7th graders took part in a 7 week on-the-job training program that was part of the regular 7th grader classes. The pupils were divided into the two learning landscapes A and B with each three different Stem Groups, which is the equivalent to a school class at the IUS.

To conclude the evaluation of the specific tools used, we used a pre-post-test on the partaking teachers, as well as a pre-post-test on the 7th grade students to evaluate the usage of videos. Furthermore, university students and teachers were interviewed regarding their view on the whole project at different times of the field study (see **Figure 2**).

The pre-post-tests regarding the evaluation of the digital creativity tools were formulated according to the rules for formulating questions for qualitative surveys (Döring et al., 2016). The questionnaires of the pre- and post-test differ in a few questions, which are only useful in each case in the pre- or post-test, for example, when first thoughts about a respective tool or experiences from the field study are asked. For each of the devices it was asked what thoughts the teachers had in each case when they saw the device for the first time. Teachers were also asked what they liked and disliked about each tool. This was intended to identify certain advantages or criticisms of the tools. Regarding possible points of criticism, the teachers could also suggest possible improvements. Also, a possible place of use away from the field study in combination with the willingness to use a digital creativity tool in the classroom was asked. This question gives first indications whether the field study has changed the willingness of the teachers to use digital creativity tools in the classroom. General desires for a digital creativity tool were also considered. This should highlight certain characteristics that digital creativity tools should have in order to be considered by teachers for use in the classroom. Teachers' responses were anonymized but coded so that pre-tests and post-tests could be matched without revealing teachers' identities.

The pre-post-tests regarding the use of videos were modeled after the *IPN Interessensstudie* (Measuring Students' interest in physics) from Häussler (1987). The original test assumed that student interest is not one-dimensional and not constant, but a complex situative variable that must be modeled along the three dimensions *topic*, *context* and *activity*. Sample item questions were "Do you want to learn more (*activity*) about how colors occur (*topic*) in the sky (*context*)" or "Do you want to

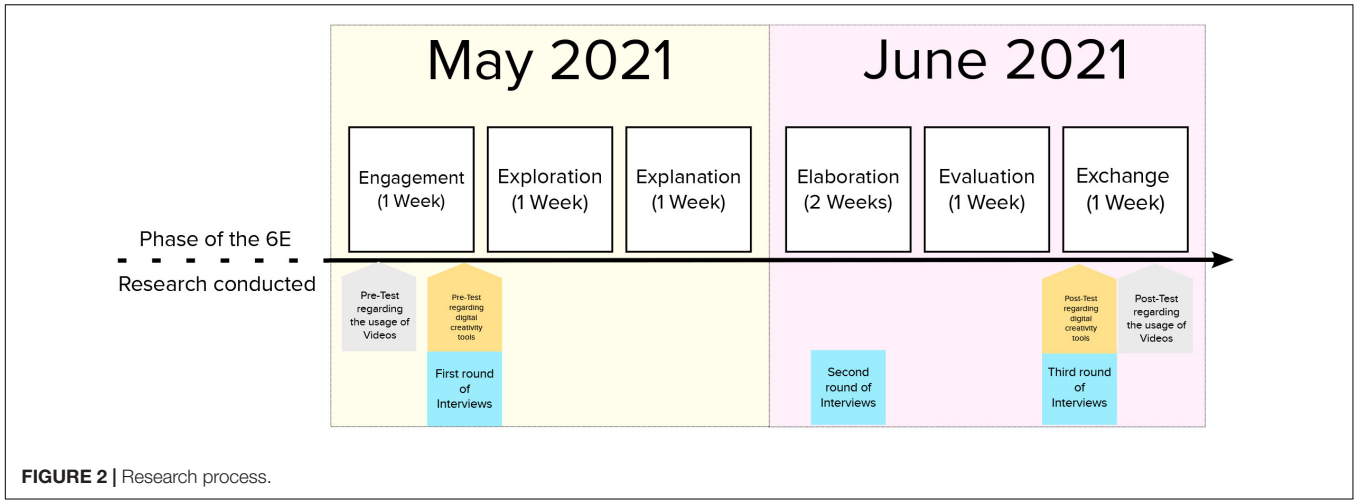


FIGURE 2 | Research process.

discuss and evaluate (*activity*) the positive and negative effect of micro-electronics (*topic*) on our lives (*context*)" (Häussler, 1987). To introduce students into the subject and to compensate for different prior knowledge of the participants, a short one-page introductory text was provided for each topic. After reading these texts, participants should indicate their interest to follow the topic in different contexts and with different activities. Interest was indicated using a five-point scale ranging from "My interest in this (item) is very high" to "My interest in this (item) is very low" (Häussler, 1987).

A test of the mathematical model conducted by Häussler on 4,034 students between 11 and 16 years revealed that the 3 dimensions are largely independent of each other, with interaction terms between the dimensions explaining only 2% of the variance (Häussler, 1987). Hence, it is reasonable to assure that the topics and contexts can be modified or exchanged independently of each other. We chose contexts that were derived from the SDGs for the pre-post-testing in this study. The one-page introductory texts of the original study were replaced by the introductory videos in the engagement phase. The proposed activities (to learn more, to construct, to discuss) of the 6E process were similar to the original study, and students could indicate on a 5-point scale if they are interested to take part in it. In addition, there were open-ended questions in which the students could independently write down activities they would take part in. These answers were clustered into *suggested* activities that are connected to the topic and context of the video, and *independent* activities that might still be connected to the context of the video (e.g., climate change) but did not have any connection to the lessons and the topic of the video (e.g., using public transportation to reduce CO_2 emission).

The questionnaire was tested with students to ensure that their understanding of the questions was comparable to the original study. Pre-testing took place immediately after the videos were shown. 91 Students took part in the pre-test (83%). The post-test that was conducted about 6 weeks later, after all activities took place. 83 students took part in the post-test (75%). The anonymous surveys ensured the privacy of the students. Since no code was generated and no socio-demographic data was collected, no conclusions can be drawn about individual students.

For the interviews, five participating teachers from the cooperating school were interviewed in three rounds each at the beginning, between the exploration and explanatory phase, in the elaboration phase, and after the end of the field study. The teachers were two women and three men. In the interviews, many open-ended questions were asked, which encouraged the persons to tell their stories freely and to follow up where, for example, dissatisfaction could be suspected. Of interest in the interviews were negative as well as positive personal experiences and aspects, learning situations, attitudes toward technology and cooperation with the school. The aim was to capture as many views as possible and to record the learning process of the individuals.

After the transcription, the qualitative data were analyzed according to Kuckartz (2018) using MAXQDA. Example main categories are praise, positive experiences, growth, learning process, attitude toward technology, criticism, negative experiences, wishes. Subcategories were then inductively derived from the data.

Since the research was conducted in German, the data is also mainly in German. Furthermore, conclusions about individuals could be drawn from the interviews despite greatest efforts to anonymize them. Therefore, the appendix of the interviews (**Appendix A**) is not distributed publicly but can be viewed on request.

RESULTS

How Can STEAM Education Based on the 5E Model Be Introduced in Schools?

One of two different videos was shown in the Engagement Phase in each of the different learning groups A or B. As a result of the survey regarding the effectiveness of the used videos, 23.8% of the students in learning group A and 26.1% in learning group B formulated ideas after watching the video on how to improve the life of people around the world. Another 21.4% of learning landscape A and 8.7% of the learning landscape B

described ideas suggested by the respective video. Only 4.8% (A) and 10.9% (B) said they would have no ideas. In each case 50% (A) and 54.3% (B) made no statement (**Appendix B**, Chapter 4, Diagram 1). Through the video analysis survey, it was found that there was a tendency for increased interest in physics among the learners at the Inclusive University School prior to the project implementation. Even though the initial interest was measured immediately after watching the respective video, the results indicated a decrease in the interest (**Appendix B**, Chapter 6, Diagram 21 and 27).

Overall, through the pre-post survey regarding the use of video, it was found that engaging videos were instrumental in generating students interest in the subject matter. A video 4–5 min in length was sufficient for the interviewed students (**Appendix B**, Chapter 8, Item 15) if all essential problems and solution ideas were presented.

Due to pandemic teaching modes, the Exploration phase could not be conducted with all students at the same time. The participating teachers did not feel that the involvement of students via distance learning was adequate, causing frustration. Students found it difficult to participate in class via video conferencing (**Appendix A**, L4, Interview 2, pos. 5). As an alternative, for example, a more targeted use of university students in online teaching could be identified by having them help develop programs, with those who are not in school, that could then be tested on site (**Appendix A**, L1, Interview 1, pos. 21).

Many of those involved in the project commented positively in connection with the playful and practical opportunities offered by the devices. It was emphasized several times that not only the students had fun with the tools, but also the adult members of the project (**Appendix A**, L4, Interview 1, pos. 25). Several teachers as well as students wished for an extension of the Exploration Phase (**Appendix A**, L1, L3, L4, B1).

From the Students' presentations and completed worksheets, it can already be concluded that through the Explore phase, they learned about many of the positive and negative features of each device and understood how to achieve possible goals with these devices (**Appendix B**, Chapter 12, Summarized Evaluation). This highlighted the simplicity and intuitiveness of the devices, as the learners only had 90 min to get to know each one, but most importantly, it reinforced the success of the previous exploration phase.

Concluding this phase, the playful introduction of the devices aroused the interest of the learners encouraging them to expand the capabilities with the device. This has been shown that they were able to recognize the advantages and disadvantages as well as potential, with help, in the short time available. Furthermore, it seems to make sense to extend this phase to give all students the opportunity to get to know each device intensively, instead of only being able to try out three of the four devices for about 90 min each, as was the case in this project.

Sharing learning outcomes across learning landscapes in the Explanation phase was seen by teachers as critical for students because learning landscapes had little contact and additional connectivity issues would have limited already difficult communication. However, the fact that the students had to explain the devices to each other was seen positively (**Appendix A**, L2, Interview 2, pos. 7–9).

It can be concluded that while mutual exchange is important, it should be limited to the known peer group and, ideally under non-pandemic conditions, should take place in person. This means, for example, that there is no inhibition of communication that could arise from speaking in front of other children. This is an aspect that could be investigated in further studies in the future.

Regarding the Elaboration phase, a teacher reported that the students did not understand why, despite being fully present in class again, they should still interact with university students in videoconferences (**Appendix A**, L2, Interview 2, pos. 13). Therefore, the help that the university students were supposed to represent was not accepted by the pupils. Which is why, from the moment when all pupils were back in class, the university students perceived the negative reaction to support via videoconferencing (**Appendix A**, B1, Interview 3, pos. 7, 26).

This phase was described as particularly stressful by both teachers and members of the university team. They were forced to deliver the intervention, manage the classroom, and provide technical support to multiple groups simultaneously (**Appendix A**, L2, Interview 2, pos. 21; M3, Interview 1, pos. 21–41). The projects the students worked on were deepened and revised by them to solve a selected problem connected to the SDGs. They organized themselves into groups and worked on their projects without further instruction from the teachers or university students. No further motivation was needed than handing out the digital creativity tools and giving them a short overview of the schedule.

Extending this phase was mentioned afterward as a possible improvement (**Appendix A**, L1, Interview 3, pos. 9), since the students only had about 12 h over a 2 week period to work on their projects.

Summing up this phase, it can be said that the students had a good opportunity to work on their own projects. In order to create a more relaxed environment for all involved, including the teachers and in our case students of the university, this phase could be extended to allow more time on the one hand and on the other hand to give the teachers more possibilities to interact.

In the Evaluation phase the presentation of the Students' projects to a public audience took place. Since the learners were free to work on their project, prepare a presentation or do both at the same time, this phase blurs with the preceding Elaboration until the day of presentation.

The participation of the learners in the oral feedback in the Exchange phase was excellent and helped us to understand their perception of the project as well as providing insight on what could be improved going forward.

As mentioned with the Evaluate Phase, this phase has been added to the 5E model to allow for sharing of the learning journey. This exchange should only refer to the learning process and explicitly not to the learning outcome, so that the students can give unevaluated feedback, whereupon the learning process can be better adapted for them in the future.

The Post-Survey regarding the usage of videos in the learning landscapes A and B showed that 38.6% (A) and 35.9% (B) of the 7th grade students had their own ideas on how they could improve the life of people, which is an increase of + 14.8% (A) and + 9.8% (B) in contrast to the pre-Survey. Further 22.7% (A) and 23.1% (B) gave ideas suggested by the respective Videos they had watched. Nevertheless 27.3 (A) and 25.6% (B) of the students said they would have no ideas, which is a drastic increase of + 22.5% (A) and + 14.7% (B). Additionally, 11.4% (A) and 15.4% (B) did not answer this question in the Post-test (**Appendix B**, Chapter 5, Diagram 14). A possible Explanation for the increase of students saying to have no ideas is the decrease in students not answering this question. They might have just answered with no intention of giving an idea but unwilling to not-answer to this question.

The results show that interest in physics decreased after the 7-week project period, which could be associated with a kind of routine and saturation that occurred among the students (**Appendix B**, Chapter 6, Diagram 21, Diagram 27).

The project was well received by the teachers involved, especially regarding the cooperation between the university and the school, the motivation that the pupils experienced through the project, the equipment used as well as the learning paths taken by the learners (see **Figure 3**). In terms of the learning process, communication with students, the use of technology in the classroom, and programming were emphasized, and the interdisciplinary teaching was, among other things, also praised (see **Figure 4**). The main points of criticism relate to the usage of video conferences and didactic decisions and content. Also, the wish to strengthen teamwork was often mentioned, as well as more transparency in terms of organization (see **Figure 5**). Due to the COVID-19 pandemic, some phases of the project had to be designed either through distance learning or hybrid learning. This is also the biggest point of criticism from those involved. Because this will (hopefully) no longer be a problem in the future, this point of criticism should not be overestimated. The individual phases of the project also suffered from distance learning and hybrid learning, especially around the Explore and Elaborate phases. In a renewed implementation or consistent further development of the project, more time for these important practical parts should be considered. Furthermore, in a renewed implementation special incentives and insights could be created through possible links with experts on the respective topic.

What Is the Attitude of Teachers Toward the Adoption of STEAM Tools in the Context of STEAM Teaching and How Does It Change in the Course of a 7-Week, On-the-Job Training Program?

In the pre-post-test, the teachers expressed confidence in their Students' ability to work with the devices prior to the project, since they had great trust in their Students' abilities (**Appendix B**, Chapter 11, pos. 6, K10). Their belief that the students had already grown up with technology and thus had a high affinity for technology served as an important factor, which is why intuitive handling was to be expected (**Appendix B**, Chapter 11, pos. 6, K101). The teachers also reported little fear of contact on the part of the students and a high degree of curiosity. (**Appendix B**, Chapter 11, pos. 6, K102). All teachers at the project school indicated in a survey that they had not previously used any of the devices used in this project, nor had they used similar devices, in the classroom (**Appendix B**, Chapter 11, pos. 5, G50). Most of the persons interviewed showed a positive attitude toward technology in school lessons and emphasized on the advantages of it; but not without mentioning the importance of critical thinking while using technology (see **Figure 6**).

When asked in the post-test whether they would use other devices in the future, other than the ones used in the project, two teachers indicated *yes*, whereas one indicated *no* (**Appendix B**, Chapter 11, pos. 5, G50). Due to the wording of the question, it remains unclear at this point whether teachers would use the devices used in the project for teaching in the future. The teachers who stated *yes* to the above question named other projects and workshops as possible reasons along with other devices (**Appendix B**, Chapter 11, pos. 5, G51). When asked why the teachers have not yet used any digital creativity tools in their lessons, a lack of experience or the lack of the necessary equipment were the main reasons (**Appendix B**, Chapter 11, pos. 5, G50). Nevertheless, in many

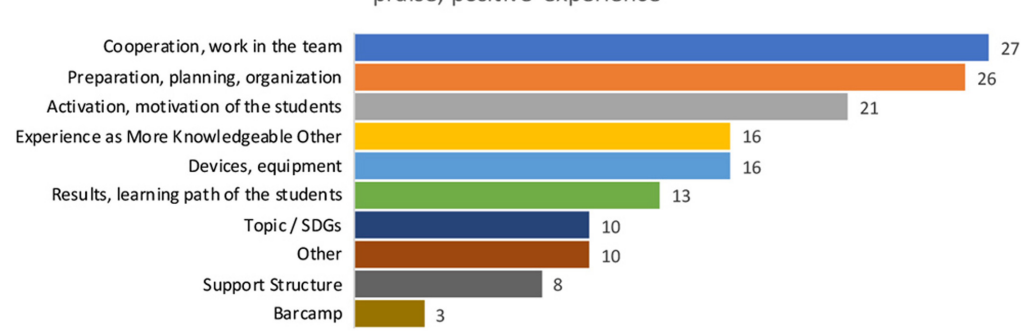


FIGURE 3 | Frequencies of interviews with mentions of the listed categories regarding praise and positive experiences in the project.



FIGURE 4 | Frequencies of interviews with mentions of the listed categories regarding growth and learning in the project.

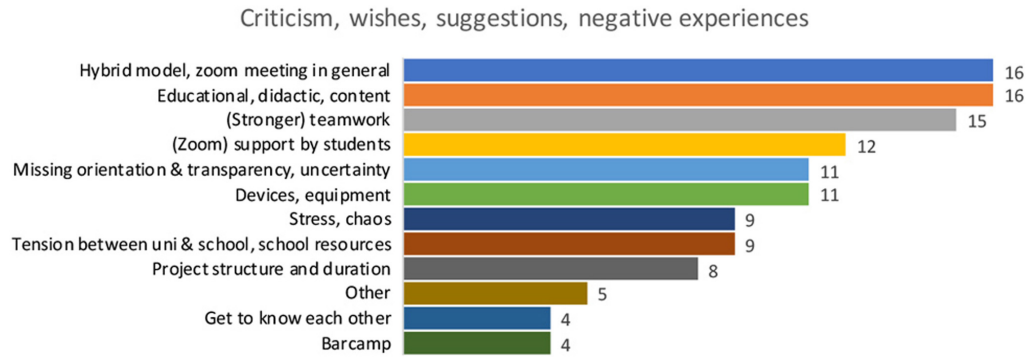


FIGURE 5 | Frequencies of interviews with mentions of the listed categories regarding criticism, wishes, suggestions and negative experiences in the project.

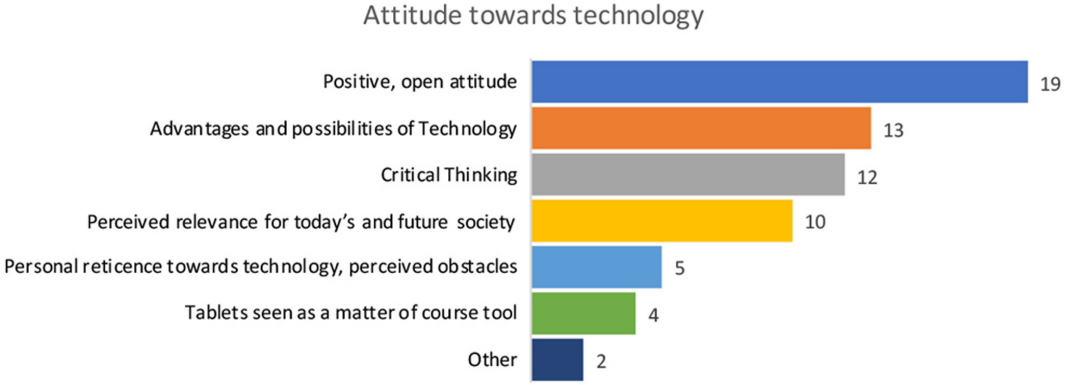


FIGURE 6 | Frequencies of interviews with mentions of the listed categories regarding the attitude toward technology.

interviews it was made clear that all participants were open toward using technology in their lessons and highlighted the advantages of it.

The digital creativity tools used in the project were generally well received by the teachers and students involved in the project. Nevertheless, from the point of view of the study participants, there are also possibilities for improving the tools, which draw attention to the disadvantages of the devices.

The easiest accessible device, the mTiny, was also rated the least popular by the participating teachers of all the devices. This can be explained with the target group (age 3 and older) that is usually addressed by this device. The mTiny can therefore only be recommended to a very limited extent for use in the seventh grade or higher, as it offers too few options for this age group, which is why students who already have experience with digital products quickly reach the limits of the device (**Appendix B**, Chapter 11, pos. 1, G10.4). It would therefore be an option to improve the

mTiny by creating the possibility of programming using a tablet PC. This would enable more complex tasks for higher grades as well as technical enhancements.

The Codey Rocky, on the other hand, is much more suitable for the project's target group according to the data available. This can be concluded from the fact that the complexity is appropriate and variable, i.e., it is very easy to get started with the tool, but at the same time very complex problems can be processed. The given robustness against falls is also a factor that can play a central role in everyday school life (**Appendix B**, Chapter 11, pos. 2, G20.4).

Some teachers suspect that it looks too childish for seventh-grade students, which could create a barrier to learning. In contrast, however, the appearance was also viewed positively by other teachers as well as students. However, it was also suggested that a neutral version be developed for adolescent learners (**Appendix B**, Chapter 11, pos. 2, G20.1). Of the teachers involved in the project, four out of five stated in the pre-test that they would not use Codey Rocky in their lessons outside of the project. The main reasons for this were uncertainty in dealing with digital creativity tools and a lack of ideas for integrating the tool in a project in a meaningful way. One teacher stated that she would use the device in grades four to seven to reduce fear of contact with technical devices. In the post-test, on the other hand, one of the teachers who could not yet imagine using the Codey Rocky in the classroom in the pre-test, stated that she would want to use this device in the sixth or seventh grade in the context of programming. Another teacher, who stated in the pre-test that she had no ideas for the usage of it, answered in the post-test that she still had no ideas, but that she would build on the Students' results from the project to see how they could be transferred into reality or what possibilities already existed. In total, four teachers responded in the post-test that they could imagine using it in school (**Appendix B**, Chapter 11, pos. 2, G20.4).

The Neuron set is perceived as very positive both individually and as an extension of the Codey Rocky. The color scheme of the individual building blocks signalizing the purpose of the blocks was also positively emphasized. According to the teachers, this reinforces the inclusive character of the set and thus makes it easier to work with. The variable complexity, as with the Codey Rocky, also ensures a wide range of applications (**Appendix B**, Chapter 11, pos. 4, G40.2).

Overall, this digital creativity tool was also well received by the subjects of this study, as already in the pre-test three of the four teachers who answered this question stated that they would use the device in their own lessons outside of the project as a toy on the one hand and as an experimental kit for learners on the other. In addition, the set is intuitive and can be used from grade six in creative contexts without prescribing concrete tasks, since the urge to discover can be acted out here. One of the teachers also stated that she did not want to limit the use of the set to one grade level but wanted to use it in all grades. She confirmed this in the post-test and added that the complexity showed great variability. In the post-test, four of the five respondents said they would use the device outside of the project. It should be added here that one person would use it in grades five to seven, and another person noted that the Neuron Set was useful in science projects on the one hand, and as a pastime during breaks on the other. One teacher seemed to be particularly enthusiastic about the Neuron Set, stating that she would choose the Neuron Set if she were allowed to choose only one device for school, as it could be used in a variety of ways in science, arts, and social studies subjects. However, this teacher emphasized that she would never buy such a set because she was convinced that technology is always developing and therefore such a set could quickly become obsolete. Only one teacher stated in the pre- or post-test that they did not know whether they would use the Neuron set. However, these are two different teachers who did not fill out the corresponding test, so that no change can be determined here. The variability of the Neuron Set was described by many teachers as a positive aspect. It was also frequently mentioned that the set promotes the urge to discover and to be creative. The possibility of combining the set with the Codey Rocky was also emphasized by the teachers as a positive aspect. The haptics of the individual blocks, complexity and yet simplicity and highlighting the individual functions of the building blocks were also mentioned. Also, the Neuron Set promotes inclusive learning opportunities and ties into learners' interests (**Appendix B**, Chapter 11, pos. 4, G40.2).

The DJI Tello drone is suitable as a means of addressing several aspects of math and science education in the classroom. However, the math and science aspects should be central to reach this purpose, as it could otherwise distract too much from the actual subject matter. (**Appendix A**, Interview Transcription: René Foellmer, pos. 30).

One example was the discussion of possible flight paths for a load of water after being dropped from the drone on a plant. This discussion resembled an item of the Force Concept Inventory (FCI) which is regularly used in physics education. Nevertheless, this discussion was observed in the *elaborate* phase of a group concerned with the SDG 2: *promote sustainable agriculture* and was initiated by the problem solving process, without intervention by teachers. The possibility to program the drone, instead of controlling it, is positively emphasized by teachers (**Appendix B**, Chapter 11, pos. 3, G30.2).

Of the teachers involved in the project, two stated in the pre-test that they would also use the drones presented outside of the project, for example to take aerial photos. The teachers considered the drones equally suitable for higher grades, since responsible handling of the drones is important, and many questions can be raised. The other person who would use the drones outside of the project would use them in a foreign language and humanities class in grades seven and eight. Another teacher stated in the pre-test that she would not use the drone in her classes outside of the project because she did not have the confidence to develop a didactic concept and also did not have the subject expertise. This opinion changed in the post-test, with this teacher now being confident enough to use the drone in her own lessons after the project. Overall, of the five teachers who answered this question in the post-test, three said they would use the drone outside of the project. One person would continue to use it in projects and only one person answered that they did not know what they would use the drone for. Again, for the

drone, grades seven and eight were indicated as possible settings (**Appendix B**, Chapter 11, pos. 3, G30.4).

CONCLUSION

This research gives a brief answer to the first research question *How can STEAM education based on the 5E model be introduced in schools?*

The 5E-Model with an additional sixth phase has proven to be a good foundation on how to implement STEAM into school lessons with the help of digital creativity tools. Adding the Exchange phase as a sixth phase to the already established 5E Model seems to be a profitable expansion. On the one hand, it allows exchange between students and students, and students and teachers. On the other hand, it allows both teachers and educational researchers to collect more insights into the Students' way of learning by examining Student's presentations and prototypes. Finally, teachers get to know their students better and can prepare their future teaching in a more adjusted way. The effectiveness of this must be proven in further studies but this and another study conducted by the university of cologne emphasizing on six instead of five phases indicates the possible impact of this addition.

The use of videos to introduce the 7th grade students into the topic proved to be extremely beneficial and it became clear through the interviews and student results that a differentiated examination of the videos can be sufficient to motivate the learners. The devices used were quickly and persistently understood by the students through the introductions designed by university students. This is supported by the observation made within the Exchange phase, seeing the students having designed intelligent examples to explain how the devices work. The fact that most of the supporting students were not on site in the hybrid situation and the pupils therefore had to learn how to use the tools on their own supports the idea that STEAM tools are easy and intuitive to handle (at least compared to typical equipment in a traditional science lab). It also supports promising ideas of STEAM tools as tools to foster creativity, and reduce the workload on teachers, since all activities were guided only by work instructions. The potential of the devices is visible in various projects the students developed. After asking the teachers what a digital creativity tool suitable for STEAM should be able to do, it became clear that the previously mentioned aspect of intuition was the most important. Also, further features like sturdiness just as a prerequisite to promote creativity were mentioned by teachers as something that should be characteristic for a digital creativity tool. Those features can all be found in the tools used in this project as well as many other tools on the market. As we have furthermore seen, the project itself as well as the used digital tools were able to expand and deepen the 4C competencies (Creativity, Cooperation, Communication, and Critical Thinking) and further competencies according to the teachers' assessments.

The research question *What is the attitude of teachers toward the adoption of STEAM tools in the context of STEM Teaching and how does it change in the course of a 7-Week, on-the-Job training program?* is difficult to answer due to the data situation.

The teachers participating in the project mentioned many different features a digital creativity tool should offer. What seems to be important to many of them is that the tool should be intuitive to use. In other words, it should be obvious at first glance how the device can be used, so that with the help of such tools, basic computer literacy can be taught in a playful manner at an early age. In contrast, it was also mentioned that a digital creativity tool should have a certain complexity so that students remain motivated not only to learn on the device but also to explore its different facets. Other frequently mentioned characteristics are that such a tool should, above all, promote creativity and explorative learning. Furthermore, it should be versatile and combinable so that it is able to implement most of the ideas and conceptions of the students. Features that are important for everyday school life, such as robustness and safety, were also mentioned by the teachers.

What furthermore seems to be important, is the possibility to individualize the devices so that the students can build up a personal relationship with them. From a special education perspective, it was also important to the school's teachers that a digital creativity tool could be used by all students in one way, that being ideally for those unable to read or write.

Overall, only a small change is observable, since only a few teachers who completed the pre-test also completed the post-test. However, a tendency toward more readiness can be observed when the tools are considered individually. In the post-test, more people indicated that they wanted to use these or similar tools in the classroom. The research question cannot be answered in general terms, but at least the described tendency can be derived from the available data, since in the pre-test, none of the teachers stated that they had previously used a digital creativity tool in the classroom, whereas in the post-test several teachers stated that they would consider doing so in the future.

In order to answer the title question *How might we raise interest in Robotics, Coding, AI, STEAM and Sustainable Development in university and on-the-job teacher training?* conclusively it requires more research in the future and could be focusing on different areas of the basis laid with this paper. One example could be more in-depth research regarding individual tools or certain activities for the respective phases of the expanded 5E-Model. It is also possible to adapt the field study for other schools and try to get more teachers to answer the research forms, to gain more insight on this concept and also avoid having to use video conferences and involve the university students in a better way.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://uni-koeln.sciebo.de/s/aWF3sU4PomZVQk4.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was provided by the participants or their legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

AB, CS, and JH contributed to the conception and design of the study, and wrote sections of the manuscript. AB led the design of the pedagogical concept. JH led the design of the problem based learning scenarios for computational thinking and selected the appropriate technological platforms, performed the analysis of the interview with domain experts, and wrote the first draft of the manuscript, acted as the main author and revised the manuscript. JH and CS organized the cooperation with the school and conducted jointly the interviews. CS performed the analysis of the interview with teachers. SM developed and evaluated questionnaires to assess the influence of the videos in the engage phase on the students. All authors approved the submitted version.

FUNDING

This project was part of the "Qualitätsoffensive Lehrerbildung," a joint initiative of the Federal Government and the Länder which aims to improve the quality of teacher training. The programme was funded by the Federal Ministry of Education and Research (No. 01JA1815).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feduc.2022.872637/full#supplementary-material

Supplementary Figure 1 | Screenshot from the video regarding SDGs 2 zero hunger and 6 clean water and sanitation.

Supplementary Figure 2 | Screenshot from the video regarding SDGs 3 good health and wellbeing and 11 sustainable cities and communities.

Supplementary Figure 3 | 7th grade students use the modular neuron-set to build their own prototype of a remote-controlled robot that carries a water pump. In later versions they included a water tank as well as a measuring device for soil moisture to determine when the pump should be activated to water certain spots.

Supplementary Figure 4 | Students elaborate on how to apply significant payload to a small drone while maintaining stable flight.

REFERENCES

Aebli, H. (1985). Das operative prinzip. *Math. Lehren* 11, 4–6.

Bybee, R. (2009). "The BSCS 5E instructional model and 21st century skills," in *Paper Prepared for the Workshop on Exploring the Intersection of Science Education and the Development of 21st Century Skills*. (Ottawa: National Research Council).

Bybee, R., Taylor, J., Gardner, A., van Scotter, P., Carlson Powell, J., Westbrook, A., et al. (2006). *The BSCS 5E Instructional Model: Origins, Effectiveness and Applications*. Available Online at: http://www.bscs.org/bscs-5e-instructional-model (accessed April 12, 2021).

Döring, N., Bortz, J., Pöschl, S., Werner, C. S., Schermelleh-Engel, K., Gerhard, C., et al. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Berlin: Springer.

Duran, L. B., and Duran, E. (2004). The 5E instructional model: a learning cycle approach for inquiry-based science teaching. *Sci. Educ. Rev.* 3, 49–58.

Etkina, E., van Heuvelen, A., White-Brahmia, S., Brookes, D. T., Gentile, M., Murthy, S., et al. (2006). Scientific abilities and their assessment. *Phys. Rev. Spec. Top. Phys. Educ. Res.* 2:020103. doi: 10.1103/PhysRevSTPER.2.020103

Harris Insights & Analytics (2019). *Confidence in Learning Poll*. Available Online at: https://mojebambino.pl/lego-education/wp-content/uploads/2022/01/ankieta-harris-poll.pdf (accessed November 15, 2021).

Häussler, P. (1987). Measuring students' interest in physics - design and results of a cross-sectional study in the Federal Republic of Germany. *Int. J. Sci. Educ.* 9, 79–92. doi: 10.1080/0950069870090109

Jackson, H., Basham, J., Thomas, K., and Hunt, C. (2020). "Designing STEAM learning environments," in *Challenges and Opportunities for Transforming from STEM to STEAM Education*, eds K. Thomas and D. Huffman (Hershey, PA: IGI Global), 1–23. doi: 10.5642/steam.20200402.04

Kalogiannakis, M., Tzagaraki, E., and Papadakis, S. (2021). "A systematic review of the use of BBC micro: bit in primary school," in *Proceedings of the 10th Virtual Edition of the International Conference New Perspectives in Science Education*, Crete, 379–384.

Kircher, E., Girwidz, R., and Häußler, P. (eds). (2014). *Physikdidaktik: Theorie und Praxis*. Berlin: Springer.

Krakower, B. (2018). *Getting Started with STEAM: Practical Strategies for the K-8 Classroom*. Milton: Routledge.

Kuckartz, U. (2018). *Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung*. Weinheim: Beltz Juventa.

Leong, D., and Bodrova, E. (1996). *Play: A Vygotskian Approach*. Available Online at: https://video.alexanderstreet.com/watch/play-a-vygotskian-approach (accessed March 2, 2021).

Makeblock. (2019). *How to Teach STEAM Education in the Classroom*. Available Online at: https://education.makeblock.com/blogs/blog-primary-education/48109/ (accessed March 4, 2021).

Namdar, B., and Kucuk, M. (2018). Preservice science teachers' practices of critiquing and revising 5E lesson plans. *J. Sci. Teach. Educ.* 29, 468–484. doi: 10.1080/1046560X.2018.1469188

Papadakis, S. (2022). Can preschoolers learn computational thinking and coding skills with scratchJr? A systematic literature review. *Int. J. Educ. Reform* 1–34. doi: 10.1177/10567879221076077

Quigley, C., Herro, D., Shekell, C., Cian, H., and Jacques, L. (2020). Connected learning in STEAM classrooms: opportunities für engaging youth in science and math classrooms. *Int. J. Sci. Math. Educ.* 18, 1441–1463. doi: 10.1007/s10763-019-10034-z

Reich, K. (2018). Chapter 1. The Helios school – inclusive university school in the city of cologne. *Int. Res. High. Educ.* 3, 4–22. doi: 10.5430/irhe.v3n1p4

Tamim, R. (2011). What forty years of research says about the impact of technology on learning: a second-order meta-analysis and validation study. *Rev. Educ. Res.* 81, 4–28. doi: 10.3102/0034654310393361

Thomas, K., and Huffman, D. (2020). "Becoming STEAM: perspectives from school leaders," in *Challenges and Opportunities for Transforming From STEM to STEAM Education*, eds K. Thomas and D. Huffman (Hershey, PA: Business Science Reference), 104–129. doi: 10.4018/978-1-7998-2517-3.ch005

Turan, S. (2021). Pre-service teacher experiences of the 5E instructional model: a systematic review of qualitative studies. *Eurasia J. Math. Sci. Tech. Ed.* 17:em1994. doi: 10.29333/ejmste/11102

United Nations Department of Economic and Social Affairs (2021). *Sustainable Development Goals Report 2021*. New York, NY: United Nations Department of Economic and Social Affairs.

United Nations Development Programme (2021). *The SDGs in action: What are the Sustainable Development Goals?*. Available Online at: https://www.undp.org/sustainable-development-goals#:~:text=The%20Sustainable%20Development%20Goals%20(SDGs)%2C%20also%20known%20as%20the,people%20enjoy%20price%20and%20prosperity (accessed January 11, 2022).

Wandari, G. A., Wijaya, A. F. C., and Agustin, R. R. (2018). The effect of STEAM-based learning on students' concept mastery and creativity in learning light and optics. *J. Sci. Learn.* 2, 26–32. doi: 10.17509/jsl.v2i1.12878

Yakman, G. (2008). "STΣ@M education: an overview of creating a model of integrative education," in *Proceedings of the Pupils Attitudes Towards Technology*, Netherlands.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Henze, Schatz, Malik and Bresges. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPENDIX

Appendix A | Interview data with teachers (L) and students (B/M) participating in the project and manufacturer Yu Hu as well as special education teacher René Foellmer.
Appendix B | Data regarding video analysis and survey data from participating teachers regarding digital creativity tools.

frontiers | Frontiers in Psychology

ORIGINAL RESEARCH
published: 29 June 2022
doi: 10.3389/fpsyg.2022.892276

Retrieval Practices Enhance Computational and Scientific Thinking Skills

Osman Yaşar[1], Jose Maliekal[1], Peter Veronesi[1], Leigh Little[1], Michael Meise[2] and Ibrahim H. Yeter[3]*

[1]The College at Brockport, State University of New York, Brockport, NY, United States, [2]Rochester City School District, Joseph Wilson Magnet School, Rochester, NY, United States, [3]National Institute of Education, Nanyang Technological University, Singapore, Singapore

The notion of teaching experts' habits of mind (e.g., computational thinking and scientific thinking) to novices seems to have inspired many educators and researchers worldwide. In particular, a great deal of efforts has been invested in computational thinking (CT) and its manifestations in different fields. However, there remain some troubling spots in CT education as far as how to teach it at different levels of education. The same argument applies to teaching scientific thinking (ST) skills. A remedy has been suggested to narrow CT and ST skillsets down to core cognitive competencies so they can be introduced in early and middle grades and continue to be nurtured during secondary and post-secondary years. Neuroscientists suggest that the *act of (computational) thinking* is strongly linked to the acts of information *storage/retrieval* by our brain. Plus, years of research have shown that retrieval practices promote not only knowledge retention but also inductive reasoning and deductive reasoning. Not surprisingly, these reasoning skills are core elements of both CT and ST skillsets. This article will mesh the findings of a teacher professional development with the existing literature to lay a claim that retrieval practices enhance CT and ST skills. The study offered training to secondary school teachers ($n = 275$) who conducted classroom action research to measure the impact of retrieval practices on teaching and learning of STEM and CT concepts. We used a quasi-experimental research design with purposeful sampling and a sequential mixed-methods approach focusing on the impact of professional development on teacher outcomes and, in turn, on student outcomes. A survey of teacher participants showed that the majority (96%) of survey respondents ($n = 232$) reported a good understanding of retrieval strategies, and how relevant ideas can be implemented and tested in the classroom. A large number of action research (target-control) studies by teachers ($n = 122$) showed that students who learned STEM and CS concepts through retrieval practices consistently scored 5–30% higher than those using the usual blocked practice. In most cases, the difference was statistically significant ($p < 0.05$). While the study contributes to retrieval practices literature, those looking for best practices to teach core CT and ST skills should benefit from it the most. The study concludes with some recommendations for future research based on the limitations of its current findings.

Keywords: retrieval practices, computational thinking, scientific thinking, inductive reasoning, deductive reasoning

OPEN ACCESS

Edited by:
Stamatios Papadakis, University of Crete, Greece

Reviewed by:
Urip Sulistiyo, University of Jambi, Indonesia
Dimitrios Stamovlasis, Aristotle University of Thessaloniki, Greece

****Correspondence:***
Osman Yaşar
oyasar@brockport.edu

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

Received: *08 March 2022*
Accepted: *27 April 2022*
Published: *29 June 2022*

Citation:
Yaşar O, Maliekal J, Veronesi P, Little L, Meise M and Yeter IH (2022) Retrieval Practices Enhance Computational and Scientific Thinking Skills. Front. Psychol. 13:892276. doi: 10.3389/fpsyg.2022.892276

INTRODUCTION

More than two decades ago, when computational science (an interdisciplinary practice incorporating modeling, simulation, visualization, and problem solving) emerged as a new workforce strategy for institutions of higher education (IHEs) and as an innovative teaching pedagogy for K-12, many had hoped that it would revolutionize the STEM education. As such, in 1998, SUNY College at Brockport launched the nation's first undergraduate degree program in computational science (Yaşar et al., 2000; Yaşar and Landau, 2003; Turner et al., 2011). In 2006, Jeannette Wing, an influential computer scientist and an assistant director at the US National Science Foundation (NSF), mobilized significant NSF resources, rebranded computational science as computational thinking (CT), and claimed in her 2006 essay that CT should be taught as a fundamental skill in public schools just like reading and writing (Wing, 2006). The notion of teaching computational thinking (CT) as a fundamental competency seems to have inspired many educators and researchers worldwide. However, teaching experts' habits of practice to novices is inherently problematic because of prerequisite content knowledge and practice skills needed to engage in the same thinking processes (Kirschner et al., 2006), not to mention the cost of providing them a similar environment to conduct inquiry and design. A remedy has been suggested to link experts' habits of practice to fundamental cognitive processes so we can narrow their skillsets down to more basic competencies that can be taught to young students.

Linking computation to cognition is not a new idea—in fact, it goes as far back as to the time of human computers during Babylonians (Denning and Tedre, 2019). Obviously, after the electronic computer age began in the 1940s, the term "computer" has often referred to electronic devices rather than human agents. What led to the design of electronic computing 80 years ago in the first place was that if thoughts (i.e., information) can be broken up into simple quantifiable constructs and algorithmic steps, then machines can add, subtract, or rearrange them as our brains do (Turing, 1936). The human brain employs a distributed network of neurons to rearrange information (Hebb, 1949). As such, information is stored into the memory *via* a specific pattern of neurons placed on a pathway and fired together. Arrival of new information lights up all related cues, neurons, and pathways in a distributive process that is similar to the top-down action in **Figure 1**, whereby a new concept is broken up into related pieces. The converse, retrieving information, involves reassembly of the original pattern of neurons and pathways in an associative process similar to the bottom-up action in **Figure 1**. Retrieval, in other words, is not an act of merely recalling facts and figures. It is a process of reassembly involving different pathways that are linked to one's knowledge. What is retrieved is not a carbon copy of the original but a re-imagined copy of the original with some holes and/or extra bits. Neuroscientists see little or no distinction between the acts of information storage/retrieval and the act of (computational) thinking (Montague, 2006; Brown et al., 2014).

Our brain's inclination to process information in an associative and distributive fashion, as well as to store and retrieve memories and concepts in a scatter and gather fashion by a distributed neural network, appears to be a manifestation of a basic duality engrained in the fabric of matter and information. Quantifiable things appear to behave in one of only two ways (as in **Figure 1**): they either unite associatively to form bigger constructs or break down distributively to smaller ones. Such a duality at the core of information processing by a computational mind carries itself up to higher-level cognitive processes, such as deductive reasoning in the form of distributive processing of information and inductive reasoning in the form of associative processing of information (Dunbar and Klahr, 2012; Yaşar, 2017, 2018).

We are all naturally inclined to employ inductive thinking and deductive thinking in everyday life. They are the two major cognitive competencies at the root of the CT skillset (Wing, 2006; Yaşar et al., 2016; Yaşar, 2018; Denning and Tedre, 2019; Mills et al., 2021), which are often cited as abstraction and decomposition skills. We all employ computational thinking by the virtue of having a computational mind. However, when used together in certain ways, the combination of inductive and deductive thinking becomes a much more powerful skill, as first described by Kant (1787) more than two centuries ago. For example, through *iterative* and *cyclical* use of inductive and deductive thinking, as depicted simplistically in **Figure 1**, does the conceptual change occur in our learning progression, all the way from childhood to the adulthood (Carey, 1985). Conceptual change is also at the heart of the scientific thinking (Vosniadou, 2013) both at the level of an individual scientist, or those who think like scientists, as well as that of the scientific progress by the scientific community (Kuhn, 1962; Thagard, 1999). Not surprisingly, imaging techniques have revealed that scientific thinking is not just thinking about the content (of sciences); it encompasses a set of cognitive processes, such as conceptual change, that transcend the field of science (Dunbar and Klahr, 2012). These processes include (a) problem solving, (b) design and modelling, (c) hypothesis testing, (d) concept formation, (e) conceptual change, and (f) reasoning

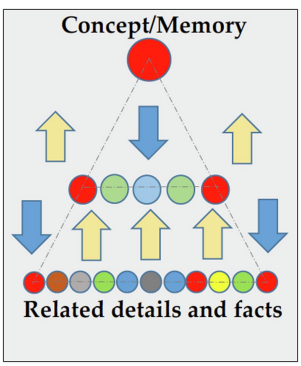


FIGURE 1 | Distributive and associative ways of information storage and retrieval. Figure © 2017 IEEE. Reprinted, with permission, from (Yaşar, 2017).

(inductive, deductive, abductive, causal, and analogical thinking). According to Thagard (2012), these ST processes are no different from those employed in everyday living by non-scientists—the difference comes from how they are used. In a sense, what distinguishes ST from everyday thinking (i.e., computational thinking) is that while CT involves any use of inductive and deductive thinking, ST involves *iterative* and *cyclical* use of these two opposite reasoning skills to accomplish conceptual change and other ST skills listed above (Yaşar, 2021).

A great deal of efforts has gone into analyzing CT as a result of recent technological advancements which have affected our professional and personal lives. These efforts include definition of CT (Papert, 1980; Wing, 2006; Guzdial, 2008; Denning, 2009; Aho, 2012), its cognitive essence (Yaşar, 2017, 2018) and manifestations in different fields and ways to teach it at different levels of education (Denning, 2017a,b; Yadav et al., 2017; Denning and Tedre, 2019; Tedre and Denning, 2021). For a literature review, see Grover and Pea (2013); Angeli and Giannakos (2019); Denning and Tedre (2019); Kakavas and Ugolini (2019) and Saqr et al. (2021). In the 1990s, the focus was on literacy and fluency issues with a push to teach programming. The arrival of easy-to-use M&S tools, which hid the underlying mathematics and programming, allowed a new way of studying scientific phenomena and teaching CS principles in the 2000s. The present decade has seen even easier tools, such as mobile apps, to support children's computational thinking and literacy skills (Papadakis, 2021).

Today, there are plenty of tools available for teaching various CT skills. However, the discourse on what it means to different stakeholders continues to this date. Some have suggested to categorize it as "CT for beginners" and "CT for professionals." The same argument applies to teaching of ST skills. There is a need for innovative practices to provide continuity in CT and ST education all the way from elementary to post-secondary. We posit that an information processing approach to cognition, as briefly explained above, allows us to teach core CT/ST cognitive competencies with appropriate grade-level challenges and skills. If indeed the acts of information storage and retrieval strongly correlate to the act of computational and scientific thinking, then all we need to do is to strengthen those information processes. Whatever practices we come up with to strengthen them, one way to measuring their effectiveness could be through the act of information retrieval itself. We are lucky in that sense because long before such correlation was made, researchers in cognitive psychology had been studying the impact of memory retrieval practices on knowledge retention and other cognitive functions as explained in the next section. This article establishes ground that retrieval practices can be used as a way of strengthening CT and ST skills. We hope that the findings from our professional development program and related action research by participating teachers will shed a light on the discourse about CT and ST education. While the practitioners would benefit from reproducing similar results from a tested and scalable strategy, the researchers could expand their efficacy studies, *via* retrieval practices, to the teaching of more basic CT and scientific thinking (ST) concepts at a variety of grades. An approach such as the retrieval practice, which causes learning to stick and promotes core CT and ST skills, could have a broad impact in STEM education.

MATERIALS AND METHODS

Teacher Professional Development

A grim situation occurring in most urban and rural school districts' math and science achievement scores (National Center for Education Statistics, 1996; Qazi et al., 2020) has drawn concerns from local and state groups as well as the higher education institutions in our area. Educators point to poverty rates, lack of resources, and poor parental involvement as its root causes. Remedies suggested by State agencies include recruitment and training of effective leaders and teachers, ongoing professional development for teachers, maintaining standards, offering a rigorous curriculum for all, improving instruction *via* new technologies and pedagogies, and involving community; some of which have been adopted in our work, including the premise that, more than anything else, improving the teacher quality profiles would help improve STEM and CT education.

To support use of retrieval practices in secondary schools as an intervention, we offered and iteratively improved a 3-tier (beginner/intermediate/advanced) professional development (PD) to STEM teachers from partnering school districts in the area. The decision to offer a multi-tier program mainly came from: (a) our experience of previous PD effort (Yaşar et al., 2014) (b) recommendations by the ESC initiative at Los Angeles (Margolis et al., 2008; Goode and Margolis, 2011) (c) questions we got from districts to assure them continuing support and training, and (d) reports published by the Urban Institute (Beatriz, 2005) and others such as L-Horsley et al. (2010). The PD was based on the Iowa Chautauqua Learning Model (Blunck and Yager, 1996), with a summer institute and a series of academic-year training and debriefing workshops and mentoring activities. Attendance was voluntary, but project funds and school principals encouraged teachers to complete all 3 levels of training. Throughout the separate components of the PD, we used an expert-teacher-student cognition cycle and well-known principles of effective PD (Guskey, 2000; D-Hammond and Bransford, 2005; L-Horsley et al., 2010), including: (1) examining student work, (2) demonstration lessons, (3) lesson plans study, and (4) case discussions.

The beginner-level training trained participants on practicing tools (e.g., Google Forms and Microsoft Forms) and *spaced-out* retrieval strategies (see the next section for details of various retrieval practices mentioned here). Given the situation during the COVID-19 pandemic, we included online training as an option to increase the number of participants. The intermediate level training focused on (a) *distributed* retrieval practice, (b) use of the *rate of change* to model and simulate problems of interest (e.g., growth of disease and motion) with Excel, and (c) basics of conducting Action Research to improve classroom instruction (Ferrance, 2000). The advanced-level training included *generative* retrieval practices with simulations (SIMs) as well as basic programming skills (Scratch and Python). At the completion

of the 3-year training, teacher participants were expected to learn and deploy various retrieval strategies, understand cognitive underpinnings of memory retrieval, CT, and ST, and conduct modeling and simulation of scientific phenomena using alternative tools (hands-on, Excel, web-based SIMs, Scratch, and Python). Summer teachers formed the backbone of the project, and their PD activity continued throughout the academic year in various forms to promote their skills to engage other teachers and help test and revise the resources for their colleagues and students. During the summer, returning teachers were asked to share evidence in the form of artifacts and presentations about a full year they left behind. Identifying parts of curriculum that are hard to teach where retrieval strategies might help was made a priority. We met the teachers where they were so they might build from their zone of proximal development (Vygotsky, 1978). Practice and peer sharing and critique throughout the school year were an integral part of this project. A sense of community and trust was nurtured *via* meetings, school visits, and online communications. We used a quasi-experimental research design with purposeful sampling and a sequential mixed-methods approach focusing on the impact of professional development on teacher outcomes and, in turn, on student outcomes. Details of the teacher PD program can be found in Yaşar et al. (2021).

Retrieval Practices

Retrieval practices are grounded in research which suggests that new concepts are retained in memory far longer if the retrieval process is effortful, spaced-out, interleaved, and generative in nature (Brown et al., 2014). The empirical evidence further shows that (low-stakes) practice tests can serve as learning tools (Karpicke et al., 2014; Agarwal et al., 2020) to help students retrieve newly taught concepts in effortful ways that will, in essence, burn new knowledge into memory through connected understandings rather than rote memorization. A cognitive and constructive effort to recall recently learned concepts through connections to what had been previously stored in long-term memory has a much greater chance of being retained longer. If memory has lapsed, each time students engage in purposeful recall, it reassembles concepts through different pathways or links to one's knowledge. One of the ways to accomplish these learning pathways is *spaced-out retrieval* (SR) practice through quizzes, self-testing, or flashcards. Spacing allows some forgetting that will trigger a cognitive effort for retrieval while repeated retrieval leads to more durable memories. Another one is *interleaving retrieval* (IR) practice that help link newly learned concepts to different contexts, changing conditions and parameters, and even multiple subjects. Often times, spaced-out and interleaving are used together and called *distributed retrieval* practice. A third one is *generative retrieval* (GR) practice; the act of trying to answer a question or attempting to solve a problem rather than being presented with an answer or the solution. The generative retrieval refers to both experiential and exploratory learning *via* trial-and-error. It is generally known to lead to complex mastery and greater knowledge of the interrelationships among parts of the problem and its solution. A learner may be able to arrive at the understanding of a phenomena on their own (Grabowski, 2004). One can practice generation by predicting an outcome or a concept before it happens while simply testing out their prediction and observing and noting the results—much like modeling and simulation (Yaşar and Maliekal, 2014).

These retrieval practices are all consistent with active learning and scaffolding strategies by which students are challenged in incremental steps as they build more skills (Mooney, 2013). The distributed retrieval (DR) practice has been tested in social sciences (Brown et al., 2014; Agarwal and Bain, 2019), math (Rohrer et al., 2014), natural sciences (Yaşar et al., 2019a,b, 2021; Samani and Pan, 2021), and computer sciences (Casanova et al., 2020) against the usual *blocked practice* whereby students learn to apply a certain method to solution of various questions of the same type on only one topic. An example of *blocked practice* would be to apply the Pythagorean Theorem to compute the hypotenuse of a right-angle triangle, $a^2 + b^2 = c^2$. Students need not learn to choose a solution method because problems within a blocked practice require the same strategy. In a *distributed* retrieval practice, two or more types of questions (and topics) are mixed as in **Table 1**, and students are faced with choosing a strategy to solve a problem. Despite the growing evidence about the impact of retrieval practices, they are yet to become prevalent in schools. The blocked practice is still the norm for many reasons, including a belief that repeated practice alone of the same drill builds up skills. The distributed practice does require a re-arrangement of topics within practices and lectures, but what is missing perhaps is a theory or framework to link retrieval practices to other educational reforms, such as CT, that are underway. This paper reports findings of a 3-tier teacher professional development, along with an extensive Action Research effort to examine the impact of distributed and generative retrieval practices on teaching and learning in secondary school STEM classrooms.

Use of retrieval practices in the teaching of CT skills is very new. The only study that we have found in the literature is the one by Casanova et al. (2020) who examined if DR practices had durable effects on retention and learning of CT and programming concepts. A total of 10 elementary schools participated in a quasi-experimental study, consisting of 6 weekly sessions on CT concepts, including an introduction to Makey Makey (week 1), inequality symbols (week 2), identification and understanding of programming concepts such as conditions in week 3, loops in week 4, and inputs/outputs in week 5. A pre-questionnaire probed student familiarity and prior knowledge of Scratch, Makey Makey, and CT in general. Each session included hands-on practices with Scratch (Resnick et al., 2009; Funke et al., 2017) or Makey Makey (Silver et al., 2012), as well as DR practices (week 1 through 4) which involved quizzes made *via* Kahoot! with interleaved questions from current and preceding sessions. Students were given a review test at the end of the fifth session and an unannounced test a week later. The control condition and sample size were affected by COVID-19, and researchers used a *t-test* to compare scores of one group ($n = 20$) that consistently participated in all activities. While a week time may be short to test durability of concepts, the average score was higher for the unannounced test than the review test (72.6 vs. 67.9).

TABLE 1 | An experimental set up to compare blocked vs. distributed practices.

Group	Assign #1	Assign #2	Time Delay	Unit Review	Time Delay	Unannounced
Blocked	Topic X	Topic Y	3	&	15–30	Test
5 days	Topic X	Topic Y	days	Test	days	
Distributed	Topic Y	Topic X				
5 days	Topic X	Topic Y				

The difference may not be all that significant ($p = 0.19$) but maintaining one's score after 1 week is even a success. Test scores often go down over time due to loss of knowledge unless there is an intervention. There may have been other confounding factors such as students' continuing exposure to topics in other units or similar courses.

RESULTS

Teacher Professional Development

In a five-year timeframe (2016–2021), we trained 275 secondary school STEM teachers from 42 regional school districts (SDs), including 33% from urban, 30% from rural, and 36% from suburban SDs. 160 teachers returned for 2nd year advanced-level training, and of those, 40 teachers returned for 3rd year expert-level training. Yearlong support was offered to help deploy these strategies and conduct action research (Ferrance, 2000) in the classroom. Two independent evaluators were hired to design and conduct quantitative surveys and focus group interviews with teacher participants. One of these was Brockport Research Institute (BRI), which had previously conducted evaluations for over 60 National Science Foundation projects. BRI assigned two evaluation experts to design, collect, and analyze both teacher and student data. The other external evaluator was an education research faculty (Dr. John Tillotson) from a Teacher Education program at Syracuse University. While the BRI staff focused mainly on new participants who attended spaced-out and interleaved retrieval PD activities, Dr. Tillotson worked mainly with those returning to attend additional training on generative retrieval. Findings from both evaluation efforts took place simultaneously 3 years in a row and their overlap served as a triangulation of results and review of participant progress from distributed retrieval to generative retrieval. The teacher PD aspect of our program, along with its quantitative findings, has been described in an earlier publication (Yaşar et al., 2021). Below, we will only give a brief account.

In quantitative teacher surveys (see **Table 2**) conducted annually between 2017 and 2018, scores above 4.0 on a 5-point scale for nearly every survey item suggested that teachers found the program to be both engaging and effective in raising their awareness of retrieval strategies and of the research base supporting their efficacies within STEM classrooms as indicated by mean. The respondents highly valued the opportunities to collaborate with other teachers in designing online retrieval practices and considering possible research designs for studies they planned to conduct during the upcoming school year. Ninety-six percent (96%) of the teacher respondents highly valued the opportunity to learn about distributed retrieval strategies and the research base supporting the effectiveness of this pedagogical approach in K-12 classrooms. A majority of the teacher participants (96%) also indicated that the time devoted to accomplishing each of the primary workshop objectives was appropriate and that the workshops were effective in helping them develop a clear understanding of their roles and responsibilities pertaining to the classroom-based research component of the initiative (93%).

In another subsequent quantitative teacher survey conducted after the 2019 summer workshop on the generative retrieval strategy and related deployment, the data shown in **Table 2** suggest that the workshop gave them a good understanding of the generative retrieval practice, enhanced their skills to use SIMs and mobile apps, increased their confidence and ability to design action research, provided a chance to interact with colleagues, and helped them design an experiment to measure the impact of SIM-based generative retrieval. All of participating teacher respondents (100%) found the workshops to be valuable overall (strongly agreed or agreed). More specifically, they all indicated that the workshops were effective in providing them with the conceptual knowledge and practical skills necessary to effectively engage students in SIM-based generative retrieval during the 2019–2020 school year. Similarly, 92% of the STEM teacher respondents highly valued the opportunity to learn about generative processes and the research base supporting the effectiveness of these pedagogical approaches in K-12 classrooms. All teacher participants indicated that the workshop prepared them with knowledge and skills to successfully design, implement, and even conduct classroom-based action research project to assess the impact of intervention on students' learning outcomes. The surveys' construct and face validities had been confirmed by their designers and the experts they hired.

The surveys were followed up by an enriched qualitative case study (focus group interviews) to explore the meaning of trends/findings in the quantitative part of the study. Two independent experts coded the responses and used an inductive process (Creswell, 2012) to form broader themes. For example, one of the broad themes was that the workshops afforded teachers deeper insight into the practical aspects of the classroom implementation of retrieval strategies and the positive effects research has shown these techniques to have on student learning outcomes. Another theme was that workshops provided ample time for the participants to delve deeply into their work in creating the required SIM-based learning exercises that would be implemented in the following school year. Evaluators used several forms of validation, including triangulation *via* data from multiple sources and member checking by asking project teachers and faculty to review the findings (Fincher and Petre, 2005). Further details of qualitative findings

TABLE 2 | Likert-scaled questions and answers from summer workshops.

The following questions apply to spaced-out (SR), interleaved (IR), and generative (GR) retrieval strategies/practices as indicated by enclosed within brackets.	Score (out of 5)		
	2017 SR $n = 10$	2018 IR $n = 26$	2019 GR $n = 18$
1. The summer workshop dates and times fit well with my schedule and commitments.	4.8	4.72	4.94
2. The goals and expectations were clearly articulated by the project leadership team.	4.9	4.16	4.94
3. The time devoted to accomplishing each of the primary objectives was appropriate.	4.9	4.36	4.88
4. After the workshop I now have a deeper conceptual understanding of the research and literature supporting the use of [SR/IR/GR] retrieval practices in the classroom.	4.9	4.56	4.82
5. The workshop enhanced my skills in using digital devices, relevant tools [SIMs], and mobile Apps to assess students' understanding of important concepts.	3.7	3.96	4.82
6. The workshop increased my confidence and ability to design classroom-based research to investigate the effectiveness of [SR/IR/GR] retrieval practices	4.5	4.24	4.60
7. The workshop provided me a chance to interact with colleagues to discuss the use of SIMs and digital mobile App development ideas and potential research projects	4.6	4.08	4.92
8. The project overview, leadership, and framework discussion for the research needs was effective in helping me understand my role and responsibilities as a participant	4.4	4.24	4.60
9. The workshop was effective in helping me design Action Research cycles to test the impact of SIM-based generative retrieval on my students' learning outcomes.	NA	NA	4.88
10. Overall, the workshop was effective in preparing me with the knowledge and skills necessary to successfully participate in the project during the upcoming school year.	4.7	4.32	4.88

are being published separately due to their nature and volume. The main focus of our article here is on classroom action research by teachers, which will be given in the following two subsections.

Distributed Retrieval Practice

A large cadre ($n = 82$) of trained teachers participated in the Action Research to explore the impact of distributed retrieval (DR) practices (the combination of spaced-out and interleaving) in teaching and learning of various topics in geometry (e.g., quadrilaterals and altitudes in right-angled triangles), biology (photosynthesis, respiration, and Punnett squares), chemistry (Le Chatelier's Principle, Potential Energy Diagrams, heat, and half-life decay), and earth sciences (erosion and planetary motion). Of these, 68 teachers conducted research 3 years in a row using different classes and student populations while improving their methods of selecting topics and students more randomly and increasing the delay time between pre- and post-assessments. Google (and Microsoft) Forms provided a framework by which students could use a mobile program to record their thinking while the assessment data could be collected by the teacher and ultimately shared with students for immediate feedback. The science teachers used a question bank (e.g., CastleLearning™) to draw questions from. Students were placed randomly into control and target groups of equal sizes ranging from 12 to 32 depending on each study. About half the cases followed a research design whereby one group (A) followed the blocked (BL) strategy while the other group (B) used a DR strategy for practices and assignments. Other cases followed a design whereby each group (A and B) was taught using both strategies (blocked and distributed) alternately, though care was taken to make sure that group placement was not visible to students—that is, all students participated in simultaneous classroom instruction. Instruction for each topic lasted for the same number of school days with both strategies. An in-class review of both topics was conducted a short while after completion of instruction. The review was concluded with a test which in some cases served as a baseline. Finally, an unannounced test was conducted later to measure student retention of the content knowledge 15–30 days after the review. In most cases, teachers conducted pre- and post-activity assessments with multiple-choice questions on all control and target groups to identify and reduce the number of confounding variables and triangulate the results as much as possible.

In 82 independently run control-target experiments, like the one in **Table 1**, students who learned STEM and CS concepts through DR practices scored better than those using blocked (BL) practice in 80 cases (95% of the time). The average of pre- and post-scores for DR groups was 70.7 (pre) and 71.1 (post), meaning that knowledge was retained, whereas the average scores were 70 (pre) and 65 (post) for BL groups, meaning information was lost ($p < 0.007$), during the blocked practice. The difference between DR & BL post-assessment averages (71.1 vs. 65) was also significant ($p < 0.015$). Standard deviation in group average scores was 15 (pre) and 16 (post) for BL and 15 (pre) and 12 (post) for the DR groups, implying more consistency in post-assessment scores by the DR groups. A few representative studies are shown below to illustrate more details. *T*-test statistics was used to examine the significance of differences both between and within groups.

- 10th Grade Chemistry (Topics: Half-life, Heat formula): While Group A ($n = 20$, blocked) and Group B ($n = 18$, distributed retrieval) scored about the same (Group A at 63.5 vs. Group B at 61.4) at the review (pre) test for half-life, their performance on an unannounced test given 30 days later differed significantly ($p = 0.014$) with Group A's average being

TABLE 3 | Comparing 9th grade chemistry classes using blocked practice vs distributed retrieval practice.

	Group A $n = 22$	Group B $n = 22$	*Is the difference statistically significant?*
Topic: Photosynthesis	**BLOCKED**	**DISTRIBUTED**	
Pre-test	53.66	45.86	No ($p = 0.19$)
Post-test	57.47	62.04	No ($p = 0.55$)
Is the difference significant?	No ($p = 0.59$)	Yes ($p = 0.001$)	
Topic: Respiration	**DISTRIBUTED**	**BLOCKED**	
Review test	15.59	20.27	No ($p = 0.14$)
Post-review test	57.72	50.90	No ($p = 0.33$)
Is the difference significant?	Yes ($p < 0.01$)	Yes ($p < 0.01$)	

40 vs. Group B's 63.3. On the topic of heat, while Group A scored significantly ($p = 0.027$) higher than Group B (81.1 vs. 68.8) at the review test, its performance (32.6) on an unannounced test given 30 days later fell substantially below Group B's (42.4). The drops in performance by both groups were statistically significant ($p < 0.0006$), yet the drop by the group with the blocked practice was twice as high as the drop by the group with the distributed retrieval practice.

- 9th Grade Chemistry (Topics: Photosynthesis, Respiration): Two groups ($n = 22$ each) experimented with alternating practices on different topics. While Group A used the blocked practice to learn about photosynthesis, Group B used the distributed retrieval practice. Similarly, while Group B used the blocked practice to learn about respiration, Group A used the distributed retrieval practice. Groups were compared to themselves (pre-test vs. post-test) to see their self-improvement and to their counterparts (post-tests) which learned the same topic *via* different practices. As shown in **Table 3**, while groups that used the blocked practice scored higher on pre-tests, they scored lower than the groups with the distributed practice. While pre- and post-test differences between groups are not statistically significant, the improvements by the distributed groups from pre-test to post-test are significant and large enough to exceed their counterpart.
- 9th Grade Technology Education (Topics: Design, Drawing, and Production). Group A ($n = 17$, blocked practice) and Group B ($n = 17$, distributed retrieval practice) scored 65.6 and 72.5 on the pre-delay test. 30 days later, the score for the group with distributed retrieval remained about the same (76.25) while the score for the group with blocked practice went down significantly ($p = 0.04$) to 56.87.
- 7th Grade Science (Unit: Motion, Topic: Acceleration). Group A ($n = 21$) and Group B ($n = 20$) were taught acceleration by the same teacher. 15 days later, both groups received a unit review, followed by a test to set a baseline for the recall later. As shown in **Table 4**, both groups scored the same right after the unit review. In an unannounced test 15 days later, however, the average score for the group (A) with blocked practice decreased significantly ($p < 0.01$) by 22.8%, whereas the average score for the group with distributed retrieval (B) decreased by only 5.5%, indicating that retrieval practices helped students retain knowledge better than the blocked practice.
- 7th Grade Biology (Topic: Punnett squares): The average score of 4 daily assignments conducted in the same week by Group A (blocked, $n = 27$) and Group B (distributed retrieval, $n = 29$) were about the same (46.79 vs. 47.34 out of 100). As shown in **Table 5**, the average scores on the review test were 55.88 (Group A) and 52.45 (Group B). When an unannounced quiz was given to both groups 15 days after the review test, Group B not only retained its knowledge of the topic but outscored Group A while improving its average significantly by 8 points ($p < 0.01$) to 60.21. Group A scored slightly better with 2-points compared to its review test. A similar trend was seen in the analysis of each student's progress for both groups: 19 students in the distributed retrieval group increased their score while 8 decreased versus 14 increasing and 10 decreasing in the blocked group.
- 7th Grade (Topics: Erosion, Planetary Motion): As shown in the **Table 6**, Group A (distributed, $n = 30$) outperformed Group B (blocked, $n = 31$) by 9% on a post-test on weathering and erosion; a difference that is statistically significant. In another experiment on planetary motion and the effect of mass on the gravity of an object, Group B (distributed, $n = 21$) outperformed Group A (blocked, $n = 29$) by 38%; a difference that is also statistically significant. In the same experiment, the distributed group outperformed the blocked group by 30% on levels of organization (progression of levels by cell(s) to reach an organism).

Generative Retrieval Practice

With a vast array of simulations (SIMs) available (in-house and on the internet) as surrogates for real phenomena, student comprehension of a STEM topic or phenomena was compared using generative retrieval practices with SIMs against regular practice with text and illustrations. Neither group had been previously introduced to topic-related concepts whose understanding was the purpose of this phenomena-first experiment. Secondary school teachers ($n = 40$) in this study selected up to 6 topics to compare learning in control-target groups using 3-level Socratic level questions (based on Bloom's Taxonomy), testing for growing complexity, interrelationships, and greater content knowledge (see **Table 7**). Level 4 and higher were optional. While students were not randomly selected (they came from each teachers' classes), the groups

TABLE 4 | Comparing 7th grade science classes using blocked practice vs distributed retrieval practice.

Topic: Motion	**Group A, *n* = 21 BLOCKED**	**Group B, *n* = 20 DISTRIBUTED**	***Is the difference statistically significant?***
Pre-test	70	72	No ($p > 0.05$)
Post-test	54	68	Yes ($p < 0.01$)
Is the difference significant?	Yes, $p < 0.01$	No, $p > 0.05$	

TABE 5 | Comparing 7th grade biology classes using blocked practice vs distributed retrieval practice.

Topic: Punnett Squares	**Group A (*n* = 27) BLOCKED**	**Group B (*n* = 29) DISTRIBUTED**	***Is the difference statistically significant?***
Review test	55.88	52.45	No ($p = 0.52$)
Post-test	57.67	60.21	No ($p = 0.71$)
Is the difference significant?	No, $p = 0.50$	Yes, $p = 0.01$	

TABLE 6 | Comparing 7th grade earth sciences classes using blocked practice vs distributed retrieval practice.

Topic ↓	**Post-test (Group A)**	**Post-test (Group B)**	***Is the difference significant?***
Erosion	**DISTRIBUTED (*n* = 30)**	**BLOCKED (*n* = 31)**	yes; $p = 0.013$
	81.66	75.16	
Planetary motion	**BLOCKED (*n* = 29)**	**DISTRIBUTED (*n* = 21)**	yes; $p = 0.008$
	51.13	70.5	

TABLE 7 | A template for socratic questions in generative retrieval practices.

Overall Goal	Construct a minimum of 12 questions per simulation (SIM), including illustrated text. NOTE: Levels 4 and up are optional.
Level 1	Explore students' first thoughts and observations; clarify student's thinking
Level 2	Challenge student thinking; have students manipulate the SIM to challenge such thinking
Level 3	Point out the evidence; ask for evidence that backs up student claims
Level 4	Point out counter thinking; ask students for conflicting issues, if any
Level 5	Explore student expertise of the concept/phenomena; ask "if/then what happens" questions
Level 6	Question the intent of questions asked; Explore the main idea of the simulation

were reversed for half of the topics, so each group had access to equal number simulations and illustrative texts. This eliminated the concern that one group was simply academically superior to the other group. T-test statistics was used to compare group scores.

In order to generalize the findings, we combined data across multiple subjects, specific topics, teachers, school districts, etc. to remove confounding variables. Of the ~14,000 questions answered by control and target group students, students in the generative SIMs practice consistently answered more questions correctly. The comparison test was run at all question levels at a significance level of 0.05. As expected, performance decreased in both groups as the level of the question increased, but the drop-off was more pronounced in the illustrative text group (18% worse on level 3 vs. level 1) than in the generative retrieval group with SIMs (7% worse on level 3 vs. level 1). Target group students answered level-1 questions 2% more correctly and the difference rose to 5% more correctly for level-2 and 8% more correctly for level-3 questions. In all levels, the *value of p* was less than 4×10^{-4}. The standard deviation was also smaller for the generative group, implying more consistency in its results. Students in the generative retrieval group not only performed better but also SIMs-based generative retrieval was superior as a delivery method to increase comprehension of STEM and CT concepts as well as critical thinking. A few representative studies are shown below to illustrate more details.

- 10th Grade Special Education/Algebra: (Topics: Graphing Quadratics, Vertex Form of Quadratic Equations, Point Slope, Slope Intercept, Solving Linear Equations with One Variable, Quadratic Solutions, $n = 81$). Research Design: Random division of the class populations from two Intro to Algebra courses and three Algebra/Common Core courses with both General Education and Special Education students. In both control and target groups, there were 81 students; of which 43 students were special education students. As shown in **Table 8**, the groups with SIM generative retrieval practice overwhelmingly outperformed the groups with text illustrations. According to the teacher report, the SIMs allowed students to use visual examples and check scenarios to come up with correct answers more quickly than other students who used text and illustrations. In particular, this kind of visual and interactive aspects of practice helped special education students more significantly because they had reading levels below their grade level.

TABLE 8 | Comparing 10th grade math classes using SIM-based generative practices vs text-based illustration practices.

(a)												
Topics →	**Point slope**						**Slope intercept**					
Students →	**All students**		**General edu.**		**Special edu.**		**All students**		**General edu.**		**Special edu.**	
Practice →	**SIM**	**Text**	**SIM**	**Text**	**SIM**	**Text**	**SIM**	**Text**	**SIM**	**Text**	**SIM**	**Text**
Level1	56	25	80	41	32	10	72	55	83	80	61	31
Level2	30	17	41	27	19	8	34	21	60	31	8	11
Level3	38	8	39	11	38	5	33	20	38	29	28	11
Total	46	20	62	32	30	8	53	37	66	54	40	21
(b)	**Graphing quadratics**						**Vertex form of quadratic functions**					
Level1	64	32	70	48	58	16	58	35	74	54	42	17
Level2	27	22	40	35	15	10	51	32	67	50	36	16
Level3	28	10	41	17	15	3	32	18	49	27	15	10
Total	51	28	60	42	42	14	54	33	70	50	38	16
(c)	**Solving linear equations**						**Quadratic solutions**					
Level1	76	62	88	74	65	49	42	25	47	31	37	18
Level2	31	22	41	30	21	14	28	16	38	21	17	11
Level3	21	10	26	18	16	2	19	5	29	6	9	4
Total	36	25	44	34	28	16	32	17	40	22	25	13

- 9th Grade Biology (Topics: Circulation, Homeostasis, Enzymes, Photosynthesis, Diffusion, Natural Selection, and Gene Expression, $n = 15$). Since students were not allowed to use the internet without supervision and since they have limited access to computers, the teacher decided to display the simulations on the Smart Board and allowed students to do the practice as a group exercise. According to the teacher report, students showed a fear of failure with the first couple of SIMs even when they were told that this would not affect their grade. Compared to the text-based exercise, they answered more questions correctly. While this was not readily apparent for level 1 questions, it became more noticeable for the level 2 and 3 questions. The group aggregate over the use of multiple SIMs is also given in **Table 9**.
- 8th Grade Biology (Topics: Circulation, Homeostasis, Enzymes, Photosynthesis, and Diffusion, $n = 30$). Research Design: Group A had SIMs for topics 1–3 and Text with Illustrations for topics 4–5 while Group B had Text with Illustrations for topics 1–3 and SIMs for topics 4–5. Each group had topic-related level-based questions, which generally included 6 to 8 level-1 questions, 2 to 3 level-2 questions, and 1 to 3 level-3 questions. As shown in **Table 10**, the groups with SIMs scored higher at all levels except for topic 2 (homeostasis). However, as questions got more difficult and complicated, even for homeostasis (#2), the group with SIMs scored higher.
- 8–11th Grades General Science and Physics: (Topics: Forces on a Ramp, States of Matter, Ideal Gas Law, Pendulums, Projectile Motion, and Hooke's Law, $n = 68$). Research Design: A heterogeneous population of 68 students, ranging in age from 12 to 18 in different classes (8th grade General Science and 11th grade Regent Physics). Students were randomly divided in half for each of the 6 modules. While one group practiced *via* phenomena first (SIMs), the other practiced *via* traditional (Text) instruction using text and illustrations. As shown in **Table 11**, groups with SIMs generally outperformed the other groups with text and illustrations. However, the difference is significant for only half of the topics.
- 6th Grade Introductory Computer Programming (Topics: Flashing Heart, Name Tag, Coin Flip, Smiley Face, Random Dice, Rock Paper Scissor, $n = 30$). Research design included having two different classes each doing 3 SIMs and 3 Text with Illustrations. As shown in **Table 12**, groups with SIMs consistently and significantly outscored others. The teacher used Microsoft MakeCode (a free open-source platform) for creating engaging CS learning experiences that support a progression path into real-world programming. He followed Project Lead the Way's Computer Science curriculum and repeated the experiment twice (Fall and Spring) with two different samples. He noted that students appeared to have put more effort into completing the SIMs during Spring 2020 COVID lockdown and focused more on the questions rather than relying on help from the teacher.

DISCUSSION

To support new pedagogical experiences, we offered professional development opportunities on memory retrieval strategies to secondary school teachers from an urban city surrounded by many suburban and rural school districts. Both quantitative and qualitative data from participating teachers pointed to the effectiveness of the spaced-out, interleaved, and generative retrieval strategies in the classroom. The multi-year quantitative and qualitative data consistently suggested that all of the participating teachers found the workshops to be valuable overall, specifically in providing them with the conceptual knowledge and practical skills necessary to effectively engage

TABLE 9 | Comparing 9th grade biology classes using SIM-based generative practices vs text-based illustration practices.

Topics →	Circulation		Homeostasis		Enzymes		Photosynthesis		Diffusion		Total	
Method →	**SIM**	**Text**	**SIM**	**Text**	**SIM**	**Text**	**SIM**	**Text**	**SIM**	**Text**	**SIM**	**Text**
Level1	100	66	100	83	100	0.0	100	100	80	80	96	64
Level2	80	50	50	50	100	0	75	100	66	0	76	58
Level3	100	50	50	0.0	66	100	100	66	75	75	80	60
Total	93	60	80	60	90	20	91	90	73	70	85	61

TABLE 10 | Comparing 8th grade biology classes using SIM-based generative practices vs text-based illustration practices.

Topics →	Circulation		Homeostasis		Enzymes		Photosynthesis		Diffusion		Total	
	SIM	**Text**	**SIM**	**Text**	**SIM**	**Text**	**SIM**	**Text**	**SIM**	**Text**	**SIM**	**Text**
Level1	86	75	72	89	89	89	78	64	87	68	85	79
Level2	45	37	94	95	79	80	83	70	91	71	75	70
Level3	21	11	59	56	51	26	82	61	63	14	54	42
Total	72	62	76	85	84	82	80	65	85	51	78	72

TABLE 11 | Comparing grades 8–11 general science and physics classes using SIM-based generative practices vs text-based illustration practices.

Topics →	Forces on a ramp		States of matter		Ideal gas law		Pendulums		Projectile motion		Hooke's law		Weighted average	
Method	**SIM**	**Text**	**SIM**	**Text**	**SIM**	**Text**	**SIM**	**Text**	**SIM**	**Text**	**SIM**	**Text**	**SIM**	**Text**
Size	**7**	**9**	**16**	**15**	**16**	**18**	**15**	**15**	**18**	**16**	**15**	**17**		
Level1	96	83	74	51	72	73	67	67	94	86	75	61	78.50	69.37
Level2	86	67	62	53	81	71	36	55	64	61	62	51	63.35	59.37
Level3	79	44	75	60	80	79	47	32	56	61	56	24	64.20	50.91
Total	89	69	71	53	77	74	53	56	76	72	65	43	70.44	60.78

TABLE 12 | Comparing 6th grade introductory computer programming classes using SIM-based generative practices vs text-based illustration practices.

Topics →	Flashing heart		Name tag		Coin flop		Smiley face		Random dice		Rock paper scissor		Weighted average	
Fall	**SIM**	**Text**	**SIM**	**Text**	**SIM**	**Text**	**SIM**	**Text**	**SIM**	**Text**	**SIM**	**Text**	**SIM**	**Text**
Level1	100	75	100	100	100	100	100	66	100	100	100	66	100	82.3
Level2	66	33	100	75	50	25	66	33	75	25	60	20	69.5	34.7
Level3	66	66	100	66	75	50	75	25	75	50	50	50	75	50
Total	80	60	100	80	70	50	80	40	80	50	70	40	80	53.3
Spring														
Level 1	100	50	100	66	100	50	100	66	100	50	100	66	100	58
Level 2	66	33	75	25	75	25	100	66	75	50	80	40	78	39
Level 3	100	33	100	0	50	50	75	25	75	75	100	50	80	40
Total	90	40	90	30	70	40	90	50	80	60	90	50	85	45

in implementation efforts during the school year. Similarly, they highly valued the opportunity to learn about generative processes and the research base supporting the effectiveness of these pedagogical approaches in K-12 classrooms.

An overall analysis of student data from classroom action research studies shows that those who learned a topic *via* the distributed retrieval strategy scored considerably better than those who learned in the traditional (blocked) way. In many cases reported, the differences were statistically significant in favor of the distributed retrieval practice. While students in the DR practice group retained knowledge of the topics, students in the BL practice group lost on the average 8% of

it after 2 weeks. For experiments with longer time delays (30 days), the knowledge loss reached up to 60% (for example, see the first bullet listed in the Distributed Retrieval Practice section above). A drop in performance after 30 days is normal for both groups because of natural information loss due to time, but in some cases the drop for the blocked practice group was twice as high as the drop by the treatment group.

A similar trend was seen with students who learned a concept *via* a phenomena-first approach through interactive and generative retrieval using modeling and simulations (M&S) versus others who were first lectured about the topic through reading text and illustrations. This is consistent with the literature on the use of modeling and simulation in science education (Stinken-Rösner, 2020). Researchers have reported that M&S supports both inductive and deductive approaches to learning (Rutten et al., 2012; Smetana and Bell, 2012; Yaşar and Maliekal, 2014). Teachers have historically used deductive pedagogies to instruct students (Bransford et al., 2000). In this approach, information flows from general to details (top-down), or from simple to more complex, as seen in **Figure 1**. The teacher introduces a concept and then shows supporting facts, applications, and details to afford students an opportunity to apply it themselves. In an inductive pedagogical approach, the flow of information is from details to more general (bottom-up) and it is the student, not teacher, at the center of action. Through experimentation and problem solving, students are led to discover their own conclusions by sorting out details and connecting the dots to arrive at more general patterns and principles. Inquiry-based teaching is one such form. Since learning invariably involves movement in both directions (Haight et al., 2007), a teaching that matches bi-directional learning could maximize the benefits (Prince and Felder, 2007).

Limitations of our study include some confounding variables such as suitability of content for distributed and generative practice, and each group's prior experience and background were possibly in play in our research. We suspect that the level of improvement in some cases depended on the nature of topics, grade level, as well as teacher experience and school environment. For example, mathematics and geometry seem to be well suited for retrieval practices. The benefits to students with special needs (as seen in **Table 8**) seemed greater in all content areas tested by one of the best math teachers in the program. However, given the large number of cases, both the distributed and generative retrieval practice outshined the traditional blocked practice.

Our findings complement those in the literature about the impact of retrieval practices on retention (Brown et al., 2014; Agarwal and Bain, 2019) as well as on comprehension of STEM (Rohrer et al., 2014) and CT concepts (Casanova et al., 2020). In almost all cases, students in the treatment groups not only seemed to retain their knowledge of the topics they were tested on, but they also improved their scores, an indication that retrieval helped them comprehend the topics better and make inductive and deductive associations with other topics taught within the same course. This is especially true for the groups with generative retrieval practices because students' ability to comprehend a topic was put to test without their exposure to the material about such topic ahead of time. The effectiveness of SIM-based generative retrieval may be partly driven by inductive and deductive cognitive processes of modeling and simulation, but other researchers have also reported that retrieval practices support inductive and deductive learning even without the use of modeling and simulation tools. While we did not directly measure the impact of SIM-based generative retrieval on basic reasoning skills (i.e., inductive and deductive thinking), recent reviews by Birnbaum et al. (2013); Brunmair and Richter (2019), and Firth et al. (2021) cite that interleaving retrieval practices improve inductive learning—a mental process of acquiring conceptual knowledge from the synthesis of exemplars (Prince and Felder, 2007) that is often known as abstraction in both CT and ST literatures (Wing, 2006; Dunbar and Klahr, 2012; Thagard, 2012; Yaşar et al., 2017; Denning and Tedre, 2019). Wissman et al. (2018) found spaced-out retrieval to improve deductive learning—a mental process of testing factual knowledge, a formula, concept, or theory to various scenarios (Prince and Felder, 2007) as described earlier in terms of distributive processing of information. Eglington and Kang (2018) also reported that retrieval practice improves deductive inference ($p = 0.013$, $d = 0.41$). Deductive thinking skills are often categorized as decomposition skills in the CT literature (Wing, 2006; Barr and Stephenson, 2011; Denning and Tedre, 2019) and as analysis skills (the opposite of synthesis) in the ST literature (Dunbar and Klahr, 2012; Thagard, 2012). According to Eglington and Kang (2018)—Kang was a collaborator of the PI on and NSF project—for the spaced-out retrieval practice to benefit students' deductive thinking, the material to draw appropriate inferences from may need to be presented together.

There is strong evidence that the combination of spaced-out and interleaving retrieval (called distributed retrieval) practices promotes both inductive and deductive thinking skills—which, when used in an iterative and cyclical fashion, constitute a major thrust of ST (Dunbar and Klahr, 2012). Additionally, Samani and Pan (2021) examined their effects on factual knowledge and problem-solving skills. Over 8 weeks, students ($n = 350$) in two lecture sections of an introductory physics course practiced interleaving in thrice-weekly homework assignments. The control group practiced one topic at a time whereas the target group practiced alternating topics. The study consisted of two stages, similar to those we reported, where in stage 1 class A used blocked practice and class B used interleaved practice, and in stage 2 class A used interleaved practice and class B used blocked practice. In each of the two stages, students completed 84 practice problems across 10 homework assignments. On two unannounced tests (one at the end of each stage) containing novel and more challenging problems than those in the homework assignments, the target group recalled more relevant information and more frequently produced correct solutions (with observed median improvements of 50% on test 1 and 125% on test 2). Effect sizes were reported in terms of Cohen's d. Interleaving retrieval yielded higher test scores than blocked practice in Stage 1 ($d = 0.40$, $p = 0.0008$) and Stage 2 ($d = 0.91$, $p < 0.0001$). When spaced-out, even restudy or cramming (blocked practice) has its benefits in terms of helping students to apply (or transfer) the same formula or facts to the solutions of various problems or situations. Karpicke and Blunt (2011) and others (Butler, 2010; Rohrer et al., 2010) showed that repeated retrieval practice of scientific concepts

could promote transfer to test questions which are related to, but different from, originally practiced materials.

IMPLICATIONS, CONCLUSION, AND RECOMMENDATIONS

The findings of classroom action research by a large number of teachers in our study support the use of retrieval practices for retention and comprehension of secondary school STEM and CT concepts. This is consistent with the literature on retrieval practices (Agarwal et al., 2020). More importantly, our study complements findings of other recent studies such as Prince and Felder (2007), Birnbaum et al. (2013), Eglington and Kang (2018); Brunmair and Richter (2019); Firth et al. (2021), and Samani and Pan (2021) to point out that retrieval practices enhance core elements of computational and scientific thinking. While other researchers had reported a favorable impact of retrieval practices on inductive and deductive thinking, a more general realization that—taken together with our findings—retrieval practices enhance core CT and ST skills is the most important conclusion of our study. This is a timely discovery because of the ongoing search in the STEM community for innovative and fundamental practices and tools to support CT and ST education across different levels of education (Denning and Tedre, 2019; Papadakis, 2021). The authors posit that use of retrieval practices is perhaps one of the most direct ways of improving core CT skills because memory retrieval is nothing but the thinking itself by a computational brain. For those who are looking for best practices to improve CT and ST skills of young children, memory retrieval practices set a natural example.

There were several limitations of our study, which can be improved in future studies. Even though a large number of action research studies ($n = 122$) by teachers consistently point to the benefit of retrieval practices, some confounding variables such as suitability of content for retrieval practices as well as prior experience and background by control and target groups were possibly in play in our research. Circumstances surrounding each experiment were obviously different because they were each run by different researchers. This may have also been good to eliminate bias. At the same time, we suspect that the level of improvement in some cases depended on the nature of topics, grade level, as well as teacher experience and school environment. The control variables can be isolated better in future studies. Also, the time delay between the unannounced test and the review test (the last time students are exposed to the topic) should be long enough (at least 30 days) to allow more contrast, if any, to surface out between control and target groups. In some cases, it appears that performance of both groups was higher at the final test than the review test, which means the examined topics were continued to be discussed either in similar classes that students attended or in the same class within the context of other topics. In studies where group exposure to the examined topic were better isolated, student retention appears to go down for both groups but more so for the group with blocked practice than the group with retrieval practices. Finally, since the current study was limited to secondary schools, the authors recommend expanding the study to lower grades. Future research should include use of retrieval practices in a wider set of CT and ST concepts. A recommendation for the computer science education community is that because of the importance of programming in CT education, there is a need for more studies such as Casanova et al. (2020) to examine effect of retrieval practices on retention and learning of programming concepts.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

OY wrote the manuscript and led the research. All authors helped to conduct the research and implement that program. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The authors acknowledge support by the US National Science Foundation through a grant (DRL # 1614847) from the Innovative Technology Experiences for Students and Teachers (ITEST) program.

REFERENCES

Agarwal, P. K., and Bain, P. M. (2019). *Powerful Teaching: Unleash the Science of Learning*. San Francisco, CA: Jossey-Bass (Wiley).

Agarwal, P. K., Roediger, H. L., McDaniel, M. A., and McDormett, K. B. (2020). *How to Use Retrieval Practice to Improve Learning*. Available at: http://retrievalpractice.org (Accessed May 7, 2022).

Aho, A. (2012). Computation and computational thinking. *Comput. J.* 55, 832–835. doi: 10.1093/comjnl/bxs074

Angeli, C., and Giannakos, M. (2019). Computational thinking: issues and challenges. *Comput. Hum. Behav.* 105:106185. doi: 10.1016/j.chb.2019.106185

Barr, V., and Stephenson, C. (2011). Bringing computational thinking to K-12: what is involved and what is the role of the computer science education community? *ACM Inroads* 2, 48–54. doi: 10.1145/1929887.1929905

Beatriz, C. (2005). *Review of Evaluation Studies of Mathematics and Science Curricula and Professional Development Models*. Washington, DC: Urban Institute.

Birnbaum, M. S., Kornell, N., Bjork, E. L., and Bjork, R. A. (2013). Why interleaving enhances inductive learning: the roles of discrimination and retrieval. *Mem. Cogn.* 41, 392–402. doi: 10.3758/s13421-012-0272-7

Blunck, S. M., and Yager, R. E. (1996). "The Iowa Chautauqua program: a proven in-service model for introducing STS in K-12 classrooms," in *Science/Technology/Society as Reform in Science Education*. ed. R. E. Yager (Albany, NY: NY State Univ. of New York Press), 298–305.

Bransford, J., Brown, A., and Cocking, R. (2000). *How People Learn: Brain, Mind, Experience, and School*. National Academy Press, Washington, D.C.

Brown, P. C., Roediger, H. L., and McDaniel, M. A. (2014). *Make, it Stick. The Science of Successful Learning*. London: The Belknap Press of Harvard University.

Brunmair, M., and Richter, T. (2019). Similarity matters: a meta-analysis of interleaved learning and its moderators. *Psychol. Bull.* 145, 1029–1052. doi: 10.1037/bul0000209

Butler, A. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *J. Exp. Psychol. Learn. Memory Cognit.* 36, 1118–1133. doi: 10.1037/a0019902

Carey, S. (1985). *Conceptual Change in Childhood. 1st Edn*. Cambridge, MA: MIT Press.

Casanova, L., Theophilou, E., Beardsley, M., Santos, P., and Hernandez-Leo, D. (2020). "Distributed retrieval practice enhances primary school students' retention of computational thinking concepts." *17th International Conference on Cognition and Exploratory Learning in Digital Age*. Lisbon, Portugal.

Creswell, J. W. (2012). *Educational Research: Planning, Conducting and Evaluating Quantitative and Qualitative Research. 4th Edn*. Boston, MA: Pearson Education, Inc.

Denning, P. (2009). Beyond computational thinking. *Commun. ACM* 52, 28–30. doi: 10.1145/1516046.1516054

Denning, P. (2017a). Computational thinking in science. *Am. Sci.* 105, 13–17. doi: 10.1511/2017.124.13

Denning, P. (2017b). Remaining trouble spots with computational thinking. *Commun. ACM* 60, 33–39. doi: 10.1145/2998438

Denning, P., and Tedre, M. (2019). *Computational Thinking*. The MIT Press, Cambridge, MA, USA.

D-Hammond, L., and Bransford, J. (2005). *Preparing Teachers for a Changing World: What Teachers Should Learn and Be Able to Do*. Jossey-Bass: John Wiley & Sons. San Francisco, CA.

Dunbar, K., and Klahr, D. (2012). "Scientific thinking and reasoning," in *The Oxford Handbook of Thinking and Reasoning*. eds. K. Holyoak and R. Morrison (London: Oxford University Press), 701–718.

Eglington, L. G., and Kang, S. H. K. (2018). Retrieval practice benefits deductive inference. *Educ. Psychol. Rev.* 30, 215–228. doi: 10.1007/s10648-016-9386-y

Ferrance, E. (2000). *Themes in Education: Action Research*. Providence, RI: Northeast and Islands Regional Educational Laboratory.

Fincher, S., and Petre, M. (2005). *Computer Science Education Research*. Taylor and Francis e-Library: London and New York.

Firth, J., Rivers, I., and Boyle, J. (2021). A systematic review of interleaving as a concept learning strategy. *Rev. Educ* 9, 642–684. doi: 10.1002/rev3.3266

Funke, A., Geldreich, K., and Hubwieser, P. (2017). "Analysis of scratch projects of an introductory programming course for primary school students." In *2017 IEEE Global Engineering Education Conference (EDUCON)*, pp. 1229–1236.

Goode, J., and Margolis, J. (2011). Exploring computer science: a case study of school reform. *Trans. Comput. Educ.* 11, 1–16. doi: 10.1145/1993069.1993076

Grabowski, B. L. (2004). "Generative learning contributions to the design of instruction and learning," in *Handbook of Research on Educational Communications and Technology*. ed. D. H. Jonassen (Mahwah, NJ: Lawrence Erlbaum Associates Publishers), 719–743.

Grover, S., and Pea, R. (2013). Computational thinking: a review of the state of the field. *Educ. Res.* 42, 38–43. doi: 10.3102/0013189X12463051

Guskey, T. R. (2000). *Evaluating Professional Development.* Thousand Oaks, CA: Corwin.

Guzdial, M. (2008). Paving the way for computational thinking. *Commun. ACM* 51, 25–27. doi: 10.1145/1378704.1378713

Haight, C. E., Herron, C., and Cole, S. P. (2007). The effects of deductive and guided inductive instructional approaches on the learning of grammar in the elementary foreign language college classroom. *Foreign Lang. Ann.* 40, 288–310. doi: 10.1111/j.1944-9720.2007.tb03202.x

Hebb, D. (1949). *The Organization of Behavior*. New York: Wiley.

Kakavas, P., and Ugolini, F. C. (2019). Computational thinking in primary education: a systematics literature review. *Res. Educ. Media*. 11, 64–94. doi: 10.2478/rem-2019-0023

Kant, I. (1787). *The Critique of Pure Reason*. (J. M. D. Meiklejohn, Trans.). eBook@Adelaide, The University of Adelaide Library, Australia.

Karpicke, J. D., and Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science* 331, 772–775. doi: 10.1126/science.1199327

Karpicke, J. D., Blunt, J. R., Smith, M. A., and Karpicke, S. S. (2014). Retrieval-based learning: the need for guided retrieval in elementary school children. *J. Appl. Res. Mem. Cogn.* 3, 198–206. doi: 10.1037/h0101802

Kirschner, P. A., Sweller, J., and Clark, R. E. (2006). Why minimal guidance during instruction does not work. *Educ. Psychol.* 41, 75–86. doi: 10.1207/s15326985ep4102_1

Kuhn, T. (1962). *The Structure of Scientific Revolutions*. University of Chicago Press, Chicago.

L-Horsley, S., Stiles, K. E., Mundry, S., Love, N., and Hewson, . (2010). *Designing Professional Development for Teachers of Science and Mathematics. Third Edition*, Thousand Oaks, CA: Corwin Press.

Margolis, J., Estrella, R., Goode, J., Holme, J., and Nao, K. (2008). *Stuck in the Shallow End: Education, Race, and Computing*. Cambridge, MA: MIT Press.

Mills, K., Coenraad, M., Ruiz, P., Burke, Q., and Weisgrau, J. (2021). Computational thinking for an inclusive world: a resource for educators to learn and lead. *Digital Promise.* doi: 10.51388/20.500.12265/138

Montague, R. (2006). *How We Make Decisions*. New York, NY: Plume Books.

Mooney, C. G. (2013). *An Introduction to Dewey, Montessori, Erikson, Piaget, and Vygotsky*. Redleaf Press: St. Paul, MN.

National Center for Education Statistics. (1996). Urban schools: the challenge of location and poverty. Report NCES 96–184. Available at: https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=96184. (Accessed February 4, 2021).

Papadakis, S. (2021). The impact of coding apps to support young children in computational thinking and computational fluency: a literature review. *Front. Educ.* 6:657895. doi: 10.3389/feduc.2021.657895

Papert, S. (1980). *Mindstorms: Children, Computers, and Powerful Ideas*. New York: Basic Books, A Division of Harper Collins Publishers, Inc.

Prince, M. J., and Felder, R. M. (2007). The many faces of inductive teaching and learning. *J. Coll. Sci. Teach* 36, 14–20.

Qazi, M. A., Gray, J., Shannon, D. M., Russell, M., and Thomas, M. (2020). "A state-wide effort to provide access to authentic computer science education to underrepresented populations." In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*. (Portland, OR), 241–246.

Resnick, M., Maloney, J., Monroy-Hernández, A., Rusk, N., Eastmond, E., Brennan, K., et al. (2009). Scratch: programming for all. *Commun. ACM* 52, 60–67. doi: 10.1145/1592761.1592779

Rohrer, D., Dedrick, R. F., and Stershic, S. (2014). Interleaved practice improves mathematics learning. *J. Educ. Psychol.* 107, 900–908.

Rohrer, D., Taylor, K., and Sholar, B. (2010). Tests enhance the transfer of learning. *J. Exp. Psychol. Learn. Mem. Cogn.* 36, 233–239. doi: 10.1037/a0017678

Rutten, N., van Joolingen, R., and van der Veen., (2012). The learning effects of computer simulations in science education. *Comput. Educ.* 58, 136–153. doi: 10.1016/j.compedu.2011.07.017

Samani, J., and Pan, S. C. (2021). Interleaved practice enhances memory and problem-solving ability in undergraduate physics. *Sci. Learn.* 6:32. doi: 10.1038/s41539-021-00110-x

Saqr, M., Ng, K., Oyelere, S. S., and Tedre, M. (2021). People, ideas, milestones: a scientometric study of computational thinking. *ACM Trans. Comput. Educ.* 21, 1–17. doi: 10.1145/3445984

Silver, J., et al. (2012). "Makey makey improvising tangible and nature-based user interfaces." in *Proceedings of the ACM tangible embedded and embodied interaction*, Kingston, Ontario, Canada, pp. 367–370.

Smetana, L. K., and Bell, R. L. (2012). Computer simulations to support science instruction and learning: a critical review of the literature. *Int. J. Sci. Educ.* 34, 1337–1370. doi: 10.1080/09500693.2011.605182

Stinken-Rösner, L. (2020). Simulations in science education – status quo. *Prog. Sci. Educ.* 3, 26–34. doi: 10.25321/prise.2020.996

Tedre, M., and Denning, P. (2021). "Computational thinking: a professional and historical perspective," in *Computational Thinking in Education: A Pedagogical Perspective*. eds. A. Yadav and U. D. Berthelsen (London: Routledge).

Thagard, P. (1999). *How Scientists Explain Disease*. Princeton, NJ: Princeton Univ. Press.

Thagard, P. (2012). *The Cognitive Science of Science*. The MIT Press, Cambridge, MA.

Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungsproblem. *Proc. Lond. Math. Soc. Ser.* 2, 230–265.

Turner, P., Petzold, L., Shiflet, A., Vakalis, I., Jordan, K., and John, S. S. (2011). Undergraduate computational science and engineering education. *SIAM Rev.* 53, 561–574.

Vosniadou, S. (2013). *International Handbook of Research on Conceptual Change. 2nd Edn.* New York and London: Routledge.

Vygotsky, L. S. (1978). *Mind in Society*. Cambridge, MA: Harvard University Press.
Wing, J. M. (2006). Computational thinking. *Comm. ACM* 49, 33–35. doi: 10.1145/1118178.1118215
Wissman, K. T., Zamary, A., and Rawson, K. A. (2018). When does practice testing promote transfer on deductive reasoning tasks? *J. Appl. Res. Mem. Cogn.* 7, 398–411. doi: 10.1016/j.jarmac.2018.03.002
Yadav, A., Stephenson, C., and Hong, H. (2017). Computational thinking for teacher education. *Commun. ACM* 60, 55–62. doi: 10.1145/2994591
Yaşar, O. (2017). The essence of computational thinking. *Comput. Sci. Eng.* 19, 74–82. doi: 10.1109/MCSE.2017.3151241
Yaşar, O. (2018). A new perspective on computational thinking. *Commun. ACM* 25, 33–39. doi: 10.1145/3214354
Yaşar, O. (2021). A computational view of scientific thinking. *Acad. Lett.* 2021:1416. doi: 10.20935/AL1416
Yaşar, O., and Landau, R. (2003). Elements of computational science and engineering education. *SIAM Rev.* 45, 787–805. doi: 10.1137/S0036144502408075
Yaşar, O., and Maliekal, J. (2014). Computational Pedagogy. *Comp. Sci. Eng.* 16, 78–88.
Yaşar, O., Maliekal, J., Little, L., and Veronesi, P. (2014). An interdisciplinary approach to professional development for math, science, and technology teachers. *J. Comput. Math. Sci. Teach.* 33, 349–374.
Yaşar, O., Maliekal, J., Veronesi, P., and Little, L. (2017). "The essence of scientific and engineering thinking and tools to promote it." Proceedings *of 2017 ASEE Annual Conference & Exposition*; Columbus, Ohio.
Yaşar, O., Rajasethupathy, K., Tuzun, R., McCoy, A., and Harkin, J. (2000). A new perspective on computational science education. *Comput. Sci. Eng.* 5, 74–79.
Yaşar, O., Veronesi, P., Maliekal, J., and Little, L. (2021). A professional development program on memory retrieval strategies for STEM teachers. *J. Comput. Math. Sci. Teaching* 40, 265–298.
Yaşar, O., Veronesi, P., Maliekal, J., Little, L., and Tillotson, J. (2019a). "Memory Retrieval Strategies to Help Retain STEM Content Knowledge." *Proceedings of ASEE Annual Conference & Exposition*. Tampa, Florida.
Yaşar, O., Veronesi, P., Maliekal, J., Little, L., and Tillotson, J. (2019b). "Computational pedagogical strategies to help learners retain STEM content knowledge," in *Proceedings of Society for Information Technology & Teacher Education International Conference*. ed. K. Graziano (Las Vegas, NV, USA, Association for the Advancement of Computing in Education (AACE)), 351–360.
Yaşar, O., Veronesi, P., Maliekal, J., Little, L., Vattana, S. E., and Yeter, I. H. (2016). Computational pedagogy: fostering a new method of teaching. *Comput. Educ. J.* 7, 51–72.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Yaşar, Maliekal, Veronesi, Little, Meise and Yeter. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Psychology

ORIGINAL RESEARCH
published: 06 July 2022
doi: 10.3389/fpsyg.2022.903252

Promoting Secondary Students' Twenty-First Century Skills and STEM Career Interests Through a Crossover Program of STEM and Community Service Education

Biyun Huang[1], *Morris Siu-Yung Jong*[1,2]*, *Ronnel B. King*[3], *Ching-Sing Chai*[2] and *Michael Yi-Chao Jiang*[1,2]

[1] Centre for Learning Sciences and Technologies, The Chinese University of Hong Kong, Hong Kong, Hong Kong SAR, China, [2] Department of Curriculum and Instruction, Faculty of Education, The Chinese University of Hong Kong, Hong Kong, Hong Kong SAR, China, [3] Centre for the Enhancement of Teaching and Learning, Academic Unit of Human Communication, Development, and Information Sciences, The University of Hong Kong, Hong Kong, Hong Kong SAR, China

OPEN ACCESS

Edited by:
Stamatios Papadakis,
University of Crete, Greece

Reviewed by:
Zhi Hong Wan,
The Education University of Hong Kong, Hong Kong SAR, China
Kani Ulger,
Cumhuriyet University, Turkey
Laurinda Leite,
University of Minho, Portugal
Irma Rahma Suwarma,
Universitas Pendidikan Indonesia, Indonesia

***Correspondence:**
Morris Siu-Yung Jong
mjong@cuhk.edu.hk

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

Received: 24 March 2022
Accepted: 03 June 2022
Published: 06 July 2022

Citation:
Huang B, Jong MS-Y, King RB, Chai C-S and Jiang MY-C (2022) Promoting Secondary Students' Twenty-First Century Skills and STEM Career Interests Through a Crossover Program of STEM and Community Service Education. Front. Psychol. 13:903252. doi: 10.3389/fpsyg.2022.903252

STEM education has been regarded as an important educational initiative for cultivating students' twenty-first century skills. The present work aimed to explore ways to promote students' twenty-first century skills through an integrated STEM-based curriculum. Specifically, we designed and implemented an 8-week crossover program of STEM and community service education. In this program, students learned about STEM domain knowledge and community service issues. They then applied the knowledge to solve authentic problems faced by Hong Kong community-housing residents from disadvantaged groups. A mixed-method approach was employed to evaluate the effectiveness of the program in enhancing students' twenty-first century skills and attitudes, including (i) creative thinking, (ii) collaboration, (iii) perseverance, as well as their (iv) STEM career interests. The research participants were 121 secondary students from a government-subsidized school. The quantitative results showed that the participants' creative thinking, collaboration, and perseverance improved alongside their STEM career interests. These findings were further supported by the data gathered through focus-group interviews. This study provides theoretical and practical insights into the integration of STEM education with community service learning.

Keywords: STEM education, community service learning, secondary students, twenty-first century skills, creativity, collaboration, perseverance, career interests

INTRODUCTION

STEM education is one of the most prominent educational initiatives in the twenty-first century (Geng et al., 2019; Jong et al., 2021; Lau and Jong, 2022). This initiative has been viewed as a promising pathway for nurturing a nation's future workforce and enhancing economic competitiveness (Honey et al., 2014; Chai et al., 2020b; Reynante et al., 2020). Meanwhile, constructivist learning paradigms have been largely promoted in school education since the past two decades (Luk et al., 2006; Jong et al., 2010; Song et al., 2017; Bower and Jong, 2020).

Instead of using conventional didactic teaching approaches, successful STEM education programs usually adopt constructivist student-centered approaches with a strong emphasis on learners' active participation (Ulger, 2018; Wan et al., 2021; Weng et al., 2022). Integrating community service learning (Yorio and Ye, 2012; Chiva-Bartoll et al., 2020) with STEM education, where learners apply what they learn to provide solutions to the problems in their communities, is considered an innovative and constructivist approach (Collins et al., 2020; Marcus et al., 2021).

Despite the promise of synergizing community service learning with STEM education, empirical research on the effectiveness of this integration is scarce (Botelho et al., 2020). To address this shortcoming, we explored the integration of community service learning with STEM education. We evaluated its influence on students' twenty-first century skills and their attitudes, including (i) creative thinking, (ii) collaboration, (iii) perseverance, and their interest in STEM careers (Fullan and Langworthy, 2014; Scherer and Gustafsson, 2015; Maiorca et al., 2021).

Twenty-first-century skills pertain to the knowledge, skills, and attitudes needed by children and the youth to fully engage in and contribute to the development of society (Lin et al., 2021). Binkley et al. (2012) summarized that twenty-first century skills could be classified into four groups, including ways of thinking (e.g., creativity and innovation, critical thinking, problem solving, and learning to learn), ways of working (e.g., communication and collaboration), tools for working (e.g., information and ICT literacy), and living in the world (e.g., citizenship, life and career, and sense of responsibility). Chai et al. (2020a) reviewed frameworks for twenty-first century skills and noted that most of the frameworks emphasized the importance of engaging students in creative thinking, critical thinking, and authentic problem solving. In addition, classroom-based and out-of-classroom learning communities should support students in building communication and collaboration skills in the problem-solving process, which contributes to growth in learning and innovation (Chai et al., 2020b). Tsai et al. (2013) expressed that the essence of education in the twenty-first century has to be epistemologically shifted from reproduction of knowledge to production of new knowledge through creativity and innovation.

In response to the prominent call for the educational shift to cultivating creative, competent, and responsible future citizens (Papadakis, 2016), we launched the integrated STEM-based community service program, and examined its impact on students' twenty-first century skills, such as creative thinking and collaboration. Specifically, we situated our study in the context of Hong Kong junior secondary setting and developed a program wherein students learned about the common living problems encountered by community-housing residents (e.g., the problem of having limited housing spaces). The students collaborated to propose solutions in teams.

Aside from these two critical twenty-first century skills, we also examined the impact of the program on students' perseverance and STEM interest. Perseverance pertains to a proactive attitude toward difficulties and the willingness to make continuous efforts to solve problems (Foster and Schleicher, 2022). Perseverance is one of the main characteristics of creative individuals and an important component of creative capacity (Amabile, 1983; Scherer and Gustafsson, 2015). It is an important variable that links to the quality of students' creative problem-solving (Foster and Schleicher, 2022). Moreover, STEM career interest is another important variable because many studies have shown that there are insufficient numbers of students who want to pursue a STEM career (Karahan et al., 2021; Maiorca et al., 2021). Hence, we also examined whether the crossover program could potentially enhance students' interests in STEM careers.

LITERATURE REVIEW

Rationale Behind Integrating STEM Education and Community Service Learning

Integrated STEM is an approach that explores teaching and learning between two or more STEM subjects within an authentic context to build connections between disciplinary knowledge and their applications (Kelley and Knowles, 2016; Nadelson and Seifert, 2017). Fostering connections across the disciplines is more likely to promote students' twenty-first century skills, develop a STEM-capable workforce, and boost interest in STEM (Honey et al., 2014; Lee et al., 2019). English (2019) integrated designing ones' own pair of shoes activity with engineering and science disciplines in a fourth-grade STEM course and found that students' design skills and trouble-shooting strategies improved. Morrin and Liston (2020) integrated visual arts with school space design activities and reported that students' attitudes and creativity skills were positively influenced. When accompanied by teamwork, the integrated approach unlocks untapped potential and could benefit learners' knowledge acquisition, higher order thinking, as well as collaboration skills (Moirano et al., 2020).

Although the integration of STEM with different subjects is promising, there is a need for more empirical endeavors to explore how and what could be done to achieve desired learning outcomes. In recent years, scholars have emphasized that STEM education is not just confined to the cognitive and technical realm (i.e., the systematic application of mathematical and scientific knowledge to develop novel solutions to complex problems) (Gunckel and Tolbert, 2018). Rather, it should also involve the cultivation of values, civic attitudes, empathy, and encourage socially responsible and human-centered solutions (Bielefeldt, 2017; Huang et al., 2022). Therefore, advocates have recently suggested that "arts" should be incorporated into STEM education, which refers to a single arts discipline or an expanded area of the liberal arts and humanities disciplines (Perignat and Katz-Buonincontro, 2019; Quigley et al., 2019). The integration of "arts" offers a number of benefits to students (Bush et al., 2020, p. 693), such as engaging more students who do not identify themselves with STEM and providing meaning for learning about STEM knowledge, through the process of "solving authentic problems to make the world a better place."

Community service learning can provide an authentic sociocultural context for STEM education. It is a form

of experiential education that engages students in human-centered service-learning activities to address community needs and develop students' value and knowledge (Jacoby, 1996; Tijsma et al., 2020). The explicit connections between learning objectives and structured community interactions promote a broad appreciation of the discipline and enhance personal growth, values, and civic attitudes (Bringle and Hatcher, 1996; Salam et al., 2019). Evidence of the positive effects of this learning approach has been reported in several studies. For example, Reed et al. (2015) reported that university students who took community service-learning courses demonstrated a higher likelihood of persistence. Gerholz et al. (2018) found that university students who collaborated with charitable organizations in problem-solving projects improved their self-efficacy and self-concept. Burton and Winter (2021) examined two service-learning courses where university students applied history or psychology knowledge to help community partners, and reported that students strengthened communication and problem-solving skills and that service-learning courses raised students' intentions to pursue a service-focused career. These studies exhibited the potent effects of community service learning (Burton and Winter, 2021). Nevertheless, most of the explorations were in higher education sectors. Though Collins et al. (2020) provided an initial exploration on integrating community service learning with STEM education through summer workshops in high school, our knowledge about how to implement crossover programs in formal K-12 classroom settings is limited.

To extend the scholarship and innovative practices in the field, there is a need to further explore other curricular and pedagogical possibilities. The present work aims to provide secondary school students with authentic learning experiences based on real-life problems faced by the Hong Kong community. In particular, the problem of limited living space was chosen to engage students in human-centered design.

Toward a Design Model of Integrated STEM-Based Community Service Learning

Model of Human-Centered Design Thinking

To facilitate human-centered product or solution design, the Hasso Plattner Institute of Design (2010) proposed a five-stage model, namely empathize, define, ideate, prototype, and test (EDIPT). The model has been widely adopted in STEM learning activities for scaffolding innovative design processes. In the empathize stage, opportunities need to be created for designer(s) to observe and interact with potential users in their lived context to understand users' needs. In the define stage, designer(s) decide the challenge to take on based on what they learned about the users in the previous phase. In the ideation stage, designer(s) brainstorm a range of potential solutions to solve the challenge. In the prototype stage, designer(s) build prototypes to get close to the final solution. In the test stage, designer(s) interact with the potential users after they test the prototype and collect feedback to refine the solution.

The EDIPT model emphasizes the procedure of empathizing with users and iterative design based on user feedback. It does not stipulate other elements necessary for scaffolding innovation, such as the environment and prior skills. It has been used either together with other models to guide the design process (e.g., Da Silva et al., 2020) or independently (e.g., Yalçin and Erden, 2021). Yalçin and Erden (2021) implemented the EDPIT model in a pre-school STEM program and engaged the children in a series of design activities, for example, designing a prototype that would help them climb high. Yalçin and Erden (2021) found that the experimental group gained higher scores in creativity, problem-solving, and persistence after completing the program.

Componential Model of Creativity

In the componential model of creativity, Amabile and Pratt (2016) summarized the social and psychological components necessary for an individual or team to produce creative work or solutions. Amabile and Pratt (2016) emphasized that there are three primary elements needed for innovation, including resources in the task domain, skills and processes for combing the recourses in new ways, and the motivation to innovate. They further illustrated the five stages of innovation as: (1) identifying the problems to be solved or goals to be achieved; (2) preparing for a successful process, such as gathering resources, collecting information, and assigning tasks; (3) generating possibilities; (4) evaluating possibilities; and (5) assessing outcomes and making decisions based on results achieved.

Amabile and Pratt's (2016) componential model of creativity has been successfully applied in K-12 contexts, particularly in secondary education. For example, Hong and Song (2020) observed the physics class of a secondary school, analyzed students' critical incidents worksheets, and interviewed student participants to identify the components influencing creativity. They found that teacher support, such as guidance in the inquiry process and positive feedback, and classroom environment are closely associated with creativity in science classes which aligns with Amabile's (1983) proposition and other recent arguments underlining the importance of teachers in STEM education (e.g., Geng et al., 2019; Jong et al., 2021; Lau and Jong, 2022). Ginns et al. (2021) surveyed students across 13 secondary schools in Australia and confirmed the predictions based on the componential model, such as a supportive environment for creativity is related to intrinsic motivation and creative self-efficacy. They further identified that intrinsic motivation is conducive to creative self-efficacy. Sun et al. (2020) investigated the relationship between students' skills in the task domain (e.g., domain knowledge) and scientific creativity, and found that students' domain knowledge positively influenced their creativity. The above studies confirmed that it is essential for schools to provide students with resources and raw materials for solving problems, equip students with domain-specific skills and creative process skills, and establish a supportive classroom environment through positive feedback.

Proposing an Integrated STEM-Based Community Service-Learning Model

Building upon the EDIPT design thinking model and the componential model of creativity, we proposed an integrated STEM-based community service-learning model for designing

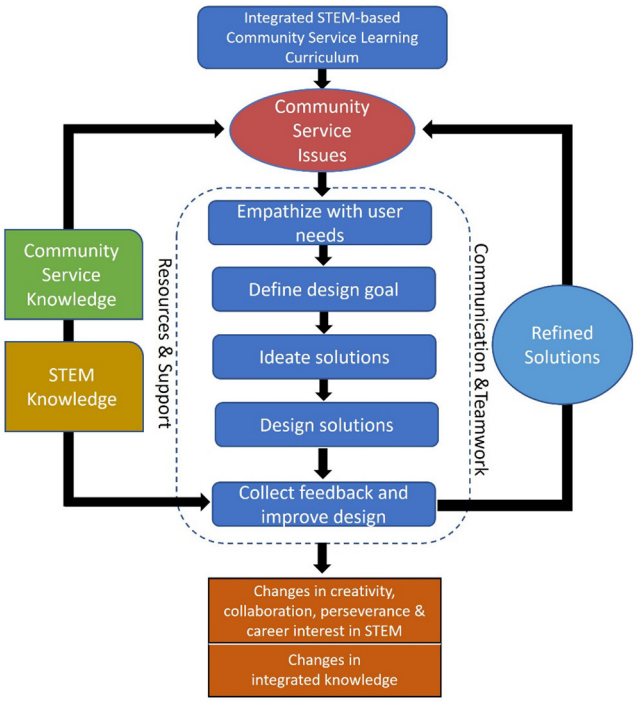


FIGURE 1 | The integrated STEM-based community service-learning model.

integrated STEM courses (see **Figure 1**). The model depicted the main components for supporting individual or team creativity in K-12 setting, such as resources and support from school level, and opportunity to learn relevant domain knowledge. In order to build students' capacity in designing solutions for community service issues, students need to learn foundational knowledge in community service and STEM (Amabile, 1983; Sun et al., 2020). Hence, community service knowledge and STEM knowledge should be included as indispensable parts of the learning model. In the meantime, students should follow a creative process to explore solutions addressing real-world challenges in the community. The exploration can be conducted by individual students or teams, but preferably in team formats, to stimulate better solutions and higher synergy (Mavri et al., 2020). Furthermore, user and expert feedback can be provided to students on improving the designs and creating user-oriented meaningful solutions (Vo and Asojo, 2021). It is assumed that after completing cross-disciplinary curriculums based on the integrated model, students' twenty-first century skills, such as creative thinking, collaboration, perseverance and STEM career interests, would be enhanced.

RESEARCH DESIGN

Design of the Integrated STEM-Based Community Service Course

In designing the interdisciplinary course, we followed the integrated STEM-based community service-learning model. In the STEM knowledge aspect, students learned concepts and skills on coding, the Internet of things (IoT), smart home devices, and the development of mini devices. In the community service knowledge aspect, students learned foundation knowledge of community services (e.g., missions and main responsibilities), the problems faced by residents in the community (e.g., inconveniences of living in community-housing), and the process and techniques for designing human-centered home devices and furniture for community residents. The learning materials of STEM knowledge and community service knowledge were developed in collaboration with industry experts, social service staff, and participating community residents. The key contents covered in STEM and community service lessons are illustrated in **Table 1**. Among them, the section covering techniques for designing home devices and furniture was delivered by an experienced interior designer. Videos on in-depth interviews with residents and social service staff were also presented to enable a comprehensive understanding of users' needs. A mixed-methods approach was employed to evaluate the effectiveness of the program for enhancing students' twenty-first century skills and attitudes, including (i) creative thinking, (ii) collaboration, (iii) perseverance, as well as their (iv) STEM career interests.

TABLE 1 | Contents in the integrated STEM-based community service-learning course.

Disciplines	Key contents
STEM knowledge	• Coding skills and computational thinking • IoT concepts and applications • Smart home devices and examples • Maker product and its development (1 & 2)
Community service knowledge	• Community services and foundational knowledge • Community housing residents and their needs • Human-centered home devices, furniture, and the design techniques • In-depth interviews with residents and social service staff • Presentation of design solutions and collection of feedback

Students completed their initial design solutions in teams in the community service lessons. Afterwards, they presented the design solutions to experts and peers in class and solicited feedback. Next, students spent 2 weeks refining and finalizing their design solutions. **Figure 2** presents some of the teaching and learning activities of the program.

Participants and Experimental Procedures

The duration of the integrated course lasted 8 weeks, including 6 weeks for learning STEM and community service knowledge and 2 weeks for finalizing the design solution. The participants were from a junior secondary school in Hong Kong ($n = 121$). Prior to the participation in the course, students completed a pre-survey adapted from the instruments on twenty-first century skills, covering the dimensions of creative thinking, collaboration, perseverance, as well as STEM career interest (e.g., Vennix et al., 2018; Chai et al., 2020a). The same survey questionnaire was



FIGURE 2 | Teaching and learning activities in the program.

completed by the end of the program to track the changes of students in the four dimensions. Correlations between students' gender and twenty-first century skills were also explored. To triangulate with the survey results and gain more understanding of students' perceptions of the integrated course, 12 students from three groups were invited to participate in post-course interviews. The survey and interview results are reported in Section Results. Consents from students were obtained prior to participating in the research.

Data Collection

Questionnaire-Based Survey

The survey questionnaire was comprised of 4 subscales and 14 items in total. The subscales were adapted from validated instruments on creative thinking (e.g., Chai et al., 2020a), collaboration (e.g., Lai and Hwang, 2014), perseverance (e.g., Toering et al., 2012), and STEM career interest (e.g., Vennix et al., 2018). The items were contextualized and translated by the authors for being adopted in the research context. To ensure the validity of the instrument, two experts in the field were invited to review and comment on the adapted items. The creative thinking subscale consisted of three items. A sample question of the creative thinking dimension is, "In STEM lessons, I can always come up with many new ideas." The collaboration subscale consisted of three items, and a sample item is "In STEM lessons, my teammates and I usually help each other." The perseverance subscale consisted of four items, and a sample item is "In STEM lessons, I always try my best to complete all the tasks assigned by the teachers." The STEM career interest subscale consisted of four items, and a sample item is "I think STEM-related jobs are interesting." Students were briefed that "in STEM lessons" refers to the lessons carried out during the integrated STEM-based community service-learning program. A six-point scale ranging from "strongly disagree" to "strongly agree" was adopted in the study. The Cronbach's alpha values for the subscales were 0.86 for creative thinking, 0.94 for perseverance, 0.94 for collaboration, and 0.94 for STEM career interest.

Focus-Group Interviews

Three focus-group interviews were conducted to understand students' perceptions of the integrated course. Each interview consisted of three to five students. The groups were selected from the eight project teams recommended by teachers, covering students of different engagement levels observed in the course (i.e., highly engaged and less engaged) to maximize the variety of sampling (Patton, 2001). Among the participants, there

TABLE 2 | Results of paired sample *t*-test for pre- and post-surveys.

Dimension	Pre-survey		Post-survey		*t*-value	*p*-value
	M	SD	M	SD		
Creative thinking	4.07	1.02	4.58	0.99	−5.00	<0.001
Collaboration	4.54	1.00	4.75	1.02	−2.16	0.033
Perseverance	4.41	1.07	4.77	1.00	−3.26	0.002
Career interest	3.52	1.12	4.19	1.19	−5.17	<0.001

n = 97. Among the 121 students, 97 completed both the pre- and post-surveys.

TABLE 3 | Results of correlation test for the main variables.

Dimension	1	2	3	4	5
Creative thinking	–				
Collaboration	0.635**	–			
Perseverance	0.691**	0.745**	–		
Career interest	0.730**	0.550**	0.643**	–	
Gender	−0.125	0.033	−0.013	−0.228*	–

*$p < 0.05$, **$p < 0.01$. For gender, 1, male; 2, female.*

were three girls (25%) and nine boys (75%). Students were asked about how the integrated STEM course helped facilitate or nurture creative thinking, collaboration, and STEM career interest. They were also asked to articulate the challenges and opportunities they encountered as well as their recommendations for improving the program. The focus groups were audiotaped, and then transcribed for analysis. Thematic analysis was conducted to identify and report the recurring patterns of meaning (themes) (Braun and Clarke, 2006). The data was first coded by one researcher of the study, and then the second researcher randomly selected 25% of the transcripts and coded them. The percent agreement between the two researchers was 81%. Discrepancies were discussed between the two researchers until agreements were reached.

RESULTS

Pre- and Post-Survey Findings

To examine whether there were any differences in students' creative thinking before and after completing the course, a paired sample *t*-test was conducted. Results indicated that there was a significant difference in the scores of post-creative thinking ($M = 4.58$, $SD = 0.99$) and pre-creative thinking ($M = 4.07$, $SD = 1.02$), $t_{(96)} = -5.00$, $p < 0.001$. See **Table 2**.

To examine whether there were any differences in students' collaboration before and after completing the course, a paired sample *t*-test was conducted. Results showed that there was a significant difference in the scores of post-collaboration ($M = 4.75$, $SD = 1.02$) and pre-collaboration ($M = 4.54$, $SD = 1.00$), $t_{(96)} = -2.16$, $p = 0.033$.

To examine whether there were any differences in students' perseverance before and after completing the course, a paired sample *t*-test was conducted. Results showed that there was a significant difference in the scores of post-perseverance ($M = 4.77$, $SD = 1.00$) and pre-perseverance ($M = 4.41$, $SD = 1.07$), $t_{(96)} = -3.26$, $p = 0.002$.

To examine whether there were any differences in students' career interest before and after completing the course, a paired sample *t*-test was conducted. Results indicated that there was a significant difference in the scores of post-career interest ($M = 4.19$, $SD = 1.19$) and pre-career interest ($M = 3.52$, $SD = 1.12$), $t_{(96)} = -5.17$, $p < 0.001$.

Moreover, the results revealed that career interest and creative thinking were the two most improved areas among the examined sub-dimensions.

Correlational Findings

Correlation analyses were conducted to further understand the relationships between the subscales. The post-survey data were used in testing the correlations. As shown in **Table 3**, the results indicated that creative thinking, collaboration, perseverance, and career interest were significantly and highly associated, with correlations ranging from 0.550 to 0.745 ($p < 0.01$).

Furthermore, the results showed that gender was not significantly related to creative thinking, collaboration, or perseverance. Boys were more likely to have a higher interest in STEM careers, whereas girls were less interested. However, there were no differences regarding creative thinking, collaboration, or perseverance between girls and boys.

Focus-Group Interviews

In the focus-group interviews, students shared their perceptions of the influences of the integrated course from three angles, including creative thinking, collaboration, and career interest. Students also shared the difficulties they encountered and how they coped with them. In general, students appreciated the positive changes brought by the program. Students also made several suggestions for further improving the program.

Regarding creative thinking, most students expressed that their creative thinking skills improved after completing the program. Several students stated that the program pushed them to think deeper and more thoroughly, and helped them to be more productive in generating new ideas (Gralewski and Karwowski, 2019). For example, one student commented that: "*This project boosted our creativity. I was not so creative in daily life. But, when I worked on this project, I threw up many new ideas.*" They also expressed that they would look at different angles to ensure the idea generated is not just "novel" but also "appropriate" (Runco et al., 2005). For example, one student elaborated that: "*In the design process, we would think deeper and consider more factors. In the past, we would not have thought about so many aspects. Now, we would consider whether there could be any errors if we put it into production, and whether existing technology is ready for implementing our design ideas, etc.*"

Moreover, students described that they used adaptation and synthesis skills to improve the design solutions, which are considered important indicators of creativity (Ramalingam et al., 2020). For example, one student shared that "*In one session, all teams presented their design solutions, from which we learned about the weaknesses of our designs and others'. So, we synthesized*

what we learnt from this experience, and enhanced our final design solution."

As for the aspect of collaboration skills, students expressed that their collaboration skills were enhanced through this learning experience. For example, one student commented, "*I felt that my collaboration skills become better, because we often work together to formulate the design concept, as well as the physical structure of the product*." The experience also improved students' eagerness to communicate and strengthened their closeness. For example, another student manifested that: "*At first, I did not like to share my viewpoints, but after participating in more group discussions, I became more willing to communicate with teammates. The collaboration and communication further enhanced the closeness between our teammates*."

Students also expressed that they were actively adapting their approach in collaboration. One student expressed that: "*I used to spend a long time conceptualizing an idea, and would only share it when it was fully developed. Now, (to be more effective) I would share all my initial ideas, and work with my peers to find the best solution*." In fact, the willingness to work with others and share ideas with peers is considered important enabler of creative thinking (Foster and Schleicher, 2022). The positive feedback on collaboration, in a sense, explained why students had higher creative thinking after completing the program.

Regarding perseverance, we wanted to find out whether students had encountered difficulties in their studies and whether they exerted continuous efforts to resolve them, i.e., an indicator of perseverance (Toering et al., 2012). Hence, we raised the following questions, "Did you meet any difficulties in completing the project, and how did you cope with them?" Most of the interviewees commented that they did not encounter many difficulties in completing the project because they functioned as a team. One student shared that they met technical problems in designing their solution, and they solved them by searching the website and reading relevant technical posts. Another student mentioned that it took them a long time to record an introduction video on the design solution. Though it was quite challenging and time-consuming, they insisted on discussing and clarifying the details of each scene and provided timely advice to the students in charge of filming. Another challenge reported by students was the time restriction. Another challenge reported by the students was the time constraint. One student repeatedly mentioned that though they had 2 weeks to refine and finalize their design solution, he felt that the time was still a bit tight and hoped that more time was assigned to refine the solutions. Although this point is not directly related to perseverance, it reminds us of what to pay attention to when designing future projects. We can allow more time to incubate, reflect and select among alternative design solutions (Sternberg, 2003).

During the interview, we found that girls were more interested in STEM subjects and felt more at ease with coding activities than they were before participating in the program, but they did not yet have a plan for their future careers. Meanwhile, the program has motivated some boys to consider pursuing STEM as their future career. One boy expressed that he was thinking about whether to work in STEM-related fields after graduation but had not decided yet. Another boy said that though he would like to work in other fields in the future, he would discover more about STEM and pursue it as a hobby. The results indicated that the program positively impacted students' interest in STEM and related careers, but further efforts are needed to foster a stronger interest in the profession.

DISCUSSION

This study proposed an integrated STEM-based community service-learning model referring to the componential model of creativity and the design thinking model. The model advocates the engagement of students in design challenges to solve real-world problems faced by community residents, while providing students with access to learning domain knowledge in STEM and community service. Moreover, it encourages the establishment of a supportive learning environment for students by offering expert or user feedback. Other resources and support from the school level can also be provided depending on the context of each study. In our program, to ensure the authenticity of the learning materials, the program team collaborated with industry experts, staff of social service organizations, and community residents. The designing home devices and furniture section was delivered by an industry expert in video format. The survey results showed that students increased their creative thinking, collaboration, perseverance, and STEM career interest, especially in creative thinking and career interest, upon completion of the program.

The interview findings are consistent with the survey results. The positive results support the assumption that when students are provided with adequate resources, domain-specific skills, and a well-structured creative design process, twenty-first century skills can be improved. The exploration of solving community service problems also benefits local community residents (Andreoletti and Howard, 2018). In the future, we can promote the model in more K-12 schools and examine its effectiveness.

As Amabile and Pratt (2016) stated, the incubation of creative thinking and innovation requires three essential elements, such as resources, domain-specific skills and process skills, and a stimulating and supportive learning environment. In this program, we developed an integrated course collaborating with industry experts, social service staff, and community residents to equip students with domain-specific knowledge and foundations for designing user-centered products. We also guided students' design in a step-by-step manner. Furthermore, to better support students, we created an opportunity to interact with industry experts and social service staff to collect feedback on the feasibility of their design solutions. These practices promoted an improvement in creative thinking. The result is consistent with the finding of Nazzal and Kaufman's (2020) work that creative thinking is connected to domain-specific knowledge. In addition, the design challenge was carried out in the form of teamwork, and the positive challenge and collaboration in teams may have encouraged creative thinking (Maiorca et al., 2021). Moreover, the authentic problem-based learning process may have sparked students' interest in STEM subjects and related careers, and consequently, promoted their creativity (Bush et al., 2020).

Perseverance drives people to achieve their goals despite obstacles and difficulties (Pury, 2009). This integrated course promoted students' perseverance, supporting the previous

finding that the authenticity of learning tasks and materials contributes to higher perseverance in learning (Mutlu and Yildirim, 2019). In addition, our result on the positive correlation between perseverance and creative thinking is consistent with Scherer and Gustafsson's (2015) finding that perseverance is positively correlated with creative problem-solving.

Collaboration is a predictor of highly effective problem-solving teams (Fullan and Langworthy, 2014). Chai et al. (2020a) examined the core dimensions of twenty-first century skills, and confirmed that collaborative learning predicts students' creative problem-solving. Hsia et al. (2021) found that promoting collaborative learning and interaction in class could enhance students' creative thinking tendency. Our correlation test result matches their findings. In the future, we can continue to encourage collaboration among students and explore the connections between group styles, creative thinking, and STEM career interest.

Our survey results indicated that students' STEM career interest increased from a low level (3.5) to a relatively higher level (4.19), which corroborates the survey finding of Vennix et al. (2018). The survey results showed that STEM courses are more likely to have a positive impact on STEM career interest when they are linked to real-life contexts. The interview results are consistent with the survey results. Most students expressed that their interest in STEM is higher than before. In addition, although gender was not correlated with creative thinking in STEM, gender was weakly correlated with career interest (i.e., girls had relatively lower career interest). The result suggests that girls are less likely to choose STEM careers in the future, despite their potential to play an equally important role in STEM fields (Ünlü and Dökme, 2020). Further work on how to promote female students' career interest in STEM is worth additional research efforts.

IMPLICATIONS, LIMITATIONS, AND FUTURE RESEARCH

This study has important theoretical and practical implications for science education. From a theoretical perspective, we proposed an integrated STEM-based community service learning model that builds on the strengths of the componential theory of creativity and the EDIPT design thinking model, one of the earliest theoretical attempts in the cross-disciplinary field of STEM education and community service. Integrating STEM with community service learning is a nascent field, and thus there is a paucity of empirical research and theoretical exploration in this area (Collins et al., 2020). This model analyzes the core elements (e.g., resources, domain knowledge, and process) that enable individual or team innovation, and illustrates the procedure for organizing the design activities. Furthermore, the model links real-life community service issues, community service learning, STEM education, and design activities, and empowers cross-disciplinary learning with meaning (i.e., solve authentic problems to improve the comfortable and convenience of community residents' lives).

From the perspective of practice, the study demonstrated the process for integrating community service learning and STEM education following the integrated model. The study also exhibited how contents can be developed building upon the multiple expertise of industry experts, teachers, and social service staff. Moreover, the pre- and post-implementation results showcased the potential outcomes that could be achieved through this sort of integrated attempt regarding students' twenty-first century skills, perseverance, and STEM career interest.

Practitioners in science education and humanities can also refer to this model for designing interdisciplinary curriculum to nurture creative thinking while strengthening attitudes, values, and sense of social responsibility. Teachers are always important stakeholders of any new educational initiatives, playing a salient role in the adoption process (Geng et al., 2019; Jong, 2019); there should be no exception in STEM education (Chai et al., 2019). Training in understanding and implementing this interdisciplinary curriculum can be organized for teachers to remove barriers to adopting new pedagogical approaches, facilitate their smooth implementation of this innovative teaching model, and ultimately improve student performance (Jong et al., 2021; Lau and Jong, 2022).

Despite its strengths, this study also has some limitations. First, a limitation of the study is that it mainly used a one-group pre- and post-survey approach to understand the impact of this integrated course. Future studies that include a control group would produce more rigorous results. In addition, we encourage future studies to collect different data sets, such as tests or artifacts to further understand the impacts of the STEM-based community service-learning model. Second, given that the program was carried out in a secondary school in Hong Kong, its generalizability to other settings needs to be tested. In the future, this integrated learning approach can be extended to more schools in different settings and see whether consistent results could be achieved. Moreover, this study mainly explored the combination of STEM with the theme of community service learning such as designing home devices and furniture for residents in community-housing. Future studies can also explore the applicability of other community service themes.

CONCLUSION

This study provides quantitative evidence that integrating STEM with community service learning is beneficial to promoting students' creative thinking, collaboration, perseverance, and STEM career interest. The qualitative interviews with the students further validated these quantitative findings. Scardamalia et al. (2012) advocated that citizen in the twenty-first century need to add value to existing knowledge by creating new knowledge or designing new artifacts that are valuable to others. Our STEM-based community service program is a small step toward that ideal and provides a concrete example of how this could be done in the educational context.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because the access of the datasets should be approved by the involved schools. Requests to access the datasets should be directed to Morris Siu-Yung Jong, mjong@cuhk.edu.hk.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Chinese University of Hong Kong. The participants/their legal guardians provided the informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

BH: conceptualization, visualization, data collection, data analysis, writing the first draft of the manuscript, and revision. MS-YJ: conceptualization, visualization, funding acquisition, project administration, supervision, validation, rewriting sections of the manuscript, review, and edit. RK: rewriting sections of the draft, visualization, review, and edit. C-SC and MY-CJ: visualization, validation, review, and edit the manuscript. All authors approved the submission.

FUNDING

The work described in this paper was supported by the Hong Kong Jockey Club Charities Trust (Project Title: Jockey Club Community Care and STEM in Action [Project S/N Ref: JC 2019/0112]).

ACKNOWLEDGMENTS

Thank you very much to all participants in this study.

REFERENCES

Amabile, T. M. (1983). The social psychology of creativity: a componential conceptualization. *J. Pers. Soc. Psychol.* 45, 357–376. doi: 10.1037/0022-3514.45.2.357

Amabile, T. M., and Pratt, M. G. (2016). The dynamic componential model of creativity and innovation in organizations: making progress, making meaning. *Res. Organ. Behav.* 36, 157–183. doi: 10.1016/j.riob.2016.10.001

Andreoletti, C., and Howard, J. L. (2018). Bridging the generation gap: intergenerational service-learning benefits young and old. *Gerontol. Geriatr. Educ.* 39, 46–60. doi: 10.1080/02701960.2016.1152266

Bielefeldt, A. (2017). "Is engineering about caring for and helping others?" in *Focus on Educational Research: Practices, Challenges, and Perspectives*, ed B. Hernandez (Hauppauge, NY: Nova Science Publishers, Inc.) 89–119.

Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., et al. (2012). "Defining twenty-first century skills," in *Assessment and Teaching of 21st Century Skills*, eds P. Griffin, B. McGaw, and E. Care (Dordrecht: Springer) 17–66. doi: 10.1007/978-94-007-2324-5_2

Botelho, J., Eddy, R. M., Galport, N., and Avila-Linn, C. (2020). Uncovering the quality of STEM service-learning course implementation and essential elements across the California state university system. *Mich. J. Community. Serv. Learn.* 26. doi: 10.3998/mjcsloa.3239521.0026.201

Bower, M., and Jong, M. S. Y. (2020). Immersive virtual reality in education. *Brit. J. Educ. Technol.* 51, 1981–1990. doi: 10.1111/bjet.13038

Braun, V., and Clarke, V. (2006). Using thematic analysis in psychology. *Qual. Res. Psychol.* 3, 77–101. doi: 10.1191/1478088706qp063oa

Bringle, R. G., and Hatcher, J. A. (1996). Implementing service-learning in higher education. *J. Higher Educ.* 67, 221–239. doi: 10.2307/2943981

Burton, C., and Winter, M. A. (2021). Benefits of service-learning for students during the Covid-19 crisis: two case studies. *Scholarsh. Teach. Learn. Psychol.* doi: 10.1037/stl0000292

Bush, S. B., Cook, K. L., Edelen, D., and Cox, R. Jr. (2020). Elementary students' STEAM perceptions: extending frames of reference through transformative learning experiences. *Element. Sch. J.* 120, 692–714. doi: 10.1086/708642

Chai, C. S., Jong, M. S. Y., Yin, H. B., Chen, M. Y., and Zhou, W. (2019). Validating and modelling teachers' technological pedagogical content knowledge for integrative science, technology, engineering and mathematics education. *Educ. Technol. Soc.* 22, 61–73. Available online at: https://www.jstor.org/stable/26896710

Chai, C. S., Liang, J. C., Tsai, C. C., and Dong, Y. (2020a). Surveying and modelling China high school students' experience of and preferences for twenty-first-century learning and their academic and knowledge creation efficacy. *Educ. Stud.* 46, 658–675. doi: 10.1080/03055698.2019.1627662

Chai, C. S., Rahmawati, Y., and Jong, M. S. Y. (2020b). Indonesian science, mathematics, and engineering preservice teachers' experiences in STEM-TPACK design-based learning. *Sustainability* 12, 1–14. doi: 10.3390/su12219050

Chiva-Bartoll, O., Moliner, M. L., and Salvador-García, C. (2020). Can service-learning promote social well-being in primary education students? A mixed method approach. *Child. Youth. Serv. Rev.* 111. doi: 10.1016/j.childyouth.2020.104841

Collins, M. A., Totino, J., Hartry, A., Romero, V. F., Pedroso, R., and Nava, R. (2020). Service-learning as a lever to support STEM engagement for underrepresented youth. *J. Exp. Educ.* 43, 55–70. doi: 10.1177/1053825919887407

Da Silva, R. H., Kaminski, P. C., and Armellini, F. (2020). Improving new product development innovation effectiveness by using problem solving tools during the conceptual development phase: integrating design thinking and TRIZ. *Creat. Innov. Manag.* 29, 685–700. doi: 10.1111/caim.12399

English, L. D. (2019). Learning while designing in a fourth-grade integrated STEM problem. *Int. J. Technol. Des. Educ.* 29, 1011–1032. doi: 10.1007/s10798-018-9482-z

Foster, N., and Schleicher, A. (2022). Assessing creative skills. *Creat. Educ.* 13, 1–29. doi: 10.4236/ce.2022.131001

Fullan, M., and Langworthy, M. (2014). *A Rich Seam: How New Pedagogies Find Deep Learning*. London: Pearson.

Geng, J., Jong, M. S. Y., and Chai, C. S. (2019). Hong Kong teachers' self-efficacy and concerns about STEM education. *Asia Pac. Educ. Res.* 28, 35–45. doi: 10.1007/s40299-018-0414-1

Gerholz, K. H., Liszt, V., and Klingsieck, K. B. (2018). Effects of learning design patterns in service learning courses. *Act. Learn. High. Educ.* 19, 47–59. doi: 10.1177/1469787417721420

Ginns, P., Freebody, K., Anderson, M., and O'Connor, P. (2021). Student experience of creativity in Australian high school classrooms: a componential model. *Learn. Individ. Differ.* 91. doi: 10.1016/j.lindif.2021.102057

Gralewski, J., and Karwowski, M. (2019). Are teachers' ratings of students' creativity related to students' divergent thinking? A meta-analysis. *Think. Skills Creat.* 33. doi: 10.1016/j.tsc.2019.100583

Gunckel, K. L., and Tolbert, S. (2018). The imperative to move toward a dimension of care in engineering education. *J. Res. Sci. Teach.* 55, 938–961. doi: 10.1002/tea.21458

Hasso Plattner Institute of Design (2010). *An Introduction to Design Thinking Process Guide*. Available online at: https://web.stanford.edu/$\sim$mshanks/MichaelShanks/files/509554.pdf (accessed January 8, 2022).

Honey, M., Pearson, G., and Schweingruber, H. (eds.). (2014). *STEM Iintegration in K-12 Education: Status, Prospects, and*

An Agenda for Research. Washington, DC: National Academies Press.

Hong, O., and Song, J. (2020). A componential model of Science Classroom Creativity (SCC) for understanding collective creativity in the science classroom. *Think. Skills Creat*. 37. doi: 10.1016/j.tsc.2020.100698

Hsia, L. H., Lin, Y. N., and Hwang, G. J. (2021). A creative problem solving-based flipped learning strategy for promoting students' performing creativity, skills and tendencies of creative thinking and collaboration. *Brit. J. Educ. Technol.* 52, 1771–1787. doi: 10.1111/bjet.13073

Huang, B., Jong, M. S. Y., and Chai, C. S. (2022). The design and implementation of a video-facilitated transdisciplinary STEM curriculum in the context of COVID-19 pandemic. *Educ. Technol. Soc*. 25, 108–123. Available online at: https://www.jstor.org/stable/48647034

Jacoby, B. (1996). "Service-learning in today's higher education," in *Service-Learning in Higher Education: Concepts and Practices*, eds B. Jacoby and Associates (Hoboken, NJ: Jossey-Bass), 3–25.

Jong, M. S. Y. (2019). Sustaining the adoption of gamified outdoor social enquiry learning in high schools through addressing teachers' emerging concerns: a three-year study. *Brit. J. Educ. Technol*. 50, 1275–1293. doi: 10.1111/bjet.12767

Jong, M. S. Y., Shang, J. J., Lee, F. L., and Lee, J. H. M. (2010). An evaluative study on VISOLE-Virtual Interactive Student-Oriented Learning Environment. *IEEE Trans. Learn. Technol*. 3, 307–318. doi: 10.1109/TLT.2010.34

Jong, M. S. Y., Song, Y., Soloway, E., and Norris, C. (2021). Teacher professional development in STEM education. *Educ. Technol. Soc*. 24, 81–85. Available online at: https://www.jstor.org/stable/48629246

Karahan, E., Kara, A., and Akçay, A. O. (2021). Designing and implementing a STEM career maturity program for prospective counselors. *Int. J. STEM Educ*. 8. doi: 10.1186/s40594-021-00281-4

Kelley, T. R., and Knowles, J. G. (2016). A conceptual framework for integrated STEM education. *Int. J. STEM Educ*. 3.doi: 10.1186/s40594-016-0046-z

Lai, C. L., and Hwang, G. J. (2014). Effects of mobile learning time on students' conception of collaboration, communication, complex problem-solving, meta-cognitive awareness and creativity. *Int. J. Mob. Learn. Organ*. 8, 276–291. doi: 10.1504/IJMLO.2014.067029

Lau, W. W. F., and Jong, M. S. Y. (2022). Typology of teachers' stages of concern for STEM education. *Res. Sci. Technol. Educ*. doi: 10.1080/02635143.2022.2064447

Lee, M. H., Chai, C. S., and Hong, H. Y. (2019). STEM education in Asia Pacific: challenges and development. *Asia Pac. Educ. Res*. 28, 1–4. doi: 10.1007/s40299-018-0424-z

Lin, P. Y., Chai, C. S., Jong, M. S. Y., Dai, Y., Guo, Y., and Qin, J. (2021). Modeling the structural relationship among primary students' motivation to learn artificial intelligence. *Comput. Educ. Artif. Intell.* 2, 100006. doi: 10.1016/j.caeai.2020.100006

Luk, E. T. H., Wong, M. K. H., Cheung, K. K. F., Lee, F. L., and Lee, J. H. M. (2006). "Design and implementation of FARMTASIA: a game designed for the VISOLE teaching style," in *Proceedings of the 1st International Conference on Technologies for e-Learning and Digital Entertainment*, eds Z. Pan, R. Aylett, H. Diener, X. Jin, S. Gobel, and L. Li (Hangzhou), 566–571. doi: 10.1007/11736639_67

Maiorca, C., Roberts, T., Jackson, C., Bush, S., Delaney, A., Mohr-Schroeder, M. J., et al. (2021). Informal learning environments and impact on interest in STEM careers. *Int. J. Sci. Math. Educ*. 19, 45–64. doi: 10.1007/s10763-019-10038-9

Marcus, V., Atan, N., Yusof, S., and Mastura, U. (2021). Students' perception towards engaging factors of extreme e-service learning design for computer network course. *Int. J. Interact. Mob. Technol.* 15, 100–115. doi: 10.3991/ijim.v15i05.20901

Mavri, A., Ioannou, A., and Loizides, F. (2020). The assessment scale for creative collaboration (ASCC) validation and reliability study. *Int. J. Hum-Comput. Int.* 36, 1056–1069. doi: 10.1080/10447318.2019.1709338

Moirano, R., Sánchez, M. A., and Štěpánek, L. (2020). Creative interdisciplinary collaboration: a systematic literature review. *Think. Skills Creat*. 35. doi: 10.1016/j.tsc.2019.100626

Morrin, A. M., and Liston, M. (2020). Engaging children with authentic STEAM learning experiences through design-based approaches. *Connect. Sci. Learn.* 2. Available online at: https://www.nsta.org/connected-science-learning/connected-science-learning-october-december-2020/engaging-children

Mutlu, G., and Yildirim, A. (2019). Learning environment perceptions and student background variables as determinants of persistence in EFL learning. *SAGE Open* 9. doi: 10.1177/2158244019898805

Nadelson, L. S., and Seifert, A. L. (2017). Integrated STEM defined: contexts, challenges, and the future. *J. Educ. Res*. 110, 221–223. doi: 10.1080/00220671.2017.1289775

Nazzal, L. J., and Kaufman, J. C. (2020). The relationship of the quality of creative problem solving stages to overall creativity in engineering students. *Think. Skills Creat*. 38. doi: 10.1016/j.tsc.2020.100734

Papadakis, S. (2016). Creativity and innovation in European education. Ten years eTwinning. Past, present and the future. *Int. J. Technol. Enhanc. Learn*. 8, 279–296. doi: 10.1504/IJTEL.2016.082315

Patton, M. Q. (2001). *Qualitative Research and Evaluation Methods, 3rd Edn*. Washington, DC: SAGE Publications.

Perignat, E., and Katz-Buonincontro, J. (2019). STEAM in practice and research: an integrative literature review. *Think. Skills Creat*. 31, 31–43. doi: 10.1016/j.tsc.2018.10.002

Pury, C. L. S. (2009). "Perseverance," in *The Encyclopedia of Positive Psychology*, ed S. J. Lopez (Hoboken, NJ: Blackwell Publishing Ltd) 678–682.

Quigley, C. F., Herro, D., and Baker, A. (2019). "Moving toward transdisciplinary instruction: a longitudinal examination of STEAM teaching practices," in *STEAM Education,* eds M. Khine and S. Areepattamannil (Cham: Springer), 143–164. doi: 10.1007/978-3-030-04003-1_8

Ramalingam, D., Anderson, P., Duckworth, D., Scoular, C., and Heard, J. (2020). *Creative Thinking: Definition and Structure*. Australian Council for Educational Research. https://research.acer.edu.au/ar_misc/4 (accessed February 10, 2022).

Reed, S. C., Rosenberg, H., Statham, A., and Rosing, H. (2015). The effect of community service learning on undergraduate persistence in three institutional contexts. *Mich. J. Commun. Serv. Learn*. 21, 22–36. Available online at: https://files.eric.ed.gov/fulltext/EJ1116299.pdf

Reynante, B. M., Selbach-Allen, M. E., and Pimentel, D. R. (2020). Exploring the promises and perils of integrated STEM through disciplinary practices and epistemologies. *Sci. Educ.* 29, 785–803. doi: 10.1007/s11191-020-00121-x

Runco, M. A., Illies, J. J., and Eisenman, R. (2005). Creativity, originality, and appropriateness: what do explicit instructions tell us about their relationships?. *J. Creat. Behav*. 39, 137–148. doi: 10.1002/j.2162-6057.2005.tb01255.x

Salam, M., Awang Iskandar, D. N., Ibrahim, D. H. A., and Farooq, M. S. (2019). Service learning in higher education: a systematic literature review. *Asia Pac. Educ. Rev*. 20, 573–593. doi: 10.1007/s12564-019-09580-6

Scardamalia, M., Bransford, J., Kozma, B., and Quellmalz, E. (2012). "New assessments and environments for knowledge building," in *Assessment and Teaching of 21st Century Skills*, eds P. Griffin, B. McGaw, and E. Care (Dordrecht: Springer), 231–300. doi: 10.1007/978-94-007-2324-5_5

Scherer, R., and Gustafsson, J. E. (2015). The relations among openness, perseverance, and performance in creative problem solving: a substantive-methodological approach. *Think. Skills Creat.* 18, 4–17. doi: 10.1016/j.tsc.2015.04.004

Song, Y., Jong, M. S. Y., Chang, M., and Chen, W. (2017). "HOW" to design, implement and evaluate the flipped classroom? - A Synthesis. *Educ. Technol. Soc*. 20, 180–183. Available online at: https://media.proquest.com/media/hms/PFT/1/VSKp7?_s=PGlkXYvtWfl8qCC5S8giPpuacac%3D

Sternberg, R. J. (2003). Creative thinking in the classroom. *Scand. J. Educ. Res.* 47, 325–338. doi: 10.1080/00313830308595

Sun, M., Wang, M., and Wegerif, R. (2020). Effects of divergent thinking training on students' scientific creativity: the impact of individual creative potential and domain knowledge. *Think. Skills Creat* 37. doi: 10.1016/j.tsc.2020.100682

Tijsma, G., Hilverda, F., Scheffelaar, A., Alders, S., Schoonmade, L., Blignaut, N., et al. (2020). Becoming productive 21st century citizens: a systematic review uncovering design principles for integrating community service learning into higher education courses. *Educ. Res*. 62, 390–413. doi: 10.1080/00131881.2020.1836987

Toering, T., Elferink-Gemser, M. T., Jonker, L., van Heuvelen, M. J., and Visscher, C. (2012). Measuring self-regulation in a learning context: reliability and validity of the self-regulation of learning self-report scale (SRL-SRS). *J. Sport Exerc. Psychol.* 10, 24–38. doi: 10.1080/1612197X.2012.645132

Tsai, C.-C., Chai, C. S., Wong, B., Hong, H.-Y., and Tan,. S. C. (2013). Positioning design epistemology and its applications in education technology. *Educ. Technol. Soc*. 16, 81–90. Available online at: https://drive.google.com/file/d/1RyEKbQNN672uga1wzBQNJC2L3culLg18/view

Ulger, K. (2018). The effect of problem-based learning on the creative thinking and critical thinking disposition of students in visual arts education. *Interdiscipl. J. Probl. Based Learn* 12. doi: 10.7771/1541-5015.1649

Ünlü, K. Z., and Dökme, I. (2020). Multivariate assessment of middle school students' interest in STEM career: a profile from Turkey. *Res. Sci. Educ*. 50, 1217–1231. doi: 10.1007/s11165-018-9729-4

Vennix, J., den Brok, P., and Taconis, R. (2018). Do outreach activities in secondary STEM education motivate students and improve their attitudes towards STEM?. *Int. J. Sci. Educ*. 40, 1263–1283. doi: 10.1080/09500693.2018.1473659

Vo, H., and Asojo, A. (2021). Feedback in interior design studio and students' creativity. *J. Creat*. 31. doi: 10.1016/j.yjoc.2021.100009

Wan, Z. H., Jiang, Y., and Zhan, Y. (2021). STEM education in early childhood: a review of empirical studies. *Early. Educ. Dev*. 32, 940–962. doi: 10.1080/10409289.2020.1814986

Weng, X., Chiu, T. K. F., and Jong, M. S. Y. (2022). Applying relatedness to explain learning outcomes of STEM maker activities. *Front. Psychol*. 12, 800569. doi: 10.3389/fpsyg.2021.800569

Yalçin, V., and Erden, S. (2021). The effect of STEM activities prepared according to the design thinking model on pre-school children's creativity and problem-solving skills. *Think. Skills Creat*. 41. doi: 10.1016/j.tsc.2021.100864

Yorio, P. L., and Ye, F. (2012). A meta-analysis on the effects of service-learning on the social, personal, and cognitive outcomes of learning. *Acad. Manag. Learn. Educ*. 11, 9–27. doi: 10.5465/amle.2010.0072

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Huang, Jong, King, Chai and Jiang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Education

ORIGINAL RESEARCH
published: 22 July 2022
doi: 10.3389/feduc.2022.872888

Teaching Preschoolers Theory of Mind Skills With Mobile Games

Mariya Nikolayev[1], Anya S. Evmenova[1], Stephanie M. Reich[2], Kevin A. Clark[1] and M. Susan Burns[1]*

[1] College of Education and Human Development (CEHD), George Mason University, Fairfax, VA, United States, [2] School of Education, University of California, Irvine, Irvine, CA, United States

This single-case research study examined whether interactive touch screen apps enriched with Theory of Mind (ToM)—enhancing language would promote ToM skills in preschoolers. Six typically developing girls between the ages of 46- and 52-months participated in multiple sessions across the three phases of the study: In baseline, participants played games without voice-overs; in the original treatment phase, participants played games with embedded voice-overs; finally, in the modified treatment phase, participants first played games with embedded voice-overs, then engaged in the researcher-led conversation. All sessions across the three phases concluded with ToM assessments: two measures based on a continuous scale. The first measure included three tasks targeting earlier-developing ToM skills (diverse desires, diverse beliefs, and knowledge access), and the other measure had two tasks that assessed a later-developing ToM competency, false belief understanding. Results showed that apps with ToM-embedded language improved children's earlier-developing ToM skills (i.e., understanding that people can have different desires, beliefs, and knowledge access) in the phase where an adult-led conversation also followed voice-over-enriched app play. Apps with ToM-embedded language without a follow-up discussion were only marginally effective in promoting the earlier-emerging ToM skills. Across the conditions, apps were not effective in promoting children's later-developing ToM skills—false belief understanding. Our findings indicate that incorporating ToM conducive language in mobile apps can promote ToM development in preschoolers, especially when supplemented by an adult-led conversation.

Keywords: theory of mind, preschoolers, educational mobile apps, intervention, single-case design, joint media engagement, digital adult, language

OPEN ACCESS

Edited by:
Stamatios Papadakis, University of Crete, Greece

Reviewed by:
Marlou Poppelaars, Radboud University, Netherlands
David Pérez-Jorge, University of La Laguna, Spain
Julie Vaiopoulou, Aristotle University of Thessaloniki, Greece

****Correspondence:***
Mariya Nikolayev
mariya.nikolayev@gmail.com

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Education

Received: *10 February 2022*
Accepted: *20 June 2022*
Published: *22 July 2022*

Citation:
Nikolayev M, Evmenova AS, Reich SM, Clark KA and Burns MS (2022) Teaching Preschoolers Theory of Mind Skills With Mobile Games. Front. Educ. 7:872888. doi: 10.3389/feduc.2022.872888

INTRODUCTION

Theory of Mind (ToM) is an essential area of social-emotional development that enables children to recognize, understand, and predict feelings, intentions, beliefs, and desires of the self and others (Astington, 2003; Keenan, 2003). The level of ToM development across different ages has been found to predict children's positive and negative behaviors. Advanced mastery of ToM skills helps to make and keep friends (e.g., Slaughter et al., 2015), promotes persuasion and leadership skills (e.g., Peterson et al., 2018), contributes to the development of prosocial behaviors (e.g., Imuta et al., 2016), correlates with being liked by teachers (e.g., Slaughter et al., 2002) and being popular among peers (e.g., Fink et al., 2014; Slaughter et al., 2015), and promotes academic achievement (e.g., Lecce et al., 2014; Dore et al., 2018; Florit et al., 2020).

Conversely, delays and poor ToM development are associated with difficulties across developmental stages, e.g., problems in social-emotional functioning in preschool (e.g., Vissers and Koolen, 2016), aggressive behaviors in kindergarten (e.g., Renouf et al., 2010), bullying during middle childhood and teenage years (e.g., Shakoor et al., 2012), and feelings of loneliness in adolescence (e.g., Bosacki et al., 2020).

During ToM development, when learning about mental states of others, most Typically Developing (TD) children across the world follow a standard sequence of skills acquisition, with early abilities serving as precursors for later skills (Wellman and Liu, 2004; Shahaeian et al., 2011), including an understanding of false beliefs (i.e., being able to infer the incorrect belief of another person). Although most TD children demonstrate roughly the same level of false belief understanding by the end of preschool (Wellman and Peterson, 2013), the timing and rate of development, conceptual elaboration, and the degree to which children apply the skills in social situations vary among individual children (Cutting and Dunn, 1999; Wellman et al., 2001; Keenan, 2003; Charness et al., 2019). The variability in ToM skills continues to exist into adolescence (e.g., Caputi et al., 2012; Fink et al., 2015; Hughes and Devine, 2015), differentially contributing to the success of children's social and academic experiences across the life span (e.g., Devine et al., 2016; Dore et al., 2018; Peterson et al., 2018).

Given that this variation in ToM skills acquisition has an important impact on well-being (Hughes and Devine, 2015; Weimer et al., 2021) and the overall effectiveness of intervention programs (Hofmann et al., 2016; Roheger et al., 2022), some researchers suggest that ToM-enhancing programs should be provided not only to remedy ToM delays but also to prevent them (e.g., Hofmann et al., 2016). Supplemental early ToM learning may be especially significant for populations of children who are known to lag behind their peers in ToM development (Holmes et al., 1996; Tarullo et al., 2007; Dessen and de Hollanda Souza, 2014; Devine and Hughes, 2018; Charness et al., 2019).

Currently, most of the existing ToM interventions are provided through face-to-face interactions, with none, to our knowledge, being implemented beyond research, clinical, or school-based settings (see Mori and Cigala, 2015; Hofmann et al., 2016 for a review of the training programs). Furthermore, children who demonstrate deficits in ToM development, such as those from disadvantaged backgrounds (Charness et al., 2019; Ebert et al., 2020), are those who could potentially benefit from intervention programs, but also lack access to these interventions and educational opportunities (Hodgkinson et al., 2017; Griffith et al., 2019).

In 2011 the Global Child Development Steering Group named educational media a promising method for promoting early child development and addressing at-risk child populations (Engle et al., 2011). School districts across the country have invested millions of dollars into new technologies for education (Blackwell et al., 2013). A decade later, the presence of mobile devices became almost universal in the homes of children of all socioeconomic backgrounds as the popularity of educational gaming also increased (Griffith and Arnold, 2019; Rideout and Robb, 2020). This resulted in the production of hundreds of thousands of apps claiming educational benefits (Apple, 2021). The educational apps' prevalence and accessibility makes them a promising scalable method for delivering ToM educational content to various populations of children in need of ToM development support.

In this paper, we sought to examine whether the ToM-stimulating language in educational games for preschoolers can improve children's ToM performance with the support of the game's interactivity and parental engagement. There is evidence that interactivity and parental engagement are beneficial for children's learning of various skills, including coding, math, literacy, and some social-emotional, from digital games, yet whether they impact the ToM skills is unclear due to the lack of research on the topic. The preschool age of the participants was selected because ToM gains are most rapid between 3 and 5 years (e.g., Tompkins et al., 2019), and by four years of age, most children are capable of playing digital games independently but still benefit from adult support (Pempek and Lauricella, 2017 as cited in Bindman et al., 2021). Furthermore, given the potential of digital games to serve as a scalable platform for delivering ToM interventions, we proposed ToM-focused design suggestions to be used by edutainment game developers.

Background and Related Work

The review goes over the existing face-to-face language-based interventions designed to improve preschoolers' ToM skills and discusses scientific literature on children's learning from digital games. Intending to create effective ToM-educational content for mobile games, we focus on the language elements in face-to-face interventions that influence ToM development and various design and contextual factors of digital games that promote children's learning.

Language-Based ToM Interventions

A large body of literature highlights the critical role of language in developing ToM (e.g., see Tompkins et al., 2019 for a review), specifically the socio-linguistic environment and a child's abilities (De Villiers and De Villiers, 2000; Astington and Baird, 2005; Taumoepeau and Reese, 2013; Lillard and Kavanaugh, 2014). The concept of a socio-linguistic environment is based on the notion that language needed for ToM is facilitated by a child's social experiences, such as the content of conversations between parents and children, home literacy environment, and the frequency and content of conversations overheard by children (Astington and Baird, 2005; Slaughter et al., 2007; Ruffman, 2014; Tompkins et al., 2018; Lecce et al., 2021). The socio-linguistic environment and children's language are closely related and contribute both jointly and independently to ToM development (Astington and Baird, 2005). For example, studies have shown that the ToM of children from disadvantaged backgrounds, such as low SES or institutionalized settings, develops slower than children from more advantaged ones (e.g., Tarullo et al., 2007; Charness et al., 2019). Further, Ebert et al. (2020) demonstrated how various SES-related aspects of the home language and literacy environment contribute to children's ToM and language development.

The natural pace of ToM development is slow; children's false belief performance without intervention typically improves only slightly between the ages of 3.5 and 4 (Amsterlaw and Wellman, 2006). However, experimental intervention studies have accelerated children's ToM acquisition in as short a time frame as two weeks (e.g., Slaughter and Gopnik, 1996; Hale and Tager-Flusberg, 2003; Wellman, 2018), demonstrating that it is possible for the interventions to help children who are behind on ToM development to catch up with their more advanced peers.

Various methods have been used to deliver training content in face-to-face language-based ToM interventions (see Mori and Cigala, 2015; Hofmann et al., 2016 for review): some teach caregivers to reminisce about past events, pose questions in a specific way, and incorporate words for feelings, desires, and beliefs to promote children's expressive vocabulary (e.g., Taumoepeau and Reese, 2013; Spruijt et al., 2020). Others have teachers and researchers read storybooks enriched with mental-state language, show videos, or demonstrate puppet shows with the mental-state-verb-laden script (e.g., Esteban et al., 2010; Gola, 2012; Tompkins, 2015; San Juan and Astington, 2017). Finally, some have children complete false-belief tasks and then follow the task with evidence-based or corrective feedback and explanations (e.g., Slaughter and Gopnik, 1996; Clements et al., 2000).

Many studies have found that using ToM-promoting language alone is not enough to accelerate ToM (e.g., Peskin and Astington, 2004; Amsterlaw and Wellman, 2006; Ornaghi et al., 2011). The interactivity seems to be crucial for ToM mastery; in conversations, children observe how their own and other people's perspectives become clear as well as inconsistencies between their own and others' mental states and realities (De Villiers and De Villiers, 2014). Several studies have tried to make the children active participants rather than passive observers by engaging them in language-based activities and discussions containing mentalistic language (Lohmann and Tomasello, 2003; Ornaghi et al., 2011). For example, language that encompasses references to emotional states (e.g., *happy, sad, excited*), mental processes (e.g., *know, think, remember, understand, fee*l), desires (e.g., *want, wish, hope*), and modulations of assertion (*e.g., guess, maybe, perhaps*) (Ruffman et al., 2002). Others use storybook readings followed by adult-led discussions and reflections about the mental states and behaviors of characters in the stories (e.g., Guajardo and Watson, 2002; Tompkins, 2015). Despite the common presence of mobile devices in the house of preschoolers and their potential in delivering educational content (e.g., Griffith et al., 2019), no studies, to our knowledge, have leveraged the interactive affordances of digital games to grant the users agency in decision-making and provide them with contingent feedback to teach ToM skills.

Educational Games and Learning

Despite the recognition by the educational and scientific communities of ToM as an essential set of social-emotional skills, it is largely overlooked by designers of educational games for young children (Nikolayev et al., 2015) and, along with other social skills, relatively understudied by digital media researchers (Flynn et al., 2019). The few existing studies of preschoolers and ToM focus on relations between ToM abilities and video content (Reiß et al., 2019; Cingel et al., 2020), but not the interactive platforms. Only one recent study, that we are aware of, has examined ToM in relation to interactive gameplay in preschoolers, not as a dependent variable, but rather as a moderator between gameplay and prosocial behaviors (Shoshani et al., 2022). Although many educational apps marketed for preschoolers do not use optimal pedagogical approaches and are not rooted in developmental science (Callaghan and Reich, 2018; Meyer et al., 2021; Nikolayev et al., 2021), a growing body of literature demonstrates that digital apps that employ developmentally appropriate content and design elements have the potential to teach preschool children (ages 3–5) a wide variety of skills (Hirsh-Pasek et al., 2015; Herodotou, 2018; Flynn et al., 2019; Griffith et al., 2020; Kim et al., 2021; Papadakis, 2021b; Callaghan and Reich, 2022) including language (Teepe et al., 2017; Neumann, 2018; Dore et al., 2019; Kirsch, 2021), computational thinking (Papadakis, 2022), and executive function skills (Huber et al., 2018).

Beyond the content and design elements, children's learning from digital games is mediated by contextual factors of play (Guernsey, 2007; Takeuchi and Levine, 2014), such as joint media engagement (Takeuchi and Stevens, 2011). Meaningful adult-child co-play supports and enhances young children's learning from educational apps (Neumann and Neumann, 2014; Radesky et al., 2015; Sweeney, 2017; Neumann, 2018; Rasmussen et al., 2019; Toh and Lim, 2021). In the process of digital co-play, adults scaffold children's learning by engaging them in dialogs and explaining complex concepts (Yelland and Masters, 2007; Bindman et al., 2021), directing children's attention to the specific content and highlighting important information (Sobel et al., 2019), providing affection and encouragement (Yelland and Masters, 2007; Wood et al., 2016), and helping with technical and physical tasks, such as logging in, typing, and touching the screen (Reich et al., 2012; Wood et al., 2016). However, studies have found that joint media engagement does not require adults to be co-playing the game for it to be beneficial for learning (Eisen and Lillard, 2020; Musick et al., 2021). For example, Reich et al. (2012) observed children narrating digital gameplay and explaining their choices to playmates who were not actively gaming, and Bers (2020) as cited in Papadakis (2021a) demonstrates how collaboration when learning coding promotes not only computational thinking, but language and social-emotional skills.

Digital Games as Social Partners

Although digital co-play is believed to be instrumental for preschoolers and recommended by experts like the American Academy of Pediatrics (AAP Council on Communications and Media, 2016), children often engage in touchscreen media independently (see Ewin et al., 2021 for review). Promisingly, researchers are increasingly finding that digital games can function as social partners or "more knowledgeable others" to young players (e.g., Richert et al., 2011; Xu et al., 2021). Similarly, Xu et al. (2021) and Russo-Johnson et al. (2017) have demonstrated the ways in which interactivity *via* an app or conversational agent can support vocabulary learning and story comprehension, suggesting that interactive technology could

also facilitate ToM abilities. Recent work finds that artificial-intelligent (AI) agents (e.g., Alexa) supports preschoolers' deeper thinking and understanding (Xu et al., 2021). Xu et al. (2021) applied a sociocultural approach to create AI-mediated experiences to support children's language development through dialogic reading. They found that AI-powered conversational agents can indeed function as "more knowledgeable others" and provide the same benefits as dialogic reading with adult human partners.

Flynn et al. (2019), applied play framework by Zosh et al. (2018) to digital spaces and theorized that when design affords a specific type of interactivity, a game can assume the role of a "digital adult" and, in turn, may provide the benefits of adult-child co-engagement. One such type is contingent interactivity, also known as full interactivity in some studies (Peebles et al., 2018), which involves meaningful reciprocal exchanges between the player and the system and includes turn-taking, responsive contingent feedback, and device control. In contingent interactivity, the game assumes the role of a "digital adult" and initiates some activities within the game or directs the play.

A digital home literacy environment, i.e., shared and independent literacy activities using a digital device is a large part of children's everyday life (Segers and Kleemans, 2020). Preschoolers across different social-economic backgrounds benefit from educational app use (Arnold et al., 2021; Rowe et al., 2021) and spend, on average, at least 40 min daily on mobile devices (Rideout and Robb, 2020). From content access to co-engaging in digital use (see Papadakis et al., 2021 for a review), caregivers shape children's interactions with technology and differentially influence learning from educational media. In recent years, studies have examined scaffolding of children's digital learning by adults and the possibility of scaffolding by "digital adults" in the form of contingent interactivity. In this paper, we built on this premise. We designed language for apps to create a socio-linguistic environment to help promote children's ToM development with the support of games' contingent interactivity and real-life adult conversation.

PRESENT STUDY

Drawing on the promise that interactive technology can serve as a social partner, the present study examines whether an interactive, touch screen app that utilizes language known to promote children's ToM skills through face-to-face interventions and real-life interactions could be effective in boosting preschoolers' ToM skills.

Specific research questions were the following:

RQ1: Is there a functional relation between the use of digital apps with ToM -promoting language and children's understanding that people have different desires, beliefs, and knowledge sources?

RQ1a: Does children's understanding that people have different desires, beliefs, and knowledge sources increase when the use of games enriched with ToM-promoting language is followed by an adult-led discussion about the games?

RQ2: Is there a functional relation between the use of digital apps with ToM-promoting language and children's understanding of false belief and knowledge sources?

RQ2a: Does children's false belief understanding increase when the use of games enriched with ToM-promoting language is followed by an adult-led discussion about the games?

MATERIALS AND METHODS

Single-case methods have been commonly employed in special education settings for many years and have been recognized as especially appropriate and valuable for identifying evidence-based practices in education research (Odom and Strain, 2002; Horner et al., 2005; Kratochwill et al., 2021). One of the benefits of a single-case research method is that it allows researchers to respond to individual differences and implement intervention modifications if needed (Ledford and Gast, 2018), making it especially valuable in researching technology-delivered personalized education. A multiple-baseline single-case design was used in this study for several reasons:

1. ToM skills are irreversible thus, withdrawal is not possible.
2. The design makes it possible to measure target responses to multiple assessments.
3. The design allows for control due to developmental maturation, which is important given that children's ToM skills improve with age.

There were three phases (conditions) in this study in the sequence of A—B—B+C, where A was the Baseline phase, B was the Voice-overs treatment phase, and B+C was Voice-overs combined with Discussion (VAD) modified treatment phase. Following the multiple-baseline across participants research logic, children were introduced to treatment in a staggered fashion to ensure that changes in the data patterns were due to the introduction of the treatment and did not have alternative explanations such as the multiple exposures to the assessment procedures or maturation (Ledford and Gast, 2018). In a multiple-baseline study, the experimental control and the functional relation between dependent and independent variables are established when participants' performance (e.g., level, trend, variability of data) changes only after they are introduced to treatment, while the performance of participants in baseline remains unchanged. The present study met the single-case design standards outlined by Kratochwill et al. (2021).

Participants

The participants for this study were selected from a preschool that serves low-income families in a Mid-Atlantic metropolitan area in the United States. Each participant was assigned a unique pseudonym, and all student identifying information was removed to maintain the confidentiality of the participants. Potentially identifying information about the preschool in which

the study was conducted was purposefully eliminated from the description. This research study was reviewed and approved by the Institutional Review Board (IRB) at George Mason University (GMU) to ensure the rights and welfare of the study participants. Parental consent/student assent was obtained prior to the beginning of the study. Permission was also obtained from the principal of the preschool.

For the preliminary selection, preschool director and classroom teachers were asked to identify typically developing children who were between 3 and 5 years of age and fluent in English from the pool of the children whose parents had consented to their child's participation in the research study. Fourteen children (eight girls, 6 boys) were recommended by the teachers for the stud. The first screening involved playing one of the five games without voice-overs and undergoing the Theory of Mind (ToM) assessment. Nine children who failed three or more of the five tasks in the ToM assessment were selected to participate in the intervention study. The data collection happened over the summer, and three children dropped out during the study for attendance reasons: two went on vacation during baseline and one of the phases of the testing, and one was sick for an extended period of time. The final sample included six girls between 42 and 54 months of age (see **Table 1**).

Mia was an only child. Social workers and teachers described Mia as social and talkative, able to explain her feelings, although with some challenges in interpersonal communications. Specifically, Mia seemed to be led by a friend who would not let her play with other children. Isabella came from a large Spanish-speaking, multigenerational family where she was "the baby" of the family. Teachers reported that Isabella appeared to be skilled in having relationships with adults but had trouble making and sustaining friendships with peers. Sienna was a very energetic, happy, and assertive girl. She was an only child of a single mother. Sienna and her mother spoke Spanish exclusively when together. Paula had an older sister, and her family spoke English and Spanish. Teachers described Paula as a well-liked girl who was very calm and patient but was constantly tired. Camilla came from a Spanish-speaking family and had an older brother. She was a friendly child who played well with others and made friends quickly. She could often be observed engaged in group play. Teachers described Emily as quiet and reserved but friendly. Emily usually played with one friend, a quiet girl, who was beginning to learn English. Emily was the youngest of five children and the only girl in her family.

TABLE 1 | Description of participants.

Name	Gender	Ethnicity	Age (in months)	Second language
Mia	Female	African American	49.5	Shona
Isabella	Female	Caucasian	51.5	Spanish
Sienna	Female	Hispanic	46	Spanish
Paula	Female	Hispanic	48.5	Spanish
Camilla	Female	Hispanic	46	Spanish
Emily	Female	Asian	48	Korean

In this study, a second language refers to any language other than English that children are exposed to on an everyday basis.Some participants exclusively spoke another language at home, whereas others only occasionally heard parents speaking another language.

Data Collection Procedures

Children were visited three to five times a week for five weeks, for a total of 19–20 sessions. All children were trained and tested individually in a private room at the preschool. Only one researcher (the first author) collected the data. The researcher brought each child to the testing room (preschool library) and briefly went over the procedures and solicited verbal assent. All sessions were video recorded. Once the child was done with the session, the researcher brought them back to their class. Due to the single-case method, each child went through the three phases of testing in the same order (Baseline, Voice-over, VAD), but started treatment phases in a staggering fashion, e.g., participants 1 and 2 started treatment on the seventh visit, whereas participants 3 and 4 started on the eighth. According to the single case design standards, each phase for each participant should have at least five data points; and it's appropriate to change to the next phase of the intervention when data show stable patterns (Kratochwill et al., 2021). Each session, including Baseline condition and both types of treatment conditions (original and modified VAD), started with the child playing a game on an iPad and concluded with face-to-face assessment procedures. The sessions were numbered, and a specific game was assigned to each session.

Baseline Procedures (Phase I)

Baseline procedures began with participants playing one of the five LEGO® DUPLO® game apps on the iPad while wearing headphones connected to the researcher's computer. The computer volume was muted, and no sounds other than those from the iPad were played to the participants. The researcher sat across from the child at an angle that allowed her to see the child's screen and be available to help with the procedures and technical aspects of the play, but she did not initiate or support discussions about the game plot.

Voice-Over Training Procedures (Phase II)

During the voice-over training procedures, the participant and researcher sat at an angle to each other, each with their own device. The researcher could see the participant's iPad's screen, but the children could not see the researcher's laptop screen. Both the researcher and the participant wore headphones connected to the researcher's computer to hear the verbal component of the game. At the beginning of the first session in the original treatment phase, the researcher explained to the participants that she had forgotten to turn on the "sound" in the games before and that from now on, children would play with the turned-on sound. None of the participants expressed any concern or suspicion that the sound came from the researcher's computer and not the games. The sound in the games did not seem to change children's enjoyment of the game. The child played the game, and the researcher followed the gameplay and started the voice-overs from her laptop at

specific times. The game's original music and sounds were not muted on the iPad so that the gameplay would feel more natural. If a child missed a step, the researcher skipped the accompanying voice-over and introduced the next one at the appropriate time.

Voice-Overs and Discussion (VAD) Training Procedures (Phase III)

The VAD training procedures were identical to the voice-over treatment procedures, with one exception; right after the gameplay and before the assessment activities, the researcher engaged the child in a semi-structured conversation about the game. Depending on the child, each discussion lasted between 3 and 7 min. Children enjoyed the conversations and were eager to participate. After the discussion was over, the researcher started the ToM assessment procedures.

Independent Variable and Materials

Games

Five LEGO® DUPLO® game apps were used: LEGO® DUPLO® Circus, LEGO® DUPLO® Ice Cream, LEGO® DUPLO® ZOO, LEGO® DUPLO® FOREST, LEGO® DUPLO® FOOD®. LEGO® DUPLO® apps are distributed internationally, and as such, use sound effects and background music, rather than words, to be accessible to children who speak different languages. The games have a storyline and a goal for completion (e.g., help deliver a package), and require children to complete several mini games focusing on prosocial behaviors, social interactions, and decision-making.

Independent Variable: Voice-Overs

The voice-overs were designed based on research from the extant literature on verbal interactions that promote ToM. They included: (a) explanatory, causal, and contrastive talk about mental states (e.g., *Everyone in the audience thinks the acrobat may fall from the swing, but she knows she won't.*); (b) an abundance of mental verbs, specifically verbs referring to mental processes (e.g., *think, know, and remember*) that scaffold preschoolers' transition to belief-based thinking, and verbs of desire (e.g., *like, want*) to accommodate younger children who are still transitioning from desire-based to belief-based explanations of behaviors; (c) mental state verbs along with embedded sentential complements structures (e.g., *Bunny and Teddy think there is a green rock, but click on it* [wait for the child to click]—*it is really a turtle!*); (d) explanations of mental states underlying characters' behaviors (e.g., *The driver did not stop to help because he thinks you can put the food away all by yourself*); (e) references to events that occurred earlier in the game (e.g., *Guess what, Giraffe. Remember the Lion didn't see you getting the package? It means he does not know that you have it)*; and (f) mental-state verbs directed at players were incorporated into statements, whereas utterances directed at other characters in the video were incorporated into questions (e.g., *Remember how the squirrel thought the box is full of candy and nuts? It turns out there was a DRUM inside*. vs *Giraffe, do you know what everyone likes?*). The voice-overs were embedded in narration: the narrator made explicit positive assumptions about children's thought process regarding false belief situations presented in the games and commented on children's and characters' performance (e.g., "*You thought these were regular stars, but they are actually musical stars"*). Additionally, voice-overs were included in contingent feedback (e.g., *We think you are such a great builder; you made an awesome forest door!*) and the dialog between the characters (e.g.,—Giraffe*, do you think everyone saw the horse jumping through the fire?—Of course, Bunny, everyone thinks the trick was awesome*).

Voice-overs were recorded using the audio editor Audacity (Audacity Team, https://www.audacityteam.org) and embedded into a presentation slideshow as audio files with captions. Each slide corresponded to a different screen in the game and contained voice-overs for all the possible game scenarios so that the researcher could observe children playing and provide contingent feedback.

Modified Independent Variable: Voice-Overs With Discussion (VAD)

Following the first treatment phase using voice-overs, an additional treatment of voice-overs with follow-up discussions was introduced. Although conversations were structured around the game narrative, the interaction between participant and researcher (the first author) resembled a naturally occurring conversation and were different for each child. The researcher asked questions to help the child reconstruct narrative plots from the games, highlighted and repeated episodes that contained mental-state references and exchanges between characters, and engaged the child in discussing instances of deception and false beliefs. For example, in one session after the child played the LEGO DUPLO Circus game, the researcher began with the initial question, asking the child to recall that the circus came to town in the game, then asked about the characters' expectations about the show and performers and whether these expectations were met. Next, the researcher asked the child about the mental states underlying the characters' behaviors (e.g., *Why did the clown run away? What did he think about the tiger*?), and finally prompted the child to describe the circus audience's thoughts. When children injected their experiences into the conversation, the researcher supported them and then returned to discussing the events in the game. The researcher did not correct the participants if they made mistakes in attributing the false belief or misremembering details. Instead, the researcher prompted children to talk more about the scenario to help them think through the conflicting perspectives between the characters' expectations and reality.

ToM Dependent Variables

ToM was assessed using variations of tasks from the five-item developmental scale created by Wellman and Liu (2004) and the Location Change task developed by Wimmer and Perner (1983). Given that children's performance on False Belief tasks remains consistent across different task presentation formats and different types of tasks (e.g., Wellman et al., 2001; Hasni et al., 2017), to keep participants engaged, half of the tasks were presented

in a digital storybook format (created with iPad drawing apps), and the other half acted out with props. Presentation order was counterbalanced across sessions for task type and format.

Following the Gola's (2012) study, tasks were grouped into two categories. In the first category, three tasks assessed children's understanding that people can have diverse desires, beliefs, and knowledge about the same thing; and in the second category, two tasks assessed False Belief understanding. All tasks corresponded to a progression of milestones in children's development of ToM (the two False Belief Tasks are of similar difficulty), thus, the first category contained conceptually easier tasks than the second one. In the study these categories were used as two separate measures.

Measure 1: Desires, Beliefs, Knowledge. Three tasks assessed children's understanding that people can have different desires, beliefs, and knowledge access. These three skills were judged as either correct or not (1 = correct, 0 = incorrect). The scores were then added together for the Desires, Beliefs, Knowledge score, with the total score ranging from 0 to 3. The Desires, Beliefs, and Knowledge scenarios were presented randomly to prevent children from expecting the same order of items across the sessions.

- In the Diverse Desires task, a child must demonstrate an understanding that someone might have a different desire about the same object. The child is presented with a doll and pictures of two different snacks. The researcher asks for the child's preference of snack and subsequently states that the doll wants a different snack than the one selected by the child. The child is then asked which snack the doll would choose; the child must provide an answer to the target question that is different than what they desire.
- In the Diverse Beliefs tasks, a child must demonstrate an understanding that someone might hold a different belief about the same thing. The child is shown a doll and pictures of a garage and bushes. The child is told that Linda is looking for her cat; the researcher then asks the child where they think the cat is, in the bushes or in the garage. The researcher then says that Linda believes her cat is in a different location than indicated by the child and asks the child where Linda would look for her cat. The child must say the opposite of their belief.
- In the Knowledge Access (Seeing-Knowing) task, a child must correctly judge the knowledge of another person who does not have access to the information available to the child. The child is presented with a small box and asked what they think is in the box. After the child guesses or says that they do not know, the researcher lets the child open the box to see the contents (a Lego piece). The researcher then introduces a doll and says a doll has never seen inside the box and asks whether the doll knows what's inside the box. The child must say "No" to be correct.

Measure 2: False Belief Three tasks, Unexpected Contents False Belief and Explicit False Belief (Wellman and Liu, 2004), and Location Change task (Wimmer and Perner, 1983) were used to measure children's false belief understanding, with two different tasks used per session.

- In the Contents False Belief task, a child must reason how another person might misjudge the contents of a container. The child is provided with a familiar, easily identifiable container (e.g., a box of crayons) and is asked to guess what the contents are. After the child answers, "crayons," the box is opened, and a small wooden hippopotamus is revealed. The researcher then puts the toy back into the box and closes the lid. A doll appears, and the researcher states that the doll has not seen inside the box and asks the child what a doll thinks is in the box. The correct response to the question is "crayons."
- In the Explicit False Belief task, the child must decide where one would look for an object given the one's incorrect belief. The child is presented with a doll and two pictures, one of a backpack and another of a closet. The researcher then explains that a doll is looking for his mittens that are really in his backpack, but he thinks they are in his closet. The researcher then asks where a doll is going to look for his mittens. The correct response is in the "closet."
- In the Location Change Task children must decide where someone will be looking for an object given agent's information about the location. The task is based on a story of character A, who places an object (e.g., a book) in a specific location (e.g., a cabinet) and then leaves. Meanwhile, unbeknownst to character A, character B moves the object to a different location (e.g., a bookshelf), and character A then reappears. The child's task is to identify where character A will look for the object first. To be correct, the child must answer that character A will look in the original location (before the move).

To prevent children from getting used to solving the same type of scenario, each session contained either an Explicit False Belief or the Location Change Task. The tasks were randomly assigned to each session. All together there were two False Belief tasks per session and children received 1 point for each correct answer and could receive 0–2 points overall.

Reliability and Scoring

Procedural Reliability

An independent observer, trained in the procedures before the data collection, monitored session activities and compared them against a preplanned checklist of expected activities. Procedural reliability data were collected for 30% of the data for all participants across all three conditions (Baseline, Voice-overs, VAD). The number of correct actions was then divided by the number of planned actions and multiplied by 100%, yielding procedural reliability of 100%.

Interrater Agreement on ToM Outcomes

To ensure scoring reliability, an independent observer scored 30% of the assessment sessions. The independent observer was a child development professional with an extensive experience in experimental research. Inter-observer agreement was assessed for 33–35% of the observations of Desires, Beliefs, Knowledge (desires, beliefs, knowledge) scores and False Belief (False Belief) scores in the Baseline, Voice-overs, and VAD phases. The Total

Agreement formula was used to calculate interrater agreement; a smaller total of correct answers recorded by each observer was divided by the larger total and multiplied by 100% (Kennedy, 2005). The mean interobserver coefficient of agreement for Desires, Beliefs, Knowledge (desires, beliefs, knowledge) was 92% (range: 87–100%) for all participants. The average agreement for False Belief was 96% (range: 75–100%) for all participants. Thus, in most individual instances as well as in the group averages, the design standard for inter-rater agreement was met (Kratochwill et al., 2021).

Analytic Plan

Visual analysis of graphed data was used to examine the functional relation between voice-overs in the games, combination of VAD, and changes in participants' performance on Desires, Beliefs, Knowledge and False Belief (the latter two values are the scores per session). Specifically, we used the following procedure outlined by Kratochwill et al. (2013) to visually examine within and across phases changes in a) level (mean of all data points within the phase), b) trend (direction of the data slope), c) data variability (instability of data), d) immediacy of effect (degree of change from the last data point in one phase and the first data point in the next phase), and e) an index of data overlap between phases – Non-overlap of All Pairs (NAP) (Parker and Vannest, 2009; Manolov et al., 2016)—for each participant's Desires, Beliefs, Knowledge and False Belief data. The visual analysis allows for determination of (a) evidence of a functional relation between dependent and independent variables; as well as (b) the magnitude of that relation (Kratochwill et al., 2021). The decision is based on the changes within- and between-phases on six components of the visual analysis. NAP was also used to calculate the percentage of data that improved across participants for each measure. According to Kratochwill et al. (2021), there is strong evidence of a functional relation if at least three demonstrations of an effect are present at different time points; moderate evidence of a functional relation if at least three demonstrations of an effect are present with at least one demonstration of a non-effect; and no evidence of a functional relation if there are not at least three demonstrations of an effect.

Non-overlap of All Pairs

Though different indices of overlap exist, several studies (Manolov et al., 2011; Parker et al., 2011) demonstrate an advantage of a NAP index. It is derived from a non-parametric assessment procedure that involves individual comparison of all A to B data points and provides a percentage of all non-overlapping data points. NAP is appropriate for many different data types and distributions and is less susceptible to outliers than some other indices of data overlap (Parker and Vannest, 2009). The NAP is equivalent to the Mann–Whitney *U* statistic and ranges from 0 to 1, with 0.50 indicating a null effect of the treatment or a complete overlap between the baseline and intervention phases (Mann and Whitney, 1947; Parker and Vannest, 2009; Michiels et al., 2018). Values above or below 0.50 indicate improvement or regress in performance in the treatment phase in comparison to the baseline, with increasing degrees of non-overlap (Parker et al., 2011; Berrett and Carter, 2018). We calculated nap with an online calculator available at http://www.singlecaseresearch.org/calculators/nap (Vannest et al., 2011).

RESULTS

Diverse Desires, Diverse Beliefs, Knowledge Access Skills

Based on the changes in such components of the visual analysis (level, trend, variability, overlap, and consistency), two of the six participants increased their Desires, Beliefs, Knowledge scores during the Voice-over phase as compared to the Baseline, and all participants increased their scores during the VAD phase as compared to Baseline (see **Figure 1**). The visual analysis of Desires, Beliefs, Knowledge data demonstrated no evidence of the functional relation between Voice-overs training and improvements in Desires, Beliefs, Knowledge skills (Kratochwill et al., 2021) and moderate evidence of the functional relation between VAD treatment and improvements in Desires, Beliefs, Knowledge skills development. Mean Non-overlap of All Pairs (NAP) across participants was calculated to be 0.66 for Voice-overs phase and 0.87 for VAD phase. Individual Desires, Beliefs, Knowledge results for both treatments are described subsequently.

Mia

In all six Baseline sessions, Mia's data demonstrated a low level (M = 1.50, SD = 0.54), accelerating trend, and moderate levels of variability (see **Figure 1**). Upon introduction of the Voice-overs training, Mia showed a small increase in level (from Baseline M = 1.50, SD = 0.54 to Voice-overs Training M = 1.8, SD = 0.84), no immediacy of effect, a flat trend, and high variability of data (**Figure 1**). NAP for Mia's Desires, Beliefs, Knowledge data was calculated to be 0.60 from Baseline to Voice-overs phase. In response to the VAD phase, Mia's scores increased from Baseline (M = 1.50, SD = 0.55) to VAD (M = 2.33, SD = 0.71), with almost half of the answers at the ceiling level (**Figure 1**). Desires, Beliefs, Knowledge data in the VAD phase had an upward trend, high variability, and no immediacy of effect. There was also an increase in level and change in trend: from the Voice-overs phase level (M = 1.8, SD = 0.83) to VAD phase level (M = 2.33, SD = 0.70), and from the flat trend in Voice-overs to an upward trend in the VAD. NAP for Mia was calculated to be 0.81 from Baseline to VAD phase.

Isabella

Across the six Baseline sessions, Isabella had mid-range scores (M = 1.83, SD = 0.4), with a flat trend and low variability of data. With the implementation of the Voice-overs training, Isabella's Desires, Beliefs, Knowledge data showed a small change in level from Baseline (M = 1.83, SD = 0.41) to Voice-overs phase (M = 2, SD = 0), no immediacy of effect, flat trend, and absence of variability (**Figure 2**). A NAP of 0.58 was calculated from the Baseline to Voice-overs treatment phase.

In the VAD phase, Isabella showed above-baseline performance with level increase from Baseline (M = 1.83, SD = 0.4) to VAD (M = 2.22, SD = 0.66), an accelerating trend,

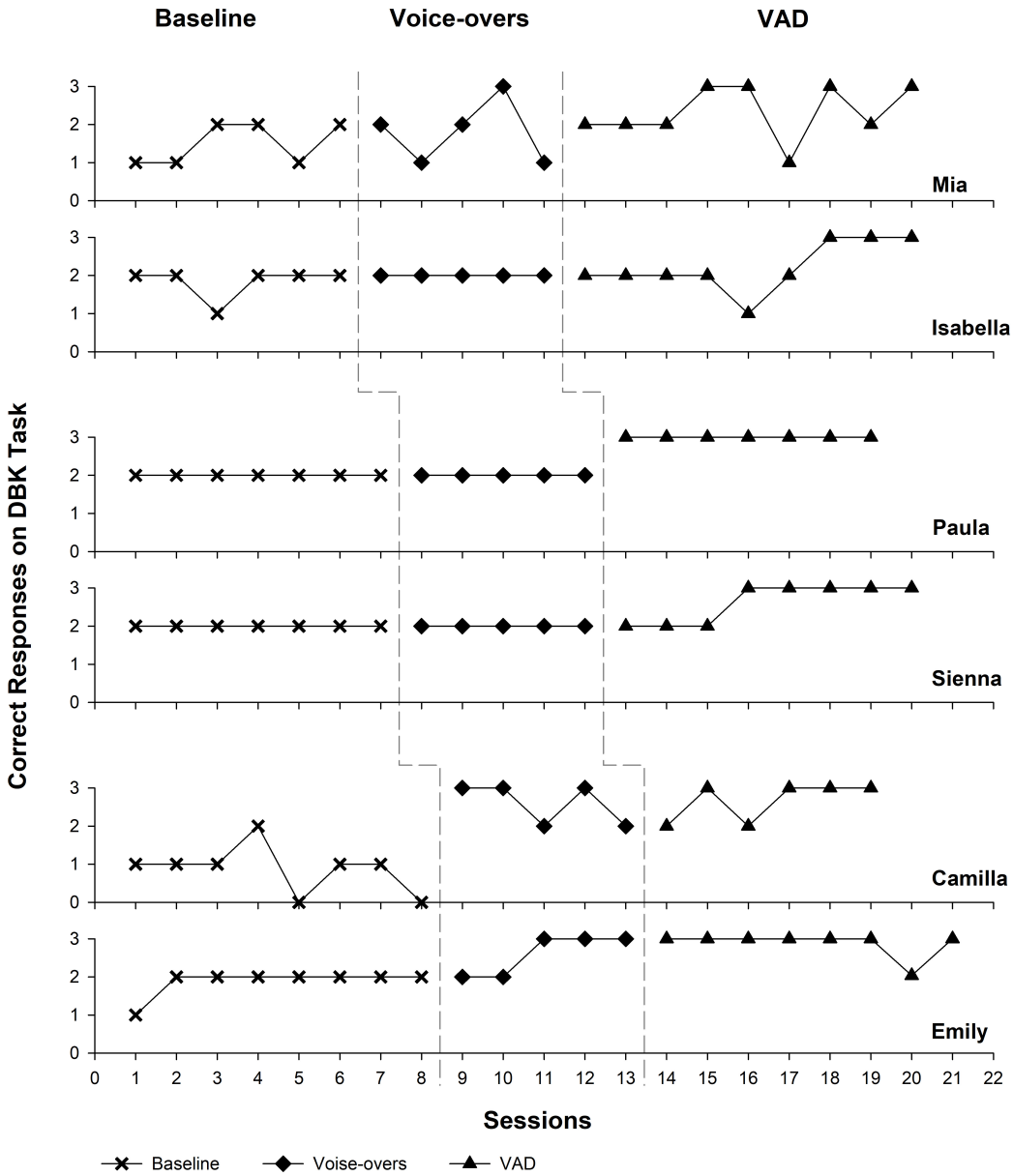


FIGURE 1 | Accuracy of responses to the three tasks in Desires, Beliefs, Knowledge measure by participants across the research phases.

and moderate variability of data (**Figure 2**). Isabella's level increased only slightly from the Voice-overs phase (from Voice-overs phase $M = 2.00$, SD = 0 to VAD phase $M = 2.33$, SD = 0.66) and showed no immediacy of effect. The greatest amount of change was observed in the trend direction that improved from being flat in the Voice-overs phase to accelerating in the VAD phase. A NAP of 0.67 from the Baseline to the VAD treatment phase was calculated for Isabella.

Paula

In seven Baseline sessions, Paula consistently solved two, Diverse Desires and Diverse Beliefs, of the three Desires, Beliefs, Knowledge questions correctly (see **Figure 1**), but not the Knowledge Access questions. Her data showed medium level ($M = 2.00$, SD = 0), no variability, and flat trend. There was no change in Paula's response to the implementation of the Voice-overs phase: level remained the same ($M = 2.00$, SD = 0), no immediacy of effect was observed, there was the absence of variability and a flat trend also remained. A Desires, Beliefs, Knowledge NAP of 0.50 from Baseline to Voice-overs treatment phase was calculated.

Upon the implementation of VAD training, Paula demonstrated a rise in level from Baseline ($M = 2.00$, SD = 0) to VAD ($M = 3.00$, SD = 0), immediacy of effect, flat trend, and no variability of data. In other words, Paula immediately reached the ceiling in her responses and remained there for all seven VAD

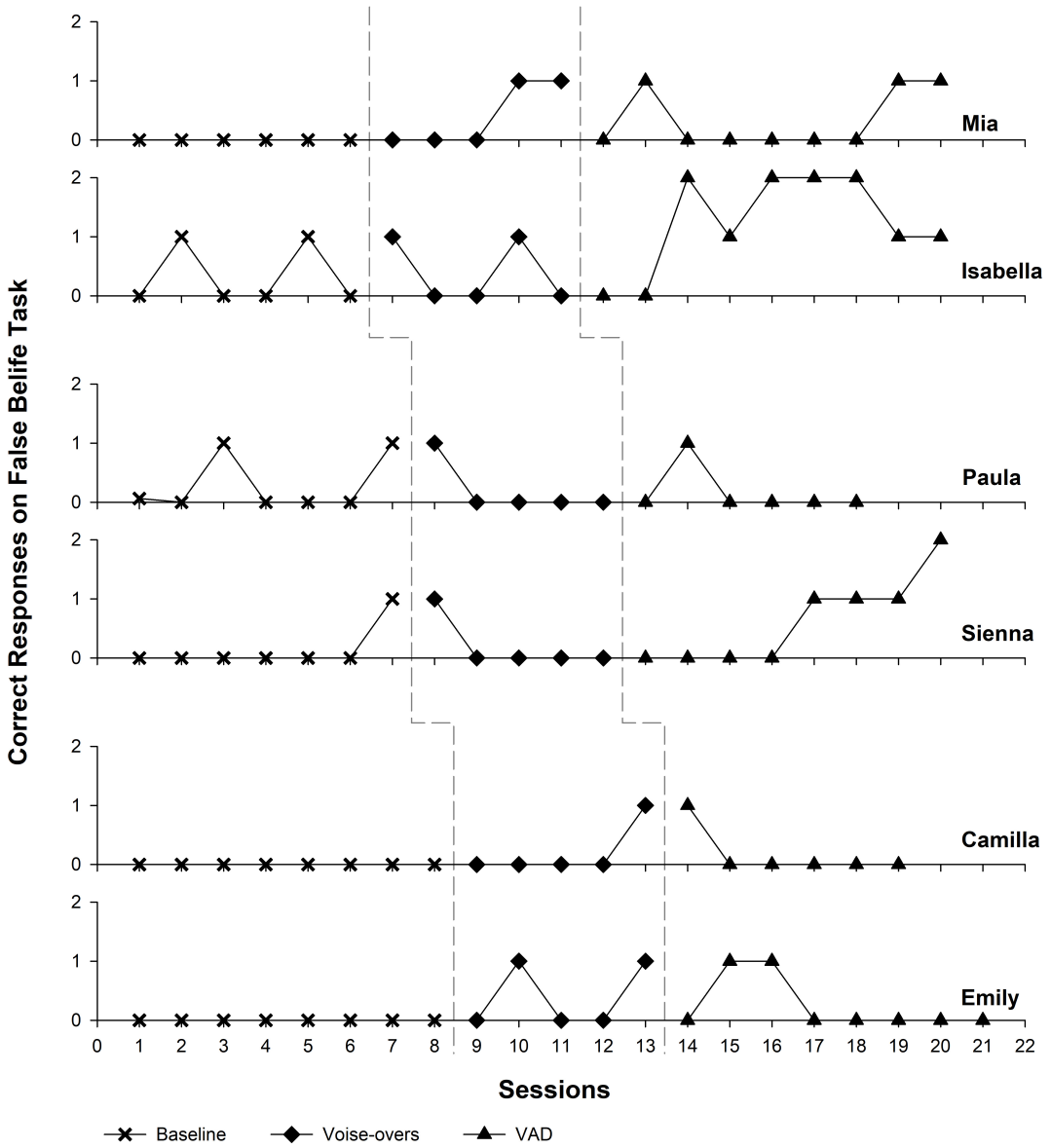


FIGURE 2 | Accuracy of responses to the two tasks in False Belief measure by participants across the research phases.

training sessions. A NAP of 1.00 from Baseline to VAD phase was calculated for Paula's Desires, Beliefs, Knowledge data.

Sienna

During seven Baseline sessions, Sienna consistently answered two questions correctly, Diverse Desires and Diverse Beliefs (see **Figure 1**) showing mid-level scores ($M = 2.00$, SD = 0), flat trend, and no variability. Sienna did not answer the Knowledge Access questions correctly, and her Desires, Beliefs, Knowledge showed no change in response during implementation of the Voice-overs phase; level remained the same ($M = 2$, SD = 0), as did the absence of variability and a flat trend. A NAP of 0.50 from the Baseline to Voice-overs treatment phase was calculated for Desires, Beliefs, Knowledge data.

In the VAD treatment phase, Sienna's data (**Figure 1**) had a increase in level from Baseline ($M = 2$, SD = 0) to VAD ($M = 2.63$, SD = 0.51), no immediacy of effect, a steep accelerating trend, and a moderate variability of data. Since Sienna's performance on the Desires, Beliefs, Knowledge measure was identical during the Baseline and Voice-overs phases, the same changes in level, trend, and data variability were observed from Baseline to VAD phases and from Voice-overs to VAD phases. A NAP 0.81 from Baseline to VAD phase was calculated on Sienna's Desires, Beliefs, Knowledge measure.

Camilla

Throughout eight Baseline sessions, Camilla had low scores (M = 0.88, SD = 0.64), with data showing a downward trend and moderate variability. Upon the introduction of the Voice-overs phase, Camilla's data (**Figure 1**) demonstrated an increase in level: from Baseline (M = 0.88, SD = 0.64) to Voice-overs Training (M = 2.60, SD = 0.55), immediacy of effect, downward trend, and moderate variability. NAP was 0.98 for Desires, Beliefs, Knowledge data from Baseline phase to Voice-overs treatment phase.

In response to the VAD training, Camilla performed at above-baseline levels [level changed from Baseline (M = 0.88, SD = 0.64) to VAD phase (M = 2.67, SD = 0.52)], showing an upward trend and moderate variability (**Figure 1**). In most of the VAD phase sessions, Camilla performed at ceiling levels on the Desires, Beliefs, Knowledge measure. Only minor changes in Camilla's VAD data were observed in comparison to the Voice-overs phase. There was almost no increase in level (from Voice-overs phase M = 2.6, SD = 0.54 to VAD phase M = 2.67, SD = 0.52), no immediacy of effect, a change in the trend from downward to upward, and less data variability. NAP from Baseline to VAD phase was 0.98.

Emily

During eight Baseline sessions, Emily's Desires, Beliefs, Knowledge data (**Figure 1**) were consistently mid-level (M = 1.88, SD = 0.35), demonstrating low variability and a slight upward trend. Specifically, Emily consistently responded correctly to two questions on Diverse Desires and Diverse Beliefs, but not the Knowledge Access task. Upon introduction of Voice-overs phase, Emily's Desires, Beliefs, Knowledge data (**Figure 1**) demonstrated increase in level: from Baseline (M = 1.87, SD = 0.35) to Voice-overs Training (M = 2.60, SD = 0.55), no immediacy of effect, steep accelerating trend, and moderate variability of data. NAP of 0.83 was calculated on Desires, Beliefs, Knowledge measure from Baseline to Voice-overs Training.

During the VAD phase, Emily's Desires, Beliefs, Knowledge data (**Figure 1**) demonstrated a rise in level from Baseline (M = 1.87, SD = 0.35) to VAD (M = 2.87, SD = 0.35), no immediacy of effect, a slightly downward trend driven by an outlier, and low variability. Emily almost always responded correctly to all three questions, except for one session. There was an increase in level from the Voice-over phase (M = 2.60, SD = 0.54) to the VAD phase (M = 2.87, SD = 0.35), less variability of data in the VAD phase, and change in the trend from steep upward to slightly downward. NAP of 0.95 was calculated on Desires, Beliefs, Knowledge measure from Baseline to VAD.

False Belief

False belief was assessed by two false belief tasks per session from a False Belief measure; each scored as either correct (1) or not (0). None of the six participants demonstrated improvement in false belief understanding in the Voice-overs phase. Visual analysis of False Belief data found no evidence of the functional relation between voice-overs training and children's false belief skills, and the mean NAP across participants was 0.63 for the Voice-overs phase. Only two participants showed improvement in the VAD phase and the mean NAP across participants was 0.59 for the VAD phase. Since fewer than three demonstrations of an effect were found by the visual analysis, per Kratochwill et al. (2021), we concluded no evidence of the functional relation between VAD treatment and early ToM skills development. Individual False Belief results for both treatment phases are presented in **Figure 2** and described below.

Mia

Mia scored 0 on all False Beliefs tasks in Baseline (**Figure 2**). Upon introduction of the Voice-overs treatment, Mia's data demonstrated some increase in level from M = 0, SD = 0 to M = 0.40, SD = 0.55, emergence of steep upward trend, moderate variability of data, and no immediacy of effect. NAP of 0.70 was calculated for Mia's False Belief data from Baseline to Voice-overs Phase.

Visual analysis did not indicate a considerable change in Mia's False Belief performance in the VAD phase (**Figure 2**) from the Baseline performance. There was a small increase in level as compared to Baseline phase (M = 0, SD = 0) to VAD (M = 0.33 SD = 0.50); there was no immediacy of effect, no obvious trend emerged in the VAD phase, and data showed moderate variability. In comparison to the Voice-overs phase, there was a small drop in level (from Voice-overs phase M = 0.40, SD = 0.55 to VAD phase M = 0.33 SD = 0.50), a change in the trend from upward to flat. Mia's NAP for False Belief was 0.67.

Isabella

Isabella's False Belief data (**Figure 2**) was at a low level with a Mean of 0.33 (SD = 0.52), showed no distinct trend, and had moderate variability. During the Voice-overs phase, Isabella's False Belief performance data remained at the low level (M = 0.40, SD = 0.55), showed no immediacy of effect, had no pronounced trend, and showed moderate variability of data. An NAP of 0.53 was calculated for Isabella's False Belief data from Baseline to Voice-overs treatment phase.

In the VAD phase, Isabella's False Belief data (**Figure 2**) showed a rise in level from Baseline (M = 0.33, SD = 0.52) to VAD (M = 1.22, SD = 0.83), no immediacy of effect, steep accelerating trend, and high variability. In a similar fashion, Isabella's False Belief data showed a rise in level from the Voice-overs phase (M = 0.4, SD = 0.54) to VAD (M = 1.22, SD = 0.83) and an emergence of upward trend. A False Belief NAP of 0.80 from Baseline to VAD phase was calculated for Isabella's False Belief data.

Paula

During the Baseline, Paula's False Belief data (**Figure 2**) was at a low level (M = 0.29, SD = 0.49) across seven sessions, and had no distinct trend, as most of Paula's False Belief scores were 0 with two spikes, when she correctly answered to one of the two False Belief tasks. In the Voice-overs treatment phase, Paula's False Belief data remained at low levels (M = 0.20, SD = 0.45), showed no immediacy of effect, exhibited a downward trend,

and had low variability. A NAP of 0.46 from Baseline to Voice-overs treatment phase was calculated for Paula's False Belief data.

Upon introduction of VAD training, Paula's False Belief data (**Figure 2**) showed no considerable change in comparison to Baseline or Voice-overs treatment phases. The data remained at low levels (M = 0.17, SD = 0.41), had no distinct trend, and variability stayed low. A NAP of 0.44 from Baseline to VAD phase was calculated for Paula's False Belief data.

Sienna

Sienna's False Belief Baseline data (**Figure 2**) showed a low level (M = 0.14, SD = 0.38), a slightly upward trend, and low variability of data across the seven sessions. Like Baseline condition, Sienna's False Belief data in the Voice-overs condition was at a low level (M = 0.20, SD = 0.45), showed no immediacy of effect, demonstrated a slightly downward trend, and exhibited low variability. An NAP of 0.53 from Baseline to Voice-overs treatment phase was calculated.

Upon the introduction of VAD training, Sienna's False Belief data (**Figure 2**) showed an increase in level when comparing Baseline (M = 0.14, SD = 0.37) to VAD (M = 0.62, SD = 0.74), no immediacy of effect, steep accelerating trend, and moderate variability of data. In comparison to the Voice-overs phase, Sienna's False Belief data also increased in level from Baseline (M = 0.2, SD = 0.44) to VAD (M = 0.62, SD = 0.74), and trend direction changed from downward to upward. An NAP of 0.66 from Baseline to VAD phase was calculated for Sienna's False Belief data.

Camilla

Camilla scored 0 on all tasks in eight sessions of the Baseline phase (**Figure 2**). During the Voice-overs phase Camilla's data slightly increased in level from Baseline (M = 0, SD = 0) to Voice-overs Training (M = 0.20, SD = 0.45), showed no immediacy of effect, had an upward trend due to one correct answer in the last session, and demonstrated low variability. NAP of 0.60 was calculated for Camilla's False Belief data from Baseline to Voice-overs Training.

Camilla's False Belief data (**Figure 2**) showed only minor changes during the VAD phase; observed were a slight increase in level from Baseline (M = 0, SD = 0) to VAD (M = 0.17, SD = 0.41), no immediacy of effect, and a downward trend that was due to one correct answer in the first session of the phase and incorrect answers in all other sessions. NAP of 0.58 from Baseline to VAD phase was calculated for Camilla's performance on False Belief.

Emily

Emily scored 0 on all False Beliefs tasks in all eight sessions of Baseline (**Figure 2**). During the Voice-overs phase Emily showed a slight increase in level from Baseline (M = 0, SD = 0) to Voice-overs Training (M = 0.40, SD = 0.55), no immediacy of effect, no distinct trend, and moderate variability of data (**Figure 2**). NAP of 0.70 was calculated for Emily from Baseline to Voice-overs Training.

Once VAD was introduced, Emily's False Belief data (**Figure 2**) showed a slight increase in level from Baseline (M = 0, SD = 0) to VAD (M = 0.25, SD = 0.46), no immediacy of effect, and no distinct trend. In comparison to Voice-overs phase, Emily's data showed some drop in level; from Voice-overs Training (M = 0.40, SD = 0.55) to VAD (M = 0.25, SD = 0.46) the trend changed from upward to downward. NAP of 0.63 from Baseline to VAD phase was calculated for Emily's False Belief data.

DISCUSSION

The purpose of this study was to investigate whether apps for preschoolers, enhanced with ToM-promoting language, could help accelerate the development of children's ToM skills when played on their own or when paired with a with a follow-up adult-led conversation. Visual analysis showed that apps with voice-over-enhancements promoted the development of children's earlier-emerging ToM skills when the voice-over play was also followed by a discussion (VAD condition). Voice-overs without discussion were not effective in accelerating the earlier-emerging ToM skills. The study was not effective in promoting children's false belief understanding.

All six children improved on the three earlier-developing skills (diverse desires, diverse beliefs, and knowledge access (Desires, Beliefs, Knowledge) in VAD conditions indicating a positive conceptual change in social-cognitive understanding. During the baseline phase, no participant showed a conceptual understanding of knowledge access. From this, we can conclude that participants' improvements appear to be due to the conceptual insights gained from our training rather than participant maturation for two reasons. First, a multiple-baseline design allowed for the control of maturation with children starting the intervention at different points in a staggered fashion, and no improvements in ToM understanding were observed prior to the training implementation (during baseline phase) for any child. Second, in the natural course of development, typically developing (TD) children tend to master knowledge access tasks at 53.4 months of age (Wellman et al., 2011), taking on average 3–6 months to progress from understanding diverse beliefs to understanding knowledge access tasks (Rhodes and Wellman, 2013). By comparison, at the end of the study, all children were younger than 53.4 months (M = 48.25, ranging from 47 to 52.5 months of age), and all had advanced to knowledge access mastery in just 2–3 weeks. Thus, it is reasonable to assume that the VAD condition helped children improve earlier-developing ToM skills.

No children showed improvement in false belief higher-order ToM tasks during any phase. These findings are not surprising: ToM development is a sequential progression of conceptual achievements and requires children to master less sophisticated concepts first to achieve more complex social cognitive understanding later (Wellman and Liu, 2004; Rhodes and Wellman, 2013). Our findings align with previous studies that found pre-test performance on knowledge access tasks to correlate with children's improvement on false belief training (Benson et al., 2013; Rhodes and Wellman, 2013). Gola (2012), whose ToM video training was effective at enhancing false belief

tasks, but not tasks related to earlier-emerging diverse desires, beliefs, and knowledge (Desires, Beliefs, Knowledge), reasoned that because the participants performed near mastery levels on the Desires, Beliefs, Knowledge baseline assessment; they already had the necessary foundation and built upon it to achieve false belief understanding. In contrast, our participants scored low during the baseline testing and thus showed improvement on diverse desires, beliefs, and knowledge access that precede false belief skills.

Although all children benefited from the combination of Voice-overs in the games and follow-up discussions for Desires, Beliefs, Knowledge, only two showed improvements under Voice-overs only condition. Several explanations for these findings are possible. Children may have individual needs in terms of language development required for ToM progress: some may have required minimal support for ToM skills (Desires, Beliefs, Knowledge only) development and thus demonstrated improved performance after being exposed to mental-state vocabulary in games. In contrast, others may need more support to achieve the same results and thus benefitted from game-based discussions. Possibly, children who improved in the VAD condition built their understanding over time, benefiting from both being exposed to voice-overs and then to VAD. Finally, VAD effectiveness in comparison findings underscores the benefit of a conversational partner in cultivating these skills, and the high need for interactivity.

Interactive language-based traditional apps in our study were not successful in promoting ToM skills without a follow-up discussion. It could be that the interactivity in the games was insufficient for the ToM development. More research is needed on the types and levels of interactivity that could act as "digital adult" in supporting ToM development. It is also possible that contingent interactivity may be sufficient to promote other social-emotional skills, such as emotion recognition and social skill literacy (Craig et al., 2016; Peebles et al., 2018), emotion regulation (Craig et al., 2016; Rasmussen et al., 2019), prosocial behaviors (Shoshani et al., 2022) and social self-efficacy (Craig et al., 2016), but not ToM, given the importance of active use of the mental state language for ToM development found in some studies (Grazzani and Ornaghi, 2011; Ornaghi et al., 2011; Guajardo et al., 2013).

Extant research on preschool learning from digital devices frequently finds a greater benefit when adults support digital use than use alone by children (Reich et al., 2016; Neumann, 2020). For instance, studies of eBook reading find greater learning from these devices when facilitated by an adult (e.g., Neumann and Neumann, 2014). Further, joint media engagement, involving adults and young children, tends to increase learning (Dore and Zimmermann, 2020). Similarly, we found that voice-over app play was effective in supporting early ToM development when it was coupled with adult conversation.

Designing for Parent-Child Co-engagement

The results of this study add to the mounting evidence of the benefits of joint media engagement with digital games for children's learning, specifically, the conversations that happen during and after the gameplay (e.g., Sobel et al., 2019; Eisen and Lillard, 2020; Musick et al., 2021). Given that conversations during gameplay are not always possible, some researchers propose that rather than expecting parents to join in digital play, it may be more practical to design games that would encourage parents to initiate post-play discussion and foster discussion of the experience during game play (Farber, 2021; Musick et al., 2021). Among the ideas to promote and improve the quality of the conversations are including conversational prompts about the games as often done in children's TV shows, providing conversation-starter guides, and designing games to support parents to act as the cheerleaders and spectators (Musick et al., 2021). Further, in mystery-solving games or treasure hunts, children can find clues by figuring out and explaining characters' false beliefs to their adult partners. Apps with "social" settings or adventure games like LEGO® DUPLO® used in this study could allow players to record and modify voice-overs to narrate the story or role-play the characters using mental state verbs and ToM-enhancing sentence structures.

A growing number of designers and researchers are building technology to enable parental support of children's learning. For example, the work by Stuckelman et al. (2021) demonstrates how an interactive app can model and encourage parents' dialogic reading and discussions. A team of Harvard researchers, in collaboration with a public media producer and educational media developer, have created a series of early literacy apps to encourage child-parent conversations and interactions and, as a result, promote children's vocabulary development and literacy skills (Rowe et al., 2021; Harvard Graduate School of Education, 2022). Lastly, newly emerging platforms, such as Amazon Glow, are being built specifically with co-engagement in mind (Amazon, 2021).

Multiple-device games could be designed for adult-child strategy building that requires mental state verbs and post-game online celebration of the wins to reminisce about the experience. Further, ToM-enhancing language embedded in game content could allow the parents to draw on and learn from specific language. This could also include discussion prompts to help children transfer and further improve ToM skills beyond the gameplay context. Finally, in addition to promoting child-parent media co-engagement, future studies should continue focusing on digital "knowledgeable others" to combat SES-related disparities in child language skills and ToM understanding. This might take the form of AI-conversation partners, such as in the study by Xu et al. (2021) on "dialogic reading" but be programmed to promote the use of mentalistic language present a promising avenue for future research. As the development and accessibility of AI, interactive digital platforms, intelligent agents, and multiplayer devices grow, so do the opportunities to use them in helping to shape children's social-linguistic environments and influencing ToM skills.

Limitations

Although the functional relation between children's understanding that people have different desires, beliefs, and knowledge sources and the use of games enriched with ToM-promoting language is followed by an adult-led conversation about the games was established, these results cannot be generalized to larger populations due to the nature of single-case research (Ledford and Gast, 2018). Additionally, the current study's design did not allow for detecting and quantifying unique contributions of different linguistic elements in the voice-overs, and individual contributions of different interactive components. Finally, whereas single-case design does not require a control group because of a baseline condition for each participant, the study could have benefited from participants who did not undergo any training or did the Voice-over or VAD phase only. Doing so would further demonstrate the absence of the maturation effect in children's performance. More research is warranted, as some of these questions would best be examined in the context of a group study.

Additional limitations concern the sample of the study. First, our sample consisted only of girls, which could be a limitation as there is evidence that girls develop Theory of Mind skills (Blijd-Hoogewys and van Geert, 2017) and some language skills (Bornstein et al., 2004) earlier than boys do. Second, all the participants in the study were exposed to a language other than English at home. Previous research suggests a positive effect of bilingualism on the rate of ToM development (see Schroeder, 2018 for a review). This study did not control for participants' mastery of a second language, and we can't say whether it had contributed to the outcome.

Further, time constraints did not allow for the implementation of the maintenance phase, which is limiting because the maintenance of participant's knowledge gains remains unknown. Moreover, even though each phase of the study met the single-subject evidence standards (Kratochwill et al., 2021), a more extended data collection period would have allowed for more sessions that could potentially allow children to improve ToM skills through repeated exposure to voice-over enriched games or VAD. A longer data collection period could have also allowed children to have more control of the procedures, such as choosing which games to play or how many to play per session. Letting children control some procedures would more closely resemble real-life gameplay.

Lastly, this study did not directly compare Voice-over and VAD methods with other strategies for facilitating ToM development, which limits our ability to conclude whether there are advantages of using games to promote ToM skills. Instead, it can be established that VAD could be an effective option available for parents and teachers to promote mental state understanding in children or content designers developing educational games to teach ToM skills. The ability to generalize the effectiveness of these results requires further investigation.

CONCLUSION

This study is the first to our knowledge to explore how educational digital apps can support children's ToM development—skills that underlie, among others, perspective taking, prosocial behaviors, and academic achievement. Our findings indicate that ToM-promoting language that is effective in face-to-face settings can be successfully implemented in digital games, especially if an adult-led conversation follows. Gameplay coupled with an adult-led conversation resulted in ToM learning, unlike gameplay alone. Although all typically developing children master ToM skills with time, there are advantages to achieving conceptual understanding sooner than later. Our findings suggest that embedding ToM language within a digital game is associated with quicker development of early ToM skills. Such results are promising, as digital games, including well-designed games are popular and may meaningfully improve children's Theory of Mind skills. To help translate the study results into practice, we have provided suggestions on how to leverage mobile apps for preschoolers to create socio-linguistic environments that promote ToM development.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Review Board (IRB) at George Mason University (GMU). The IRB at GMU reviewed all the methods and procedures to ensure the rights and welfare of the study participants. Permission was also obtained from the principal of the preschool. Written informed consent to participate in this study was provided by the participants' legal guardians/next of kin.

AUTHOR CONTRIBUTIONS

MN, SR, AE, and KC contributed to the conception and design of the study. AE suggested the methodology for the study and consulted on data analysis. MN, AE, and KC developed data collection procedures. MN collected the data. KC helped with game selection and prototype development. MB reviewed and consulted on the language for voice-over and solicited teachers' feedback. MN and SR wrote the first draft of the manuscript. All authors reviewed the manuscript, read, and approved the submitted version.

REFERENCES

AAP Council on Communications and Media (2016). Media and young minds. *Pediatrics* 138:e20162591.

Amazon (2021). *Introducing Amazon Glow—A Whole New Way to Bring Families Together for Fun and Learning [Press Release]*. Available online at: https://press.aboutamazon.com/news-releases/news-release-details/introducing-amazon-glow-whole-new-way-bring-families-together [Accessed September 28, 2021].

Amsterlaw, J., and Wellman, H. M. (2006). Theories of mind in transition: a microgenetic study of the development of false belief understanding. *J. Cogn. Dev.* 7, 139–172. doi: 10.1207/s15327647jcd0702_1

Apple (2021). *Apps for Education*. Available online at: https://www.apple.com/education/products/#learning-with-ipad [Accessed January 7, 2021].

Arnold, D. H., Chary, M., Gair, S. L., Helm, A. F., Herman, R., Kang, S., et al. (2021). A randomized controlled trial of an educational app to improve preschoolers' emergent literacy skills. *J. Child. Media* 15, 457–475.

Astington, J. W. (2003). "Sometimes necessary, never sufficient: false-belief understanding and social competence," in *Individual Differences In Theory Of Mind: Implications For Typical And Atypical Development*, eds B. Repacholi and V. Slaughter (New York, NY: Psychology Press), 13–38.

Astington, J. W., and Baird, J. (2005). "Introduction: why language matters," in *Why Language Matters For Theory Of Mind*, eds J. W. Astington and J. Baird (New York, NY: Oxford University Press), 3–25.

Benson, J. E., Sabbagh, M. A., Carlson, S. M., and Zelazo, P. D. (2013). Individual differences in executive functioning predict preschoolers' improvement from theory-of-mind training. *Dev. Psychol.* 49, 1615–1627. doi: 10.1037/a0031056

Berrett, A. N., and Carter, N. J. (2018). Imagine math facts improves multiplication fact fluency in third-grade students. *J. Behav. Educ.* 27, 223–239.

Bers, M. U. (2020). *Coding as a Playground: Programming and Computational Thinking in the Early Childhood Classroom*. New York, NY: Routledge, doi: 10.4324/9781003022602

Bindman, S. W., Sobel, K., Behbakht, S. M., Divanji, R. A., Shahn, E., and Scanlon, M. (2021). ""Let's draw chocolate milk together": supports and challenges for families' joint engagement with artistic tablet apps," in *Proceedings of the Interaction Design and Children (IDC'21)*, Athens, 183–195.

Blackwell, C. K., Lauricella, A. R., Wartella, E., Robb, M., and Schomburg, R. (2013). Adoption and use of technology in early education: the interplay of extrinsic barriers and teacher attitudes. *Comput. Educ.* 69, 310–319. doi: 10.1016/j.compedu.2013.07.024

Blijd-Hoogewys, E., and van Geert, P. L. (2017). Non-linearities in theory-of-mind development. *Front. Psychol.* 7:1970. doi: 10.3389/fpsyg.2016.01970

Bornstein, M. H., Hahn, C. S., and Haynes, O. M. (2004). Specific and general language performance across early childhood: Stability and gender considerations. *First Lang.* 24, 267–304.

Bosacki, S., Moreira, F. P., Sitnik, V., Andrews, K., and Talwar, V. (2020). Theory of mind, self-knowledge, and perceptions of loneliness in emerging adolescents. *J. Genet. Psychol.* 181, 14–31. doi: 10.1080/00221325.2019.1687418

Callaghan, M. N., and Reich, S. M. (2018). Are educational preschool apps designed to teach? An analysis of the app market. *Learn. Media Technol.* 43, 280–293.

Callaghan, M. N., and Reich, S. M. (2022). "Applying a developmental lens to educational game designs for preschoolers," in *Research Anthology on Developments in Gamification and Game-Based Learning* (Hershey, PA: IGI Global), 471–487.

Caputi, M., Lecce, S., Pagnin, A., and Banerjee, R. (2012). Longitudinal effects of theory of mind on later peer relations: the role of prosocial behavior. *Dev. Psychol.* 48, 257–270. doi: 10.1037/a0025402

Charness, G., List, J. A., Rustichini, A., Samek, A., and Van De Ven, J. (2019). Theory of mind among disadvantaged children: evidence from a field experiment. *J. Econ. Behav. Org.* 166, 174–194.

Cingel, D. P., Sumter, S. R., Stoeten, E., and Mann, S. (2020). Can television help to decrease stigmatization among young children? The role of theory of mind and general and explicit inserts. *Media Psychol.* 23, 342–364.

Clements, W. A., Rustin, C. L., and McCallum, S. (2000). Promoting the transition from implicit to explicit understanding: a training study of false belief. *Dev. Sci.* 3, 81–92. doi: 10.1111/1467-7687.00102

Craig, A. B., Brown, E. R., Upright, J., and DeRosier, M. E. (2016). Enhancing children's social emotional functioning through virtual game-based delivery of social skills training. *J. Child Fam. Stud.* 25, 959–968.

Cutting, A. L., and Dunn, J. (1999). Theory of mind, emotion understanding, language, and family background: Individual differences and interrelations. *Child Dev.* 70, 853–865. doi: 10.1111/1467-8624.00061

De Villiers, J. G., and De Villiers, P. A. (2000). "Linguistic determinism and the understanding of false," in *Children's Reasoning And The Mind*, eds P. Mitchell and J. Riggs (Hove: Psychology Press), 191–228. doi: 10.1111/1460-6984.12317

De Villiers, J. G., and De Villiers, P. A. (2014). The role of language in theory of mind development. *Topics Lang. Disord.* 34, 313–328.

Dessen, M. A., and de Hollanda Souza, D. (2014). Theory-of-mind development in Brazilian low-income children. *Int. Educ. Res.* 2, 41–52. doi: 10.12735/ier.v2i3p41

Devine, R. T., and Hughes, C. (2018). Family correlates of false belief understanding in early childhood: a meta-analysis. *Child Dev.* 89, 971–987. doi: 10.1111/cdev.12682

Devine, R. T., White, N., Ensor, R., and Hughes, C. (2016). Theory of mind in middle childhood: longitudinal associations with executive function and social competence. *Dev. Psychol.* 52, 758–771. doi: 10.1037/dev0000105

Dore, R. A., Amendum, S. J., Golinkoff, R. M., and Hirsh-Pasek, K. (2018). Theory of mind: a hidden factor in reading comprehension? *Educ. Psychol. Rev.* 30, 1067–1089.

Dore, R. A., and Zimmermann, L. (2020). "Coviewing, scaffolding, and children's media comprehension," in *The International Encyclopedia Of Media Psychology*, eds J. van den Bulck and D. R. Roskos-Ewoldsen (Hoboken, NJ: Wiley), 1–8.

Dore, R. A., Shirilla, M., Hopkins, E., Collins, M., Scott, M., Schatz, J., et al. (2019). Education in the app store: using a mobile game to support U.S. preschoolers' vocabulary learning. *J. Child. Media* 13, 452–471.

Ebert, S., Lehrl, S., and Weinert, S. (2020). Differential effects of the home language and literacy environment on child language and theory of mind and their relation to socioeconomic background. *Front. Psychol.* 11:2321. doi: 10.3389/fpsyg.2020.555654

Eisen, S., and Lillard, A. S. (2020). Learning from apps and objects: the human touch. *Mind Brain Educ.* 14, 16–23. doi: 10.3389/fpsyg.2021.621372

Engle, P. L., Fernald, L. C., Alderman, H., Behrman, J., O'Gara, C., Yousafzai, A., et al. (2011). Strategies for reducing inequalities and improving developmental outcomes for young children in low-income and middle-income countries. *Lancet* 378, 1339–1353. doi: 10.1016/S0140-6736(11)60889-1

Esteban, M., Sidera, F., Serrano, J., Amadó, A., and Rostan, C. (2010). Improving social understanding of preschool children: evaluation of a training program. *Electron. J. Res. Educ. Psychol.* 8, 841–860.

Ewin, C. A., Reupert, A. E., McLean, L. A., and Ewin, C. J. (2021). The impact of joint media engagement on parent–child interactions: a systematic review. *Hum. Behav. Emerg. Technol.* 3, 230–254.

Farber, M. (2021). *Gaming SEL: Games As Transformational to Social and Emotional Learning*. New York, NY: Peter Lang Publishing Group.

Fink, E., Begeer, S., Hunt, C., and de Rosnay, M. (2014). False-belief understanding and social preference over the first 2 years of school: a longitudinal study. *Child Dev.* 85, 2389–2403. doi: 10.1111/cdev.12302

Fink, E., Begeer, S., Peterson, C. C., Slaughter, V., and Rosnay, M. (2015). Friendlessness and theory of mind: a prospective longitudinal study. *Br. J. Dev. Psychol.* 33, 1–17. doi: 10.1111/bjdp.12060

Florit, E., De Carli, P., Giunti, G., and Mason, L. (2020). Advanced theory of mind uniquely contributes to children's multiple-text comprehension. *J. Exp. Child Psychol.* 189:104708. doi: 10.1016/j.jecp.2019.104708

Flynn, R. M., Richert, R. A., and Wartella, E. (2019). Play in a digital world: how interactive digital games shape the lives of children. *Am. J. Play* 12, 54–73.

Gola, A. A. H. (2012). Mental verb input for promoting children's theory of mind: a training study. *Cogn. Dev.* 27, 64–76.

Grazzani, I., and Ornaghi, V. (2011). Emotional state talk and emotion understanding: a training study with preschool children. *J. Child Lang.* 38, 1124–1139.

Griffith, S. F., and Arnold, D. H. (2019). Home learning in the new mobile age: parent–child interactions during joint play with educational apps in the US. *J. Child. Media* 13, 1–19.

Griffith, S. F., Hagan, M. B., Heymann, P., Heflin, B. H., and Bagner, D. M. (2020). Apps as learning tools: a systematic review. *Pediatrics* 145:e20191579.

Griffith, S. F., Hanson, K. G., Rolon-Arroyo, B., and Arnold, D. H. (2019). Promoting early achievement in low-income preschoolers in the United States with educational apps. *J. Child. Media* 13, 328–344.

Guajardo, N. R., and Watson, A. C. (2002). Narrative discourse and theory of mind development. *J. Genet. Psychol.* 163, 305–325. doi: 10.1080/00221320209598686

Guajardo, N. R., Petersen, R., and Marshall, T. R. (2013). The roles of explanation and feedback in false belief understanding: a microgenetic analysis. *Journal of Genet. Psychol.* 174, 225–252. doi: 10.1080/00221325.2012.682101

Guernsey, L. (2007). *Into The Minds Of Babes: How Screen Time Affects Children From Birth To Age Five*. New York, NY: Basic Books.

Hale, C., and Tager-Flusberg, H. (2003). The influence of language on theory of mind: a training study. *Dev. Sci.* 6, 346–359. doi: 10.1111/1467-7687.00289

Harvard Graduate School of Education (2022). *Cultivating Early Literacy*. Cambridge, MA: Harvard Graduate School of Education.

Hasni, A. A., Adamson, L. B., Williamson, R. A., and Robins, D. L. (2017). Adding sound to theory of mind: Comparing children's development of mental-state understanding in the auditory and visual realms. *J. Exp. Child Psychol.* 164, 239–249. doi: 10.1016/j.jecp.2017.07.009

Herodotou, C. (2018). Young children and tablets: a systematic review of effects on learning and development. *J. Comput. Assist. Learn.* 34, 1–9.

Hirsh-Pasek, K., Zosh, J. M., Golinkoff, R. M., Gray, J. H., Robb, M. B., and Kaufman, J. (2015). Putting education in "educational" apps lessons from the science of learning. *Psychol. Sci. Public Interest* 16, 3–34. doi: 10.1177/1529100615569721

Hodgkinson, S., Godoy, L., Beers, L. S., and Lewin, A. (2017). Improving mental health access for low-income children and families in the primary care setting. *Pediatrics* 139:e20151175.

Hofmann, S. G., Doan, S. N., Sprung, M., Wilson, A., Ebesutani, C., Andrews, L. A., et al. (2016). Training children's theory-of-mind: a meta-analysis of controlled studies. *Cognition* 150, 200–212.

Holmes, H. A., Black, C., and Miller, S. A. (1996). A cross-task comparison of false belief understanding in a Head Start population. *J. Exp. Child Psychol.* 63, 263–285. doi: 10.1006/jecp.1996.0050

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., and Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Except. Child.* 71, 165–179.

Huber, B., Yeates, M., Meyer, D., Fleckhammer, L., and Kaufman, J. (2018). The effects of screen media content on young children's executive functioning. *J. Exp. Child Psychol.* 170, 72–85. doi: 10.1016/j.jecp.2018.01.006

Hughes, C., and Devine, R. T. (2015). Individual differences in theory of mind from preschool to adolescence: achievements and directions. *Child Dev. Perspect.* 9, 149–153.

Imuta, K., Henry, J. D., Slaughter, V., Selcuk, B., and Ruffman, T. (2016). Theory of mind and prosocial behavior in childhood: a meta-analytic review. *Dev. Psychol.* 52:1192. doi: 10.1037/dev0000140

Keenan, T. (2003). "The preschool years and beyond," in *Individual Differences In Theory Of Mind: Implications For Typical And Atypical Development*, eds B. Repacholi and V. Slaughter (London: Psychology Press), 122–143.

Kennedy, C. H. (2005). *Single-Case Designs For Education Research*. Boston, MA: Allyn and Bacon.

Kim, J., Gilbert, J., Yu, Q., and Gale, C. (2021). Measures matter: A meta-analysis of the effects of educational apps on preschool to grade 3 children's literacy and math skills. *AERA Open* 7:23328584211004183.

Kirsch, C. (2021). "9 Promoting multilingualism and multiliteracies through storytelling: a case study on the use of the app iteo in preschools in luxembourg," in *Multilingual Literacy*, eds E. O. Breuer, E. Lindgren, A. Stavans, and E. Van Steendam (Bristol: Multilingual Matters), 187–210.

Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., et al. (2013). Single-case intervention research design standards. *Remedial Spec. Educ.* 34, 26–38.

Kratochwill, T. R., Horner, R. H., Levin, J. R., Machalicek, W., Ferron, J., and Johnson, A. (2021). Single-case design standards: an update and proposed upgrades. *J. Sch. Psychol.* 89, 91–105. doi: 10.1016/j.jsp.2021.10.006

Lecce, S., Caputi, M., and Pagnin, A. (2014). Long-term effect of theory of mind on school achievement: the role of sensitivity to criticism. *Eur. J. Dev. Psychol.* 11, 305–318. doi: 10.1080/17405629.2013.821944

Lecce, S., Ronchi, L., and Devine, R. T. (2021). Mind what teacher says: teachers' propensity for mental-state language and children's theory of mind in middle childhood. *Soc. Dev.* 31, 303–318.

Ledford, J. R., and Gast, D. L. (2018). *Single Case Research Methodology*. New York, NY: Routledge.

Lillard, A. S., and Kavanaugh, R. D. (2014). The contribution of symbolic skills to the development of an explicit theory of mind. *Child Dev.* 85, 1535–1551. doi: 10.1111/cdev.12227

Lohmann, H., and Tomasello, M. (2003). The role of language in the development of false belief understanding: a training study. *Child Dev.* 74, 1130–1144. doi: 10.1111/1467-8624.00597

Mann, H. B., and Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.* 18, 50–60.

Manolov, R., Losada, J. L., Chacón-Moscoso, S., and Sanduvete-Chaves, S. (2016). Analyzing two-phase single-case data with non-overlap and mean difference indices: illustration, software tools, and alternatives. *Front. Psychol.* 7:32. doi: 10.3389/fpsyg.2016.00032

Manolov, R., Solanas, A., Sierra, V., and Evans, J. J. (2011). Choosing among techniques for quantifying single-case intervention effectiveness. *Behav. Therapy* 42, 533–545. doi: 10.1016/j.beth.2010.12.003

Meyer, M., Zosh, J. M., McLaren, C., Robb, M., McCafferty, H., Golinkoff, R. M., et al. (2021). How educational are "educational" apps for young children? App store content analysis using the four pillars of learning framework. *J. Child. Media* 15, 526–548. doi: 10.1080/17482798.2021.1882516

Michiels, B., Heyvaert, M., and Onghena, P. (2018). The conditional power of randomization tests for single-case effect sizes in designs with randomized treatment order: a monte carlo simulation study. *Behav. Res. Methods* 50, 557–575. doi: 10.3758/s13428-017-0885-7

Mori, A., and Cigala, A. (2015). Perspective taking: training procedures in developmentally typical preschoolers: different intervention methods and their effectiveness. *Educ. Psychol. Rev.* 28, 267–294. doi: 10.1007/s10648-015-9306-6

Musick, G., Freeman, G., and McNeese, N. J. (2021). Gaming as family time: digital game co-play in modern parent-child relationships. *Proc. ACM Hum. Comput. Interact.* 5, 1–25.

Neumann, M. M. (2018). Using tablets and apps to enhance emergent literacy skills in young children. *Early Child. Res. Q.* 42, 239–246.

Neumann, M. M. (2020). Teacher scaffolding of preschoolers' shared reading with a storybook app and a printed book. *J. Res. Child. Educ.* 34, 367–384.

Neumann, M. M., and Neumann, D. L. (2014). Touch screen tablets and emergent literacy. *Early Child. Educ.* 42, 231–239. doi: 10.1007/s10643-013-0608-3

Nikolayev, M., Clark, K., and Reich, S. M. (2015). "Social-emotional learning opportunities in online games for preschoolers," in *Emotions, Technology And Games*, eds S. Tettegah and W. Huang (New York, NY: Academic Press), 211–227.

Nikolayev, M., Reich, S. M., Muskat, T., Tadjbakhsh, N., and Callaghan, M. N. (2021). Review of feedback in edutainment games for preschoolers in the USA. *J. Child.Media* 15, 358–375.

Odom, S. L., and Strain, P. S. (2002). Evidence-based practice in early intervention/early childhood special education: single-subject design research. *J. Early Interv.* 25, 151–160.

Ornaghi, V., Brockmeier, J., and Gavazzi, I. G. (2011). The role of language games in children's understanding of mental states: a training study. *J. Cogn. Dev.* 12, 239–259. doi: 10.1080/15248372.2011.563487

Papadakis, S. (2021a). The impact of coding apps to support young children in computational thinking and computational fluency. A literature review. *Front. Educ.* 6:657895. doi: 10.3389/feduc.2021.657895

Papadakis, S. (2021b). Advances in Mobile Learning Educational Research (AMLER): mobile learning as an educational reform. *Adv. Mobile Learn. Educ. Res.* 1, 1–4.

Papadakis, S. (2022). "Apps to promote computational thinking concepts and coding skills in children of preschool and pre-primary school age," in *Research Anthology on Computational Thinking, Programming, and Robotics in the Classroom*, ed. Information Resources Management Association (Hershey, PA: IGI Global), 610–630.

Papadakis, S., Alexandraki, F., and Zaranis, N. (2021). Mobile device use among preschool-aged children in Greece. *Educ. Inf. Technol.* 27, 2717–2750. doi: 10.1007/s10639-021-10718-6

Parker, R. I., and Vannest, K. (2009). An improved effect size for single-case research: nonoverlap of all pairs. *Behav. Therapy* 40, 357–367.

Parker, R. I., Vannest, K. J., and Davis, J. L. (2011). Effect size in single-case research: a review of nine nonoverlap techniques. *Behav. Modif.* 35, 303–322. doi: 10.1177/0145445511399147

Peebles, A., Bonus, J. A., and Mares, M. L. (2018). Questions+ answers+ agency: interactive touchscreens and Children's learning from a socio-emotional TV story. *Comput. Hum. Behav.* 85, 339–348.

Pempek, T. A., and Lauricella, A. R. (2017). "The effects of parent-child interaction and media use on cognitive development in infants, toddlers, and preschoolers," in *Cognitive Development In Digital Contexts*, eds F. C. Blumberg and P. J. Brooks (Cambridge, MA: Academic Press), 53–74.

Peskin, J., and Astington, J. W. (2004). The effects of adding metacognitive language to story texts. *Cogn. Dev.* 19, 253–273.

Peterson, C. C., Slaughter, V., and Wellman, H. M. (2018). Nimble negotiators: How theory of mind (ToM) interconnects with persuasion skills in children with and without ToM delay. *Dev. Psychol.* 54:494. doi: 10.1037/dev0000451

Radesky, J. S., Schumacher, J., and Zuckerman, B. (2015). Mobile and interactive media use by young children: the good, the bad, and the unknown. *Pediatrics* 135, 1–3.

Rasmussen, E. E., Strouse, G. A., Colwell, M. J., Russo Johnson, C., Holiday, S., Brady, K., et al. (2019). Promoting preschoolers' emotional competence through prosocial TV and mobile app use. *Media Psychol.* 22, 1–22.

Reich, S. M., Korobkova, K. A., Black, R. W., and Sumaroka, M. (2012). "Hey! Can you show me how to do this?" Digital games mediating family interactions," in *Children's Virtual Play Worlds: Culture, Learning And Participation*, eds A. Burke and J. March (New York, NY: Peter Lang), 133–150.

Reich, S. M., Yau, J. C., and Warschauer, M. (2016). Tablet-based ebooks for young children: what does the research say? *J. Dev. Behav. Pediatr.* 37, 585–591. doi: 10.1097/DBP.0000000000000335

Reiß, M., Krüger, M., and Krist, H. (2019). Theory of mind and the video deficit effect: video presentation impairs children's encoding and understanding of false belief. *Media Psychol.* 22, 23–38.

Renouf, A., Brendgen, M., Parent, S., Vitaro, F., David Zelazo, P., Boivin, M., et al. (2010). Relations between theory of mind and indirect and physical aggression in kindergarten: evidence of the moderating role of prosocial behaviors. *Soc. Dev.* 19, 535–555.

Rhodes, M., and Wellman, H. (2013). Constructing a new theory from old ideas and new evidence. *Cogn. Sci.* 37, 592–604. doi: 10.1111/cogs.12031

Richert, R. A., Robb, M. B., and Smith, E. I. (2011). Media as social partners: the social nature of young children's learning from screen media. *Child Dev.* 82, 82–95. doi: 10.1111/j.1467-8624.2010.01542.x

Rideout, V., and Robb, M. B. (2020). *The Common Sense Census: Media Use By Kids Age Zero To Eight, 2020*. San Francisco, CA: Common Sense Media.

Roheger, M., Hranovska, K., Martin, A. K., and Meinzer, M. (2022). A systematic review and meta-analysis of social cognition training success across the healthy lifespan. *Sci. Rep.* 12, 1–21. doi: 10.1038/s41598-022-07420-z

Rowe, M. L., Turco, R. G., and Blatt, J. H. (2021). Can interactive apps promote parent-child conversations? *J. Appl. Dev. Psychol.* 76:101326.

Ruffman, T. (2014). To belief or not belief: children's theory of mind. *Dev. Rev.* 34, 265–293. doi: 10.1016/j.dr.2014.04.001

Ruffman, T., Slade, L., and Crowe, E. (2002). The relation between children's and mothers' mental state language and theory of mind understanding. *Child Dev.* 73, 734–751. doi: 10.1111/1467-8624.00435

Russo-Johnson, C., Troseth, G., Duncan, C., and Mesghina, A. (2017). All tapped out: touchscreen interactivity and young children's word learning. *Front. Psychol.* 8:578. doi: 10.3389/fpsyg.2017.00578

San Juan, V., and Astington, J. W. (2017). Does language matter for implicit theory of mind? The effects of epistemic verb training on implicit and explicit false-belief understanding. *Cogn. Dev.* 41, 19–32.

Schroeder, S. R. (2018). Do bilinguals have an advantage in theory of mind? A meta-analysis. *Front. Commun.* 3:36. doi: 10.3389/fcomm.2018.00036

Segers, E., and Kleemans, T. (2020). The impact of the digital home environment on kindergartners' language and early literacy. *Front. Psychol.* 11:538584. doi: 10.3389/fpsyg.2020.538584

Shahaeian, A., Peterson, C. C., Slaughter, V., and Wellman, H. M. (2011). Culture and the sequence of steps in theory of mind development. *Dev. Psychol.* 47, 1239–1247. doi: 10.1037/a0023899

Shakoor, S., Jaffee, S. R., Bowes, L., Ouellet-Morin, I., Andreou, P., Happé, F., et al. (2012). A prospective longitudinal study of children's theory of mind and adolescent involvement in bullying. *J. Child Psychol. Psychiatry* 53, 254–261. doi: 10.1111/j.1469-7610.2011.02488.x

Shoshani, A., Nelke, S., and Girtler, I. (2022). Tablet applications as socializing platforms: The effects of prosocial touch screen applications on young children's prosocial behavior. *Comput. Hum. Behav.* 127:107077.

Slaughter, V., and Gopnik, A. (1996). Conceptual coherence in the child's theory of mind: training children to understand belief. *Child Dev.* 67, 2967–2988. doi: 10.2307/1131762

Slaughter, V., Dennis, M. J., and Pritchard, M. (2002). Theory of mind and peer acceptance in preschool children. *Br. J. Dev. Psychol.* 20, 545–564. doi: 10.1348/026151002760390945

Slaughter, V., Imuta, K., Peterson, C. C., and Henry, J. D. (2015). Meta-analysis of theory of mind and peer popularity in the preschool and early school years. *Child Dev.* 86, 1159–1174. doi: 10.1111/cdev.12372

Slaughter, V., Peterson, C. C., and Mackintosh, E. (2007). Mind what mother says: narrative input and theory of mind in typical children and those on the autism spectrum. *Child Dev.* 78, 839–858. doi: 10.1111/j.1467-8624.2007.01036.x

Sobel, K., Yen, K., Cheng, Y., Chen, Y., and Hiniker, A. (2019). "No touch pig! Investigating child-parent use of a system for training executive function," in *Proceedings of the 18th ACM International Conference on Interaction Design and Children*, (New York, NY: ACM), 339–351.

Spruijt, A. M., Ziermans, T. B., Dekker, M. C., and Swaab, H. (2020). Educating parents to enhance children's reasoning abilities: a focus on questioning style. *J. Appl. Dev. Psychol.* 66:101102.

Stuckelman, Z. D., Strouse, G. A., and Troseth, G. L. (2021). Value added: digital modeling of dialogic questioning promotes positive parenting during shared reading. *J. Fam. Psychol.* doi: 10.1037/fam0000932 [Epub ahead of print].

Sweeney, S. (2017). Joint parent-child app play can bolster language development: slps can show parents how to harness apps to get their children talking at home. *ASHA Leader* 22. [Epub ahead of print].

Takeuchi, L., and Levine, M. H. (2014). "Learning in a digital age: towards a new ecology of human development," in *Media And The Wellbeing Of Children And Adolescents*, eds A. Jordan and D. Romer (New York, NY: Oxford University Press), 20–43.

Takeuchi, L., and Stevens, R. (2011). *The New Coviewing: Designing For Learning Through Joint Media Engagement*. New York, NY: The Joan Ganz Cooney Center at Sesame Workshop.

Tarullo, A. R., Bruce, J., and Gunnar, M. R. (2007). False belief and emotion understanding in post-institutionalized children. *Soc. Dev.* 16, 57–78. doi: 10.1111/j.1467-9507.2007.00372.x

Taumoepeau, M., and Reese, E. (2013). Maternal reminiscing, elaborative talk, and children's theory of mind: an intervention study. *First Lang.* 33, 388–410. doi: 10.1177/0142723713493347

Teepe, R. C., Molenaar, I., and Verhoeven, L. (2017). Technology-enhanced storytelling stimulating parent–child interaction and preschool children's vocabulary knowledge. *J. Comput. Assist. Learn.* 33, 123–136.

Toh, W., and Lim, F. V. (2021). Let's play together: ways of parent–child digital co-play for learning. *Interact. Learn. Environ.* 1–11. doi: 10.1080/10494820.2021.1951768

Tompkins, V. (2015). Improving low-income preschoolers' theory of mind: a training study. *Cogn. Dev.* 36, 1–19. doi: 10.1016/j.cogdev.2015.07.001

Tompkins, V., Benigno, J. P., Kiger Lee, B., and Wright, B. M. (2018). The relation between parents' mental state talk and children's social understanding: a meta-analysis. *Soc. Dev.* 27, 223–246. doi: 10.1111/cdev.12682

Tompkins, V., Farrar, M. J., and Montgomery, D. E. (2019). Speaking your mind: Language and narrative in young children's theory of mind development. *Adv. Child Dev. Behav.* 56, 109–140. doi: 10.1016/bs.acdb.2018.11.003

Vannest, K. J., Parker, R. I., and Gonen, O. (2011). *Single Case Research: Web Based Calculators for SCR Analysis. (Version 1.0) [Web-based Application]*. College Station, TX: Texas A&M University. Available online at: http://www.singlecaseresearch.org/calculators/nap

Vissers, C., and Koolen, S. (2016). Theory of mind deficits and social emotional functioning in preschoolers with specific language impairment. *Front. Psychol.* 7:1734. doi: 10.3389/fpsyg.2016.01734

Weimer, A. A., Warnell, K. R., Ettekal, I., Cartwright, K. B., Guajardo, N. R., and Liew, J. (2021). Correlates and antecedents of theory of mind development during middle childhood and adolescence: an integrated model. *Dev. Rev.* 59:100945.

Wellman, H. M. (2018). Theory of mind: the state of the art. *Eur. J. Dev. Psychol.* 15, 728–755.

Wellman, H. M., and Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Dev.* 75, 523–541. doi: 10.1111/j.1467-8624.2004.00691.x

Wellman, H. M., and Peterson, C. C. (2013). "Theory of mind development, and deafness," in *Understanding Other Minds: Perspectives From Developmental Social Neuroscience*, eds S. Baron-Cohen, M. V. Lombardo, and H. Tager-Flusberg (London: Oxford University Press), 51–71.

Wellman, H. M., Cross, D., and Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Dev.* 72, 655–684. doi: 10.1111/1467-8624.00304

Wellman, H. M., Fang, F., and Peterson, C. C. (2011). Sequential progressions in a theory-of-mind scale: longitudinal perspectives. *Child Dev.* 82, 780–792. doi: 10.1111/j.1467-8624.2011.01583.x

Wimmer, H., and Perner, J. (1983). Beliefs about beliefs: representation and the containing function of wrong beliefs in young children's understanding of deception. *Cognition* 13, 103–128. doi: 10.1016/0010-0277(83)90004-5

Wood, E., Petkovski, M., De Pasquale, D., Gottardo, A., Evans, M. A., and Savage, R. S. (2016). Parent scaffolding of young children when engaged with mobile technology. *Front. Psychol.* 7:690. doi: 10.3389/fpsyg.2016.00690

Xu, Y., Aubele, J., Vigil, V., Bustamante, A. S., Kim, Y. S., and Warschauer, M. (2021). Dialogue with a conversational agent promotes children's story comprehension via enhancing engagement. *Child Dev.* 93, e149–e167. doi: 10.1111/cdev.13708

Yelland, N., and Masters, J. (2007). Rethinking scaffolding in the information age. *Comput. Educ.* 48, 362–382.

Zosh, J. M., Kathy, H.-P., Emily, J. H., Hanne, J., Claire, L., Dave, N., et al. (2018). Accessing the inaccessible: redefining play as a spectrum. *Front. Psychol.* 9:1124. doi: 10.3389/fpsyg.2018.01124

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Nikolayev, Evmenova, Reich, Clark and Burns. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Education

TYPE Correction
PUBLISHED 01 February 2023
DOI 10.3389/feduc.2022.1116174

Check for updates

OPEN ACCESS

APPROVED BY
Frontiers Editorial Office,
Frontiers Media SA, Switzerland

*CORRESPONDENCE
Mariya Nikolayev
✉ mariya.nikolayev@gmail.com

SPECIALTY SECTION
This article was submitted to Educational Psychology, a section of the journal Frontiers in Education

RECEIVED 05 December 2022
ACCEPTED 16 December 2022
PUBLISHED 01 February 2023

CITATION
Nikolayev M, Evmenova AS, Reich SM, Clark KA and Burns MS (2023) Corrigendum: Teaching preschoolers theory of mind skills with mobile games. *Front. Educ.* 7:1116174. doi: 10.3389/feduc.2022.1116174

COPYRIGHT
© 2023 Nikolayev, Evmenova, Reich, Clark and Burns. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Corrigendum: Teaching preschoolers theory of mind skills with mobile games

Mariya Nikolayev[1]*, Anya S. Evmenova[1], Stephanie M. Reich[2], Kevin A. Clark[1] and M. Susan Burns[1]

[1]College of Education and Human Development (CEHD), George Mason University, Fairfax, VA, United States, [2]School of Education, University of California, Irvine, Irvine, CA, United States

KEYWORDS

theory of mind, preschoolers, educational mobile apps, intervention, single-case design, joint media engagement, digital adult, language

A corrigendum on
Teaching preschoolers theory of mind skills with mobile games

by Nikolayev, M., Evmenova, A. S., Reich, S. M., Clark, K. A., and Burns, M. S. (2022). Front. Educ. 7:872888. doi: 10.3389/feduc.2022.872888

In the published article, there was an error in the article title. Instead of "Improving preschoolers' theory of mind skills with mobile games," it should be "Teaching preschoolers theory of mind skills with mobile games."

The authors apologize for this error and state that this does not change the scientific conclusions of the article in any way. The original article has been updated.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Psychology

SYSTEMATIC REVIEW
published: 16 September 2022
doi: 10.3389/fpsyg.2022.911057

A Systematic Review of Technologies to Teach Control Structures in Preschool Education

Ewelina Bakala[1,2]*, *Anaclara Gerosa*[2], *Juan Pablo Hourcade*[3], *Gonzalo Tejera*[1,2], *Kerry Peterman*[3] and *Guillermo Trinidad*[1]

[1] Facultad de Ingeniería, Instituto de Computación, Universidad de la República, Montevideo, Uruguay, [2] Centro Interdisciplinario en Cognición para la Enseñanza y el Aprendizaje, Universidad de la República, Montevideo, Uruguay, [3] Department of Computer Science, The University of Iowa, Iowa City, IA, United States

There is growing interest in teaching computational thinking (CT) to preschool children given evidence that they are able to understand and use CT concepts. One of the concepts that is central in CT definitions, is the concept of control structures, but it is not clear which tools and activities are successful in teaching it to young learners. This work aims at (1) providing a comprehensive overview of tools that enable preschool children to build programs that include control structures, and (2) analyzing empirical evidence of the usage of these tools to teach control structures to children between 3 and 6. It consists of three parts: systematic literature review (SLR) to identify tools to teach CT to young children, analysis of tools characteristics and the possibilities that they offer to express control structures, and SLR to identify empirical evidence of successful teaching of control structures to young children using relevant tools. This work provides an understanding of the current state of the art and identifies areas that require future exploration.

Keywords: control structures, young children, computational thinking, technology, systematic literature review, preschoolers

OPEN ACCESS

Edited by:
Stamatios Papadakis, University of Crete, Greece

Reviewed by:
Cucuk Budiyanto, Sebelas Maret University, Indonesia
Anastasios Theodoropoulos, University of Peloponnese, Greece

***Correspondence:**
Ewelina Bakala
ebakala@fing.edu.uy

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

Received: *01 April 2022*
Accepted: *06 May 2022*
Published: *16 September 2022*

Citation:
Bakala E, Gerosa A, Hourcade JP, Tejera G, Peterman K and Trinidad G (2022) A Systematic Review of Technologies to Teach Control Structures in Preschool Education. Front. Psychol. 13:911057. doi: 10.3389/fpsyg.2022.911057

1. INTRODUCTION

In 2006, Jeanette Wing popularized the term "Computational thinking" as a universal set of skills which could allow everyone to use computer science concepts for problem solving (Wing, 2006, 2011). Grover (2018) defined two viewpoints on CT: one is that CT is the cognitive or "thinking" counterpart to practicing computer science in CS classrooms. This means CT is a specific characteristic of practicing computer science and is bound to this discipline. The other viewpoint is that CT is a skill to be integrated by other disciplines and it is a way to approach sense-making in different subjects. Wing's original definition of CT was broad enough that it ignited educators and policy-makers' interest in CT (Bocconi et al., 2016). Thus, over the past decade there has been an increase in research around CT interventions targeted at most levels of formal education (Grover and Pea, 2013; Hsu et al., 2018; Yadav et al., 2018; Lyon and Magana, 2020; Stamatios, 2022), its inclusion within other disciplines (Orton et al., 2016; Weintrop et al., 2016; Hickmott et al., 2018), its association with other well-established cognitive skills (Román-González et al., 2017; Robertson et al., 2020; Gerosa et al., 2021; Tsarava et al., 2022), and focusing on creating reliable and valid assessment methods (Tang et al., 2020), amongst others. Moreover, both public and privately-led initiatives have been successfully implemented to foster CT in children

and adolescents (Brackmann et al., 2016; Williamson, 2016), as it is regarded as a valuable twenty-first century skill (Yadav et al., 2016).

Several of the most widely accepted and cited definitions of CT propose the use and understanding of control structures such as loops and conditionals as an integral part of CT. For example, Brennan and Resnick (2012) named loops, conditionals and events as central computational concepts in their framework; Grover and Pea (2013) highlighted the use of conditional logic and iteration as well as Shute et al. (2017). In some cases there is no direct reference to control structures in CT definitions, but algorithm design (Khoo, 2020; Saxena et al., 2020) is considered as an essential part of CT. Control structures are basic building components for algorithms (Perkovic, 2015), and therefore an integral part of CT. Moreover, several of the assessments created for evaluating students' CT in formal education include the evaluation of loops and conditionals, such as Román-González (2015) and collaborators' CTt; Relkin et al.'s (2020) TechCheck or the CT sections that were incorporated to the PISA mathematics testing in OECD (2019).

Authors such as Bers (2019, 2020) have argued for the inclusion of CT skills in early childhood education, particularly through the use of robots as an embodied, tangible tool which would be intuitive and developmentally appropriate for young children. Teaching young children CT related concepts prepares them to solve real-life challenges in a logical and systematic way, and some authors consider CT as relevant as reading, writing and mathematics (Sanford and Naidu, 2016). The early exposure to computing has potential to engage both boys and girls mitigating gender-related barriers (Manches and Plowman, 2017; Martin et al., 2017).

This work aims at presenting the current state of the art of teaching control structures to preliterate children between 3 and 6 years of age using electronic tools (physical, virtual and hybrid systems) that allow users to construct explicit programs. Our work consists of three parts (see **Figure 1**): (1) review 1: a systematic literature review (SLR) of reviews aimed at identifying technology used to promote CT in young children; (2) technology overview: an analysis of the characteristics of these tools based on information we found in tool websites and user manuals; (3) review 2: a SLR of empirical evidence related to the use of the tools in teaching control structures to preliterate children between the ages of 3 and 6.

The research questions that guide this study are the following:

- What electronic tools exist to support the development of CT in young children? (review 1)
- Which tools are appropriate for preliterate children between the ages of 3 and 6? (technology overview)
- How can children introduce control structures into their programs using electronic tools? (technology overview)
- What tools have been reported to be successful for teaching control structures to young children? (review 2)

In the remainder of the paper we present related works that systematize the knowledge about existing tools that support the development of CT, next we present the methodology and findings of the first SLR that aims to identify existing tools for teaching CT to young children (see **Figure 1**). In the following step we analyze the tools to identify those that are electronic-based and appropriate for preliterate children between 3 and 6 years old, and provide details related to their price and possibilities that they offer to introduce control structures in children's code. The resulting list of appropriate tools is used in the second SLR to search for empirical evidence related to teaching control structures to young children. The limitations and results are discussed in the final section of the article and conclusions are laid down.

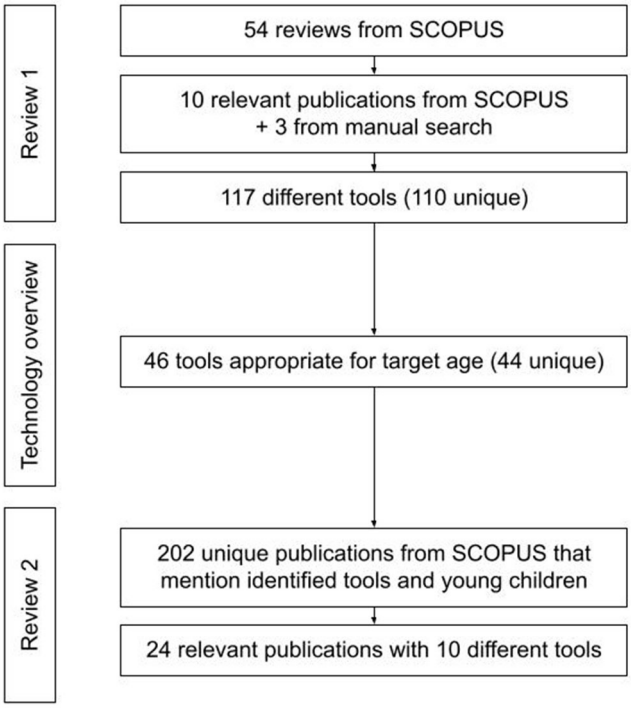


FIGURE 1 | An overview of the research pipeline.

1.1. Related Work

Previous work has focused on reviewing technological and unplugged tools to promote CT in young children. However, most of the available reviews on this topic focus on the broad aspects of CT and robotics without specifically analyzing the affordances of particular technological tools for learning a specific concept, such as control structures. For example, Silva et al. (2021) focused on describing the available technology for 2–8 year old children as well as curricula implemented for these ages, while Kakavas and Ugolini (2019) focused on they way the teaching of CT has evolved in primary education in the last decades and was successful in identifying the context in which the technology was implemented and in which way CT was assessed. In a recent review (Bakala et al., 2021) we also analyzed the characteristics of robots and activities used

in preschool education to promote CT skills with a focus on empirical research, however the specific ways in which each concept encompassed by CT was targeted was not part of our scope. Recent work by Taslibeyaz et al. (2020) shed light into the way studies with young children considered the concept of CT by analyzing its definitions, which skills were targeted and which variables were assessed and included the technological tools used to promote these skills. However, the implications as to how a specific technology causes this improvement and what are the nuances of using different technological tools were not discussed. Similarly, a recent review by Toh et al. (2016) on the use of robots for young children provided context on the type of study conducted and on the effects of robotics on children's cognitive outcomes as well as parents', educators' and children's opinions regarding the use of these tools. However, the possible benefits are discussed generally regarding robotics and this work does not focus on the outcomes of specific tools. Yu and Roque (2019) provide a comprehensive review of computational toys and kits for young children (7 and under) describing their design features, which computational concepts and practices they target and how they relate to other domains in knowledge. In particular, they analyzed the way conditionals were presented in the technological tools and argued that most of the time conditionals were implemented in an implicit way (thus not represented using explicit if-then statements). In addition, the authors explored the presentation of loops, pointing out many of the available tools used repeat blocks which encapsulated a given sequence, whether digital or concrete. In order to expand upon these findings, this review will focus specifically on the ways technology has implemented control structures and provide an overview of the evidence surrounding these implementations with young children. In this sense, our review will provide a summary of the empirical experiences in which these control structures have been taught to young children and analyze these results. To our knowledge, there isn't thus far a systematic review of literature which focuses on the implementation of control structures and provides a thorough analysis of how technological tools aimed at early childhood allow its users to learn them. In addition, we conducted a SLR on the existing empirical evidence in which control structures have been taught to children, shedding light into which practices and tools are supported by evidence and thus favorable for its inclusion in the classroom.

2. SLR OF EXISTING TOOLS (REVIEW 1)

We used a systematic literature review (Kitchenham et al., 2015) to answer our first research question: What tools exist to support the development of CT in young children?

2.1. Methodology

Systematic literature review (SLR) is a method that allows identifying relevant material to a given topic using an objective, analytical, and repeatable approach (Kitchenham et al., 2015). We carried out our literature review following the PRISMA guidelines (Moher et al., 2009). Four reviewers participated in the review process. Firstly, they defined the search term, inclusion and exclusion criteria, and data to extract from the publications. Secondly, two reviewers analyzed the publications to identify the relevant articles. One reviewer extracted the tools from relevant

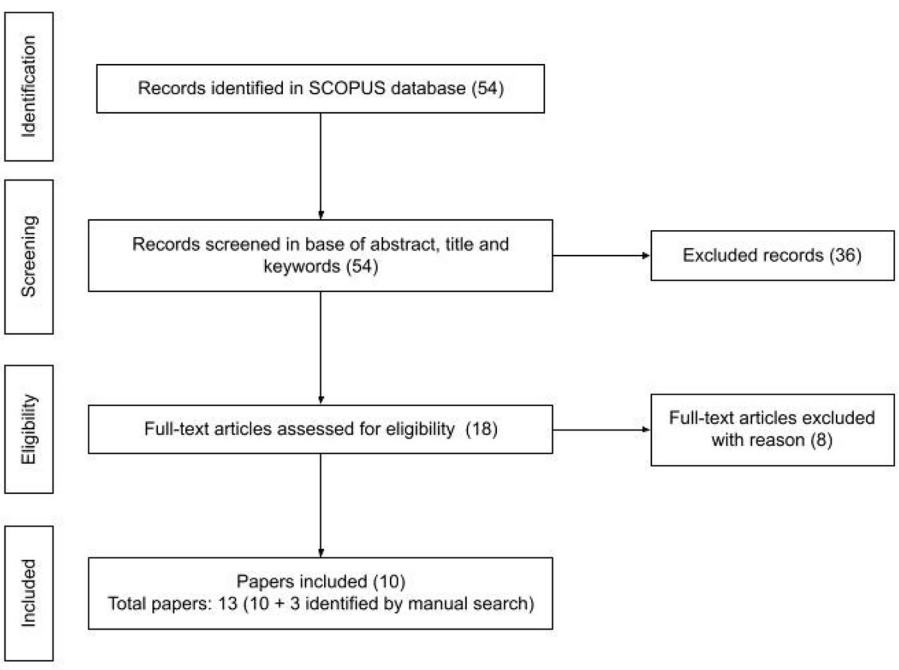


FIGURE 2 | Steps of the selection process of the first SLR. Reported in line with the PRISMA statement (Moher et al., 2009).

articles. A quality assessment stage was not included, as we were not interested in filtering out low quality studies since we still reviewed each tool or investigating changes in quality over time.

2.1.1. Search Strategy

To identify reviews of technology to support the development of CT in young children we applied an automated search (Kitchenham et al., 2015) in the Scopus search engine (Elsevier Scopus, 2022). The search term was the following:

TITLE-ABS-KEY (((review AND {computational thinking} AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learners} OR {primary school} OR {primary education} OR k-6 OR k-8 OR childhood))))

We used three keywords: review, computational thinking and childhood (and synonyms) to search in the title, abstract, and keywords.

2.1.2. Study Selection

We defined the following inclusion criteria for the studies' selection:

- Articles that review electronic-based tools to promote the development of CT in young children.
- Publications focused on children between 3 and 5 years old, including 6 years old, if attending pre-primary school educational level.

Exclusion criteria were:

- Articles written in a language other than English or Spanish.
- Publications that target children older than 6 years.
- Articles limited to unplugged tools.
- Case studies.
- Conference proceedings.

The first round of the selection was made based on the information available in the abstract. Two researchers applied the criteria independently and filter out publications that do not review tools focused on promoting CT in young children. The articles were tagged as "relevant" or "irrelevant." If an article was classified differently by the reviewers, the full text was reviewed. If there were doubts about an article, they were discussed with two other reviewers that supervised this revision step. Also the articles that were considered relevant by both reviewers were analyzed in detail to confirm or reject their relevance.

2.1.3. Data Extraction

We used a spreadsheet to extract tools found in the publications and articles that mention each tool. We sorted each tool using categories that we developed (see Section 2.4).

2.2. Findings

2.3. Relevant Articles

The search was conducted on 6th of August 2021 and we obtained 54 articles to review (see **Figure 2**). In the screening phase the reviewers tagged identically 51 of 54 articles reaching an agreement of 0.94%. In the selection process we identified 10 articles relevant for this study. We added to our analysis 3 articles (Kakavas and Ugolini, 2019; Papadakis, 2020; Silva et al., 2021) that were identified by the manual search and that we considered a valuable source of information-Kakavas and Ugolini (2019) that was not indexed by Scopus, Papadakis (2020) that does not contain the word "review" in title, abstract and keywords and Silva et al. (2021) that is a preprint submitted to Elsevier.

A total of 13 articles (see **Table 1**) were used to elaborate the list of relevant tools. All the articles were published between 2018 and 2021.

2.4. Categories to Classify the Tools

To classify the tools we adapted and expanded categories proposed by Yu and Roque (2019). We obtained 4 main categories (see **Figure 3**): Physical, Virtual, Hybrid and No information. We divided Physical, Virtual and Hybrid into sub-categories and obtained 9 categories which we used to classify existing tools: Robots with tangible programming interface, Construction kits with no explicit program, Unplugged, Virtual with explicit program, Virtual with no explicit program, Robots with virtual programming interface, Construction kits with virtual programming interface, Virtual tools with tangible programming interface, No information. In the **Figure 3** there

TABLE 1 | 13 relevant publications that we identified in the first SLR.

References	Title
Bakala et al. (2021)	Preschool children, robots, and computational thinking: A systematic review
Papadakis (2021)	The Impact of Coding Apps to Support Young Children in Computational Thinking and Computational Fluency. A Literature Review
Fagerlund et al. (2021)	Computational thinking in programming with Scratch in primary schools: A systematic review
Yang et al. (2020)	The influence of robots on students' computational thinking: A literature review
Pedersen et al. (2020)	The effect of commercially available educational robotics: A systematic review
Umam et al. (2019)	Literature review of robotics learning devices to facilitate the development of computational thinking in early childhood
Isnaini et al. (2019)	Robotics-based learning to support computational thinking skills in early childhood
Yu and Roque (2019)	A review of computational toys and kits for young children
Ching et al. (2018)	Developing Computational Thinking with Educational Technologies for Young Learners
Ioannou and Makridou (2018)	Exploring the potentials of educational robotics in the development of computational thinking: A summary of current research and practical proposal for future work
Silva et al. (2021)	A Systematic Review of Computational Thinking in Early Ages
Papadakis (2020)	Robots and Robotics Kits for Early Childhood and First School Age
Kakavas and Ugolini (2019)	Computational thinking in primary education: a systematic literature review

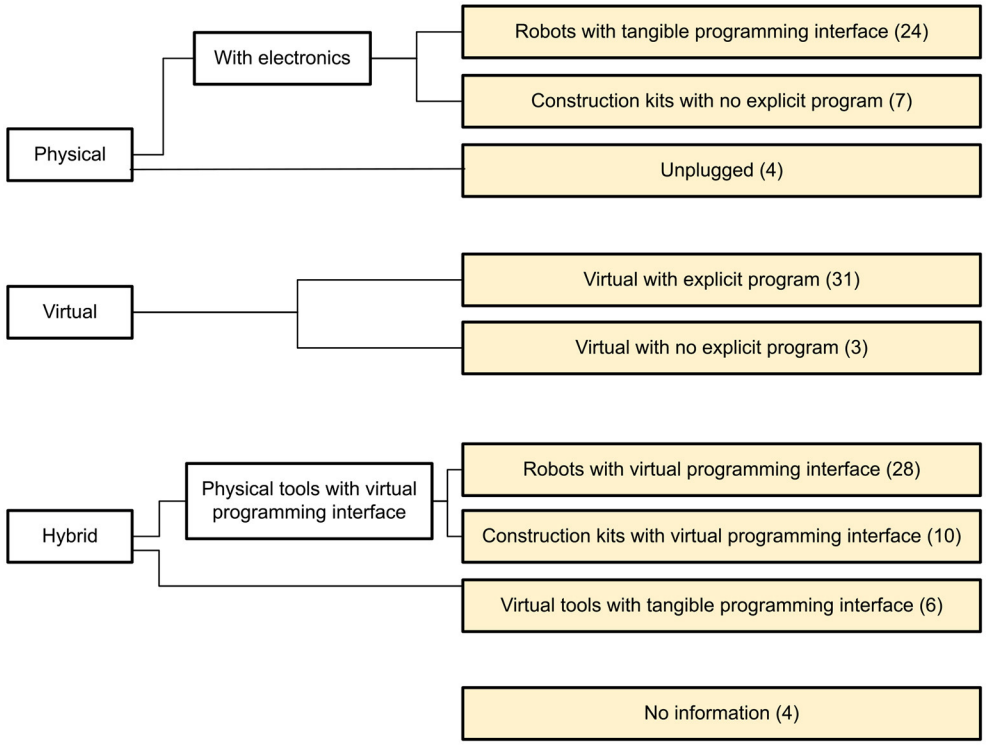


FIGURE 3 | Categories developed to classify the physical aspect of the tools.

are more than 8 categories, but only those highlighted in yellow were used to classify the tools.

We used the category Physical for tools that are fully tangible and do not require screen-based applications. We divided it into Unplugged and Physical tools with electronics. The last category was composed of Robots with tangible programming interface and Construction kits with no explicit program. The category Construction kits with no explicit program contains electronic building blocks that can be connected together to cause certain behavior of the system, but do not require the user to write an explicit program.

Virtual contains fully screen-based tools, such as desktop, mobile, or web apps. This category encompasses tools that do not require the user to write an explicit program (e.g., tools like CompThink App where the user has to solve logical problems without writing code) and those which need an explicit program.

Hybrid tools combine physical and virtual parts. We divided them into Virtual tools with tangible programming interface or Physical tools with virtual programming interface. The first category consists of applications with tangible programming interfaces. The second category is composed of Robots with virtual programming interface and Construction kits with virtual programming interface. The last category contains embedded systems like Arduino that can be programmed using a virtual programming interface.

The "No information" category was used if there was no information about the tool that could be used to classify it.

2.5. Tools

From the 13 relevant publications we extracted 110 tools (106 unique tools). In the case of Code & Go Robot Mouse, we found three different names that referred to this tool: Robot Mouse (Yu and Roque, 2019; Pedersen et al., 2020), Colby robotic mouse (Papadakis, 2020; Bakala et al., 2021) and Code & Go Robot Mouse (Ching et al., 2018; Silva et al., 2021), and we analyzed it as one single tool.

While reviewing the tools mentioned in the articles we found in external sources 4 more tools that we considered relevant for our work. We added Qobo (Physical and Hybrid), VEX 123 (Physical and Hybrid), Sphero indi (Physical and Hybrid), Scottie Go (Virtual) and ended up with a total of 117 tools (110 unique tools, see **Table 2**).

We classified 35 as Physical, 34 as Virtual, 44 as Hybrid and 4 as No information (see **Figure 3**).

It is important to say that seven tools were present in more than one category (Blue-Bot, Qobo, VEX 123, Sphero indi, VBOT, Makeblock Neuron, Tuk Tuk). For example, Blue-Bot is a robot that can be programmed using buttons on its back and because of that it belongs to the category Robots with tangible programming interface, but there is also a possibility to program it using an application, so it was also classified as a Robot with a virtual programming interface. That is why we refer to 110 unique tools, although we analyzed 117 relevant tools that included duplicated items.

In three cases (Ozobot, LEGO, Robotis and roboplus software) the names that we found in publications were names of brands,

not names of specific tools, so it was impossible to classify them, and they were categorized as No information. One publication mentioned Robo Cup Junior as a tool. As far as we know RoboCup Junior (RoboCupJunior, 2022) is an educational initiative, not one particular technology, so we categorized this item as No information as well.

3. TECHNOLOGY OVERVIEW

The first aim of this part of our study was to identify how young, preliterate children can introduce conditionals and iterations into their programs using existing tools. This section is motivated by the following research questions:

- Which tools are appropriate for preliterate children between the ages of 3 and 6?
- How can children introduce control structures into their programs using electronic tools?

3.1. Methodology

Four reviewers participated in the revision of existing tools. Two of them reviewed the available online information and extracted the information of interest. The other two participated in the definition of the categories to classify tools' characteristics and helped to classify doubtful cases.

3.1.1. Tools Selection

We were interested in electronic tools that allow users to construct explicit programs, so we did not further analyze the tools classified as Unplugged, Construction kits with no explicit program, Virtual with no explicit program, and No information.

We identified the relevant tools by filtering out those not appropriate for children between 3 and 6 - tools that target children older than 6 years old or that should be programmed using interfaces that require reading skills (see **Table 2**). During tool selection we first analyzed the target age of each tool. If the information of the target age was expressed using educational levels like "elementary school" or "kindergarten" we translated this information into age using the United States educational system as reference. If the tool was designed for children older than 6, we tagged it as inappropriate and did not analyze its programming interfaces. If the age was of our interest, we proceeded with the inspection of the user interface. In many cases hybrid tools offered different programming languages/interfaces to cover a wide age spectrum of users, for example, Finch Robot can be programmed using 8 different programming languages and its promotional video states that it is suitable for users from "from kindergarten to college." In those cases we evaluated only programming languages appropriate for preliterate children. If there was no interface suitable for preschoolers, we marked it as a tool that requires reading skills.

3.1.2. Data Extraction

To collect the information about the tools we reviewed the official websites, video material provided by the manufacturer, online manuals, as well as, youtube videos and amazon websites.

During data extraction we were interested in classifying different types of control structures that can be used with each tool, so we defined categories that we present in Sections 3.2.2.1, 3.2.2.2.

3.2. Findings

3.2.1. Tools Selection

We identified 46 tools (44 unique) appropriate for preliterate children (see **Table 3**). Twenty Robots with tangible programming interface, 11 Virtual with explicit program and 15 Hybrid tools: 8 Robots with virtual programming interface, 1 Construction kit with virtual programming interface and 6 Virtual tools with tangible programming interface. Two tools (Blue Bot and Sphero indi) were classified as both: Robots with tangible programming interface and Robots with virtual programming interface.

There were three tools that we analyzed together: KIBO, KIWI and CHERP. KIBO is a robot currently available in the market, formerly known as "KIWI" or Kids Invent with Imagination (Tufts University, 2022). CHERP is a programming language that is used to program KIBO and KIWI, so evaluating CHERP is equivalent to evaluating KIBO and KIWI.

In the case of some tools, the programming interface contained images which made it accessible for preliterate children, but we had the impression that the systems were designed for children older than our target age. They contained text-based challenges (Scottie Go) and menus (BOTS, Neuron App, Move the turtle, RoboZZle), design that we consider unattractive for young children (RoboZZle, BOTS), text-based options with no associated image ("tap" event in Roberto), or comparisons involving high numeric values (Neuron App). Although these tools raised some doubts, we decided to include them in our analysis as we wanted to provide an inclusive overview of the existing tools.

3.2.2. Categories to Classify Control Structures

We developed categories related to the use of control structures to classify tools suitable for young children (see **Table 3**) that we identified during tools selection step (see Section 3.2.1).

3.2.2.1. Conditionals

To identify how the children can introduce decision making based on certain conditions into their programs we reviewed the programming interfaces and classified the existing tools with categories that we defined in an iterative process. Introducing conditions in the code was typically based on conditional branches (e.g., if-else structures) or based on events (e.g., blocking the program execution until some event occurs). From now on we will refer to those two forms of incorporation of conditions into the code as "conditionals."

To classify the degree of liberty that the children have while using and building conditionals in their programs, we propose three levels, ordered by increasing complexity for the user:

1. Predefined connection of condition and action: it is possible to use a predefined programming statement that connects an event with an action. For example, the Qobo robot detects coding cards below it and acts according to the statement stored in the card. It has a specific card for conditional turning

TABLE 2 | 117 tools extracted from 13 relevant publications that we identified in the first SLR.

Tool type	Name	Target age	Exclusion reason [Age, RRS (require reading skills), Unplugged, No info, No program]	Source
Robots with tangible programming interface	Bee Bot	3+		Umam et al. (2019), Yu and Roque (2019), Papadakis (2020), Pedersen et al. (2020), Silva et al. (2021), Bakala et al. (2021), Yang et al. (2020)
	Blue Bot	3–11		Yu and Roque (2019), Papadakis (2020), Pedersen et al. (2020), Silva et al. (2021)
	Botley	5+		Papadakis (2020)
	Code-a-Pillar	3–6		Ching et al. (2018), Yu and Roque (2019), Papadakis (2020)
	Cubetto	3–9		Isnaini et al. (2019), Ching et al. (2018), Yu and Roque (2019), Papadakis (2020), Umam et al. (2019), Pedersen et al. (2020)
	Dr. Wagon	6–12	RRS	Yu and Roque (2019)
	Edison robot	4–16	No program	Papadakis (2020), Pedersen et al. (2020)
	KIBO	4–7		Ching et al. (2018), Umam et al. (2019), Yu and Roque (2019), Papadakis (2020), Pedersen et al. (2020), Silva et al. (2021), Bakala et al. (2021), Yang et al. (2020);
	KIWI	5–7		Bakala et al. (2021)
	KUBO robot	4–10		Papadakis (2020), Pedersen et al. (2020)
	Matatalab Coding Set	4–9		Papadakis (2020)
	mTiny	4+		Papadakis (2020)
	Ozobot Evo	5–18		Papadakis (2020)
	Ozobot Bit	6+		Papadakis (2020), Bakala et al. (2021)
	Plobot	4+		Yu and Roque (2019)
	Pro-bot	3+		Yu and Roque (2019), Papadakis (2020), Pedersen et al. (2020), Silva et al. (2021)
	Qobo	3–8		Manual
	Roamer	4–13	No program	Papadakis (2020)
	Robot Mind Designer	7+	Age	Papadakis (2020)
	Code and Go Robot Mouse	4–9		Ching et al. (2018), Yu and Roque (2019), Papadakis (2020), Pedersen et al. (2020), Silva et al. (2021), Bakala et al. (2021);
	Robotito	4–6		Silva et al. (2021)
	Sphero indi	4–8		Manual
	TurtleBot	No info		Bakala et al. (2021)
	VEX 123	4–9		Manual
Construction kits with no explicit program	Cubelets	4+	No program	Papadakis (2020), Pedersen et al. (2020)
	Curlybot	No info	No program	Yu and Roque (2019)
	Electronic Blocks	4–6	No program	Yu and Roque (2019)
	LittleBits	8+	No program	Kakavas and Ugolini (2019), Pedersen et al. (2020)
	Makeblock Neuron	6+	No program	Pedersen et al. (2020)
	roBlocks	9+	No program	Yu and Roque (2019)
	Romibo	No info	No program	Pedersen et al. (2020)
Unplugged	Code Monkey Island	6+	Unplugged	Ching et al. (2018)
	Happy Maps	No info	Unplugged	Silva et al. (2021)
	Hello Ruby	5+	Unplugged	Yu and Roque (2019)
	Robot Turtles	4+	Unplugged	Ching et al. (2018), Yu and Roque (2019)
Virtual with explicit program	AgentCubes	8+	Age	Kakavas and Ugolini (2019)
	AgentSheets	11–13	Age	Kakavas and Ugolini (2019)
	Alice	11+	Age	Kakavas and Ugolini (2019)

(Continued)

TABLE 2 | Continued

Tool type	Name	Target Age	Exclusion reason [Age, RRS (require reading skills), Unplugged, No info, No program]	Source
	BOTS	5–18		Kakavas and Ugolini (2019)
	Cargo-Bot	10–18	Age	Ching et al. (2018), Yu and Roque (2019)
	Codeable Crafts	4+		Yu and Roque (2019)
	Code.org	4+		Ching et al. (2018), Silva et al. (2021)
	CodyColor	0+	No program	Silva et al. (2021)
	CTSiM	5–18	RRS	Kakavas and Ugolini (2019)
	Daisy the Dinosaur	7+	Age	Papadakis (2021)
	FormulaT Racing	7–13	Age	Kakavas and Ugolini (2019)
	Hopescotch	10–16	Age	Ching et al. (2018)
	Kodable	4–10		Ching et al. (2018), Papadakis (2021), Silva et al. (2021)
	Kodetu	9–17	Age	Kakavas and Ugolini (2019)
	Kodu	9+	Age	Kakavas and Ugolini (2019)
	Legato	4–11	No program	Ching et al. (2018), Silva et al. (2021)
	LightBot	9+	Age	Ching et al. (2018), Yu and Roque (2019), Kakavas and Ugolini (2019), Papadakis (2021), Silva et al. (2021)
	LightBotJr	4–8		Ching et al. (2018), Silva et al. (2021)
	MiniColon game	8–9	Age	Kakavas and Ugolini (2019)
	Move the turtle	5+		Yu and Roque (2019)
	RoboZZle	6–7		Yu and Roque (2019)
	Run Marco!	4+		Yu and Roque (2019)
	Scratch	8–16	Age	Ching et al. (2018), Isnaini et al. (2019), Kakavas and Ugolini (2019), Fagerlund et al. (2021)
	ScratchJr	5–7		Kakavas and Ugolini (2019), Yu and Roque (2019), Ching et al. (2018), Papadakis (2021), Silva et al. (2021)
	Story-Writing-Coding engine	5–11	RRS	Kakavas and Ugolini (2019)
	The Foos	5+		Yu and Roque (2019), Silva et al. (2021)
	Tuk Tuk (standard)	5–14	RRS	Silva et al. (2021)
	Tynker: Coding for Kids	5–14		Ching et al. (2018)
	VBOT	14+	Age	Ioannou and Makridou (2018), Yang et al. (2020)
	ViMAP	8–10	Age	Kakavas and Ugolini (2019)
	Zoombinis game	8+	Age	Kakavas and Ugolini (2019)
Virtual with no explicit program	CompThink App	5–11	No program	Kakavas and Ugolini (2019)
	PhysGramming	6–7	No program	Silva et al. (2021)
	Tuk Tuk (junior)	5–6	No prgram	Silva et al. (2021)
Robots with virtual programming interface	Blue Bot	3–11		Yu and Roque (2019), Papadakis (2020), Pedersen et al. (2020), Silva et al. (2021)
	CHERP	5–6		Ioannou and Makridou (2018), Kakavas and Ugolini (2019)
	Codey Rocky	5–11	RRS	Pedersen et al. (2020)
	COJI	6+		Yu and Roque (2019), Papadakis (2020)
	Cozmo	8–11	Age	Pedersen et al. (2020)
	Dash and/or Dot	6+		Ching et al. (2018), Yu and Roque (2019), Papadakis (2020), Pedersen et al. (2020)
	Finch	5+		Papadakis (2020)
	LEGO Boost	7–12	Age	Pedersen et al. (2020)
	LEGO Education WeDo	7+	Age	Kakavas and Ugolini (2019), Isnaini et al. (2019), Ching et al. (2018), Papadakis (2020), Silva et al. (2021), Pedersen et al. (2020), Umam et al. (2019), Bakala et al. (2021)

(Continued)

TABLE 2 | Continued

Tool type	Name	Target Age	Exclusion reason [Age, RRS (require reading skills), Unplugged, No info, No program]	Source
	LEGO Mindstorm	10+	Age	Kakavas and Ugolini (2019), Ching et al. (2018), Ioannou and Makridou (2018), Pedersen et al. (2020), Bakala et al. (2021)
	Max Tobo coding robot	6+	RRS	Papadakis (2020)
	mBot	8+	Age	Pedersen et al. (2020), Silva et al. (2021)
	MeeperBots	5–12	RRS	Yu and Roque (2019), Papadakis (2020)
	Mind designer robot	7+	Age	Papadakis (2020)
	MiP	8–15	Age	Pedersen et al. (2020)
	MU Spacebot	8+	Age	Pedersen et al. (2020)
	NAO	5–18	RRS	Kakavas and Ugolini (2019), Pedersen et al. (2020)
	Qobo	3–8	RRS	Manual
	ROBOTC Graphical	No info	RRS	Kakavas and Ugolini (2019)
	Scribbler	14+	Age	Pedersen et al. (2020)
	Sphero Ollie	8–14	Age	Pedersen et al. (2020)
	Sphero indi	4–8		Manual
	Sphero mini	8+	Age	Papadakis (2020)
	The Coffee Platform	No info	RRS	Ioannou and Makridou (2018)
	Thymio	6+		Yu and Roque (2019), Papadakis (2020), Pedersen et al. (2020)
	Tinkerbots	5+		Papadakis (2020)
	VEX 123	4–9	RRS	Manual
	VBOT	11–18	Age	Ioannou and Makridou (2018), Yang et al. (2020)
Construction kits with virtual programming interface	An ultra-low cost line follower Robotic	16–18	Age	Yang et al. (2020)
	Arduino+scratch	7–13	Age	Yang et al. (2020)
	CyberPLAYce	8–12	Age	Kakavas and Ugolini (2019)
	GoGo Board	10–18	Age	Ioannou and Makridou (2018)
	Hummingbird Robotics Kit	9–18	RRS	Pedersen et al. (2020)
	Makeblock Neuron	6+		Pedersen et al. (2020)
	micro:bit	8–14	Age	Pedersen et al. (2020)
	Scratch 4 Arduino, S4A)	8–17	Age	Kakavas and Ugolini (2019)
	ultimate	12+	Age	Pedersen et al. (2020)
	VEX IQ	11+	Age	Pedersen et al. (2020)
Virtual tools with tangible programming interface	Puzzlets Starter Pack	6+		Yu and Roque (2019)
	Roberto	4+		Yu and Roque (2019)
	Scottie Go	4–15		Manual
	Coding Awbie	5–11		Ching et al. (2018), Papadakis (2020), Silva et al. (2021), Yu and Roque (2019)
	Tabletop puzzle block system	4–5		Yu and Roque (2019)
	T-Maze	5–9		Kakavas and Ugolini (2019), Silva et al. (2021)
No info	LEGO	No info	No info	Yang et al. (2020), Bakala et al. (2021)
	Ozobot	No info	No info	Kakavas and Ugolini (2019), Pedersen et al. (2020)
	Robo Cup Junior	no info	No info	Isnaini et al. (2019)
	Robotis and roboplus software	No info	No info	Ioannou and Makridou (2018)

TABLE 3 | An overview of 46 relevant tools considering their price and possibilities to incorporate control structures into the code.

Classification	Tool name	Conditionals [Predefined connection, Free connection, Free condition building]	Integration with the main program [Integrated if, Blocking event, Interruption, Parallel execution]	Number of repetitions [Fixed number of repetitions, Configurable number of repetitions, Infinite loop]	Number of repeated commands [Single command repetition, Multiple command repetition]	Price (USD)
Robots with tangible programming interface	Bee Bot	–	–	–	–	85
	Blue Bot	–	–	–	–	104
	Botley	Free connection	Interruption	Configurable number of repetitions	Multiple command repetition	47
	Code–a–Pillar	–	–	Configurable number of repetitions	Single command repetition	148 (new version) or 35 (old)
	Cubetto	–	–		–	225
	KIBO	Free connection + Free condition building	Blocking event + Integrated if	Configurable number of repetitions + Infinite loop	Multiple command repetition	230 to 610
	KIWI	Free connection + Free condition building	Blocking event + Integrated if	Configurable number of repetitions + Infinite loop	Multiple command repetition	Unavailable
	KUBO robot	–	–	Configurable number of repetitions	Multiple command repetition	310 to 396
	Matatalab Coding Set	Free connection + Free condition building	Blocking event	Configurable number of repetitions	Multiple command repetition	169
	mTiny	–	–	Configurable number of repetitions	Multiple command repetition	120
	Ozobot Evo	Predefined connection	Integrated if	–	–	175
	Ozobot Bit	Predefined connection	Integrated if	–	–	Unavailable
	Plobot	Free connection	Blocking event	–	–	Unavailable
	Pro–bot	Free connection	Interruption	Configurable number of repetitions	Multiple command repetition	150
	Qobo	Predefined connection	Blocking event + Integrated if	Fixed number of repetitions	Multiple command repetition	60
	Robot Mouse	–	–	–	–	60
	Robotito	Predefined connection	Interruption	–	–	Unavailable
	Sphero indi	Predefined connection	Interruption	–	–	100
	TurtleBot	Predefined connection	Integrated if	–	–	105
	VEX 123	Free connection	Integrated if	Fixed number of repetitions + Configurable number of repetitions + Infinite loop	Single command repetition + Multiple command repetition	119
Virtual with explicit program	BOTS	Free condition building	Integrated if	Configurable number of repetitions	Multiple command repetition	Unavailable
	Codeable Crafts	Free connection	Parallel execution	Configurable number of repetitions + Infinite loop	Single command repetition + Multiple command repetition	Free
	Code.org	Free condition building	Interruption	Configurable number of repetitions	Multiple command repetition	Free
	Kodable	Free connection	Interruption	Configurable number of repetitions	Multiple command repetition	Free–2000 yearly
	LightBotJr	–	–	Configurable number of repetitions + Infinite loop	Multiple command repetition	2.99
	Move the turtle	Free condition building	Integrated if	Configurable number of repetitions	Multiple command repetition	3.99
	RoboZZle	Free connection	Interruption	Configurable number of repetitions + Infinite loop	Multiple command repetition	Free
	Run Marco!	Free condition building	Integrated if	Configurable number of repetitions	Multiple command repetition	Free

(Continued)

TABLE 3 | Continued

Classification	Tool name	Conditionals [Predefined connection, Free connection, Free condition building]	Integration with the main program [Integrated if, Blocking event, Interruption, Parallel execution]	Number of repetitions [Fixed number of repetitions, Configurable number of repetitions, Infinite loop]	Number of repeated commands [Single command repetition, Multiple command repetition]	Price (USD)
	ScratchJr	Free connection	Parallel execution	Configurable number of repetitions + Infinite loop	Single command repetition + Multiple command repetition	Free
	The Foos	Free condition building	Integrated if	Configurable number of repetitions + Infinite loop	Multiple command repetition	Free
	Tynker: Coding for Kids	Free connection	Integrated if + Interruption	Configurable number of repetitions	Single command repetition + Multiple command repetition	Free
Robots with virtual programming interface	Blue Bot	–	–	Configurable number of repetitions	Multiple command repetition	104
	CHERP	Free connection + Free condition building	Blocking event + Integrated if	Configurable number of repetitions + Infinite loop	Multiple command repetition	Unavailable
	COJI	Free connection	Interruption	–	–	32
	Dash and/or Dot	Free connection + Free condition building	Blocking event	Infinite loop	Multiple command repetition	150
	Finch	Free connection	Parallel execution	Configurable number of repetitions	Multiple command repetition	139
	Sphero indi	Free connection	Interruption	–	–	100
	Thymio	Free connection	Interruption	–	–	160
	Tinkerbots	–	–	Configurable number of repetitions	Single command repetition + Multiple command repetition	149
Construction kits with virtual programming interface	Makeblock Neuron	Free condition building	Integrated if	–	–	Unavailable
Virtual tools with tangible programming interface	Puzzlets Starter Pack	–	–	Configurable number of repetitions	Single command repetition	147
	Roberto	Free condition building	Blocking event	Infinite loop	Multiple command repetition	Unavailable
	Scottie Go	Free condition building	Integrated if	Configurable number of repetitions + Infinite loop	Single command repetition + Multiple command repetition	45–74
	Coding Awbie	Free connection	Integrated if	Configurable number of repetitions	Single command repetition + Multiple command repetition	99
	Tabletop puzzle block system	–	–	–	–	Unavailable
	T–Maze	Predefined connection	Blocking event	Configurable number of repetitions	Multiple command repetition	Unavailable

- if the robot passes over a card with a banana before passing over a bifurcation card, it turns left, but if it passes over a card with an apple, it turns right. Neither the condition nor the resulting action can be modified by the user.

2. Free connection of predefined condition and predefined action: it is possible to combine predefined conditions with predefined actions to build custom conditionals. For example, the Sphero Edu Jr application (see **Table 5**) allows users to associate a color sensed by the robot (predefined condition) with an action involving movement, light, and/or sound of the Sphero indi robot (predefined actions). The user needs at least two programming statements (condition and action)

to build a conditional. In the case of Kodable and RoboZZle these two statements are combined in one coding block: the background color of the block defines the condition (e.g., "if the tile is pink") and the arrow, the action (e.g., "go right"). The user is able to modify both: the background color and the arrow direction (see **Table 5**).

3. Free condition building: there are blocks that have to be combined with condition and action. In these cases the user has to use at least three components (bridge-block, condition, and action) to define a conditional. For example, the Matatalab Coding Set contains a "wait until" block that should be combined with a condition (e.g., dark or light) and a sequence of actions in order to build conditionals.

We provide the description and graphical example for each tool that supports conditionals in three tables: **Table 4** gathers tools that implement the first level, **Table 5** corresponding to the second level, and **Table 6** corresponding to the last one.

The only tools that enable the definition of conditionals using logical operators (e.g., AND, OR) were Makeblock Neuron and Thymio. Neuron online mode allows users to program behaviors using Neuron App, which supports multiple conditions. In the case of Thymio, the user has to associate events sensed by the robot with its behavior. It is possible to combine the sensing and internal state of the robot (e.g., if Thymio touched AND internal state equal to 1) to program advanced robot responses.

In the case of BOTS, Move the turtle, and Makeblock Neuron + Neuron (app) conditionals are based on numerical variables (e.g., a > 5) which makes them more complex than conditionals with non-numerical conditions (e.g., "if the sensed color is red"), as the children have to understand the concept of variable.

In the case of Coding Awbie, the Caution Block is the only means to introduce conditionals into the code, and is a phased out feature as the block is not included in new kits (Getting Started with Osmo Coding Awbie Manual, 2022).

We also analyzed how the code related to a certain condition interacts with the main program, and identified that they occur within either event-based or procedural programming paradigms. Within event-based programming, we identified the following categories:

- Blocking event: the main program contains a condition that blocks the execution until the condition is fulfilled. For example, KIBO contains a "wait for clap" block that makes the robot wait for a clap before executing commands stored in the following blocks.
- Interruption: the main program is interrupted when a certain event occurs. For example, in the case of Pro-bot the main program is interrupted if the sound sensor is triggered and the procedure associated with this event is executed.
- Parallel execution: It is possible for an event to lead to actions to occur in parallel or in addition to those already occurring. For example, an event in Scratch Jr. could generate a sound while a sprite continues moving on the screen.

Using a procedural programming paradigm, we identified the following category:

- Integrated if: the main program contains conditions expressed using the "if" structure that is evaluated during the program's execution. For example, KIBO allows to incorporate an if-statement into the sequence of commands. If the condition that is evaluated is true, the conditional code is executed and then, the remaining statements.

3.2.2.2. Loops

Another control structure that was relevant for us to analyze was the availability of loops enabling the iteration of commands.

We observed two modalities of implementing the iteration of commands:

- Single command repetition: the tool does not provide the possibility to repeat a sequence of commands, it allows only the repetition of a single action.
- Multiple command repetition: it is possible to repeat multiple commands. In this category we find tools that, due to the design of loop structure, limit the number of pieces that can be repeated (e.g., in Kodable the user is allowed to repeat only two commands) and tools that do not have this restriction.

We also analyzed how the amount of repetitions can be expressed:

- Fixed number of repetitions: the number of repetitions is fixed and cannot be changed by the user.
- Configurable number of repetitions: the amount of repetitions can be defined by the user.
- Infinite loop: it is possible to build infinite loops.

We provide an example for each category in the **Table 7**.

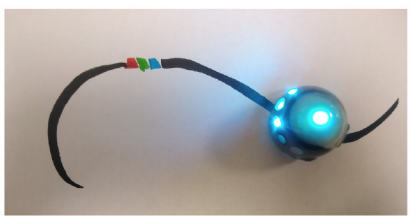
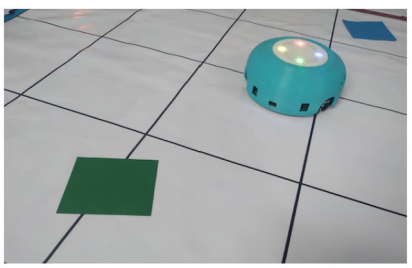
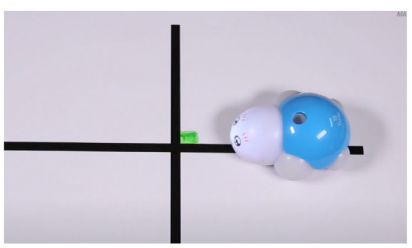
In most cases the amount of repetitions was expressed by associating the number of repetitions with a sequence of statements (similar to a for loop in more advanced programming languages), only BOTS uses exclusively conditions to stop the iteration process (similar to a while loop). KIBO, Finch, Run Marco!, Tynker: Coding for Kids, Scottie Go and VEX 123 offer both types ("repeat X times" and "repeat while") of repetition statements.

We found many different ways to implement infinite loops: using repeat forever (ScratchJr) or "go to start" command (VEX 123) at the end of the program, elements that contain pieces of code equivalent to "repeat forever" command (Roberto, Code.org), by building circular transitions between states (Dash and Dot), or by calling auxiliary functions (LightbotJr, RoboZZle).

3.3. Cost and Availability

Some tools that we analyzed are currently not available for sale: Plobot is a Kickstarter project that finished in Kickstarter (2022), Robotito, BOTS, Roberto, and T-Maze are academic developments, KIWI is KIBO's predecessor and is no longer manufactured, Makeblock Neuron and Puzzlets Starter Pack do not appear in online stores and CHERP is a programming language for KIBO and is not sold separately. All these tools were tagged as "unavailable."

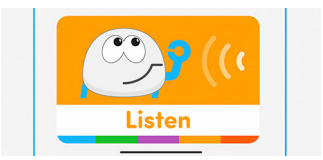
TABLE 4 | Tools that allow building conditionals categorized as "Predefined connection of condition and action".

Tool	Description	Reference image
Qobo	Specific card for conditional turning - if the robot passes over a card with a banana before passing over a bifurcation card, it turns left, but if it passes over a card with an apple, it turns right.	
Sphero indi	Color cards that the robot senses in the environment code robots' actions. Image provided by Sphero (2022).	
Ozobot Bit and Evo	Color lines that the robot senses in the environment code robots' actions.	
Robotito	Color cards that the robot senses in the environment code robots' actions.	
TurtleBot	Color codes that the robot senses in the environment code robots' actions.	

4. SLR OF EMPIRICAL EVIDENCE (REVIEW 2)

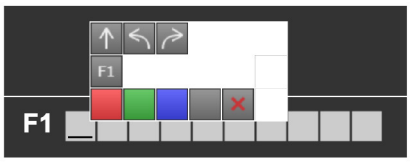
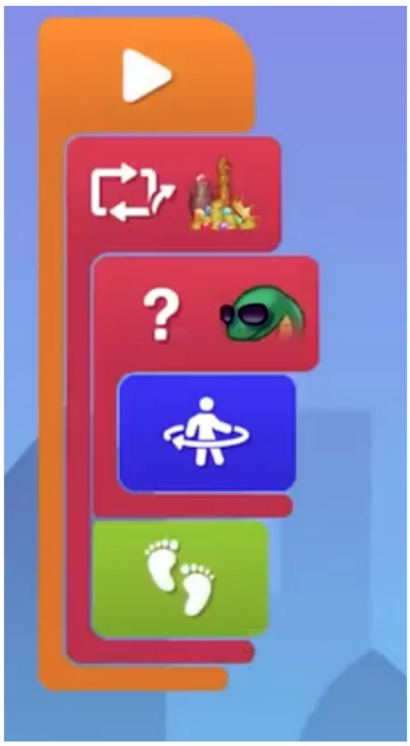
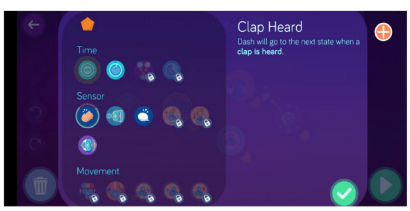
We conducted a second SLR (see **Figure 4**) to identify literature that reports empirical studies with tools that we considered relevant (see **Table 3**), in which control structures were taught and/or evaluated in order to respond the following research question: What tools have been reported to be successful for teaching control structures?

TABLE 5 | Tools that allow building conditionals categorized as "Free connection of predefined condition and predefined action".

Tool	Description	Reference image
KIBO	"Wait for clap" block stops the program execution until the clap is sensed.	
Botley	Botley's control provides an "object detection" button that is used to store the program that is executed when an obstacle is detected in front of the robot.	
Matatalab Coding Set	Two robots can send messages to each other. "Message received" block is used to define the robot's action when a message is received. The block is available in Matatalab Sensor Add-on (2022).	
Plobot	"Listen" card blocks the program execution until Plobot detects a sound louder than a soft clap.	
Pro-bot	Procedure numbers 33 to 37 are activated with sensors. For example, the procedure associated with a light sensor runs when the light sensor goes from dark to light.	
VEX 123	Control cards make use of sensors to check conditions.	
ScratchJr and Codeable Crafts	Events related to characters like "on bump" or "on tap" can be associated with actions.	

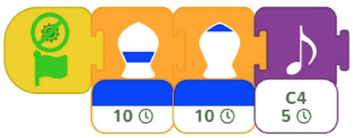
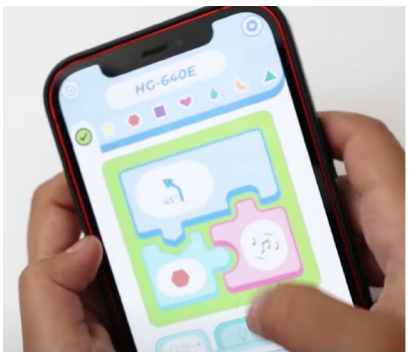
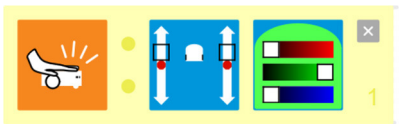
(Continued)

TABLE 5 | Continued

Tool	Description	Reference image
Kodable	The background color of the block defines the condition (e.g., "if the tile is pink") and the arrow, the action (e.g., "go right"). Image used with permission of Kodable (2022).	
RoboZZle	The background color of the block defines the condition (e.g., "if the tile is red") and the arrow, the action (e.g., "turn right").	
Tynker: Coding for Kids	Predefined condition (e.g., "if snake") can be combined with an action.	
COJI + COJI robot app	Predefined events can be associated with actions, for example, if the head is touched (event that activates procedure 1) - turn and sing (actions defined by the user).	No authorization for the image use.
Dash and Dot + Wonder for Dash & Dot Robots	Robot's actions are defined as states and the transition between can be fired based on conditions like "clap heard."	

(Continued)

TABLE 5 | Continued

Tool	Description	Reference image
Finch + Finchblox	Blocks attached to the "start when dark" block will be executed when the Finch detects that it is dark.	
Sphero indi + Sphero Edu Jr	Sphero Edu Jr application allows users to associate a color sensed by the robot with an action involving movement, light, and/or sound.	
Thymio + Thymio VPL	The user can associate events with actions.	
Coding Awbie	Caution Block enables a choice between two sets of sequences based on if there's an obstacle. Image can be found in Getting Started with Osmo Coding Awbie Manual (2022).	No authorization for the image use.
T-maze	"In a program execution, when the avatar reaches one of these squares in the maze, the child must do something with the sensors (e.g., cover a light sensor) to allow the avatar to proceed" Wang et al. (2014).	

4.1. Methodology

Two reviewers reviewed abstracts and tagged them as "irrelevant" or "relevant." The last category was used in the cases of publications that meet inclusion criteria or when it was impossible to evaluate the article relevance based on the information available in the abstract. One reviewer reviewed studies that were classified differently among reviewers in the previous step and tried to resolve the doubtful cases. If it was impossible, the articles were considered as "relevant" cases. One reviewer reviewed full-texts of relevant publications and took the final decision about their relevance for this study. We decided not to carry out any quality assessment of the studies as we wanted to provide a broad view of the existing empirical evidence. Two reviewers extracted the data.

4.1.1. Search Strategy

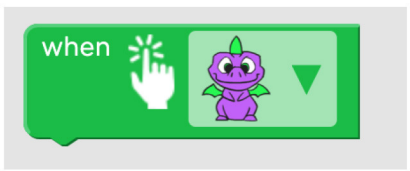
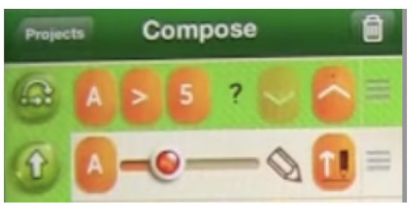
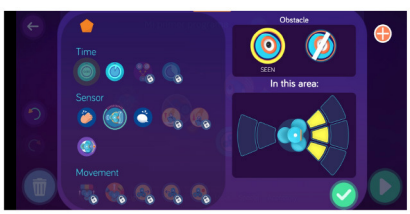
We used an automated search (Kitchenham et al., 2015) in Scopus search engine (Elsevier Scopus, 2022) to identify empirical studies with preschoolers that were developed using tools that we considered relevant (see **Table 3**). The search term was the following:

TITLE-ABS-KEY ((({**Tool name**} AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))

It had two keywords: tool name and young learners (and synonyms) and was used to search in title, abstract and keywords.

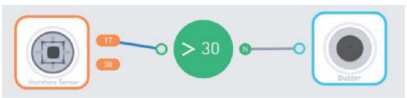
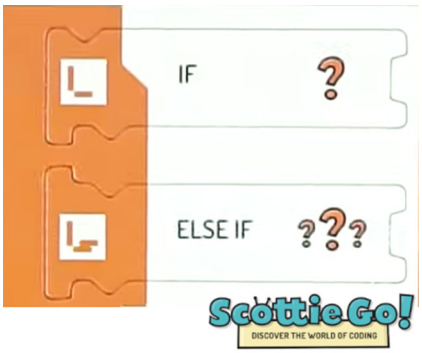
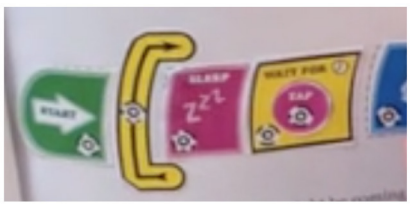
In some cases we used curly brackets, that limit the search to exact words, ignoring spelling variation or plurals, around the name of the tool ({Tool name}) to avoid false positive results. For example, in the case of "Coffee Platform" when we used Coffee AND Platform instead of Coffee Platform, the results contained irrelevant publications that did not target the robotic platform. In some cases we excluded publications from areas related to medicine, as some tools' names were equal to terms used in medicine and also brought false positive results (as in the case of

TABLE 6 | Tools that allow building conditionals categorized as "Free condition building".

Tool	Description	Reference image
KIBO	"If" block provides place to add a condition (e.g., far, near, dark, light).	
Matatalab Coding Set	"Wait until" can be connected with conditions like: dark, light, obstacle, etc. The block is available in Matatalab Sensor Add-on (2022).	
BOTS	"If" block should be associated with variable comparison (e.g., a > 5).	
Code.org	The condition in "when tapped" can be modified.	
Move the turtle	Condition block evaluates the value of a variable (A > 5).	
Run Marco!	"If" block can be modified.	
The Foos	The condition is variable and can be changed by the user. A video reference of the implementation can be found on CodeSpark Academy Youtube Channel (2022).	No authorization for the image use.
Dash and Dot + Wonder for Dash & Dot Robots	Robot's actions are defined as states and the transition between can be fired based on conditions like "obstacle detection" that can be customized (obstacle seen vs no obstacle, obstacle seen close vs far).	

(Continued)

TABLE 6 | Continued

Tool	Description	Reference image
Makeblock Neuron + Neuron (app)	Users can define conditions to establish relations between sensors and actuators.	
Scottie Go	"If" block should be associated with a specific condition.	
Roberto	"Wait for" can be combined with "tap" event.	

T-Maze). The search term used and the amount of publications found with each tool can be consulted in **appendix**.

4.1.2. Study Selection

The inclusion criteria for the studies' selection were the following:

- Articles that report empirical studies with young children using an electronic-based tool that enables activities with control structures.
- Publications that report activities or evaluations focused on control structures.
- Publications focused on children between 3 and 5 years old, including 6 years old, if attending pre-primary school educational level.

Exclusion criteria were:

- Publications that target children older than 6 years.
- Publications that do not report activities or evaluations focused on control structures.
- Off topic articles.
- Articles that describe experiences with users with neurodevelopmental disorders.
- Articles written in a language other than English or Spanish.
- Conference proceedings.

4.1.3. Data Extraction

In the data extraction step we used a spreadsheet to collect information related to the age of participants, number of participants, type of the study, learning outcome, activities aimed at programming conditions, activities that incorporate iterations. Based on the extracted data, two researchers conducted a thematic analysis to summarize study results.

4.2. Findings

4.2.1. Scopus Search Result

The Scopus search for all tools was conducted on 13th of October 2021. In many cases the search brought no results. Only 26 tools of 44 unique tools that we identified, counted with Scopus entries (see **Appendix**). A total of 205 (202 unique) publications were analyzed. Three publications appeared as repeated because the research that they described involved two relevant tools, for example, Pugnali et al.'s research involved KIBO and ScratchJr, so it was found under the search query for KIBO and ScratchJr. We identified 24 unique publications (see **Table 8**) that met all inclusion criteria. In the screening phase the reviewers identically tagged 152 of 202 unique articles reaching an agreement of 75%.

The 24 relevant publications reported experiences with 10 different tools that we identified as relevant: ScratchJr (evaluated in 7 articles), KIBO (8), KIWI (2), CHERP (2), Code.org (2), BOTS (1), Kodable (1), Move the turtle (1), Strawbies (1) and T-maze (1). Strawbies is an alternative name for Coding Awbie that was used for the search, as the search term with "Coding Awbie" brought no results. Daisy the Dinosaur was mentioned in a study related to Kodable (Pila et al., 2019), but it targets older children (see **Table 2**). We also found one case of a custom tool

TABLE 7 | Examples of tools for categories developed to classify code iteration.

Category	Description	Reference image
Single command repetition	ScratchJr direction blocks can be modified to make more than one step using single block.	
Multiple command repetition	Kodable allows to repeat two commands.	
Fixed number of repetitions	Qobo coding card with fixed number of repetitions.	
Configurable number of repetitions	Finchblox allows to modify the number of repetitions.	
Infinite loop	KIBO allows to associate the repeat block with an infinity symbol.	

(Rose et al., 2017): a game with both ScratchJr-like and Lightbot style programming interface.

4.2.2. Thematic Analysis

4.2.2.1. KIBO/CHERP/KIWI Articles

The only set of technologies for which control structures have been evaluated multiple times was KIBO/CHERP/KIWI, developed by Marina Bers' group at Tufts University. Of the articles we identified using this technology, five evaluated children's use of control structures while separating the performance of young children from that of older children, or only including children within our inclusion criteria. All these evaluations used the Solve-It assessments, which were developed by the same research group. Through these assessments, in four of the publications, children who fit our inclusion criteria demonstrated proficiency when programming repeat loops (with a given number of repetitions) and "wait for clap" programs, but were not tested on or were unable to be proficient in the use of sensor loops or conditionals (Strawhacker and Bers, 2015; Elkin et al., 2016; Sullivan and Bers, 2016b; Bers et al., 2019). There was one outlying study where children in Kindergarten were able to demonstrate proficiency across all Solve It assessment areas, including repeat loops, sensor loops, "wait for clap" programs, and conditionals (Sullivan and Bers, 2018). Four other evaluations of this tool did not include specific evaluations of control flow (Kazakoff and Bers, 2014; Sullivan et al., 2017; Bers, 2019; Jurado et al., 2020) while two others did not separate children in our age range of interest from older children.

TABLE 8 | 24 relevant publications that we identified in the second SLR.

References	Title	Tool name	Type of tool	Age of participants	Number of participants
Jurado et al. (2020)	Social steam learning at an early age with robotic platforms: A case study in four schools in Spain	KIBO	Physical	4–6	65
Bers (2019)	Coding as another language: a pedagogical approach for teaching computer science in early childhood	KIBO, Scratch Jr	Physical, virtual	4–7	at least 9
Sullivan and Bers (2019)	Investigating the use of robotics to increase girls' interest in engineering during early elementary school	KIBO	Physical	5–7	105
Bers et al. (2019)	Coding as a playground: Promoting positive learning experiences in childhood classrooms	KIBO	Physical	3–5	172
Sullivan and Bers (2018)	Dancing robots: integrating art, music, and robotics in Singapore's early childhood centers	KIBO	Physical	3–6	98
Sullivan et al. (2017)	Imagining, playing, and coding with kibo: Using robotics to foster computational thinking in young children	KIBO	Physical	3–7	322
Pugnali et al. (2017)	THE impact of user interface on young children's computational thinking	KIBO, Scratch Jr	Physical, virtual	4–7	28
Elkin et al. (2016)	Programming with the KIBO Robotics Kit in Preschool Classrooms	KIBO	Physical	3–5	64
Sullivan and Bers (2016b)	Robotics in the early childhood classroom: learning outcomes from an 8–week robotics curriculum in pre–kindergarten through second grade	KIWI	Physical	4–7	60
Sullivan and Bers (2016a)	Girls, boys, and bots: Gender differences in young children's performance on robotics and programming tasks	KIWI, BOTS	Physical, virtual	4–7	45
Strawhacker and Bers (2015)	"I want my robot to look for food": Comparing Kindergartner's programming comprehension using tangible, graphic, and hybrid user interfaces	CHERP	Hybrid	5–6	35
Kazakoff and Bers (2014)	Put your robot in, put your robot out: Sequencing through programming robots in early childhood	CHERP	Hybrid	4–6	34
Arfé et al. (2020)	The effects of coding on children's planning and inhibition skills	Code.org	Virtual	5–6	179
Çiftci and Bildiren (2020)	The effect of coding courses on the cognitive abilities and problem–solving skills of preschool children	Code.org	Virtual	4–5	28
Pila et al. (2019)	Learning to code via tablet applications: An evaluation of Daisy the Dinosaur and Kodable as learning tools for young children	Kodable, Daisy the Dinosaur	Virtual	4–6	28
Jung et al. (2019)	TurtleTalk: An educational programming game for children with voice user interface	Move the turtle	Virtual	6–9	8
Strawhacker and Bers (2019)	What they learn when they learn coding: investigating cognitive domains and computer programming knowledge in young children	ScratchJr	Virtual	5–8	57
Pinto and Osório (2019)	Learn to program in preschool: Analysis with the participation scale [Aprender a programar en educación infantil: Análisis con la escala de participación]	ScratchJr	Virtual	3–6	71
Strawhacker et al. (2018)	Teaching tools, teachers' rules: exploring the impact of teaching styles on young children's programming knowledge in ScratchJr	ScratchJr	Virtual	5–7	222
Rose et al. (2017)	An exploration of the role of visual programming tools in the development of young children's computational thinking	Game with ScratchJr–and Lightbot–like programming interface	Virtual	6–7	40
Portelance et al. (2016)	Constructing the ScratchJr programming language in the early childhood classroom	ScratchJr	Virtual	5–7	62
Papadakis et al. (2016)	Developing fundamental programming concepts and computational thinking with ScratchJr in preschool education: A case study	ScratchJr	Virtual	4–6	43
Hu et al. (2015)	Strawbies: Explorations in tangible programming	Strawbies	Hybrid	4–10	No info
Wang et al. (2014)	A tangible programming tool for children to cultivate computational thinking	T–maze	Hybrid	5–9	20

4.2.2.2. Scratch Jr and Others

Most of the other evaluations involved Scratch Jr. (Papadakis et al., 2016; Portelance et al., 2016; Strawhacker et al., 2018; Pinto and Osório, 2019) and did not evaluate children's use or understanding of control structures, even though the tool enables the use of control structures. The same happened with evaluations of other systems (Wang et al., 2014; Hu et al., 2015; Rose et al., 2017; Jung et al., 2019; Pila et al., 2019; Arfé et al., 2020; Çiftci and Bildiren, 2020). The evaluations that did include reports on the use of control

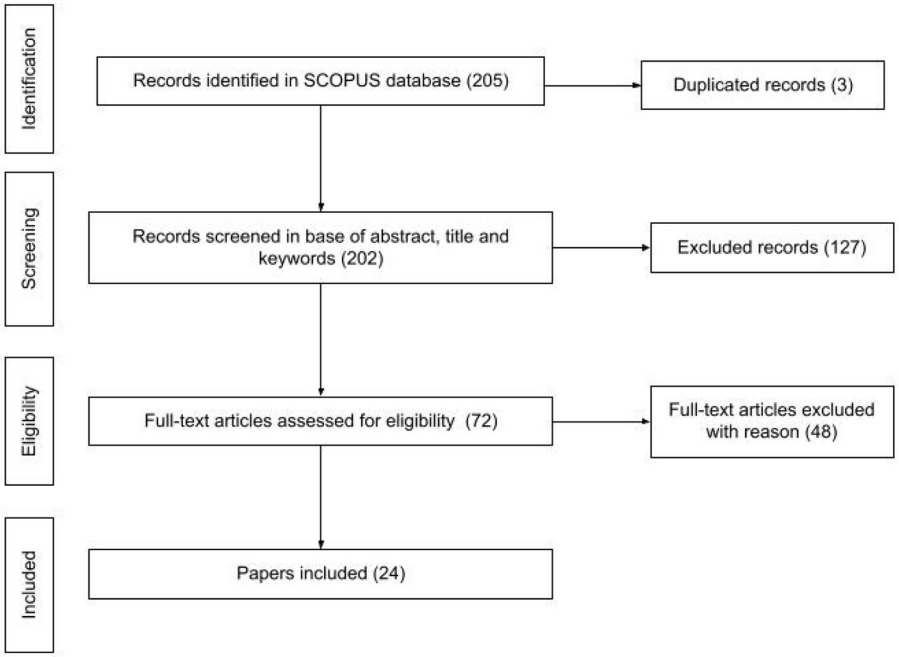


FIGURE 4 | Steps of the selection process of the second SLR. Reported in line with the PRISMA statement (Moher et al., 2009).

structures, without an evaluation, involving Scratch Jr., reported either little use or difficulty with control flow blocks (Pugnali et al., 2017; Strawhacker and Bers, 2019). Another included children in our target age, but also older children without separating their performance (Pugnali et al., 2017). One evaluation of LEGO WeDo found some success with repeat loops, but greater success with CHERP (Strawhacker and Bers, 2015).

4.2.2.3. Bottom Line

Only one study (Sullivan and Bers, 2018) provides evidence of children in Kindergarten mastering conditionals and sensor loops. Multiple studies provide evidence of children in our target age group mastering the use of simple repeat loops (repeat a given # of times) or wait for clap programs. The caveat with all these studies is that they are all from the same research group, use the same system, and the same assessment.

With other tools, except for a study of Lego WeDo which also included CHERP (Strawhacker and Bers, 2015), there are no specific assessments of control flow, other than reports of low use or difficulty with using control flow structures for children in our target age range. In other words, in spite of the great diversity of options for children in our target age range to learn about control flow structures, in our review we found only one technology for which there have been multiple empirical studies to understand whether these children can learn how to use these features.

5. LIMITATIONS

Although we tried to carry out our study in a systematic way, document all the decisions, and report doubtful cases, the current study still has certain limitations. To complement the tools characteristics related to control structures and cost, we had to appeal to online information. We firstly reviewed official websites and online user manuals, but in some cases the information contained in these sources was not sufficient to answer our research questions. In those cases we reviewed unofficial sources such as youtube videos, blogs and private web pages to complete the missing information. We understand that these are not the most convenient information sources, but we used them if there was no available information through official channels. Another limitation related to our online search is that we reported information that we were able to find, which does not ensure that it is the complete existing information. For example, we reported that the application The Foos allows users to build conditionals of "Free condition building" type based on a youtube video that we found, but we cannot ensure that the tool does not allow building other types of conditionals. There is no free online manual that could provide required information, so to confirm that "Free condition building" is the only type that the tool supports it is necessary to pass all the levels that the game offers, and it was impossible for our team to acquire and personally analyze all the relevant tools. Also, our initial list of tools for young children is limited to the tools reported in scientific publications. It is possible that there are valuable tools that were not mentioned in reviews that we analyzed. We tried to address this issue by adding

3 publications that were not found by SLR and by adding four tools that we found in external sources.

6. DISCUSSION

The present study reviewed the state of the art in the teaching of control structures to young children, specifically preliterate children 3 to 6 years of age. While many of the definitions of CT for young children which gather large amounts of consensus amongst academics describe control structures such as conditionals and loops amongst central aspects of CT (Brennan and Resnick, 2012; Grover and Pea, 2013), how this aspect of CT should be developmentally adapted for young children remains unclear. Our findings suggest there is still a large knowledge gap regarding how children acquire early notions about control structures and what the best tools are to introduce children to these concepts. Despite this, these concepts are often included in the interventions targeted at young children and assessed through specific items in the validated CT tests available for young children (Relkin et al., 2020; Zapata-Cáceres et al., 2020).

Our findings demonstrate that there is a wide variety of technological tools which include robots, virtual applications and hybrids, which aim to teach control structures and are targeted to children of these ages. Thus, we infer it is considered relevant that children acquire these concepts early on. Despite this, our findings regarding the reported classroom based research shows that the specifics of how children learn these concepts through the available tools remains unexplored. None of the systematic review articles we identified presented results that were specific to control structures, instead focusing on broader concepts such as CT (Sullivan et al., 2017), programming literacy (Bers, 2019), or engagement (Pinto and Osório, 2019). Given that CT is an umbrella term which encompasses a wide variety of components such as sequencing, using control structures, abstraction, debugging, amongst others (Shute et al., 2017) we must focus on the specifics of each of them in order to have a better sense of the concept as a whole. This is especially relevant for younger children, as the learning curves for each specific skill might differ with age. So far, we found most of the studies focus on several concepts at once but do not further explore learning outcomes for each activity. Thus, the assessments used were more holistic and successful in detecting general learning and engagement outcomes but lacked information on each of the specific tasks and concepts encompassed. An exception to this general approach was the study reported by Kazakoff and Bers (2014) where they focused specifically on sequencing skills, however we did not find any similar study for the learning of control structures, even though our search targeted this term specifically.

Exploring these aspects is also necessary to determine which approaches provide the adequate affordances to enhance learning of each aspect of CT. For example, in our technology overview we observed several approaches to including the use of control structures in tools, such as interrupting events, active wait, or procedural conditions, however there are currently no studies contrasting the strengths and weaknesses of each of these approaches and whether they produce different results in children's understanding of the concepts. As a result, there is only evidence of one tool successfully enabling children to learn some aspects of control structures, mainly due to a lack of studies on the use of other tools by young children that include an assessment of control structure use or understanding.

Moreover, future studies on specific tools should focus on the feasibility of their inclusion in the classrooms in a scalable way. Specifically, our findings regarding the cost of several robots suggest some of them are simply too expensive to be available to all children in a given school or classroom. In addition, some of these tools are more adequately design for individual at-home use, which hinders group based-activities thus elevates the cost of its use even more. Thus, so far the use of robots in education at a large-scale would a entail substantial investment for administrators and policy makers, a problem which could be partially subsided through the design of tools with a group-based focus.

The results of our systematic reviews therefore are encouraging in terms of the wide range of approaches designed for young children to learn about control structures, but also identify a large gap in that we know very little about which of these approaches may work better, or how to structure their use. There is therefore a need for future research to further explore the strengths and weaknesses of the available approaches and understand the feasibility of their use in a variety of contexts (e.g., individual vs. shared, home vs. school).

7. CONCLUSION

The present work demonstrates that there are many diverse tools to support the development of CT in young children. It seems that both academia and industry have interest in designing approaches to enable young children to develop this so-called twenty-first century skill, as we found through our systematic reviews. Although many existing tools allow children to approach advanced programming concepts such as control structures, it is not clear which tools and activities are the most appropriate for teaching them to the youngest programmers. In order to lay the basis for the future research that targets this gap, we provide a systematic overview of existing tools for preliterate children between the ages of 3 and 6. We developed categories that classify the type and complexity of conditionals and iteration structures and used them to categorize each tool. We also provided graphical examples of conditionals that the tools provide.

The analysis of empirical evidence showed that KIBO/CHERP/KIWI is the only tool that consistently demonstrates positive results in teaching control structures to young children. Other tools in our review have not gone through similar evaluations, making it difficult to reach conclusions about their appropriateness for introducing these concepts. The contrast between the diversity of approaches available and the scarcity of evaluations focused on control structures calls for more research, ideally by groups independent

of the tools being evaluated, to compare and contrast these approaches in a variety of contexts (e.g., home, preschool).

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

EB, AG, JH, and GTe contributed to conception and design of the study. EB organized the first draft of the manuscript and extracted the tools. EB, AG, and JH wrote sections of the manuscript. EB and AG analyzed the articles involved in the first SLR and reviewed tools characteristics. GTe and JH supervised the revision process. GTr and KP analyzed the articles involved in the second SLR. KP and JH extracted the data and conducted the thematic analysis. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This review is a part of a Ph.D. project of the EB that was supported by Comisión Académica de Posgrado (Uruguay).

REFERENCES

Arfé, B., Vardanega, T., and Ronconi, L. (2020). The effects of coding on children's planning and inhibition skills. *Comput. Educ.* 148, 103807. doi: 10.1016/j.compedu.2020.103807

Bakala, E., Gerosa, A., Hourcade, J. P., and Tejera, G. (2021). Preschool children, robots, and computational thinking: a systematic review. *Int. J. Child Comput. Interact.* 29, 100337. doi: 10.1016/j.ijcci.2021.100337

Bers, M. U. (2019). Coding as another language: a pedagogical approach for teaching computer science in early childhood. *J. Comput. Educ.* 6, 499–528. doi: 10.1007/s40692-019-00147-3

Bers, M. U. (2020). *Coding as a Playground: Programming and Computational Thinking in the Early Childhood Classroom*. New York, NY: Routledge.

Bers, M. U., González-González, C., and Armas-Torres, M. B. (2019). Coding as a playground: promoting positive learning experiences in childhood classrooms. *Comput. Educ.* 138, 130–145. doi: 10.1016/j.compedu.2019.04.013

Bocconi, S., Chioccariello, A., Dettori, G., Ferrari, A., Engelhardt, K., et al. (2016). *Developing computational thinking in compulsory education-implications for policy and practice*. Technical report, Joint Research Centre (Seville site).

Brackmann, C., Barone, D., Casali, A., Boucinha, R., and Muñoz-Hernandez, S. (2016). "Computational thinking: panorama of the americas," in *2016 International Symposium on Computers in Education (SIIE)* (Salamanca: IEEE), 1–6.

Brennan, K., and Resnick, M. (2012). "New frameworks for studying and assessing the development of computational thinking," in *Proceedings of the 2012 annual meeting of the American Educational Research Association, Vancouver, Canada (AERA 2012)* (Vancouver, BC), 1–25.

Ching, Y. H., Hsu, Y. C., and Baldwin, S. (2018). Developing computational thinking with educational technologies for young learners. *TechTrends* 62, 563–573. doi: 10.1007/s11528-018-0292-7

Çiftci, S., and Bildiren, A. (2020). The effect of coding courses on the cognitive abilities and problem-solving skills of preschool children. *Comput. Sci. Educ.* 30, 3–21. doi: 10.1080/08993408.2019.1696169

CodeSpark Academy Youtube Channel (2022). *codeSpark Youtube Channel*. Available online at: https://www.youtube.com/watch?v=mSv9tsF3G2M (accessed September 6, 2022).

Elkin, M., Sullivan, A., and Bers, M. U. (2016). Programming with the kibo robotics kit in preschool classrooms. *Comput. Sch.* 33, 169–186. doi: 10.1080/07380569.2016.1216251

Elsevier Scopus (2022). *Elsevier Scopus: The Largest Database of Peer-Reviewed Literature*. Available online at: https://www.elsevier.com/en-in/solutions/scopus (accessed March 20, 2022).

Fagerlund, J., Häkkinen, P., Vesisenaho, M., and Viiri, J. (2021). Computational thinking in programming with Scratch in primary schools: a systematic review. *Comput. Appl. Eng. Educ.* 29, 12–28. doi: 10.1002/cae.22255

Gerosa, A., Koleszar, V., Tejera, G., Gómez-Sena, L., and Carboni, A. (2021). Cognitive abilities and computational thinking at age 5: evidence for associations to sequencing and symbolic number comparison. *Comput. Educ. Open* 2, 100043. doi: 10.1016/j.caeo.2021.100043

Getting Started with Osmo Coding Awbie Manual (2022). *Getting Started with Osmo Coding Awbie*. Available online at: https://assets.playosmo.com/static/downloads/GettingStartedWithOsmoCodingAwbie.pdf (accessed March 20, 2022).

Grover, S. (2018). *A Tale of Two CTS (and a Revised Timeline for Computational Thinking)*. Blog@ CACM.

Grover, S., and Pea, R. (2013). Computational thinking in k-12: a review of the state of the field. *Educ. Res.* 42, 38–43. doi: 10.3102/0013189X12463051

Hickmott, D., Prieto-Rodriguez, E., and Holmes, K. (2018). A scoping review of studies on computational thinking in k-12 mathematics classrooms. *Digit. Exp. Math. Educ.* 4, 48–69. doi: 10.1007/s40751-017-0038-8

Hsu, T. C., Chang, S. C., and Hung, Y. T. (2018). How to learn and how to teach computational thinking: suggestions based on a review of the literature. *Comput. Educ.* 126, 296–310. doi: 10.1016/j.compedu.2018.07.004

Hu, F., Zekelman, A., Horn, M., and Judd, F. (2015). "Strawbies: explorations in tangible programming," in *Proceedings of the 14th International Conference on Interaction Design and Children* (New York, NY), 410–413.

Ioannou, A., and Makridou, E. (2018). Exploring the potentials of educational robotics in the development of computational thinking: a summary of current research and practical proposal for future work. *Educ. Inf. Technol.* 23, 2531–2544. doi: 10.1007/s10639-018-9729-z

Isnaini, R., Budiyanto, C., and Widiastuti, I. (2019). "Robotics-based learning to support computational thinking skills in early childhood," in *AIP Conference Proceedings* (Surakarta).

Jung, H., Kim, H. J., So, S., Kim, J., and Oh, C. (2019). "Turtletalk: an educational programming game for children with voice user interface," in *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow), 1–6.

Jurado, E., Fonseca, D., Coderch, J., and Canaleta, X. (2020). Social steam learning at an early age with robotic platforms: a case study in four schools in spain. *Sensors* 20, 3698. doi: 10.3390/s20133698

Kakavas, P., and Ugolini, F. C. (2019). Computational thinking in primary education: a systematic literature review. *Res. Educ. Media* 11, 2019. doi: 10.2478/rem-2019-0023

Kazakoff, E. R., and Bers, M. U. (2014). Put your robot in, put your robot out: sequencing through programming robots in early childhood. *J. Educ. Comput. Res.* 50, 553–573. doi: 10.2190/EC.50.4.f

Khoo, K. Y. (2020). A case study on how children develop computational thinking collaboratively with robotics toys. *Int. J. Educ. Technol. Learn.* 9, 39–51. doi: 10.20448/2003.91.39.51

Kickstarter (2022). *Plobot | Physical Coding for Ages 4 and up*. Available online at: https://www.kickstarter.com/projects/1365116576/plobot-a-robot-coding-companion (accessed March 20, 2022).

Kitchenham, B. A., Budgen, D., and Brereton, P. (2015). *Evidence-based software engineering and systematic reviews, volume 4*. Boca Raton, FL: CRC Press.

Kodable (2022). *Kodable.* Available online at: https://www.kodable.com/ (accessed September 6, 2022).

Lyon, J. A., and Magana, A. J. (2020). Computational thinking in higher education: a review of the literature. *Comput. Appl. Eng. Educ*. 28, 1174–1189. doi: 10.1002/cae.22295

Manches, A., and Plowman, L. (2017). Computing education in children's early years: a call for debate. *Br. J. Educ. Technol*. 48, 191–201. doi: 10.1111/bjet.12355

Martin, C. K., Pinkard, N., Erete, S., and Sandherr, J. (2017). "Connections at the family level: supporting parents and caring adults to engage youth in learning about computers and technology," in *Moving Students of Color From Consumers to Producers of Technology* (IGI Global), 220–244.

Matatalab Sensor Add-on (2022). *Matatalab*. Available online at: https://matatalab.com/en/sensor-add-on (accessed September 6, 2022).

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., and the, P. R. I. S. M. A., Group (2009). Reprint–preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Phys. Ther*. 89, 873–880. doi: 10.1093/ptj/89.9.873

OECD (2019). *Oecd Education and Skills Today*. Available online at: https://oecdedutoday.com/computer-science-and-pisa-2021/ (accessed March 20, 2022).

Orton, K., Weintrop, D., Beheshti, E., Horn, M., Jona, K., and Wilensky, U. (2016). *Bringing Computational Thinking Into High School Mathematics and Science Classrooms*. Singapore: International Society of the Learning Sciences.

Papadakis, S. (2020). Robots and robotics kits for early childhood and first school age. *Int. J. Interact. Mobile Technol*. 14, 34–56. doi: 10.3991/ijim.v14i18.16631

Papadakis, S. (2021). The impact of coding apps to support young children in computational thinking and computational fluency. A literature review. *Front. Educ*. 6, 657895. doi: 10.3389/feduc.2021.657895

Papadakis, S., Kalogiannakis, M., and Zaranis, N. (2016). Developing fundamental programming concepts and computational thinking with scratchjr in preschool education: a case study. *Int. J. Mobile Learn. Organ*. 10, 187–202. doi: 10.1504/IJMLO.2016.077867

Pedersen, B. K. M. K., Larsen, J. C., and Nielsen, J. (2020). The effect of commercially available educational robotics: a systematic review. *Adv. Intell. Syst. Comput*. 1023, 14–27. doi: 10.1007/978-3-030-26945-6_2

Perkovic, L. (2015). *Introduction to Computing Using Python: An Application Development Focus*. John Wiley & Sons.

Pila, S., Aladé, F., Sheehan, K. J., Lauricella, A. R., and Wartella, E. A. (2019). Learning to code via tablet applications: an evaluation of daisy the dinosaur and kodable as learning tools for young children. *Comput. Educ*. 128, 52–62. doi: 10.1016/j.compedu.2018.09.006

Pinto, M. S. M., and Osorio, A. (2019). Aprender a programar en Educacion Infantil: analisis con la escala de participacion. *Editorial Universidad de Sevilla* 55, 133–156. doi: 10.12795/pixelbit.2019.i55.08

Portelance, D. J., Strawhacker, A. L., and Bers, M. U. (2016). Constructing the scratchjr programming language in the early childhood classroom. *Int. J. Technol. Design Educ*. 26, 489–504. doi: 10.1007/s10798-015-9325-0

Pugnali, A., Sullivan, A., and Bers, M. U. (2017). The impact of user interface on young children's computational thinking. *J. Inf. Technol. Educ*. 16, 171–193. doi: 10.28945/3768

Relkin, E., de Ruiter, L., and Bers, M. U. (2020). TechCheck: development and validation of an unplugged assessment of computational thinking in early childhood education. *J. Sci. Educ. Technol*. 29, 482–498. doi: 10.1007/s10956-020-09831-x

Robertson, J., Gray, S., Martin, T., and Booth, J. (2020). The relationship between executive functions and computational thinking. *Int. J. Comput. Sci. Educ. Sch*. 3, 35–49. doi: 10.21585/ijcses.v3i4.76

RoboCupJunior (2022). *Robocupjunior About*. Available online at: https://junior.robocup.org/about/ (accessed March 20, 2022).

Román-González, M. (2015). "Computational thinking test: design guidelines and content validation," in *Proceedings of EDULEARN15 Conference* (Barcelona), 2436–2444.

Román-González, M., Pérez-González, J. -C., and Jiménez-Fernández, C. (2017). Which cognitive abilities underlie computational thinking? criterion validity of the computational thinking test. *Comput. Hum. Behav*. 72, 678–691. doi: 10.1016/j.chb.2016.08.047

Rose, S., Habgood, M. J., and Jay, T. (2017). An exploration of the role of visual programming tools in the development of young children's computational thinking. *Electron. J. e-Learn*. 15, 297–309. doi: 10.34190/ejel.15.4.2368

Sanford, J. F., and Naidu, J. T. (2016). Computational thinking concepts for grade school. *Contemporary Issues Educ. Res*. 9, 23–32. doi: 10.19030/cier.v9i1.9547

Saxena, A., Lo, C. K., Hew, K. F., and Wong, G. K. W. (2020). Designing unplugged and plugged activities to cultivate computational thinking: an exploratory study in early childhood education. *Asia-Pacific Educ. Res*. 29, 55–66. doi: 10.1007/s40299-019-00478-w

Shute, V. J., Sun, C., and Asbell-Clarke, J. (2017). Demystifying computational thinking. *Educational Research Review* 22, 142–158. doi: 10.1016/j.edurev.2017.09.003

Silva, E. F., Dembogurski, B. J., and Semaan, G. S. (2021). A systematic review of computational thinking in early ages. *ArXiv, abs/2106.10275*. doi: 10.48550/arXiv.2106.10275

Sphero (2022). *Sphero brandfolder Indi*. Available online at: https://brandfolder.com/indi (accessed September 6, 2022).

Stamatios, P. (2022). Can preschoolers learn computational thinking and coding skills with scratchjr? A systematic literature review. *Int. J. Educ. Reform*. doi: 10.1177/10567879221076077

Strawhacker, A., and Bers, M. U. (2015). "I want my robot to look for food": comparing Kindergartner's programming comprehension using tangible, graphic, and hybrid user interfaces. *Int. J. Technol. Design Educ*. 25, 293–319. doi: 10.1007/s10798-014-9287-7

Strawhacker, A., and Bers, M. U. (2019). What they learn when they learn coding: investigating cognitive domains and computer programming knowledge in young children. *Educ. Technol. Res. Dev*. 67, 541–575. doi: 10.1007/s11423-018-9622-x

Strawhacker, A., Lee, M., and Bers, M. U. (2018). Teaching tools, teachers' rules: exploring the impact of teaching styles on young children's programming knowledge in scratchjr. *Int. J. Technol. Design Educ*. 28, 347–376. doi: 10.1007/s10798-017-9400-9

Sullivan, A., and Bers, M. U. (2016a). Girls, boys, and bots: Gender differences in young children's performance on robotics and programming tasks. *J. Inf. Technol. Educ.* 15, 145–165. doi: 10.28945/3547

Sullivan, A., and Bers, M. U. (2016b). Robotics in the early childhood classroom: learning outcomes from an 8-week robotics curriculum in pre-kindergarten through second grade. *Int. J. Technol. Design Educ*. 26, 3–20. doi: 10.1007/s10798-015-9304-5

Sullivan, A., and Bers, M. U. (2018). Dancing robots: integrating art, music, and robotics in singapore's early childhood centers. *Int. J. Technol. Design Educ*. 28, 325–346. doi: 10.1007/s10798-017-9397-0

Sullivan, A., and Bers, M. U. (2019). Investigating the use of robotics to increase girls' interest in engineering during early elementary school. *Int. J. Technol. Design Educ*. 29, 1033–1051. doi: 10.1007/s10798-018-9483-y

Sullivan, A. A., Bers, M. U., and Mihm, C. (2017). *Imagining, Playing, and Coding With Kibo: Using Robotics to Foster Computational Thinking in Young Children*. Hong Kong: Siu-cheung KONG The Education University of Hong Kong.

Tang, X., Yin, Y., Lin, Q., Hadad, R., and Zhai, X. (2020). Assessing computational thinking: a systematic review of empirical studies. *Comput. Educ*. 148, 103798. doi: 10.1016/j.compedu.2019.103798

Taslibeyaz, E., Kursun, E., and Karaman, S. (2020). How to develop computational thinking: a systematic review of empirical studies. *Inform. Educ*. 19, 701–719. doi: 10.15388/infedu.2020.30

Toh, L. P. E., Causo, A., Tzuo, P. W., Chen, I. M., and Yeo, S. H. (2016). A review on the use of robots in education and young children. *Educ. Technol. Soc*. 19, 148–163.

Tsarava, K., Moeller, K., Román-González, M., Golle, J., Leifheit, L., Butz, M. V., et al. (2022). A cognitive definition of computational thinking in primary education. *Comput. Educ*. 179, 104425. doi: 10.1016/j.compedu.2021.104425

Tufts University. (2022). *Kibo Robot*. Available online at: https://sites.tufts.edu/devtech/research/kibo-robot/ (accessed March 20, 2022).

Umam, M. U. K., Budiyanto, C., and Rahmawati, A. (2019). "Literature review of robotics learning devices to facilitate the development of computational thinking in early childhood," in *AIP Conference Proceedings*.

Wang, D., Wang, T., and Liu, Z. (2014). A tangible programming tool for children to cultivate computational thinking. *Scientific World J*. 2014, 428080. doi: 10.1155/2014/428080

Weintrop, D., Beheshti, E., Horn, M., Orton, K., Jona, K., Trouille, L., et al. (2016). Defining computational thinking for mathematics and science classrooms. *J. Sci. Educ. Technol*. 25, 127–147. doi: 10.1007/s10956-015-9581-5

Williamson, B. (2016). Political computational thinking: policy networks, digital governance and 'learning to code'. *Crit. Policy Stud*. 10, 39–58. doi: 10.1080/19460171.2015.1052003

Wing, J. (2011). Research notebook: computational thinking–what and why. *Link Mag*. 6, 20–23.

Wing, J. M. (2006). Computational thinking. *Commun. ACM*. 24, 33. doi: 10.1145/1118178.1118215

Yadav, A., Hong, H., and Stephenson, C. (2016). Computational thinking for all: pedagogical approaches to embedding 21st century problem solving in k-12 classrooms. *TechTrends* 60, 565–568. doi: 10.1007/s11528-016-0087-7

Yadav, A., Sands, P., Good, J., and Lishinki, A. (2018). "Computer science and computational thinking in the curriculum: research and practice," in *Handbook of Information Technology in Primary and Secondary Education*, 1–18.

Yang, K., Liu, X., and Chen, G. (2020). The influence of robots on students' computational thinking: a literature review. *Int. J. Inf. Educ. Technol*. 10, 627–631. doi: 10.18178/ijiet.2020.10.8.1435

Yu, J., and Roque, R. (2019). A review of computational toys and kits for young children. *Int. J. Child Comput. Interact*. 21, 17–36. doi: 10.1016/j.ijcci.2019.04.001

Zapata-Cáceres, M., Martín-Barroso, E., and Román-González, M. (2020). "Computational thinking test for beginners: design and content validation," in *2020 IEEE Global Engineering Education Conference (EDUCON)* (Porto: IEEE), 1905–1914.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Bakala, Gerosa, Hourcade, Tejera, Peterman and Trinidad. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPENDIX

TABLE A1 | Search term used with each tool to search in SCOPUS.

Tool name	Search term	Search results	Relevant results
Bee Bot	TITLE-ABS-KEY (((Bee bot AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))	23	
Blue Bot	TITLE-ABS-KEY (((Blue bot AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))	3	
Botley	TITLE-ABS-KEY (((botley AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))	1	
Code-a-Pillar	TITLE-ABS-KEY ((({Code-a-Pillar} AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))	0	
Cubetto	TITLE-ABS-KEY ((({Cubetto} AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))	7	
KIBO	TITLE-ABS-KEY ((({KIBO} AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))	20	8
KIWI	TITLE-ABS-KEY (((kiwi AND robot AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))	2	2
KUBO robot	TITLE-ABS-KEY (((kubo AND robot AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))	0	
Matatalab Coding Set	TITLE-ABS-KEY (((Matatalab AND robot AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))	0	
mTiny	TITLE-ABS-KEY (((mtiny AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))	0	
Ozobot Evo	TITLE-ABS-KEY (((ozobot AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))	6	
Ozobot Bit	considered above		
Plobot	TITLE-ABS-KEY (((plobot AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))	0	
Pro-bot	TITLE-ABS-KEY (((pro-bot AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))	0	
Qobo	TITLE-ABS-KEY (((qobo AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))	0	
Robot Mouse	TITLE-ABS-KEY ((({Robot Mouse} AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))	0	
Robotito	TITLE-ABS-KEY (((robotito AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))	2	
Sphero indi	TITLE-ABS-KEY ((({sphero} AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))	15	
TurtleBot	TITLE-ABS-KEY (((Turtlebot AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))	4	

(Continued)

TABLE A1 | Continued

Tool name	Search term	Search results	Relevant results
VEX 123	TITLE-ABS-KEY ((({vex 123} AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))	0	
BOTS	TITLE-ABS-KEY ((({BOTS} AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8)))) AND (EXCLUDE (SUBJAREA , "MEDI"))	39	1
Codeable Crafts	TITLE-ABS-KEY ((({Codeable Crafts} AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))	0	
Code.org	TITLE-ABS-KEY ((({code.org} AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))	19	2
Kodable	TITLE-ABS-KEY ((({Kodable} AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))	3	1
LightBotJr	TITLE-ABS-KEY ((({LightBotJr} AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))	0	
Move the turtle	TITLE-ABS-KEY ((({move the turtle} AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))	1	1
RoboZZle	TITLE-ABS-KEY ((({RoboZZle} AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))	0	
Run Marco!	TITLE-ABS-KEY ((({Run Marco} AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))	2	
ScratchJr	TITLE-ABS-KEY ((({ScratchJr} AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))	28	7
The Foos	TITLE-ABS-KEY ((({The Foos} AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))	0	
Tynker: Coding for Kids	TITLE-ABS-KEY ((({Tynker} AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))	1	
Blue Bot	Repeated tool		
CHERP	TITLE-ABS-KEY (((cherp AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))	4	2
COJI	TITLE-ABS-KEY ((({coji} AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))	0	
Dash and/or Dot	TITLE-ABS-KEY ((({Dash} AND robot AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))	3	
Finch	TITLE-ABS-KEY ((({finch} AND robot AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))	1	
Sphero indi	Repeated tool		
Thymio	TITLE-ABS-KEY ((({Thymio} AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))	13	
Tinkerbots	TITLE-ABS-KEY (((tinkerbots AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))	0	

(Continued)

TABLE A1 | Continued

Tool name	Search term	Search results	Relevant results
Makeblock Neuron	TITLE-ABS-KEY (((makeblock AND neuron AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))	0	
Puzzlets Starter Pack	TITLE-ABS-KEY ((({Puzzlets} AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))	0	
Roberto	TITLE-ABS-KEY ((({Roberto} AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8)))) AND (EXCLUDE (SUBJAREA,"MEDI") OR EXCLUDE (SUBJAREA,"NURS") OR EXCLUDE (SUBJAREA,"NEUR") OR EXCLUDE (SUBJAREA,"PHAR") OR EXCLUDE (SUBJAREA,"IMMU") OR EXCLUDE (SUBJAREA,"BIOC")) AND (EXCLUDE (SUBJAREA,"ARTS") OR EXCLUDE (SUBJAREA,"SOCI"))	5	
Scottie Go	TITLE-ABS-KEY ((({Scottie Go} AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))	0	
Coding Awbie	TITLE-ABS-KEY ((({strawbies} AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))	1	1
Tabletop puzzle block system	TITLE-ABS-KEY ((({Tabletop puzzle} AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8))))	1	
T-Maze	TITLE-ABS-KEY ((({t-maze} AND (preschool OR child OR {early age} OR kindergarten OR {lower education} OR {early years} OR {elementary education} OR {young learner} OR {primary school} OR {primary education} OR k-6 OR k-8)))) AND (EXCLUDE (SUBJAREA , "BIOC") OR EXCLUDE (SUBJAREA , "MEDI") OR EXCLUDE (SUBJAREA , "PHAR")) AND (EXCLUDE (SUBJAREA , "NEUR"))	1	1

frontiers | Frontiers in Psychology

ORIGINAL RESEARCH
published: 20 September 2022
doi: 10.3389/fpsyg.2022.887280

Exploring Gender Differences in Coding at the Beginning of Primary School

Chiara Montuori[1], *Lucia Ronconi*[2], *Tullio Vardanega*[3] *and Barbara Arfé*[1]*

[1] Department of Developmental Psychology and Socialization, University of Padova, Padova, Italy, [2] School of Psychology, University of Padova, Padova, Italy, [3] Department of Mathematics, University of Padova, Padova, Italy

The gender gap in Computer Science (CS) is widely documented worldwide. Only a few studies, however, have investigated whether and how gender differences manifest early in the learning of computing, at the beginning of primary school. Coding, seen as an element of Computational Thinking, has entered the curriculum of primary school education in several countries. As the early years of primary education happen before gender stereotypes in CS are expected to be fully endorsed, the opportunity to learn coding for boys and girls at that age might in principle help reduce the gender gap later observed in CS education. Prior research findings however suggest that an advantage for boys in coding tasks may begin to emerge already since preschool or the early grades of primary education. In the present study we explored whether the coding abilities of 1st graders, at their first experience with coding, are affected by gender differences, and whether their presence associates with gender differences in executive functions (EF), i.e., response inhibition and planning skills. Earlier research has shown strong association between children's coding abilities and their EF, as well as the existence of gender differences in the maturation of response inhibition and planning skills, but with an advantage for girls. In this work we assessed the coding skills and response inhibition and planning skills of 109 Italian first graders, 45 girls and 64 boys, before an introductory coding course (pretest), when the children had no prior experience of coding. We then repeated the assessment after the introductory coding course (posttest). No statistically significant difference between girls and boys emerged at the pretest, whereas an advantage in coding appeared for boys at the posttest. Mediation analyses carried out to test the hypothesis of a mediation role of EF on gender differences in coding show that the gender differences in coding were *not* mediated by the children's EF (response inhibition or planning). These results suggest that other factors must be accounted for to explain this phenomenon. The different engagement of boys and girls in the coding activities, and/or other motivational and sociocognitive variables, should be explored in future studies.

Keywords: gender gap, coding, STEM, Computer Science, computational thinking, executive function, primary school

OPEN ACCESS

Edited by:
Stamatios Papadakis, University of Crete, Greece

Reviewed by:
Andrea Paula Goldin, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina
Giovanni Fulantelli, Istituto per le Recnologie Didattiche (ITD), Italy
Arianna Boldi, University of Turin, Italy

***Correspondence:**
Barbara Arfé
barbara.arfe@unipd.it

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

Received: *01 March 2022*
Accepted: *07 June 2022*
Published: *20 September 2022*

Citation:
Montuori C, Ronconi L, Vardanega T and Arfé B (2022) Exploring Gender Differences in Coding at the Beginning of Primary School. Front. Psychol. 13:887280. doi: 10.3389/fpsyg.2022.887280

INTRODUCTION

With gender differences, or gap, in Science, Technology, Engineering, and Mathematics (STEM) researchers refer to the disparity between boys and girls, or men and women, in performance, achievements, interests or beliefs in the STEM domains. As these gender disparities lead to an underrepresentation of women in higher STEM education and careers (Wang and Degol, 2017), their emergence is considered of high theoretical and practical (societal) relevance. Gender differences in STEM have been addressed extensively from secondary school onwards (Fisher and Margolis, 2003; Zweben and Aspray, 2004; Frieze, 2005; Anderson et al., 2008; Maloney et al., 2012; Spearman and Watt, 2013; Beyer, 2014; Charles et al., 2014; Charlesworth and Banaji, 2019; Alonso et al., 2021; Gnambs, 2021). Comparatively less studies, instead, have investigated the emergence of early gender differences in preschool or elementary school (e.g., Cvencek et al., 2011; Aesaert and van Braak, 2015; Kersey et al., 2018; Master et al., 2021). In this paper we address this particular angle of the problem, considering the emergence of gender-ability differences in the learning of coding and programming at school entrance, in grade one.

Researchers have offered two main explanations to the emergence of gender differences in STEM. The first explanation maintains that the gender gap originates from (innate) sex-differences in the cognitive abilities underpinning performance in STEM (Halpern and LaMay, 2000; Miller and Halpern, 2014; Girelli, 2022). Sex-related differences in cognitive abilities underpinning performance in STEM are indeed reported in some studies (Halpern and LaMay, 2000; Maloney et al., 2012; Miller and Halpern, 2014). These gender disparities can determine differences in students' achievements (Maloney et al., 2012) and consequently affect motivation for the pursuit of studies and careers in STEM (Wang and Degol, 2017). The second explanation has it that the gender gap originates from sociocultural factors, such as inequalities in the social and educational systems, and gender role stereotypes that determine explicit and implicit biases of boys and girls in how they perceive and evaluate their and others' performance and abilities in STEM (Charlesworth and Banaji, 2019; Girelli, 2022). For instance, exposure to role models, prior experiences with STEM and the expectations of others (e.g., parents) can contribute to the emergence of biased explicit and implicit (i.e., less conscious) beliefs on boys and girls abilities in STEM, which can influence individuals' behaviors, performance and learning experience (Miller and Halpern, 2014; Flore and Wicherts, 2015; Master et al., 2017; Charlesworth and Banaji, 2019), with possible long-lasting effects on girls' motivation to pursue studies or careers in the STEM domain (Charlesworth and Banaji, 2019). Master et al. (2021), for example, show how the stereotype that girls have lower interest in CS and Engineering than boys can cause gender disparities in motivation for CS education and in engaging in novel activities in this field. This link between stereotypes and interest in CS persists throughout high school, an age at which students typically make choices about their higher education. Thus, early elementary school can be a critical period to introduce children to counter-stereotypical examples, before stereotypes are firmly endorsed.

Recently, neuroscientific research has suggested that gender-related ability differences are actually the product of biopsychosocial interactions between biological predispositions and sociocultural experience (Miller and Halpern, 2014; Wierenga et al., 2019). In that interpretation, sex-differences in brain maturation can interact with sociocultural factors such as children's experiences, determining differences in cognitive performance in specific domains, like in language or spatial tasks (Miller and Halpern, 2014; Wierenga et al., 2019). These relative cognitive strengths or weaknesses may in turn affect students' perception to be able to perform STEM tasks and mediate the relationship between gender and task anxiety (Maloney et al., 2012), with possible consequences on students' motivation toward STEM.

Research shows that gender differences in basic cognitive skills underpinning STEM achievement can be observed from early childhood (Wang and Degol, 2017). However, as STEM gender-related stereotypes develop from children's experience of STEM activities and role models, they may emerge at a different age in different domains, depending on children's opportunity to be exposed to those activities and models. For instance, gender stereotypes on science and scientists do not seem to emerge before late primary school, because formal science instruction is sporadic in early grades of it (Miller et al., 2018). Likewise, gender-interest stereotypes about Engineering being more suited for boys are evident from grade 1, whereas in CS children seem to endorse gender-interest stereotypes only later, from grade 3 (Master et al., 2021).

All along child development, various sociocultural effects, among which the influence of stereotypes, may thus cause gender differences in STEM to appear that are not simply intrinsically sex-related, or associated to original predispositions. Master (2021) observe that gender stereotypes channeled via membership in social groups influence children's interest and motivation toward CS, their ability beliefs and their sense of belonging, which are prodromic to task avoidance and failure in STEM disciplines. As such effects may cause considerable distortion in the child's learning experience, it is important to examine the emergence of gender differences at early ages, when gender-ability stereotypes, i.e., the belief that boys are better at performing certain tasks than girls, are not yet strongly endorsed by children.

Gender Differences in STEM

Studies exploring gender differences in STEM achievements show that they are most frequently observed in older students. Investigating the acquisition of mathematical abilities, Kersey et al. (2018) reports that boys and girls from 6 months to 8 years do not differ in early mathematical abilities. Stoet and Geary (2013) find them instead in favor of boys, at older age ranges, among higher-performing 15-year-old students. The latter finding seems to correlate with the observation that a lowering of self-efficacy beliefs in mathematics by girls and a

parallel increase in boys occur between the fourth and the ninth grades (Reilly et al., 2019; Mejía-Rodríguez et al., 2020).

As noted, only a few studies address the issue of gender differences in CS as yet, in spite of the fact that women are extremely underrepresented in it, for education and career (Schmidt, 2011; Beyer, 2014; Denner et al., 2014; UNESCO, 2017). A widespread belief has it that male students have greater natural inclination to and ability with Information and Communication Technology, ICT (Jackson et al., 2008). This conjecture aligns with a meta-analysis in Cai et al. (2017), which shows boys in the age range between secondary school and college to have higher self-efficacy in ICT and more positive attitudes toward it than peer girls. These findings, however, measure bias-susceptible self-perception instead of actual skills: they may be predictive of attitude, but not of actual performance (Honicke and Broadbent, 2016). Moreover, those studies address a population of higher-education students likely exposed to well-structured and robust gender stereotypes. In younger children, these strong self-beliefs and gender-ability stereotypes may be not yet fully formed (Master et al., 2021). Although young children (from kindergarten to second grade education) already begin to form opinions about which technologies and tools would be better suited for boys and girls, gender attitudes toward technologies are still mild at this age (Sullivan and Bers, 2016). For instance, Master et al. (2017) report that children as young as 6 years already hold emergent gender stereotypes regarding computing, believing that boys should be more interested and better at coding and robotics than girls. The latter (gender-ability) beliefs however are less strong than the former (gender-interest) stereotypes (Master et al., 2021), and their effects on children's performance likely depend on children's prior experiences with digital technologies and coding (Gerson et al., 2022). Other findings seem to support the hypothesis of lesser influence of gender stereotypes in ICT activities at an early age. Aesaert and van Braak (2015) report primary school girls to have better technical ICT skills and higher-order ICT competences than boys. A subsequent meta-analysis corroborates that view by reporting girl-favoring gender differences in ICT literacy, with effect sizes larger in primary than secondary schools (Siddiq and Scherer, 2019).

Gender Differences in Computational Thinking

Few studies zoom from broad ICT into the specifics of CS. When they do, they look for gender differences in coding as part of Computational Thinking (CT) activities. CT is a set of thinking skills, precursor of CS education, generally understood to comprise four constituents: (1) problem analysis via abstraction and decomposition, to distill core patterns from the original problem, to break it into smaller parts and systematically tackle each of them; (2) algorithmic thinking, to enable the development of predefined re-usable executable procedural tools for solving the given problem and classes of them; (3) evaluation of the outcomes of the solution plan, correcting it where it fails (also known as debugging), feeding all of that into (4) generalization, to lift problem-solving methods and solutions to application to similar problems (Wing, 2006; Resnick et al., 2009; Roman-Gonzalez et al., 2017; Shute et al., 2017; Yasar, 2017; Nardelli, 2019). Coding is a concrete way of practicing CT skills that consists in generating instructions (program's code) in a way that yields executable plans, whose effect to the problem can be empirically ascertained.

The studies that explored gender-related differences in CT or coding have produced contrasting findings. A study by Kožuh et al. (2018) reports finding no gender differences in problem solving skills involved in programming for fourth to sixth graders. Similarly, Papavlasopoulou et al. (2020), who used eye-tracking measures to assess the performance during coding workshops of 8- to 17-year-old students new to coding, a larger age range than Kožuh et al. (2018)'s study, report finding no statistically significant difference in gaze behaviors or learning gains between boys and girls. However, some qualitative gender-related differences emerged between girls and boys in the strategies used and in the perceptions of the coding activities. Price and Price-Mohr (2021), who explored the performance of 32 children between 10- and 11-year-old in an exercise aimed at animating stories with text-based coding, do not find gender differences in the process of coding or in the quality of the produced animations. Jiang and Wong (2021), find gender differences to be insignificant across fourth to sixth graders in the approach to conditionals, logical operators, pattern recognition, and generalization.

Other studies, involving older, fifth to tenth grade, students (Roman-Gonzalez et al., 2017; Statter and Armoni, 2017) have found significant gender differences in coding, although their findings are inconsistent regarding the direction of the gender effect. Statter and Armoni (2017) report results on seventh-grade students showing some advantage for girls in the learning of CS abstraction, and a greater effect of a learning intervention on girls, causing girls to regard CS as more than just programming, which boys did not. Conversely, Roman-Gonzalez et al. (2017) found a significant difference in CT tasks in favor of boys, with statistically significant differences emerging from grade 7, and a further increase of the gender gap between girls and boys in older, ninth to tenth grade, students. Also Yücel and Rizvanoglu (2019) found gender differences in coding among 11- to 14-year-old students, observing that they were associated with girls' lower self-confidence in performing the coding task and greater perception of task difficulty in comparison with boys. However, even in older, 14 to 19-year-old, high school students, gender differences in performance on coding tasks do not always emerge (Lau and Yuen, 2009).

As exposure to digital technologies and coding may significantly affect gender differences in perception, beliefs and motivations toward it (Master et al., 2017; Gerson et al., 2022), assessing children's coding skills as early as their first experience with coding becomes especially important. A recent systematic review (Bati, 2022) of experimental evidence on programming (i.e., coding applied to the creation of true computer programs), and CT in early childhood education, found that girls and boys from 3 to 5 years perform similarly in them.

To the best of our knowledge, however, only two studies have explored gender differences in CT and coding among children

(4–7 years) exposed to it for the first time (Sullivan and Bers, 2013, 2016). Both studies report finding significant gender effects in favor of boys. The former study (Sullivan and Bers, 2013) found kindergarten boys to be better than girls in CT activities involving building with robotic materials. The latter (Sullivan and Bers, 2016), which focused on children aged 4 to 7 years, found similar performance across boys and girls in CT tasks involving basic coding skills, but a significantly better performance of boys in the use of more advanced coding constructs, such as repeat loops dependent on sensor readings.

A possible interpretation of these findings is that explicit or implicit gender stereotypes may affect actual performance even at a young age (Steele, 1997; Spencer et al., 1999). An alternative hypothesis, tested in this study, is that gender differences at this young age are the result of other influences, such as differences in the cognitive skills that underpin coding and perhaps CT in general (Halpern and LaMay, 2000; Miller and Halpern, 2014; Grissom and Reyes, 2019). As noted earlier, in young children approaching coding for the first time, gender stereotypes are not yet fully structured or endorsed and, if present, they have most likely mild effects (Martin et al., 1990; Sullivan and Bers, 2016). Thus, if gender differences in CT are observed at this early age, other (e.g., cognitive) factors could account for such differences. Our study tests this particular hypothesis, which was not explored in the cited works by Sullivan and Bers (2016). Where Sullivan and Bers (2016) use robotics on the grounds of it giving a playful and engaging touch to the learning ground, in the study presented in this paper we used the Code.org platform under similar premises. Much like Sullivan and Bers, we looked at whether the coding abilities of young (first-grade) children, all novice to coding, are affected by gender differences. The additional angle we brought into this study is to determine whether any such emerging gender differences associate with gender differences in children's executive functioning (EF), in particular planning and response inhibition, cognitive abilities closely related to problem-solving and CT (Arfé et al., 2019, 2020).

Gender Differences in Executive Functioning

To date, the hypothesis that gender differences in coding, where they occur, can be mediated by gender differences in EF has not been tested yet. Executive functioning involves cognitive abilities used by individuals to focus on task, override automatic or impulsive responses and organize their behavior toward a goal. EF include response inhibition skills, working memory, switching, or the ability to flexibly adapt to different tasks, and more complex abilities like planning, which are involved in goal-directed behaviors and problem-solving (Miyake et al., 2000; Zelazo et al., 2003; Diamond, 2013; Viterbori et al., 2017). As noted earlier, coding tasks involve problem-solving processes that make significant demands on several levels of EF, including response inhibition, working memory (Shute et al., 2017; Di Lieto et al., 2020), and planning (Arfé et al., 2019, 2020).

When CT skills are practiced, the cited EF processes are also set in motion. Besides showing a strong association between coding abilities and first graders' planning skills (Arfé et al., 2019, 2020), between 5- and 6-year-old children's coding abilities and response inhibition (Arfé et al., 2019, 2020; Di Lieto et al., 2020), and between 5- and 6-year-old children's coding abilities and working memory (Di Lieto et al., 2020), prior research has also shown the existence of gender differences in the maturation of EF (Unterrainer et al., 2013; Grissom and Reyes, 2019; Wierenga et al., 2019). Although gender differences in executive functioning are not overwhelming (Grissom and Reyes, 2019), they are indeed observed in some domains, such as response inhibition and control over impulsive responses. For instance, males are found to be more impulsive and have more reduced reaction times than female (Grissom and Reyes, 2019). Inhibition and impulse control seem to mature earlier in girls than in boys. Indeed, between the age of 3 and 5, girls are reported to have better inhibition skills. Boys seem to catch up with girls only later, around the age of 6 (Klenberg et al., 2001). There also is empirical evidence that girls show better planning skills than boys during preschool years (Unterrainer et al., 2013) as well as that this advantage is maintained also during school years (Warrick and Naglieri, 1993; Naglieri and Rojahn, 2001). In a large-scale study involving 2,200 participants aged 5–7 to 11–17 years, Naglieri and Rojahn (2001) have shown a consistent advantage in planning skills for girls over boys across those age groups. The seemingly faster maturation of inhibition and planning in girls is of particular interest, as the ability to inhibit impulsive responses is an important prerequisite to an analytic approach to problem solving, and thus, by extension, to coding tasks. Likewise, planning is a core component skill of algorithmic thinking (the ability to define a sequence of steps to get to an objective) (Arfé et al., 2019, 2020). The existence of gender differences favoring girls in response inhibition and planning from as early as 5–6 years of age would cause expecting advantage for girls over boys to emerge also in the coding tasks that involve algorithmic thinking. Notably, this expectation goes in an opposite direction to what found by Sullivan and Bers (2016).

The Study

The study presented in this paper explored the manifestation of gender differences in coding in young (5–7 year-old) children exposed to coding for the first time (goal 1), assessing whether any such observed differences were mediated by gender differences in planning or response inhibition, measured by standardized planning and inhibition tests (goal 2). Based on prior studies that demonstrate an association between coding skills and children's inhibition and planning abilities, we regarded inhibition and planning abilities as a potential mediator of the effects of gender on coding skills.

METHOD

Participants

One-hundred and nine first-graders aged from 5 to 7 years from schools located in northern Italy were enrolled in the study (45 girls, 41%, and 64 boys, 59%). Children had no prior experience with coding. They all took part in one-month introductory course

TABLE 1 | Gender differences in age, SES (means, standard deviations, and *t*-test), and use of digital devices (Chi-square test).

	Girls ($n = 45$)		Boys ($n = 64$)				
Variable	***M***	***SD***	***M***	***SD***	***t***	χ^2	***p***
Age	6	0.48	6.05	0.41	−0.55		0.59
SES	5.87	1.57	5.58	1.41	1.00		0.32
Computer						0.64	0.42
Tablet						0.05	0.81
Smartphone						0.73	0.39

to coding at the beginning of the school year, as part of a larger research project (Arfé et al., 2020).

The mean age of participants in this study was 6.00 for girls and 6.05 for boys. None of the participants had certified developmental disabilities or attentional problems. Demographic data are reported separately for girls and boys in **Table 1**. As children's prior exposure to digital technologies and parents' socioeconomic status can be associated to the development of coding abilities in children (Chiazzese et al., 2017; Gerson et al., 2022), these two factors were considered in this study. To assess such factors we used a short socio-demographic questionnaire that children's parents returned with written informed consent to participation in the study.

Socioeconomic status (SES): Children's socioeconomic status was estimated based on the level of education of the child's parents (both mother and father), on a scale from 0 (less than primary school) to 4 (college or above), and on the level of parents' occupation, from 1 (unemployed) to 4 (professional roles). A composite score was calculated as the non-weighed sum of the highest education and occupation score obtained by either parent (mother or father), with maximum score 8.

Familiarity with technology: This indicator was gauged by asking parents about children's daily use of personal computer, smartphone, or tablet devices in their home environment. The number of girls and boys that were reported to make daily of any such device was computed and compared by chi-square analyses.

The study was approved by the Ethical Committee of the Department of Developmental Psychology at the authors' institution.

Study Design

Following Sullivan and Bers (2016), in this paper we examined children's ability to code at their first approach with coding, i.e., after an introductory course to coding at the beginning of grade one. Children's planning and inhibition skills were also assessed, and their mediation role in explaining gender differences in coding was tested. Coding, planning and response inhibition abilities were examined both *before* (time 1, T1) and *after* (time 2, T2) the coding course. This allowed to ascertain whether gender differences in coding or EF were present at any of those moments.

Procedure and Materials

Coding Introductory Course

At the start of the school year all participants received a 1-month introductory course to coding through a selected choice of coding games from Code.org (https://code.org/). The coding environment and tasks proposed by the Code.org platform propose visual block-based programming tasks for beginners. Individual children write their code on that platform by moving code blocks from a toolbox panel into a programming panel, to generate code sequences (programs) whose execution should achieve predefined results. All coding games in Code.org involve the use of coding blocks to instruct a sprite (angry bird, bee, zombie) so that they can reach a target or perform expected actions. Visual and textual informative feedback is provided at every execution in order that children can easily monitor their progress on the screen. Likewise, programming errors are immediately visible to the child. The target not being reached manifests by the sprite crashing against a wall or failing to find a route to the target (see **Figure 1**). Children were progressively introduced to coding blocks of increasing logical difficulty, for example, from simple sequences to repetitions (loops).

Examining children's needs is crucial to the design of effective instructional coding activities (Ronsivalle et al., 2019). Since our participants were unfamiliar to coding as well they were beginning readers, the training was structured on course 1 of the Code.org platform "Programma il Futuro"[1], the most basic and initial one. The children took part in coding lessons in the school's laboratory and always used a computer to carry out the assigned exercises. Although peer-based collaborative environments are thought to aid in the development of coding and CT skills (Flórez et al., 2017), the results of a pilot study suggested that the children were easily distracted by peers when working in pairs, and that the workload was not evenly shared. Classroom activities were thus designed to allow children to work individually at an assigned computer post in the school's laboratory. As children of this age are typically more familiar to touch-screen devices than to the use of mouse devices (Papadakis, 2021), a familiarization lesson took place before the course began, to accustom children to the use of the mouse. During the course, children were exposed to four main components of CT involved in coding (analysis of the problem space; decomposing problems; algorithmic thinking; evaluating and revising plans) and were allowed to practice with games that involved all such components. A group of post-graduate students (all female), trained to teach coding to children conducted each training session in collaboration with the first author of the study. During the lessons, the class teachers were present but did not intervene, except if children explicitly requested it. Postgraduate students, supported by the study's first author, were the same throughout the whole course. In turn, one post-graduate student led the lesson by explaining at the

[1]https://programmailfuturo.it/come/primaria/vecchie-lezioni-tecnologiche/corso-1

"Programma il futuro" is an initiative of the Ministry of Education, University and Research (MIUR) in collaboration with the Consorzio Interuniversitario Nazionale per l'Informatica (CINI), aimed at giving schools a series of tools to provide students with the scientific and cultural bases of computer science.

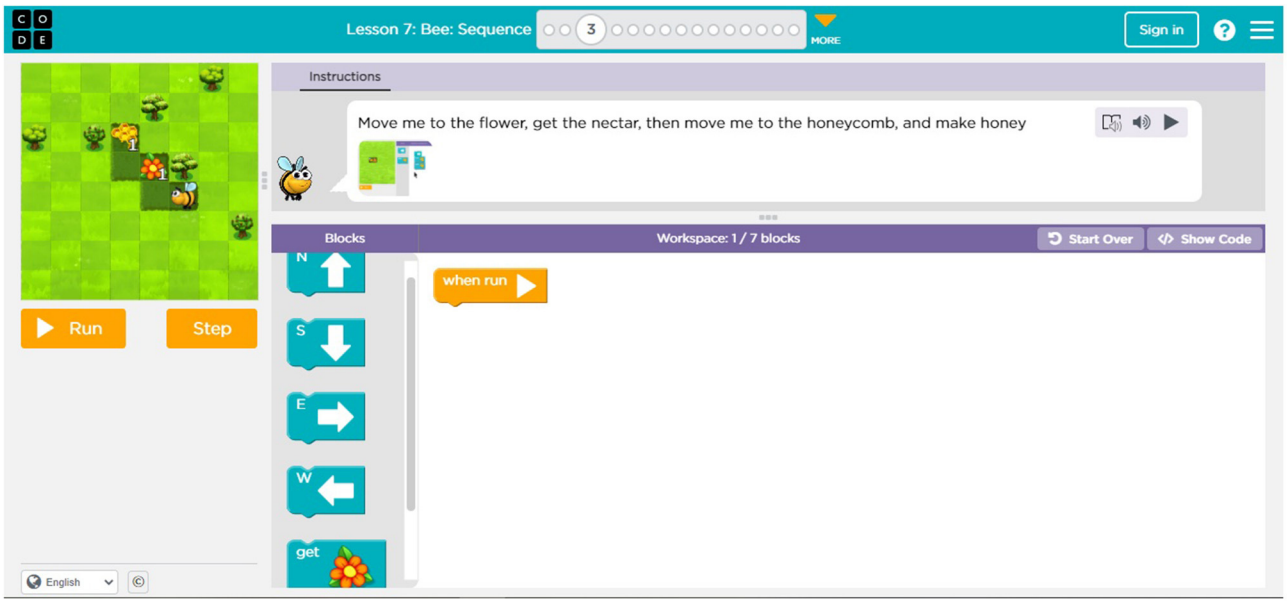


FIGURE 1 | Lesson 7, course 1 (https://studio.code.org/s/course1/lessons/7/levels/3?lang=en-US) .

beginning of each new exercise the objective and functions to solve the problem. Other students supported the class and stood ready to answer questions from the children. One student was assigned the role of observer to point out any deflection from the planned protocol in each lesson. Such observers checked for example whether the support strategies were always the same and were balanced for each child, what difficulties did children encounter in carrying out the exercises, what function (such as repetition loop) was the hardest to learn. Children performed all games individually in the classroom and were requested to signal when they completed each task, to confront and discuss their solution. Group-wise corrections ended each exercise. The support students used scaffolding strategies to support children during both individual task performance and group correction. For example, they used questions-and-hints to stimulate children's approach to the solution of the problem at hand. All children received 60 min bi-weekly coding lessons for 4 weeks (eight lessons in total). Girls and boys received the same coding lessons and practiced coding by playing the same sequence of coding games. They took part in all coding lessons together. Participation in the coding course was not mandatory for children, who were free to withdraw at any time. **Table 2** reports the full lesson plan. Children's actual liking of coding activities was not assessed through self-reports or systematic observations. However, all children appeared to be actively engaged and to enjoy the proposed activities thoroughly. No child asked to withdraw.

The data presented in this paper were collected as part of a larger study aimed to evaluate the effects of coding on children's EF. Thus, neither the instructors nor the children involved in the data collection were informed of the goal of the present study, that is to say, they did not know that gender differences in coding would also be assessed.

TABLE 2 | Lessons plan. Selected coding games from programma il futuro, Course 1 (https://programmailfuturo.it/come/primaria/vecchie-lezioni-tecnologiche/corso-1).

Coding sessions	Course 1	Trial number	Content
Session 1	Lesson 3	1, 6	Jigsaw: Drag and drop
	Lesson 4	2, 5, 6, 7	Maze: Sequence
Session 2	Lesson 4	8, 10	Maze: Sequence
	Lesson 5	3, 4, 5, 6, 7	Maze: Debugging
Session 3	Lesson 8	4, 5, 6, 7, 8	Artist: Sequence
	Lesson 5	8, 9,10	Maze: Debugging
Session 4	Lesson 8	9, 10, 11	Artist: Sequence
	Lesson 10	4, 5, 6, 7, 8	Artist: Shapes
Session 5	Lesson 13	1, 2, 3, 4	Maze: Loops
	Lesson 13	5, 6, 7	Maze: Loops
Session 6	Lesson 13	8, 9, 10, 11, 12	Maze: Loops
Session 7	Lesson 14	3, 5, 6, 7, 8, 9	Bee: Loops
Session 8	Lesson 18	2, 4, 5, 6, 7	Artist: Loops
Closing session	Classroom discussion	What have we learned?	Metacognitive reflection on the goals of computational thinking and the meaning of programming

Assessment of Coding, Planning and Response Inhibition Skills

Children's coding, planning and response inhibition skills were assessed both before (T1) and after (T2) the coding lessons. Coding skills were assessed thorough children's ability to solve four coding games on Code.org. Two standardized neurocognitive tests, the Tower of London

(Fancello et al., 2013), and a numerical Stroop test (Marzocchi et al., 2010) were used to assess their general planning and inhibition skills.

Coding Skills

Before starting the assessment, children were invited to practice with two coding games from Code.org, under guidance by the experimenter. The assessment started after the practice phase, in which the child familiarized with the Code.org platform and the mouse-based drag-and-drop mechanics necessary to perform the coding tasks. All children were asked to solve four coding problems individually and autonomously. To solve each coding trial, a maximum of three attempts were allowed, after which the trial was counted as failed. Specifically, the assessment involved solving trials 9 (lesson 4), 2 (lesson 5), 3 (lesson 8), 4 (lesson 14) from Code.org (Course 1, Italian platform, https://programmailfuturo.it/come/primaria/vecchie-lezioni-tecnologiche/corso-1). Trial 9 required guiding the Angry bird sprite to proceed in successive steps to reach a given target. Trial 2 involved debugging. Trial 3 required placing blocks in a sequence apt to instruct an artist sprite to draw a target geometric shape. Trial 4 consisted of using repetition loops. All of these exercises were of the same type as those in the coding introductory course.

For each trial, two scores were recorded as measures of children's coding skills:

(1) Accuracy: a score of 2 was given if the child successfully solved the problem at the first attempt, 1 when solving it at the second attempt, 0 otherwise;
(2) Time spent planning: the seconds elapsed from the moment the child was presented the trial to the moment s/he moved the first code block (which corresponded to starting to write the program). Time spent planning reflects the children's ability to plan their responses in advance and inhibit less mature trial-and-error strategies that are typical of younger children's approach to problem solving (Harter, 1930). Planning time was calculated on all trials, whether solved successfully or not, as in standardized planning measures (i.e., Tower of London, below). Planning time may reflect children's exploration of the coding platform, planning a sequence of steps for solving the coding problem, as well as pauses and hesitations (as, for example, holding the mouse) during the task.

Planning Skills

Planning ability was assessed by the Tower of London test (ToL; Luciana et al., 2009). In this study we used a version standardized for a population aged 4–13 years (Fancello et al., 2013).

The test requires the child to reproduce a configuration of three colored (blue, red, and green) small balls on three vertical sticks of different heights, according to a precise set of rules (e.g., moving one ball at a time; not holding the ball or placing it on the table, after picking it up). The entire test consists of 12 trials of increasing difficulty. All 12 trials were presented with no interruption criteria. As for the coding games, children's performance was scored for:

(1) Accuracy: each attempt was scored 1 if the child performed the trial correctly within 1 min, without breaking any rule; 0 otherwise.
(2) Planning time: counting from when the trial is shown to the child until when s/he makes the first move.

Reliability indices for this test are 0.57 for accuracy scores and 0.71 for planning times (Fancello et al., 2013).

Response Inhibition Skills

The Numerical Stroop test of the Batteria Italiana ADHD (BIA, Marzocchi et al., 2010) was used to assess children's ability to inhibit automatic responses. The Numerical Stroop test of the BIA, standardized for children aged 6–11, assesses response inhibition. The child is presented with a table that displays in each cell, from left to right, a digit from 1 to 5 (e.g., the digit 5), repeated n times (e.g., 3 times). The child is instructed to say as quickly and accurately as possible how many times the given digit (in the example, "5") is shown in the cell (in the example, "three" times). To succeed in the task, the child must suppress automatic digit recognition (i.e., inhibiting the automatic response "5"). Performance is scored for:

(1) Accuracy: number of errors and self-corrections.
(2) Inhibition time: the seconds required to complete the task.

Test-retest reliability and validity are not provided by the manual. Arfé et al. (2020) report moderate test-retest reliability of this test for accuracy, $r = 0.34$, and adequate reliability for inhibition time, $r = 0.62$. Concurrent validity, computed by correlating the performance on the numerical Stroop and the NEPSY-II verbal response inhibition subtest, is $r = 0.44$ for accuracy and $r = 0.48$ for inhibition time.

Differences between girls and boys in age and SES were examined by independent-samples t tests. The different distribution between girls' and boys' use of digital devices (daily use of computer, smartphone, and tablet) was explored by chi-square tests.

The data analysis was performed in three steps.

1. Independent-samples t tests were used to test gender differences in coding, planning and response inhibition abilities before and after the coding lessons. Statistical significance was set as p-value < 0.05.
2. Pearson's correlations were run to explore the association between children's coding, planning (i.e., performance on the ToL), and response inhibition skills (performance on the Stroop task) before and after the introductory course to coding. Significant correlations between these abilities is indeed a condition necessary for assuming a mediation effect of children's planning and response inhibition skills on coding (Kraemer et al., 2008).
3. Mediation analyses were run to assess direct and indirect gender effects on children's coding abilities after the course (at T2). Mediation analyses allow exploring both direct and indirect effects of one variable (A) on another variable (B), where indirect effects refer to the underlying mechanism by which variable A influences variable B through a third (C) mediator variable (MacKinnon, 2008). In the present

study, the use of mediation models allowed testing both the hypothesis of a direct influence of gender on children's coding ability (direct effect model) and the hypothesis that gender effects on coding are mediated by children's planning and inhibition skills (indirect effect models).

Mediation analyses were conducted using IBM SPSS Statistics 23.0 and Hayes' process model 4, with gender (dummy variable: girls or boys) as predictor, planning time or accuracy in coding after the course (T2) as criterion variables, and T2 planning time or accuracy on the Tower of London test (ToL), and T2 response inhibition time or errors on the Stroop test as mediators. SES and children's performance on coding, planning (ToL) and inhibition (Stroop) tasks before the course (at T1) were covariates. Confidence interval for each indirect effect was estimated considering both mediators and criterion variables at time 1 and at time 2 (95% confidence, 5000 bootstrap samples).

To examine mediation effects, four mediation models were tested: *Models a* and *b,* represented in **Figure 2**, assessed the direct and indirect effects of gender, the predictor variable, on children's planning time at coding games at T2, the outcome variable. In *Model a*, indirect effects were assessed considering the mediation of planning time on the ToL. Thus, T2 planning time on the ToL was the mediating variable (M). In *Model b*, the indirect effects were assessed by considering the mediating role of response inhibition time. Thus, T2 inhibition time on the Stroop test was the mediating variable. Similarly, *Models c* and *d*, in **Figure 3**, assessed the direct and indirect effects of gender (predictor) on children's accuracy in coding (outcome variable). In *Model c,* T2 accuracy on the ToL was the mediating variable. In *Model d,* errors in the Stroop test was the mediating variable. The covariates for all models were SES and the children's performance on coding, and planning (ToL) or inhibition (Stroop) tasks at T1. This arrangement allowed testing for the different impact of the learning experience for boys and girls, and to control for SES-related differences.

RESULTS

Between-group differences in age, SES, and familiarity with technology are reported in **Table 1**. The *t*-tests and chi-square analyses showed that the two groups were equivalent for age, $t_{(107)} = -0.55$, $p = 0.59$, SES, $t_{(107)} = 1.00$, $p = 0.32$, and familiarity with digital devices: use of computer, $\chi^2 = 0.64$, $p =$

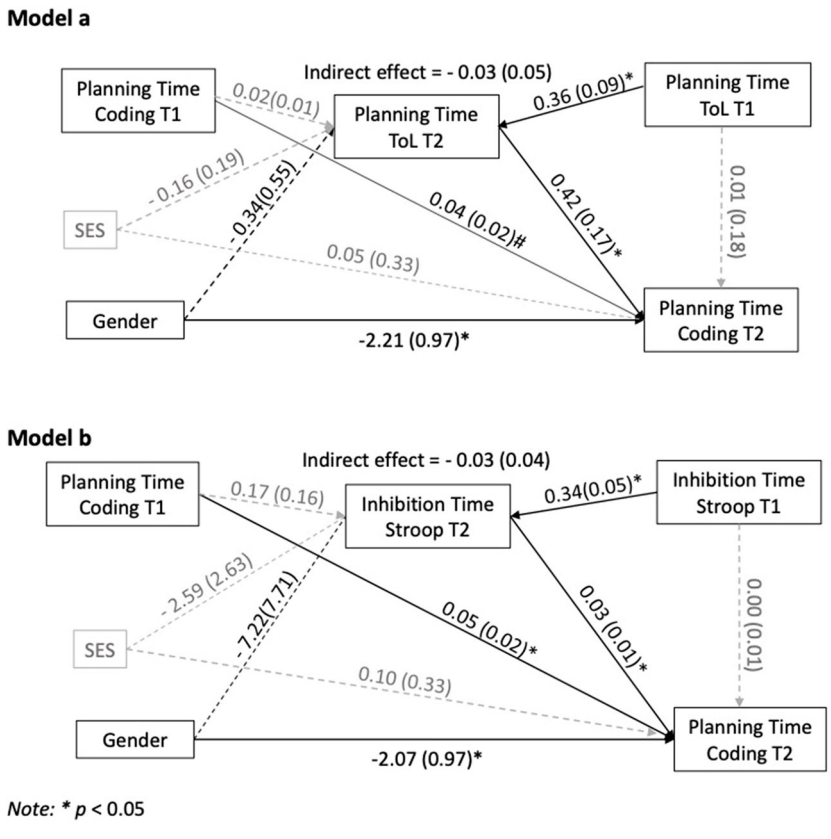


FIGURE 2 | Mediation models. Gender effects on coding planning time. *$p < 0.05$.

TABLE 3 | Differences between girls and boys in coding, planning, and inhibition skills at time 1 (T1, Before) and Time 2 (T2, After) the coding course.

	Girls ($n = 45$)		Boys ($n = 64$)				
Variable	***M***	***SD***	***M***	***SD***	***t***	***p***	**Cohen's *d***
Before the coding course (T1)							
Planning time coding	34.02	25.15	28.49	23.18	1.18	0.24	0.23
Accuracy coding	3.38	1.89	3.38	1.85	0.01	0.99	0.00
Planning time ToL	6.20	2.68	6.27	3.23	−0.13	0.90	−0.02
Accuracy ToL	5.38	2.74	5.77	2.75	−0.73	0.47	−0.14
Inhibition time Stroop	215.39	52.04	210.58	83.91	0.34	0.10	0.07
Inhibition errors Stroop	7.51	7.66	9.42	11.65	−0.96	0.25	−0.19
After the coding course (T2)							
Planning time coding	11.44	5.02	8.82	5.28	2.60	0.01*	0.50
Accuracy coding	5.87	0.87	6.27	1.03	−2.12	0.04*	−0.41
Planning time ToL	6.68	2.55	6.30	3.40	0.64	0.52	−0.12
Accuracy ToL	9.36	1.65	8.94	1.79	1.24	0.22	0.24
Inhibition time Stroop	180.26	40.65	171.20	49.97	1.00	0.70	0.19
Inhibition errors Stroop	2.00	2.34	2.39	2.44	−0.84	0.93	−0.16

*$p < 0.05$; *Planning time coding, Seconds spent planning on the coding games; Accuracy coding, Accuracy on the coding games; Planning time ToL, Seconds spent planning on the Tower of London test; Accuracy ToL, Accuracy on the Tower of London test; Inhibition time Stroop, Seconds required to complete the task on the Numerical Stroop test; Inhibition errors Stroop, Number of errors and self-corrections on the Numerical Stroop test.*

0.42, tablet, $\chi^2 = 0.05$, $p = 0.81$, and smartphone, $\chi^2 = 0.73$, $p = 0.39$.

Between-Group Differences in Coding, Planning and Response Inhibition

Table 3 reports the results of the between-group (girls, boys) comparisons in coding, planning and response inhibition abilities. After the introductory course to coding, significant differences in coding were observed. Girls spent significantly more time planning, $t_{(107)} = 2.60$, $p = 0.01$. However, their results were significantly less accurate, $t_{(107)} = -2.12$, $p = 0.04$, than those of boys. The effect sizes, reported in **Table 3**, are moderate. The performance of the two groups in coding *before* the coding course was equivalent. No significant differences emerged between girls and boys in planning or response inhibition, neither before nor after the coding course.

Associations Between Coding, Planning and Inhibition Skills

The correlational analyses, reported in **Table 4**, revealed a significant, although moderate, association between children's planning time in coding and planning time on the ToL both *before* (T1) the coding course, $r = 0.30$, $p = 0.001$, and *after* it (T2), $r = 0.31$, $p = 0.001$. Planning time in coding correlated significantly with response inhibition times *after* the coding course, $r = 0.29$, $p < 0.005$, but did not show significant associations with children's performance in response inhibition *before* the course.

Accuracy scores on the coding games and the ToL were significantly associated both *before* the coding course (T1), $r = 0.36$, $p < 0.001$, and *after* the course (T2), $r = -0.20$, $p = 0.04$. The negative association between planning accuracy (ToL) and accuracy in coding at T2 indicates that after the coding course the children with a better performance on the ToL performed worse in coding. We shall return to this finding in the discussion. Accuracy on the coding games correlated significantly with inhibition errors *before* the course, at T1, $r = -0.25$, $p < 0.01$, and with response inhibition times *after* the course (T2), $r = -0.26$, $p < 0.01$.

Mediation Analyses: Direct and Indirect Gender Effects on Coding Skills

Although no gender-related differences in planning and response inhibition emerged from the *t*-test analyses, it was still possible that children's planning and inhibition skills mediated the effects of gender on coding. The hypothesized mediation relationship was examined by testing the significance of the indirect effects of gender on coding planning time (*Models a* and *b*) and on coding accuracy (*Models c* and *d*) with ToL planning time or Stroop inhibition time (*Models a* and *b*) and accuracy on the ToL or on the Stroop test (*Models c* and *d*) as potential mediators.

The effect, whether direct or indirect, is considered significant if the interval between the upper and lower confidence bounds does not include zero. The four mediation models are reported in **Figures 2**, **3**. **Table 5** reports the direct, indirect and total model effects.

Gender Effects on Coding Planning Time

An inspection of **Table 5** and of *Model a,* reported in **Figure 2**, shows that gender has direct effects on the time spent planning in the coding games at T2 (B = −2.21, $p = 0.02$). Planning time in coding tasks *before* the coding course (T1) has direct effects on children's planning time in coding *after* the course (T2, B = 0.04, $p = 0.06$). Planning time on the ToL at T2 is predicted by

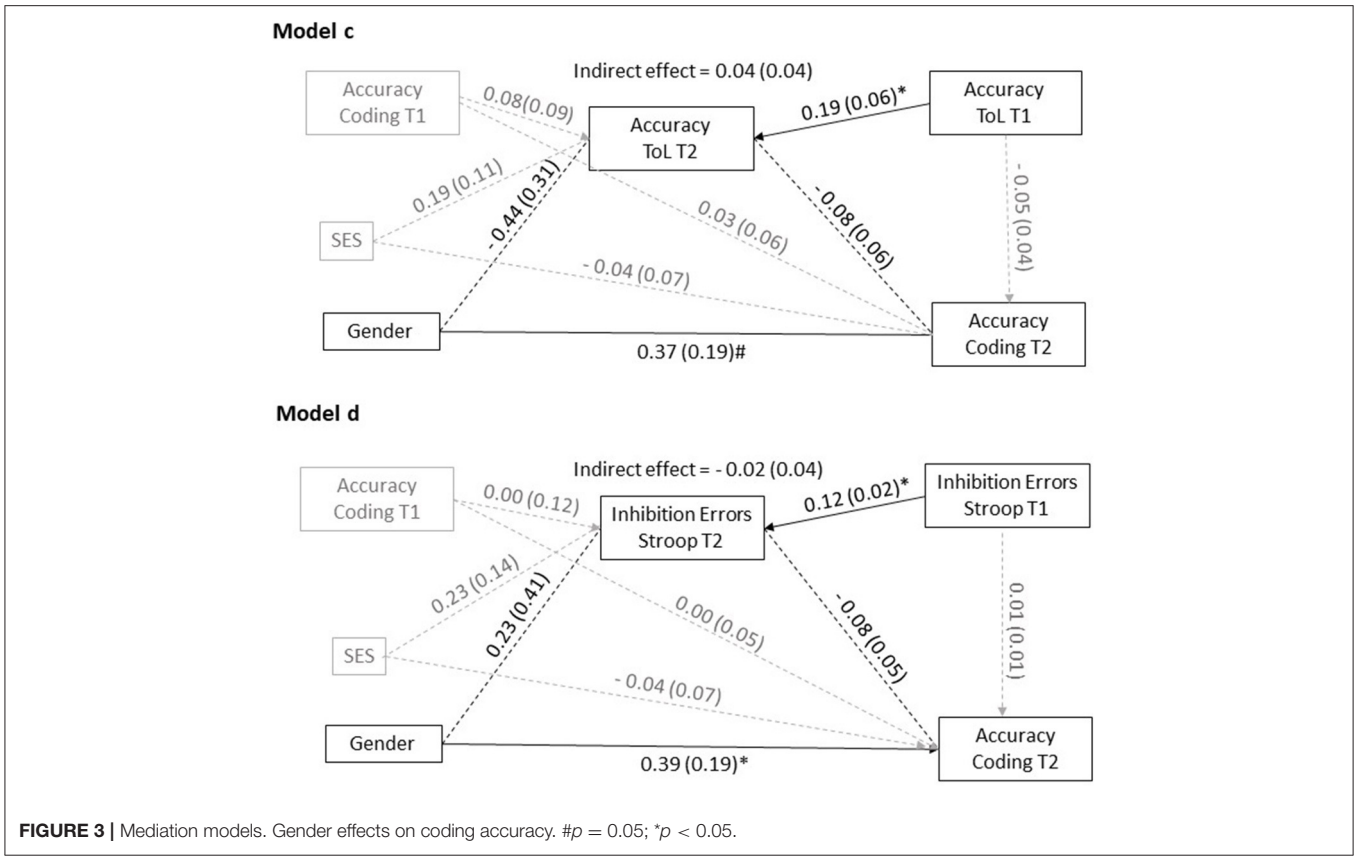


FIGURE 3 | Mediation models. Gender effects on coding accuracy. #$p = 0.05$; *$p < 0.05$.

TABLE 4 | Means, standard deviations and intercorrelations between measures before and after the coding course.

Variable	Before the coding course (T1)						After the coding course (T2)					
	1	2	3	4	5	6	1	2	3	4	5	6
1 Planning time coding	1						1					
2 Accuracy coding	0.042	1					−0.346**	1				
3 Planning time ToL	0.302**	0.219	1				0.308**	−0.347**	1			
4 Accuracy ToL	0.308**	0.361**	0.448**	1			0.126	−0.20*	0.341**	1		
5 Inhibition time	0.072	−0.152	0.174	−0.046	1		0.287**	−0.264**	0.146	0.017	1	
6 Inhibition errors	−0.093	−0.249**	−0.172	−0.318**	0.235	1	0.273**	−0.116	−0.051	−0.172	0.329**	1
M	30.78	3.38	6.24	5.61	212.57	8.63	9.90	6.10	6.46	9.11	174.94	2.23
SD	24.06	1.85	3.00	2.74	72.22	10.2	5.31	0.98	3.07	1.74	46.37	2.40

*$p < 0.05$; **$p < 0.01$; *Planning time coding, Seconds spent planning on the coding games; Accuracy coding, Accuracy on the coding games; Planning time ToL, Seconds spent planning on the Tower of London test; Accuracy ToL, Accuracy on the Tower of London test; Inhibition time, Seconds required to complete the task on the Numerical Stroop test; Inhibition errors, Number of errors and self-corrections on the Numerical Stroop test.*

planning time on the ToL at T1 and has significant effects on planning time in coding at T2 (B = 0.42, $p = 0.02$). However, the indirect effect of gender through this mediator is not significant.

Model b (**Figure 2**), shows that gender and coding planning time *before* the coding course (T1) have direct effects on children's planning time in coding *after* the course (T2), respectively B = −2.07, $p = 0.04$ and B = 0.05, $p = 0.02$. Inhibition time *after* the course (T2) is predicted by inhibition time at T1, *before* the course, and has effects on planning time in coding at T2 (B = 0.03, $p = 0.04$). However, the indirect effect of gender through this mediator is insignificant (see also **Table 5**).

Gender Effects on Coding Accuracy

In *Model c,* reported in **Figure 3**, the direct effect of gender on coding accuracy at T2 approaches statistical significance (B = 0.37, $p = 0.05$). Planning accuracy on the ToL at T1 predicts accuracy on the ToL at T2 (B = 0.19, $p < 0.01$), which does not

TABLE 5 | Direct, indirect, and total effects of the mediation models.

Mediator	Effect	Path	Point estimate	95% CI	
				Lower	Upper
Planning time ToL	Direct effect	Gender → Planning time coding	−2.21*	−4.13	−0.29
		Gender → Planning time ToL	−0.34	−1.44	0.75
		Planning time ToL → Planning time coding	0.42*	0.08	0.76
	Indirect effect	Gender → Planning time ToL → Planning time coding	−0.15	−0.69	0.32
	Total effect	Gender → Planning time coding	−2.35*	−4.32	−0.39
Accuracy ToL	Direct effect	Gender → Accuracy coding	0.37	−0.01	0.75
		Gender → Accuracy ToL	−0.44	−1.06	0.18
		Accuracy ToL → Accuracy coding	−0.08	−0.20	0.04
	Indirect effect	Gender → Accuracy ToL → Accuracy coding	0.03	−0.05	0.13
	total effect	Gender → Accuracy coding	0.40*	0.03	0.78
Inhibition time	Direct effect	Gender → Planning time coding	−2.07*	−3.99	−0.15
		Gender → Inhibition time	−7.22	−22.50	8.07
		Inhibition time → Planning time coding	0.03*	0.00	0.05
	Indirect effect	Gender → Inhibition time → Planning time coding	−0.18	−0.60	0.31
	Total effect	Gender → Planning time coding	−2.25*	−4.20	−0.31
Inhibition errors	Direct effect	Gender → Accuracy coding	0.39*	0.02	0.77
		Gender → Inhibition errors	0.23	−0.58	1.05
		Inhibition errors → Accuracy coding	−0.08	−0.17	0.01
	Indirect effect	Gender → Inhibition errors → Accuracy coding	−0.02	−0.11	0.05
	Total effect	Gender → Accuracy coding	0.38*	0.00	0.76

*$p < 0.05$: Confidence interval for indirect effect was estimated with bootstrap method (95% confidence, 5,000 bootstrap samples). Planning time coding, Seconds spent planning on the coding games at T2 (after the course); Accuracy coding, Accuracy on the coding games at T2; Planning time ToL, Seconds spent planning on the Tower of London test at T2; Accuracy ToL, Accuracy on the Tower of London test at T2; Inhibition time, Seconds required to complete the task on the Numerical Stroop test at T2; Inhibition errors, Number of errors and self-corrections on the Numerical Stroop test atT2; CI, confidence interval.

All parameters are estimated controlling for SES, coding, planning and response inhibition (time and accuracy measures) at T1, before the coding course.

have significant effects on coding accuracy at T2. The indirect effect of gender through planning accuracy (ToL) is insignificant.

Model d shows a significant effect of inhibition errors at T1, *before* the course, on inhibition errors *after* the course (T2) ($B = 0.12$, $p < 0.01$), and a direct effect of gender on coding accuracy after the course (T2) ($B = 0.39$, $p = 0.04$). Again, the indirect effect of gender, through the mediator (inhibition errors at T2) is insignificant.

As shown in **Figures 2** and **3**, the covariates, T1 accuracy in coding and SES, do not account for significant variance in the mediator or in the outcome variable.

Overall, gender effects on children's coding emerged immediately after their first experience with coding; yet, although children's planning and response inhibition skills and coding skills were significantly related, gender effects on coding abilities were not mediated by children's planning or response inhibition skills.

DISCUSSION

The study presented in this paper was inspired by prior research (Sullivan and Bers, 2013, 2016) that reported gender differences in children's coding to emerge since very early experience with it. In this study we further tested the hypothesis that gender differences may already exist among early-age (5–7 year-old) children at their first experience with coding. Moreover, we investigated whether any such gender differences were mediated by gender differences in cognitive abilities underpinning CT and coding.

There is evidence that girls and boys differ in the maturation of some cognitive functions that are known to be involved in coding: in particular, impulsive responses inhibition and planning (Warrick and Naglieri, 1993; Klenberg et al., 2001; Naglieri and Rojahn, 2001; Unterrainer et al., 2013; Grissom and Reyes, 2019). Yet, no studies have directly tested the hypothesis that the gender differences observed between girls and boys in coding could be related to differences in these underlying cognitive abilities. The original contribution of this study was to address this research question.

Prior research conducted with young children has shown that girls develop response inhibition skills and planning skills earlier than boys (Klenberg et al., 2001; Naglieri and Rojahn, 2001; Unterrainer et al., 2013). However, these gender-related developmental differences in executive functioning did not appear in our study. In contrast to our expectations and to the extant literature (Naglieri and Rojahn, 2001; Unterrainer et al., 2013), we did not find differences in planning between girls and boys on the ToL test. Indeed, boys and girls performed equally well on the ToL (planning), and on the Stroop task (response inhibition). Participants in our study ranged in age between 5

and 7 years, with a mean age of 6.03. It may be that at this age, and with school entrance, gender differences in response inhibition and planning have been leveled already (Klenberg et al., 2001; Unterrainer et al., 2013). In contrast with that, however, we did find significant gender differences in the coding tasks *after* the coding course, both in accuracy and planning time. Girls spent significantly more time planning on the coding tasks, without however achieving better performance. Conversely, they were significantly less accurate than boys in them. These gender differences emerged only *after* children experienced coding activities in class: No significant differences were indeed observed between boys and girls *before* the coding course.

The results of the mediation analyses clarified that the gender differences observed in coding *after* the course were not mediated by children's planning abilities (i.e., performance on the ToL) or response inhibition skills (i.e., performance on the Stroop task). Remarkably, the mediational analyses accounted for the effects of covariates like SES, and performance on the coding and cognitive tasks (planning and response inhibition) before the course. Thus, the observed differences cannot be attributed to these factors either.

Some studies have suggested that gender differences in coding may be due to the different strategies or approach to coding problems characteristic of girls and boys (e.g., Sullivan and Bers, 2016), which could be also related to a different use of children's own cognitive abilities. While these findings confirm boys' advantage observed in coding by Sullivan and Bers (Sullivan and Bers, 2013, 2016), they do not support the hypothesis that these differences are accounted for by cognitive predispositions to coding. We can hypothesize that sociocultural factors, such as gender-ability stereotypes, may have affected children's performance. Master et al. (2017) showed that first graders may already have embraced stereotypes that boys are better at programming than girls. If present in our participants, these stereotypes did not affect their performance at their pretest, but did so after the instructional experience with coding. That is, they were likely induced by this early experience and instructional activity. This observation contrasts with Master et al. (2017) finding that introducing girls to programming at this early age may induce counter-stereotypical beliefs and higher self-efficacy for programming. A difference between our study and Master et al.'s study is however that in Master et al.'s study children performed the programming activities in solo sessions, where boys and girls interacted individually with the experimenter. In this study, instead, coding activities were performed in the classroom, thus exposing children to making implicit or explicit comparisons among their performance, with differential effects on boys' and girls' self-confidence. We return on this hypothesis in the conclusion.

It is particularly worrying that the boys' advantage in performance on coding tasks emerged after the children were exposed to counter-stereotypical role models by being shepherded by (young) female trainers, observers, assistants, and experimenters. Other studies have shown that STEM gender stereotypes from early childhood to adolescence are not always influenced by the opportunity to interact with counter-stereotypical educators (McGuire et al., 2020). The mechanisms through which counter-stereotypical models may affect children's self-perception and performance in coding can be thus more complex. They could for instance depend on the nature of the interaction between children and the counter-stereotypical models: e.g., how much children are engaged with them, which activities are performed by these counter-stereotypical models, which is their attitude toward girls and boys. These factors could be the focus of future investigations.

Some more words should be spent on the negative association we found between children's accuracy in coding and the ToL after the coding course. This finding was unexpected, particularly considering that the same two measures were significantly and positively associated *before* the coding course. Although the statistical tests did not reveal significant differences between boys and girls in the performance on the planning task (ToL), girls scored slightly better than boys after the course and spent slightly more time planning on the ToL. However, their performance on the coding tasks was lower than for boys. This may suggest that they made a worse use of their cognitive resources than boys, in what would amount to a gender-related effect. Indeed, by-group correlations reveal that the association between coding and planning skills, albeit non-significant, is negative only for girls.

Planning time and accuracy in coding at T2 were also negatively correlated. Girls spent significantly more time planning in the coding tasks, and yet – as noted above – their accuracy was lower than that of boys. The longer time spent by girls in planning aligns with the findings of other studies, confirming a tendency of girls to be less impulsive and to control more their responses than boys (Grissom and Reyes, 2019). In this study however, while boys and girls showed equivalent response inhibition skills, girls showed a lower coding performance than boys at T2. This observation suggests an alternative explanation: the longer time spent planning in the coding tasks by girls could reflect hesitations in planning more than greater control over their act. Such interpretation would align with findings showing that girls have lower self-confidence in performing coding tasks and greater perception of task difficulty in comparison with boys (Yücel and Rizvanoglu, 2019), which is associated to gender biases. Although the participants in our study were much younger than the participants in Yücel and Rizvanoglu's study, early emerging gender-ability beliefs could have influenced their perception of the coding tasks and, consequently, their performance (Master et al., 2021).

CONCLUSIONS

The results replicated those of prior studies (Sullivan and Bers, 2016), showing that boy-favoring gender differences in coding can emerge at early age (5–7 years). Remarkably, gender differences in coding do not seem to be mediated by differences in cognitive abilities (e.g., planning or response inhibition) and emerged only after children had experienced coding in their classroom. This finding is particularly worrying on account of the young age of our study participants and of the fact that the two groups had equal experience of technological devices and similar cognitive abilities (i.e., similar planning and

inhibition skills). Moreover, all of the personnel that conducted this study and interacted with the children in the training and the experiments were female, providing role models that should have been especially motivating (or at least reassuring) for girls. A possible interpretation of the gender effects observed in this study is that boys and girls matured different self-confidence in coding or different beliefs about it in consequence of their coding experience *in the classroom*. Although performing all games individually, children were requested to signal when they completed each task, to confront and discuss their solution. This individual working format coupled with classroom discussion of the individual outcomes may have stimulated a competitive approach to the coding task, which is often present in classrooms when learning tasks are performed individually and results are compared class-wide. Sullivan and Bers (2016) suggest that competitive learning environments could favor boys. In summary, boys could have been more motivated by interpreting the coding tasks as a competition, whereas girls could have been negatively influenced by this social environment. Different results could have emerged if we requested boys and girls to work in teams and collaboratively (see Sullivan and Bers, 2016).

Limitations of The Study

The data presented in this study were collected as part of a larger project focused on the cognitive effects of coding. In the project we did *not* consider relevant sociocultural factors, such as children's implicit or explicit gender-related beliefs, which may have had significant influence on their performance in coding. In hindsight we have learned that future studies in this direction should explore the influence of motivational and sociocognitive variables (e.g., self-efficacy beliefs) and their interaction with children's cognitive abilities. We are taking this lesson learned home in the design of future interventions directed to digging further into this important gender-difference problem space.

A second limitation of this study is the lack of a qualitative analysis of children's strategies in solving the coding games. As demonstrated by other studies, gender differences could emerge in the way girls and boys approach the coding task (Sullivan and Bers, 2016). Exploring such differences and how they relate to children's cognitive abilities and the type of instruction children have received is important to inform the design of a coding curriculum. Children can be introduced to coding through individual or group activities, playing with robots and tangible environments or via virtual environments (as in our study). The way coding activities are structured, the instrument used, the social partners involved (peers or experts, within gender or across genders) could influence the way children approach coding, the coding skills they develop and the ideas they form about their coding skills (i.e., their self-efficacy beliefs). A more qualitative and finer-grained analysis of children's performance would be possible with the use of additional methodological tools, such as interviews, self-reports, or even behavioral observations and learning analytics. We are working in this very direction at the time of this writing.

A third and final limitation of the study concerns the lack of a fine-grained analysis of children's engagement in the coding course and of their relationship with digital technology. Engagement is known to be an important component of learning, which may be affected by self-perception over and above ability. Measuring engagement before, during, after the intervention should shed more light on the reasons for the emergence of the surprising results from this research. By the token, measuring familiarity with technologies calls for a finer-grained spectrum than merely considering how often children use digital devices, or which digital devices they use, examining also *how* they use them, for what goal, learning or leisure, in which way, alone and unassisted or with friends or seniors. Acquiring and analyzing such additional information would help gain a better understanding of children's prior experience with digital technologies in the way to anticipating engagement. The dimensions of engagement entail a social element and a learning element. Both are very important aspects of a wider study into the nature, origin, and mitigations of gender differences in relation to coding and, prospectively to Computer Science.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethical Committee for the Psychological Research of the University of Padova. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

BA: conception and design of the project, project coordination, paper drafting and revising, and data interpretation. CM: data gathering, coordination of the data gathering, article drafting, and data analyses. LR: guiding the data analysis. TV: article drafting and revising and coordination of the study. All authors contributed to the article and approved the submitted version.

REFERENCES

Aesaert, K., and van Braak, J. (2015). Gender and socioeconomic related differences in performance based ICT competences. *Comput. Educ.* 84, 8–25. doi: 10.1016/j.compedu.2014.12.017

Alonso, M. T., Barba-Sánchez, V., López Bonal, M. T., and Macià, H. (2021). Two perspectives on the gender gap in computer engineering: from secondary school to higher education. *Sustainability* 13, 10445. doi: 10.3390/su131810445

Anderson, N., Lankshear, C., Timms, C., and Courtney, L. (2008). 'Because it's boring, irrelevant and I don't like computers': Why high school girls avoid professionally-oriented ICT subjects. *Comput. Educ.* 50, 1304–1318. doi: 10.1016/j.compedu.2006.12.003

Arfé, B., Vardanega, T., Montuori, C., and Lavanga, M. (2019). Coding in primary grades boosts children's executive functions. *Front Psychol.* 10, 2713. doi: 10.3389/fpsyg.2019.02713

Arfé, B., Vardanega, T., and Ronconi, L. (2020). The effects of coding on children's planning and inhibition skills. *Comput. Educ.* 148, 103807. doi: 10.1016/j.compedu.2020.103807

Bati, K. (2022). A systematic literature review regarding computational thinking and programming in early childhood education. *Educ. Inf. Technol.* 27, 2059–2082. doi: 10.1007/s10639-021-10700-2

Beyer, S. (2014). Why are women underrepresented in Computer Science? Gender differences in stereotypes, self-efficacy, values, and interests and predictors of future CS course-taking and grades. *Comput. Sci. Educ*. 24, 153–192. doi: 10.1080/08993408.2014.963363

Cai, Z., Fan, X., and Du, J. (2017). Gender and attitudes toward technology use: A meta-analysis. *Comput. Educ.* 105, 1–13. doi: 10.1016/j.compedu.2016.11.003

Charles, M., Harr, B., Cech, E., and Hendley, A. (2014). Who likes math where? Gender differences in eighth-graders' attitudes around the world. *Int. Studies Sociol. Educ.* 24, 85–112. doi: 10.1080/09620214.2014.895140

Charlesworth, T. E. S., and Banaji, M. R. (2019). Gender in science, technology, engineering, and mathematics: issues, causes, solutions. *J. Neurosci.* 39, 7228–7243. doi: 10.1523/JNEUROSCI.0475-18.2019

Chiazzese, G., Fulantelli, G., Pipitone, V., and Taibi, D. (2017). "Promoting computational thinking and creativeness in primary school children," in *Proceedings of the 5th International Conference on Technological Ecosystems for Enhancing Multiculturality*. p. 1–7. doi: 10.1145/3144826.3145354

Cvencek, D., Meltzoff, A. N., and Greenwald, A. G. (2011). Math–gender stereotypes in elementary school children. *Child Develop.* 82, 766–779. doi: 10.1111/j.1467-8624.2010.01529.x

Denner, J., Werner, L., O'Connor, L., and Glassman, J. (2014). Community college men and women: a test of three widely held beliefs about who pursues computer science. *Commun. College Rev*. 42, 342–362. doi: 10.1177/0091552114535624

Di Lieto, M. C., Pecini, C., Castro, E., Inguaggiato, E., Cecchi, F., Dario, P., Cioni, G., and Sgandurra, G. (2020). Empowering executive functions in 5- and 6-year-old typically developing children through educational robotics: An RCT study. *Front. Psychol.* 10, 3084. doi: 10.3389/fpsyg.2019.03084

Diamond, A. (2013). Executive functions. *Ann. Rev. Psychol.* 64, 135–168. doi: 10.1146/annurev-psych-113011-143750

Fancello, G. S., Vio, C., and Cianchetti, C. (2013). *ToL, Torre di Londra. Test di Valutazione delle Funzioni Esecutive (Pianificazione e Problem Solving).* Trento: Erickson.

Fisher, A., and Margolis, J. (2003). "Unlocking the clubhouse: Women in computing," in *Proceedings of the 34th SIGCSE Technical Symposium on Computer Science Education*. p. 23. doi: 10.1145/611892.611896

Flore, P. C., and Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis. *J. School Psychol.* 53, 25–44. doi: 10.1016/j.jsp.2014.10.002

Flórez, B. F., Casallas, R., Hernández, M., Reyes, A., Restrepo, S., and Danies, G. (2017). Changing a generation's way of thinking: teaching computational thinking through programming. *Rev. Educ. Res.* 87, 834–860. doi: 10.3102/0034654317710096

Frieze, C. (2005). Diversifying the images of computer science: Undergraduate women take on the challenge! *ACM SIGCSE Bull.* 37, 397–400. doi: 10.1145/1047124.1047476

Gerson, S. A., Morey, R. D., and van Schaik, J. E. (2022). Coding in the cot? Factors influencing 0–17s' experiences with technology and coding in the United Kingdom. *Comput. Educ.* 178, 104400. doi: 10.1016/j.compedu.2021.104400

Girelli, L. (2022). What does gender has to do with math? Complex questions require complex answers. *J. Neurosci. Res.* doi: 10.1002/jnr.25056

Gnambs, T. (2021). The development of gender differences in information and communication technology (ICT) literacy in middle adolescence. *Comput. Human Behav.* 114, 106533. doi: 10.1016/j.chb.2020.106533

Grissom, N. M., and Reyes, T. M. (2019). Let's call the whole thing off: evaluating gender and sex differences in executive function. *Neuropsychopharmacology* 44, 86–96. doi: 10.1038/s41386-018-0179-5

Halpern, D. F., and LaMay, M. L. (2000). The smarter sex: a critical review of sex differences in intelligence. *Educ. Psychol. Rev.* 12, 229–246. doi: 10.1023/A:1009027516424

Harter, G. L. (1930). Overt trial and error in the problem solving of preschool children. *Pedagogical Semin. J. Genetic Psychol.* 38, 361–372. doi: 10.1080/08856559.1930.10532270

Honicke, T., and Broadbent, J. (2016). The influence of academic self-efficacy on academic performance: A systematic review. *Educ. Res. Rev.* 17, 63–84. doi: 10.1016/j.edurev.2015.11.002

Jackson, L. A., Zhao, Y., Kolenic, A., Fitzgerald, H. E., Harold, R., and Von Eye, A. (2008). Race, gender, and information technology use: the new digital divide. *CyberPsychol. Behav.* 11, 437–442. doi: 10.1089/cpb.2007.0157

Jiang, S., and Wong, G. K. W. (2021). Exploring age and gender differences of computational thinkers in primary school: A developmental perspective. *J. Computer Assisted Learning*. 38, 60–75. doi: 10.1111/jcal.12591

Kersey, A. J., Braham, E. J., Csumitta, K. D., Libertus, M. E., and Cantlon, J. F. (2018). No intrinsic gender differences in children's earliest numerical abilities. *Npj Sci. Learn.* 3, 1–10. doi: 10.1038/s41539-018-0028-7

Klenberg, L., Korkman, M., and Lahti-Nuuttila, P. (2001). Differential development of attention and executive functions in 3- to 12-year-old finnish children. *Develop. Neuropsychol.* 20, 407–428. doi: 10.1207/S15326942DN2001_6

Kožuh, I., Krajnc, R., Hadjileontiadis, L. J., and Debevc, M. (2018). Assessment of problem solving ability in novice programmers. *PLoS ONE* 13, e0201919. doi: 10.1371/journal.pone.0201919

Kraemer, H. C., Kiernan, M., Essex, M., and Kupfer, D. J. (2008). How and why criteria defining moderators and mediators differ between the baron and kenny and macarthur approaches. *Health Psychol.* 27, S101–S108. doi: 10.1037/0278-6133.27.2(Suppl.).S101

Lau, W. W. F., and Yuen, A. H. K. (2009). Exploring the effects of gender and learning styles on computer programming performance: Implications for programming pedagogy. *Br. J. Educ. Technol.* 40, 696–712. doi: 10.1111/j.1467-8535.2008.00847.x

Luciana, M., Collins, P. F., Olson, E. A., and Schissel, A. M. (2009). Tower of London performance in healthy adolescents: the development of planning skills and associations with self-reported inattention and impulsivity. *Develop. Neuropsychol.* 34, 461–475. doi: 10.1080/87565640902964540

MacKinnon, D. P. (2008). *Introduction to Statistical Mediation Analysis*. Routledge.

Maloney, E. A., Waechter, S., Risko, E. F., and Fugelsang, J. A. (2012). Reducing the sex difference in math anxiety: The role of spatial processing ability. *Learn. Individual Differ.* 22, 380–384. doi: 10.1016/j.lindif.2012.01.001

Martin, C. L., Wood, C. H., and Little, J. K. (1990). The development of gender stereotype components. *Child Develop.* 61, 1891–1904. doi: 10.2307/1130845

Marzocchi, G. M., Re, A. M., and Cornoldi, C. (2010). *BIA. Batteria Italiana per l'ADHD*. Trento: Erickson.

Master, A. (2021). Gender Stereotypes Influence Children's STEM Motivation. *Child Develop. Perspect.* 15, 203–210. doi: 10.1111/cdep.12424

Master, A., Cheryan, S., Moscatelli, A., and Meltzoff, A. N. (2017). Programming experience promotes higher STEM motivation among first-grade girls. *J. Exper. Child Psychol.* 160, 92–106. doi: 10.1016/j.jecp.2017.03.013

Master, A., Meltzoff, A. N., and Cheryan, S. (2021). Gender stereotypes about interests start early and cause gender disparities in computer science and engineering. *Proc. Nat. Acad. Sci.* 118, e2100030118. doi: 10.1073/pnas.2100030118

McGuire, L., Mulvey, K. L., Goff, E., Irvin, M. J., Winterbottom, M., Fields, G. E., Hartstone-Rose, A., and Rutland, A. (2020). STEM gender stereotypes from early childhood through adolescence at informal science centers. *J. Appl. Develop. Psychol.* 67, 101109. doi: 10.1016/j.appdev.2020.101109

Mejía-Rodríguez, A. M., Luyten, H., and Meelissen, M. R. M. (2020). Gender differences in mathematics self-concept across the world: an exploration of student and parent data of TIMSS 2015. *Int. J. Sci. Mathemat. Educ.* 19, 1229–1250. doi: 10.1007/s10763-020-10100-x

Miller, D. I., and Halpern, D. F. (2014). The new science of cognitive sex differences. *Trends Cogn. Sci.* 18, 37–45. doi: 10.1016/j.tics.2013.10.011

Miller, D. I., Nolla, K. M., Eagly, A. H., and Uttal, D. H. (2018). The development of children's gender-science stereotypes: a meta-analysis of 5 decades of U.S. Draw-A-Scientist Studies. *Child Develop.* 89, 1943–1955. doi: 10.1111/cdev.13039

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., and Wager, T. D. (2000). The unity and diversity of executive functions and their

contributions to complex 'Frontal Lobe' tasks: A latent variable analysis. *Cogn. Psychol.* 41, 49–100. doi: 10.1006/cogp.1999.0734

Naglieri, J. A., and Rojahn, J. (2001). Gender differences in planning, attention, simultaneous, and successive (PASS) cognitive processes and achievement. *J. Educ. Psychol.* 93, 430–437. doi: 10.1037/0022-0663.93.2.430

Nardelli, E. (2019). Do we really need computational thinking? *Commun. ACM* 62, 32–35. doi: 10.1145/3231587

Papadakis, S. (2021). The impact of coding apps to support young children in computational thinking and computational fluency. a literature review. *Front. Educ.* 6, 657895. doi: 10.3389/feduc.2021.657895

Papavlasopoulou, S., Sharma, K., and Giannakos, M. N. (2020). Coding activities for children: Coupling eye-tracking with qualitative data to investigate gender differences. *Comput. Human Behav.* 105, 105939. doi: 10.1016/j.chb.2019.03.003

Price, C. B., and Price-Mohr, R. (2021). Exploring gender differences in primary school computer programming classes: A study in an English state-funded urban school. *Education* 3–13, 1–14. doi: 10.1080/03004279.2021.1971274

Reilly, D., Neumann, D. L., and Andrews, G. (2019). Investigating Gender Differences in Mathematics and Science: Results from the 2011 Trends in Mathematics and Science Survey. *Res. Sci. Educ.* 49, 25–50. doi: 10.1007/s11165-017-9630-6

Resnick, M., Maloney, J., Monroy-Hernandez, A., Rusk, N., Eastmond, E., Brennan, K., et al. (2009). Scratch: Programming for all. *Commun. ACM* 52, 60–67. doi: 10.1145/1592761.1592779

Roman-Gonzalez, M., Perez-Gonzalez, J. C., and Jimenez-Fernandez, C. (2017). Which cognitive abilities underlie computational thinking? Criterion validity of the computational thinking test. *Comput. Hum. Behav.* 72, 678–691. doi: 10.1016/j.chb.2016.08.047

Ronsivalle, G. B., Boldi, A., Gusella, V., Inama, C., and Carta, S. (2019). How to implement educational robotics' programs in italian schools: a brief guideline according to an instructional design point of view. *Technol. Knowl. Learn.* 24, 227–245. doi: 10.1007/s10758-018-9389-5

Schmidt, F. L. (2011). A theory of sex differences in technical aptitude and some supporting evidence. *Perspect. Psychol. Sci.* 6, 560–573. doi: 10.1177/1745691611419670

Shute, V. J., Sun, C., and Asbell-Clarke, J. (2017). Demystifying computational thinking. *Educ. Res. Rev.* 22, 142–158. doi: 10.1016/j.edurev.2017.09.003

Siddiq, F., and Scherer, R. (2019). Is there a gender gap? A meta-analysis of the gender differences in students' ICT literacy. *Educ. Res. Rev.* 27, 205–217. doi: 10.1016/j.edurev.2019.03.007

Spearman, J., and Watt, H. M. G. (2013). "Women's Aspirations Towards "STEM" Careers," in W. Patton (Ed.), *Conceptualising Women's Working Lives: Moving the Boundaries of Discourse*. Rotterdam: SensePublishers. p. 175–191. doi: 10.1007/978-94-6209-209-9_10

Spencer, S. J., Steele, C. M., and Quinn, D. M. (1999). Stereotype Threat and Women's Math Performance. *J. Exper. Soc. Psychol.* 35, 4–28. doi: 10.1006/jesp.1998.1373

Statter, D., and Armoni, M. (2017). "Learning abstraction in computer science: a gender perspective," in *Proceedings of the 12th Workshop on Primary and Secondary Computing Education*. p. 5–14. doi: 10.1145/3137065.3137081

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *Am. Psychol.* 52, 613–629. doi: 10.1037/0003-066X.52.6.613

Stoet, G., and Geary, D. C. (2013). Sex differences in mathematics and reading achievement are inversely related: within- and across-nation assessment of 10 years of PISA data. *PLOS ONE* 8, e57988. doi: 10.1371/journal.pone.0057988

Sullivan, A., and Bers, M. U. (2013). Gender differences in kindergarteners' robotics and programming achievement. *Int. J. Technol. Design Educ.* 23, 691–702. doi: 10.1007/s10798-012-9210-z

Sullivan, A., and Bers, M. U. (2016). Girls, Boys, and Bots: Gender Differences in Young Children's Performance on Robotics and Programming Tasks. *J. Inf. Technol. Educ.* 15, 145–165. doi: 10.28945/3547

UNESCO (2017). Cracking the code: Girls' and women's education in science, technology, engineering and mathematics (STEM)—UNESCO Digital Library. Available online at: https://unesdoc.unesco.org/ark:/48223/pf0000253479 (accessed August 23, 2022).

Unterrainer, J. M., Ruh, N., Loosli, S. V., Heinze, K., Rahm, B., and Kaller, C. P. (2013). Planning Steps Forward in Development: In Girls Earlier than in Boys. *PLOS ONE* 8, e80772. doi: 10.1371/journal.pone.0080772

Viterbori, P., Traverso, L., and Usai, M. C. (2017). The role of executive function in arithmetic problem-solving processes: A study of third graders. *J. Cogn. Develop.* 18, 595–616. doi: 10.1080/15248372.2017.1392307

Wang, M. T., and Degol, J. L. (2017). Gender gap in science, technology, engineering, and mathematics (STEM): current knowledge, implications for practice, policy, and future directions. *Educ. Psychol. Rev.* 29, 119–140. doi: 10.1007/s10648-015-9355-x

Warrick, P., and Naglieri, J. (1993). Gender differences in planning, attention, simultaneous, and successive (PASS) cognitive processes. *J. Educ. Psychol.* 85, 693–701. doi: 10.1037/0022-0663.85.4.693

Wierenga, L. M., Bos, M. G. N., van Rossenberg, F., and Crone, E. A. (2019). Sex effects on development of brain structure and executive functions: greater variance than mean effects. *J. Cogn. Neurosci.* 31, 730–753. doi: 10.1162/jocn_a_01375

Wing, J. M. (2006). Computational thinking. *Commun. ACM* 49, 33–35. doi: 10.1145/1118178.1118215

Yasar, O. (2017). The essence of computational thinking. *Comput. Sci. Eng.* 19, 74–82. doi: 10.1109/MCSE.2017.3151241

Yücel, Y., and Rizvanoglu, K. (2019). Battling gender stereotypes: A user study of a code-learning game, "Code Combat," with middle school children. *Comput. Human Behav.* 99, 352–365. doi: 10.1016/j.chb.2019.05.029

Zelazo, P. D., Müller, U., Frye, D., Marcovitch, S., Argitis, G., Boseovski, J., Chiang, J. K., Hongwanishkul, D., Schuster, B. V., and Sutherland, A. (2003). The development of executive function in early childhood. *Monographs Soc. Res. Child Develop.* 68, vii−137. doi: 10.1111/j.0037-976X.2003.00261.x

Zweben, S., and Aspray, W. (2004). Undergraduate enrollments drop; department growth expectations moderate. *Computing Research News*. Available online at: http://archive.cra.org/Activities/workshops/broadening.participation/cra/Taulbee/2002-2003.pdf (accessed August 23, 2022).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Montuori, Ronconi, Vardanega and Arfé. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 21 September 2022
DOI 10.3389/fpsyg.2022.1010033

Check for updates

OPEN ACCESS

EDITED BY
Stamatios Papadakis,
University of Crete,
Greece

REVIEWED BY
Nyet Moi Siew,
Universiti Malaysia Sabah,
Malaysia
Effransia Tzagkaraki,
University of Crete,
Greece

*CORRESPONDENCE
Tianxiang Zheng
zheng_tx@jnu.edu.cn

SPECIALTY SECTION
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

RECEIVED 02 August 2022
ACCEPTED 05 September 2022
PUBLISHED 21 September 2022

CITATION
Zheng Y, Liu P, Yang X, Guo Y, Qiu X, Jin X, Luo X and Zheng T (2022) K–12 Science, Technology, Engineering, and Math characteristics and recommendations based on analyses of teaching cases in China.
Front. Psychol. 13:1010033.
doi: 10.3389/fpsyg.2022.1010033

COPYRIGHT
© 2022 Zheng, Liu, Yang, Guo, Qiu, Jin, Luo and Zheng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

K–12 Science, Technology, Engineering, and Math characteristics and recommendations based on analyses of teaching cases in China

Yunxiang Zheng[1], Panpan Liu[1], Xinru Yang[1], Yidong Guo[1], Xinxin Qiu[1], Xiunan Jin[1], Xianfei Luo[1] and Tianxiang Zheng[2]*

[1]School of Educational Information Technology, South China Normal University, Guangzhou, Guangdong, China, [2]Department of E-Commerce, Jinan University, Shenzhen, Guangdong, China

Science, Technology, Engineering, and Math (STEM) education emphasizes solving problems in authentic contexts and developing 21st-century skills. It also helps to cultivate individuals possessing scientific curiosity and innovative abilities. These capacities align with China's core literacy training. Recent years have seen K–12 STEM cases flourish nationally. However, little attention has been paid to the shared characteristics of these practices, and suggestions for implementing STEM in primary and secondary schools are scarce. This paper presents commonalities in STEM practices within China from a curriculum perspective and offers recommendations for implementation based on these attributes. Specifically, this study first constructed analytical metrics *via* the Delphi method to assess STEM cases. Next, 51 typical STEM teaching cases in different regions of China were analyzed using these metrics. Based upon the statistical results, five characteristics of STEM cases were summarized: China's STEM education has an unbalanced geographical distribution; current practices benchmark the need for innovative talent training; most instructional content is drawn from real-world problems, but interdisciplinary integration deserves closer focus; the cases featured rich teaching activities and were conducted in a project-based learning fashion with insufficient emphasis on mathematical applications; and China seems to be holistically promoting STEM education, especially through new technologies and supplementary materials. Findings should allow instructors to better understand the intricacies of STEM implementation and to promote successful STEM cases. Recommendations are also provided to optimize the localization of STEM education in China in order to cultivate innovative and interdisciplinary talent.

KEYWORDS

STEM education, STEM characteristics, teaching cases, analytical metrics, case analysis

Introduction

Science, Technology, Engineering, and Math (STEM) education, a domain conceived by the U.S. National Science Foundation, is a student-centered approach that encourages people to problem solve *via* scientific methods. It emphasizes tackling real-world problems (Honey et al., 2014) using 21st-century skills (e.g., critical and creative thinking, research and questioning, communication, and teamwork) in school settings (Falloon et al., 2020). This type of education prepares students to view issues through an interdisciplinary lens. Students also learn to apply scientific and technical knowledge in multiple life domains and to innovatively solve problems. Ultimately, STEM training enables students to cultivate skills they will need in the future.

STEM education enjoys worldwide popularity; its practices are of great importance within primary, secondary, and higher education (Bolatli and Korucu, 2018; Bryan and Guzey, 2020). STEM is especially useful in K–12 settings—it lays a foundation for a nation's future science and engineering workforce, prosperity, and even security. Although various views exist regarding the nature of K–12 STEM education, stakeholders have come to focus on STEM integration. The term "integrated STEM" refers to the deliberate combination of core disciplinary content from STEM disciplines (Guzey et al., 2016; Bryan and Guzey, 2020). STEM is no longer considered four isolated disciplines that are implemented individually; the domain is instead treated as a single unified discipline. For the purposes of this paper, "STEM" refers to integrated STEM unless stated otherwise.

Several studies have documented the advantages of integrated STEM education and the benefits of students' engagement in STEM activities in K–12 classrooms (English, 2017; Means et al., 2017; Kim et al., 2018; Gardner and Tillotson, 2019). Based on these success stories, together with the idea that STEM education can meet scholastic needs for talent training, China's STEM education has flourished. A growing number of STEM cases have thus emerged nationally, especially in primary and secondary schools. To further advance STEM education, China's "Ministry of Education of the People's Republic of China (2022)" proposes fostering students' interdisciplinary literacy by explaining interdisciplinary concepts based on core aspects of each subject area and then applying these core elements to real situations. Instructors are now encouraged to implement STEM in their teaching. However, involving K–12 students in STEM education calls for reforms to curricula, pedagogy, and the learning environment to ensure a focus on disciplinary knowledge as well as creativity, reasoning, and innovation (Freeman et al., 2019). Most K–12 schools in China, especially those in underdeveloped rural areas, have little experience with such practices (Xu et al., 2021; Wang, 2022). Instructors in these schools may face obstacles when selecting topics, designing activities, and combining related disciplines. To address these concerns, the "China STEM Education Innovation Action Plan 2029" recommends highlighting successful models of STEM education and sharing examples of sound practices nationwide (National Institute of Education Sciences, 2017). Yet sporadic experience cannot be replicated, and a single case only spreads across a small area. A dearth of analysis on the characteristics of K–12 STEM practices in China precludes the smooth development of such education. A literature review spotlighted the lack of attention to common characteristics from a curriculum perspective. Scholars have also rarely offered suggestions for implementing STEM in primary and secondary schools based on these attributes. As such, government officials are unfamiliar with the constitution of STEM education. Instructors in most schools possess a limited sense of what STEM practice entails from a curriculum perspective. These considerations necessitate a systematic analysis of STEM cases to reveal how STEM is implemented.

This paper outlines common STEM practices in China from a curriculum perspective and offers corresponding guidance on implementation. Findings are expected to help instructors better understand how to implement STEM, to promote success in this area, and to foster the localization of STEM education throughout the country. It is first necessary to determine how to evaluate K–12 STEM practices from an instructional design standpoint. To date, relevant studies in China have either covered individual STEM projects or assumed a non-curriculum view. Suitable analytical metrics must therefore be developed. These metrics will lay the groundwork for case analysis and relevant recommendations. It is similarly necessary to discuss typical STEM cases and discern their commonalities based on these metrics. The results of this endeavor are expected to inform suggestions for ways to incorporate STEM into the school curriculum and implement it through teaching reform to promote its localization in China.

The following questions drive this research:

RQ1: How can a K–12 STEM case be systematically analyzed from a curriculum perspective?

RQ2: What characteristics do typical STEM cases in China share?

RQ3: What suggestions can be made for implementing STEM in primary and secondary schools in China based on these characteristics?

Related work

Cross-case analysis or evaluation of K–12 STEM practices

Cross-case analysis or evaluation of K–12 STEM practices has garnered extensive interest. For example, STEMworks at Wested (Stemworks at Wested, 2017) proposed a rubric for STEM projects that aligns with a set of common "Design Principles for Effective STEM Philanthropy." The group sought to create a framework for corporate engagement that improves youth's STEM performance. Lynch et al. (2018) conducted a cross-case analysis of case studies describing the design and implementation of eight "exemplar" inclusive STEM high schools. The authors identified several

critical components and painted a picture of how these high schools achieved their goals (e.g., administrative structure, college-preparatory STEM-focused curriculum, well-prepared STEM teachers). Falloon et al. (2021) performed a cross-case analysis to evaluate factors contributing to the development of four contrasting schools' STEM profiles. The schools were found to have delivered STEM curricula that met students' learning needs in the local context. Guzey et al. (2016) judged 20 STEM integration units using the STEM Integration Curriculum Assessment tool and compared them. Łukaszczyk and Grebski (2020) carried out a comparative analysis of the curriculum by selecting two STEM-oriented high schools in Poland and the United States.

In China, Li and Xu (2018) scrutinized eight outstanding science, technology, engineering, arts, and mathematics (STEAM) education cases in the United States. Their analysis addressed educational goals, characteristics, practice, and evaluation. Results indicated that STEAM education can uniquely nurture innovative talent by integrating art and STEM. Wang et al. (2021) considered 45 typical STEM curriculum cases in China and the United States to explore trends in both countries and put forward actionable suggestions for optimal STEM curricula in China. Wang et al. (2019) examined eight typical STEM education programs in the United States from four dimensions: evaluation subject, evaluation object, evaluation content/indicator, and supervision and feedback. The authors discovered that American STEM education displayed a relatively complete closed-loop evaluation mechanism. Chen K. et al. (2021) conducted content analysis on 78 STEM teaching cases in major journals in the fields of physics, chemistry, biology, and geography. They observed clear disciplinary attributes of STEM instructional design in middle schools; however, the use of some key pedagogical strategies was lacking, and a detailed evaluation instrument was absent. Several master's students (Miao, 2018; Yin, 2019; Lv, 2020; Qiao, 2020) respectively analyzed prototypical STEM courses from the United States and China.

Characteristics of K–12 STEM practices

Many researchers have examined the attributes of STEM practices. Hansen (2014) generated value-added estimates in math and science to categorize schools into performance levels and evaluated differences in school-attributed STEM outcomes using longitudinal data on students in the United States. states of Florida and North Carolina. Scott (2009) considered 10 STEM-focused high schools in the United States and identified key design components. Schools were selected from various regions across the country. He noted that half of the high schools used a lottery system to select students; in addition to coursework requirements, students also needed to complete internships and/or a capstone project. At the curriculum level, Okulu and Oguz-Unver (2021) developed a measure to evaluate STEM activities using a case study method. The STEM activity assessment form was developed based on a literature review and experts' opinions. Nite et al. (2017) explicated the characteristics of STEM teaching and learning in middle and high school and in informal settings by examining 58 research sources between 2005 and 2012. Associated themes included reform-based teaching and learning, informal education, teacher factors, and technology use. Sources in different categories were compared based on their features.

As described, limited research has entailed STEM analysis. Non-Chinese scholars appear especially interested in the holistic assessment of STEM projects. They have especially focused on the return on investment, which can help schools and governments more thoroughly evaluate the impact of STEM in K–12 education. Chinese scholars have tended to compare STEM cases from different countries and to investigate the teaching modes or instructional design of these curricula. Far less is known about the shared elements of these practices from a curriculum perspective. Scholars have also rarely offered suggestions for implementing STEM in primary and secondary schools based on these attributes. Given the scope of STEM practices throughout China, a closer analysis of several cases can highlight typical characteristics at the curriculum level. Findings are intended to inform teachers' design, implementation, and refinement of local STEM courses.

Methodology

Similar to prior work (Carter, 2013; Hansen, 2014; Okulu and Oguz-Unver, 2021), metrics were created in this study based on a literature review and experts' opinions when analyzing selected cases. Accordingly, to address the three questions raised in Section "Introduction," we first constructed a set of analytical metrics. We next chose several cases for fine-grained investigation using the designed metrics and distilled cases' common features. Then, we identified areas for improvement.

Instrumentation

To answer RQ1, we extracted core elements of a K–12 course based on China's Primary and Secondary School Curriculum Standard, elicited key aspects of STEM, and compared several STEM design models *via* a literature review. Quantitative content analysis was used to clarify the items of interest for these cases, from which primary indicators were obtained. We then identified secondary indicators to construct preliminary analytical metrics. Next, the metrics were finalized *via* the Delphi method (Rowe and Wright, 1999; Green, 2014): a team of experts, researchers, and front-line teachers in STEM were invited to assess the metrics' rationality and give suggestions for revision. The Delphi technique is grounded in a series of questionnaires and iterations. The first survey may include general questions. In each subsequent stage, the questions become more specific in relation to responses on previous questions. Questionnaires in this study were distributed *via* email for experts to complete (and offer feedback) to ensure participant anonymity. We, as the researchers, integrated their

comments and then quantitatively interpreted the collected data. The expert panel was given a summary of expert judgments and then pondered their opinions in light of this information. A consensus was reached after two rounds of iteration. The resultant metrics consisted of six primary indicators (background information, teaching objectives, knowledge content, teaching activities, teaching support, and teaching evaluations) and 30 secondary indicators (see Section "Analytical metrics" for details).

To resolve RQ2, we selected and analyzed several STEM cases in China based on our metrics. Secondary indicators related to background information and teaching evaluations were measured qualitatively through content analysis and quantitatively *via* statistics (e.g., average, standard deviation, word frequency). The other indicators were measured quantitatively on 5-point Likert-type scales ranging from 1 (*non-compliant*) to 5 (*completely compliant*). We discerned commonalities across these sample cases. Finally, we made several recommendations for improvement when implementing STEM in primary and secondary schools in China by combining the above attributes with national conditions; in doing so, we responded to RQ3.

Sampling

Each profiled K–12 STEM teaching case met several criteria. First, each case was implemented using project-based learning (PjBL); that is, it was independently carried out in line with the PjBL principles of planning, creating, processing, and evaluating. Second, each case was interdisciplinary (i.e., it conformed to the characteristics of STEM education and was distinguishable from general teaching in its interdisciplinary nature). Third, each case was whole-class-oriented—teaching objects were students from a whole class in a primary or secondary school (vs. small groups in a school club). Fourth, each case had recently been completed (i.e., between 2018 and 2021) to provide a timely sense of China's STEM practices. Fifth, each case featured abundant resource support: rich materials were available for analysis (e.g., instructional handouts, photos, or videos; archives of student worksheets or other student work). Our formal dataset consisted of 51 K–12 STEM teaching cases in China.

Analysis procedure

To improve the accuracy and objectivity of our results, several analysts were asked to review the same STEM teaching case. Results were subjected to a Kappa test for consistency (*via* SPSS). Analysts appeared to agree in their case judgments (Kappa = 0.679). The analysts later shared their opinions of the cases, resolved disagreements through discussion, and evaluated the 51 cases according to the established metrics. Results of the analysis were visualized through radar charts and bar charts in Excel. Some indicators (e.g., topics and keywords) were further imported into SmartAnalyze to build word cloud diagrams. The characteristics of all STEM cases were summarized to devise corresponding suggestions.

Analytical metrics

Theoretical background

Core elements of a K–12 course

According to China's Primary and Secondary School Curriculum Standard, each K–12 course includes a purpose, objectives, content, and implementation. These core elements were deconstructed into indicators or sub-indicators for analysis.

Key features of STEM

The "White Paper on STEM Education in China," issued by the National Institute of Education Sciences (2017), frames science and technology innovation education as a lifelong learning activity. STEM is an interdisciplinary and interprofessional domain that serves as a carrier for inclusive student training. This type of education also requires joint participation from society to realize educational innovation, exemplifying STEM's interdisciplinary nature.

Besides interdisciplinarity, STEM education partly relies on PjBL; that is, presented problems are usually authentic based on the curriculum. Associated PjBL activities include the presentation of context (importing), problem identification, group-based problem exploration, group-based manual engineering, displays of achievement, and self-reflection. Put simply, projects require students to conduct research similarly to a scientist, test with technology like an engineer, and think as a mathematician would. PjBL dominates STEM learning models (Khotimah et al., 2021). The features of interdisciplinarity and PjBL should thus be captured in an analytic model of STEM teaching cases. Kang (2019) noticed that more than half of all students, both elementary and secondary, identified either subject integration or group work as the most distinctive feature of STEAM classes. This pattern further supports the above two attributes.

STEM design models

Li and Li (2019) argued that a STEM curriculum includes four basic elements: the STEM topic, learning objectives, learning activities, and supporting materials. Yu and Hu (2015) built a STEM design model based on constructivism; it comprises teaching analysis, a learning task, learning scaffolds, learning activities, tools and resources, learning evaluations, summary and exercises, and experimentation and improvement. Learning objectives, theme selection, learning activities, teaching evaluations, and learning support should hence be considered when designing a STEM project.

To provide authentic learning settings and to enable students to make connections among STEM disciplines, teachers should provide interdisciplinary activities to engage students rather than merely lecturing to impart knowledge. Such activities are integral

TABLE 1 Components of three STEM activity models.

Model	Components
5E	Engagement, Exploration, Explanation, Elaboration, Evaluation
6E	Engage, Explore, Explain, Engineer, Enrich, Evaluate
5EX	EQ (Scenario Entering and Question Raising), EM (Scientific Exploration and Mathematical Application), ET (Engineering Design and Technical Making), EC (Knowledge Expansion and Creative Design), ER (Multi-evaluation and Learning Reflection)

to STEM education—they shape how students learn and communicate. These activities also inform students' thinking abilities, collaboration skills, presentation skills, and problem-solving approaches. Common STEM activity models include the 5E inquiry STEM teaching model (Zhao et al., 2018), the 6E learning design model (Barry, 2014), and the 5EX design model (Li and Li, 2019). These models' main components are listed in Table 1; as indicated, engagement, exploration, design and explanation, and engineering evaluation should be considered when designing STEM activities.

Construction of analytical metrics

Analysis items for STEM cases

By referencing the scoring criteria for 2021 outstanding national cases of STEM education in primary and secondary schools, we examined 13 indicator systems (i.e., models/frameworks) related to STEM evaluation (National Research Council, 2009; Change the Equation, 2014; Maryland State Department of Education, 2014; Guzey et al., 2016; Holmlund et al., 2018; Li and Xu, 2018; Miao, 2018; Yin, 2019; Lv, 2020; Qiao, 2020; Lakanukan et al., 2021; Wang et al., 2021). The systems are summarized in Table 2 (items with a frequency of <3 were omitted). Items such as content, evaluation, objectives, implementation, and cooperation appeared frequently and were adopted to construct analytical metrics.

Each case included background information such as the number of students, number of class hours per week, and related subjects (respectively falling under the indicators of "teaching," "student," and "implementation" in Table 2). We combined several items and extracted six for STEM case analysis: (1) background information (covering "teaching," "course," "student," and "implementation"); (2) teaching objectives (covering "objectives"); (3) knowledge content (covering "content"); (4) learning activities (covering "activities" and "cooperation"); (5) teaching evaluations (covering "evaluation"); and (6) supporting materials (covering "resources"). These categories account for the first 10 items in Table 2 and served as primary indicators among our analytical metrics. Specifically, a case's background information reflects teaching objectives, based on which appropriate STEM content can be chosen. Learning activities can then be designed accordingly. Teaching evaluations convey whether teaching objectives have been achieved. Tools, technology, and resources (common supporting materials) facilitate STEM teaching. These six items were confirmed as primary indicators among our metrics.

TABLE 2 Ranking of relevant items for STEM evaluation.

Item	Frequency	Item	Frequency
Content	13	Design	4
Teaching	12	Processes	4
Course	11	Method	4
Student	11	Skills	4
Evaluation	9	Technology	3
Objectives	9	Explore	3
Implementation	8	Theme	3
Activities	7	Strategy	3
Resources	6		
Cooperation	6		

Preliminary metrics and optimization

We created several sub-indicators (also called "secondary indicators") to explore the chosen cases. In accordance with Sections "Theoretical background" and "Key features of STEM," and coupled with relevant studies (Clarke, 2015; Jiang and Cai, 2017; Lin, 2017; National Institute of Education Sciences, 2017; Shaw, 2018; Kang, 2019; Office of Educational Technology, 2019; Durovic, 2020; Falloon et al., 2020; Geesa et al., 2021; Lakanukan et al., 2021; Sirajudin and Suratno, 2021), we decomposed the six primary indicators into secondary indicators as displayed in Table 3.

We adopted the Delphi method to verify the scientific soundness and feasibility of our analytical metrics. Specifically, we designed a consultation form and invited 10 experts in the field to complete it *via* email. They were asked to give their opinions on indicators' rationality and language. In each round, experts considered every indicator (primary and secondary) separately. If an expert deemed an indicator reasonable, it was assigned a score of 1; otherwise, it was assigned a score of 0 with suggested amendments. Experts returned their feedback *via* email. We subsequently calculated each indicator's mean and standard deviation, integrated experts' recommended modifications, and updated the indicators. We then distributed another questionnaire to gather additional feedback on issues up for debate. The metrics were optimized during this repetitive process. In brief, experts' opinions revolved around the following points:

1. The wording of indicators: for example, "topic selection" was replaced with "topic types," and the expression of primary indicators was unified.

TABLE 3 Decomposition of six primary indicators.

Primary Indicators	Secondary Indicators	Description
Background Information	Topic selection	Options include validating, exploring, designing, manufacturing, and creating
	Related subjects	Options: Chinese, math, English, science, physics, chemistry, biology, geography, information technology, music, arts, history
	Class hours	Number of teaching hours per week and total number of teaching hours
	Grade	Students' grade(s)
	Class size	Number of students per class
Teaching Objectives	Interdisciplinary knowledge and skills	Students master basic principles and skills to solve problems in an interdisciplinary manner
	Scientific spirit	Students can think rationally; raise and analyze questions; and solve problems by formulating hypotheses, exploring, and interpreting data to draw conclusions
	Innovative ability	Students can develop innovative solutions or optimize existing solutions using technology
	Cooperative spirit and ability	Students work in teams to communicate, collaborate, and share with others
Knowledge Content	Real-world scenario	Problems relate to real life
	Conforms to curriculum standard	Knowledge content conforms to the curriculum standard and students' cognitive level
	Interdisciplinary integration	The content of each associated subject is well integrated
	Targeted content	Content is well organized and topic-specific
Learning Activities	Real situation introduction	Teacher demonstrates real-world problems for students from the outset
	Scientific exploration	Teacher encourages students to think rationally, pose questions, and solve them by formulating hypotheses and presenting/evaluating evidence to engage in scientific argumentation
	Handcrafting with technology	Teacher encourages students to choose appropriate technology/tools/materials to complete their work in a hands-on way
	Engineering design	Teacher helps students define engineering tasks and encourages them to complete tasks like an engineer: by drafting, assembling, testing, and optimizing
	Math application	Teacher encourages students to measure, collect, and analyze data to describe the objective world in a mathematical way
	Creative expansion	Teacher encourages students to improve their work creatively according to practical needs
	Evaluation and reflection	Teacher uses multiple evaluation methods to test the learning effect and encourages students to engage in self-reflection
Supporting Materials	Software	Software required for learning (e.g., programming tools, drawing tools)
	Hardware	Hardware that supports students' cooperative work (e.g., hammer, wooden slats, scissors, robots)
	Multimedia resources	Multimedia resources to facilitate teaching and learning (e.g., PowerPoint, micro-video, reading material, animation)
	Manual or instruction	Operation manual or activity instruction that guides students through group tasks or participation in self-regulated learning
	Learning logs	Records of how students conduct their learning process
	Evaluation tools	Tools that help teacher and students complete individual or collaborative assessments
Teaching Evaluations	Diagnostic evaluation	Records of students' existing knowledge and skills gained through pre-test(s)
	Formative evaluation	Records of how students conduct their learning process as evidenced by classroom observations, worksheet assessments, self-reflection reports, and peer-review reports
	Summative evaluation	Tests, quizzes, or other criterion-referenced assessments where a score is assigned based on learner-supplied evidence of having mastered desired knowledge or skills

2. The overlap between certain indicators: for instance, the secondary indicator "evaluation and reflection" overlapped with the primary indicator "teaching evaluations"; the secondary indicators "real-world scenario" and "real situation introduction," which initially appeared under different primary indicators, shared content.
3. The scope of related subjects: experts pointed out that STEM education could cover all subjects in primary school, including comprehensive practice, ethics, and the rule of law.
4. The consideration of the implementation effects of teaching cases: experts suggested including interviews in the analysis to clarify teachers' perceptions of their instruction and what students gained.
5. Missing secondary indicators: for example, "venue support" should be considered in STEM practice—many cases capitalized on local resources (e.g., Dan et al., 2017) to add a humanistic slant or used smart classrooms to offer students opportunities to learn by integrating new technologies.

TABLE 4 Finalized metrics for STEM teaching cases.

Primary indicators	Secondary indicators	Description
Background information	Topic types	Options: validating, exploring, designing, manufacturing, and creating
	Related subjects	Options: Chinese, math, English, science, physics, chemistry, biology, geography, information technology, music, arts, history, ethics and the rule of law, comprehensive practice
	Class hours	Number of teaching hours per week and total number of teaching hours
	Grade of students	Students' grade(s)
	Class size	Number of students per class
Teaching objectives	Interdisciplinary knowledge and skills	Students master basic principles and skills of how to solve problems in an interdisciplinary manner
	Scientific spirit	Students can think rationally, raise and analyze questions, and solve problems by formulating hypotheses, exploring, and interpreting data to summarize knowledge
	Innovative ability	Students can develop innovative solutions or to optimize existing solutions using technology
	Cooperative spirit and ability	Students work in teams to communicate, collaborate, and share with others
Knowledge content	Based on real-world problems	Problems originate from the objective world and are related to real life
	Conforms to curriculum standard	Knowledge content conforms to the curriculum standard and students' cognitive level
	Interdisciplinary integration	The content of each associated subject is well integrated
	Targeted content	Content is well organized and topic-specific
Teaching activities	Scenario startup	Teacher illustrates a scenario at the beginning of the lesson
	Scientific exploration	Teacher encourages students to think rationally, pose questions, and solve them by formulating hypotheses and presenting/evaluating evidence to engage in scientific argumentation
	Handcrafting with technology	Teacher encourages students to choose appropriate technology/tools/materials to complete their work in a hands-on way
	Engineering design	Teacher helps students define engineering tasks and encourages them to complete tasks like an engineer: by drafting, assembling, testing, and optimizing
	Math application	Teacher encourages students to measure, collect, and analyze data to describe the objective world in a mathematical way
	Creative expansion	Teacher encourages students to improve their work creatively according to practical needs
	Presentation and reflection	Teacher encourages students to make presentation to share their work in public and engage in self-reflection.
Teaching support	Software	Software required for learning (e.g., programming tools, drawing tools)
	Hardware	Hardware that supports students' cooperative work (e.g., hammer, wooden slats, scissors, robots)
	Venue support	On-campus and/or off-campus venues that support teaching and learning
	Multimedia resources	Multimedia resources to facilitate teaching and learning (e.g., PowerPoint, micro-video, reading material, animation)
	Manual or instruction	Operation manual or activity instruction that guides students through group tasks or participation in self-regulated learning
	Learning logs	Records of how students conduct their learning process
	Evaluation tools	Tools that help teacher and students complete individual or collaborative assessments
Teaching evaluations	Diagnostic evaluation	Records of students' existing knowledge and skills gained through pre-test(s)
	Formative evaluation	Records of how students conduct their learning process as evidenced by classroom observations, worksheet assessments, self-reflection reports, and peer-review reports
	Summative evaluation	Tests, quizzes, or other criterion-referenced assessments where a score is assigned based on learner-supplied evidence of having mastered desired knowledge or skills

Finalized metrics

After two rounds of iteration, the indicators were approved by all experts. We then finalized the analytical metrics (six primary indicators and 30 secondary indicators as presented in Table 4). Weights were not assigned to each indicator; our intention was not to evaluate the advantages and drawbacks of teaching cases but instead to synthetize their characteristics.

Analysis results

Regional distribution

Figure 1 indicates that the 51 STEM teaching cases spanned 13 regions including Hong Kong, Macau, Shanghai, Beijing, Guangzhou, Zhuhai, and others. Most cases were based in

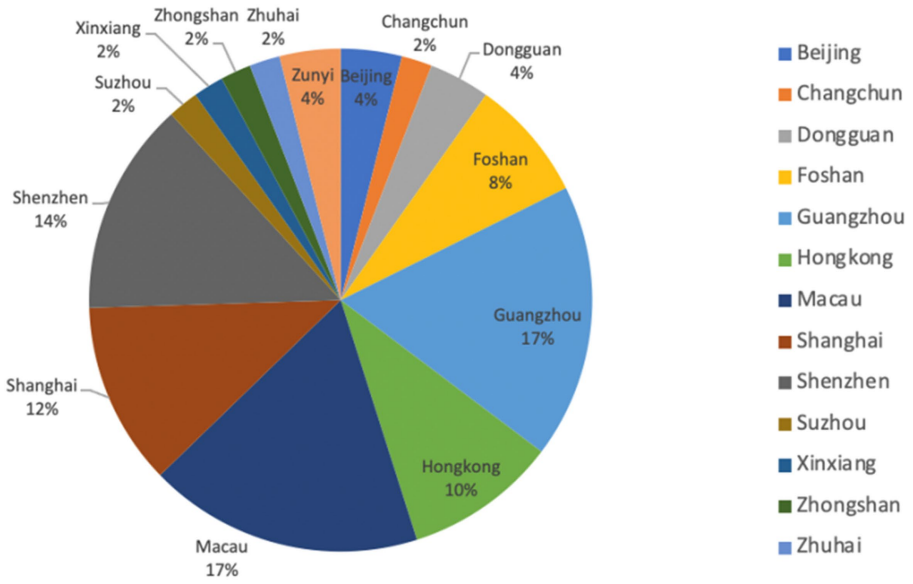


FIGURE 1
Regional distribution of sample cases.

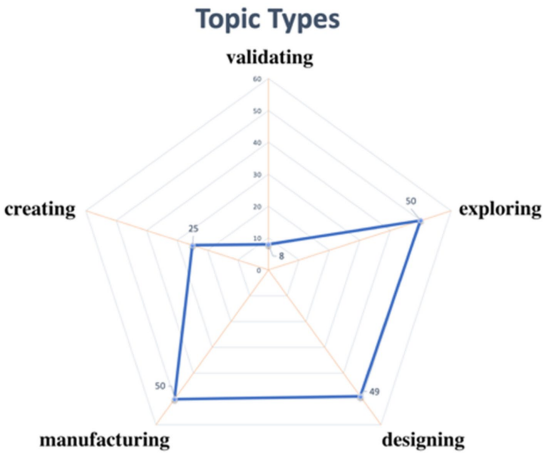


FIGURE 2
Topic types from sample cases.

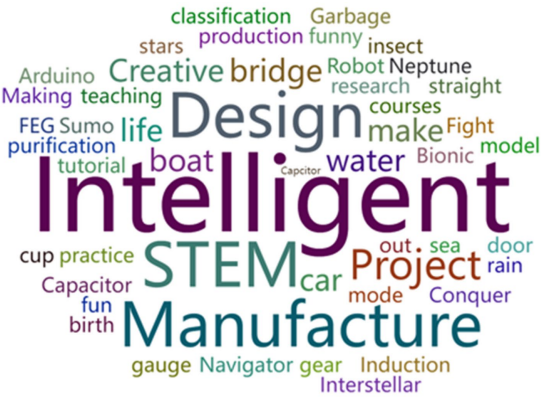


FIGURE 3
Topic keywords from sample cases.

Guangzhou (17%), Macau (17%), Shenzhen (14%), and Shanghai (12%), which is proportional to these regions' economic power.

Background information

Topic types and topic keywords

Figure 2 indicates "exploring," "manufacturing," and "designing" as main topic types. The word cloud diagram in Figure 3 shows that "intelligent," "design," and "manufacture" were frequently used in naming cases, which mirrors the topic-based keywords. For instance, "Design and Manufacture of Intelligent Nursery," "Design and Manufacture of Intelligent Fire Alarm," and "Design of Intelligent Catapult" were popular STEM case names; all originated from real life and have practical significance.

Related subjects and class hours

Interdisciplinarity is the cornerstone of STEM. Of the sample cases, each was related to an average of 4.3 subjects, with "math," "science," and "physics" accounting for the highest proportion. The most common combination was "math," "science," "physics," and "information technology"; 21 (out of 51) cases fell into this category. In terms of class hours, the cases were generally carried out over one semester, with an average of 15.2 h per case (for a rough duration of 11 weeks when holding two classes per week).

10.3389/fpsyg.2022.1010033

TABLE 5 Teaching objectives in sample cases.

Teaching objectives	*M*	*SD*
Interdisciplinary knowledge and skills	3.90	0.76
Scientific spirit	4.20	0.66
Innovative practical ability	4.29	0.76
Cooperative spirit and ability	4.33	0.71

TABLE 6 Knowledge content in sample cases.

Knowledge content	*M*	*SD*
Based on real-world problems	4.16	0.73
Conforms to curriculum standard	4.39	0.57
Interdisciplinary integration	4.06	0.61
Targeted content	4.45	0.61

Teaching objects

The statistical results demonstrated that these STEM cases were mainly designed and implemented for students in Grades 4–7, with 31 students per class on average.

Teaching objectives

According to Table 5, the overall objectives of these STEM teaching cases were appropriate, but imbalances persisted: high scores accompanied "innovative practical ability" and "cooperative spirit and ability," whereas the other two indicators earned relatively low scores. As such, teachers paid more attention to students' innovative and cooperative abilities than to interdisciplinarity and a scientific spirit. Realization of the interdisciplinary objective was not as noteworthy as anticipated despite each case relating to 4.3 subjects on average.

Knowledge content

Table 6 shows that most cases highlighted their intended topics, and relevant knowledge content conformed to students' curriculum standards.

Teaching activities

Based on Table 7, the design and implementation of teaching activities were unbalanced. Scores on "handcrafting with technology" and "presentation and reflection" were 4.33 and 4.35, respectively—each relatively high. "Math application" was scored the lowest, indicating that it was not fully applied in some cases.

TABLE 7 Teaching activities in sample cases.

Teaching activities	*M*	*SD*
Scenario startup	4.14	0.69
Scientific exploration	4.18	0.68
Handcrafting with technology	4.33	0.65
Engineering design	4.10	0.76
Math application	3.96	0.82
Creative expansion	4.12	0.68
Presentation and reflection	4.35	0.63

TABLE 8 Teaching support in sample cases.

Teaching support	*M*	*SD*
Software	4.06	0.83
Hardware	4.24	0.62
Venue support	3.86	0.66
Multimedia resources	4.02	0.62
Manual or instruction	3.94	0.68
Learning logs	3.94	0.76
Evaluation tools	3.90	0.78

Teaching support

As for STEM teaching support, "hardware" was scored highest (see Table 8), whereas "venue support," "evaluation tools," "manual or instructions," and "learning logs" were scored low. We observed that 3D modeling and 3D printing were used in several cases, with some teachers even leveraging intelligent technologies such as sensors or programming with Arduino. Students were also given various physical materials, such as acrylic plates, motors, batteries, and scissors. Statistical forms or diagrams were widely used to help students record experimental data and conduct further analysis.

Teaching evaluations

All cases included summative evaluations, and most cases included formative evaluations. Only two cases included diagnostic evaluations. Instructors tended to use teacher evaluations and student self-evaluations, with some also using group evaluations (as an example of multivariate assessment). In most cases, multidimensional evaluation forms were provided to assess students' learning, including their performance on plan proposals, group discussions, handcrafting, and presentations.

Discussion and recommendations

The above analysis demonstrates that China's typical STEM education possesses several common characteristics. First, it

presented an uneven geographical distribution: many excellent cases emerged in economically developed areas (e.g., Beijing, Guangzhou, Shanghai), whereas effective STEM education seemed rare in less developed areas. Such education relies heavily on numerous types of hardware and software support, which greatly increases expenses. Economically developed areas therefore tend to outperform less developed areas in STEM education (Xu et al., 2021). Globally, the countries best known for excelling in STEM are those with small populations and a relatively well-developed economy (English, 2019). Our finding is consistent with this trend.

Second, current STEM practices in China benchmark the need for innovative talent training as advocated in China's Primary and Secondary School Curriculum Standard (2022). Our cases related to more than four subjects per case on average. Secondary indicators such as "interdisciplinary knowledge and skills," "scientific spirit," "innovative practical ability," "cooperative spirit and ability," and "interdisciplinary integration" contribute to the goal of innovative talent cultivation. We noted high scores on these indicators, revealing that they served the aim of innovative talent training. Many schools in our sample offered STEM courses for students in Grades 4–7, and classes contained 31 students on average. These outcomes aligned with those identified by Batdi et al. (2019) and echoed circumstances in the United States (Committee on Highly Successful Schools or Programs in K-12 STEM Education, N. R. C., 2011).

Third, most instructional content was drawn from real-world problems, but interdisciplinary integration deserves more focus. Most cases in our sample showed adequate content selection. Course material usually stemmed from actual problems, enhancing the student experience and exposing students to realistic scenarios. The YouthInsight survey report (YouthInsight for the Department of Industry, S., Energy and Resources of Australian Government, 2021) came to similar conclusions: 72% of teachers reported feeling very confident in connecting STEM content with real-world applications. Yet in our cases, each discipline became fairly self-contained and independent over time, hampering integration. This consequence corroborates that of a prior study (Chen K. et al., 2021). Previous work (Banilower et al., 2018) documented that 35% of middle schools offered single-discipline science courses for students in Grade 6 while 80% offered single-subject mathematics courses. As an interdisciplinary teaching method, STEM mandates teaching across disciplines using strategies that support knowledge integration *via* authentic tasks. However, when instructors who are accustomed to single-discipline content teach STEM courses cooperatively, they tend not to adhere to top-level interdisciplinary design. Interdisciplinary integration was clearly inadequate in these STEM cases.

Fourth, these cases featured rich teaching activities and were conducted in a PjBL fashion with insufficient emphasis on mathematical applications. PjBL enables students to pursue solutions to problems in the same way that professional scientists do (Office of Educational Technology, 2019). The process normally includes brainstorming, planning, discussing, measuring, assembling, experience sharing, and assessing. As an innovative talent cultivation mode, STEM is typically combined with PjBL, a method with demonstrated efficacy (Lou et al., 2017; Kartini et al., 2021; Lakanukan et al., 2021). As shown in Table 7, the teaching activities in these cases were detailed enough to cover those associated with PjBL. These activities are distinct from traditional instructional methods and are expected to cultivate innovative talent. Among the seven teaching activities in this study, mathematical applications received insufficient consideration: in many cases, teachers preferred to simply inform students of the results instead of inviting students to measure, calculate, or make comparisons on their own. The relative lack of mathematical applications in these activities is contradictory to the premise of STEM education (i.e., in ignoring "M").

Fifth, China seems to be holistically promoting STEM education, especially through new technologies (e.g., hardware and software support) and supplementary materials such as instruction manuals. Different from Wan et al. (2021), where deficient resources were frequently mentioned, the cases in this study featured a wide range of support when implementing STEM courses (e.g., software, hardware, and multimedia resources). These circumstances coincide with a YouthInsight survey (YouthInsight for the Department of Industry, S., Energy and Resources of Australian Government, 2021) highlighting the websites Teachers Pay Teachers, Scootle, and Khan Academy as the most popular online resources for teachers. The Information Resources Management Association of the United States also stressed the roles of digital resources in promoting STEM literacy (Management Association, I. R, 2018). In addition to hardware and software empowered by new technologies (e.g., robots, 3D printing, sensors), supplementary materials such as learning logs can aid students' collaborative interdisciplinary learning. Worksheets can help bolster students' critical thinking skills (Hartini et al., 2020). In a related vein, the newest National Survey of Science & Mathematics Education report from the United States (Banilower et al., 2018) outlined various instructional resources for STEM courses, including supplementary materials (e.g., laboratory handouts). Students can accordingly plan, revise, implement, and test solutions to problems *via* engineering design processes and appropriate support technologies. Assessments are embedded in these courses to solicit students' reflections on the quality of their explanations, models, or problem solutions, calling for a strong record of their learning processes.

Several recommendations arose from our analysis of STEM teaching cases in primary and secondary schools in China. First, the country should harness the development of STEM education in multiple regions. Underdeveloped regions will otherwise struggle to develop first-class STEM education on a large scale. Quality STEM education requires vast investment: funding, venues with high-tech equipment, and highly qualified teachers. These features exacerbate the financial burden for schools—especially those in second- and third-tier cities. To leverage the development of STEM education in different regions while promoting educational balance, China should seek to support

STEM practices in these cities. The government should allocate educational resources to second- and third-tier cities as needed and rearrange the distribution of education to give full play to the impact of STEM education. For example, the government should consider increasing investment and promulgating a post-service training plan for STEM teachers in primary and secondary schools in western China (Chen T. et al., 2021). Teaching reform projects and teaching achievement evaluations related to STEM education could also be created (Xu et al., 2021). Stakeholders could additionally refer to farming culture to realize the localization reform and innovation of rural STEM education (Yang, 2020).

Second, primary and secondary schools should continue teaching STEM based on curriculum standards (especially the version released in 2022) and gradually expand to cover all students in the same grade instead of only one or two classes. As reflected in the STEM2026 report (The U.S. Department of Education, A. I. f. R, 2016) and the white paper on STEM education in China (National Institute of Education Sciences, 2017), STEM education entails lifelong learning. This process can stimulate students' enthusiasm for scientific exploration and innovation. STEM education offers a way to improve all students' core literacy instead of selectively nurturing exceptional talent. Limited classes were chosen for pilot studies in most cases, contrary to the above goal. We recommend that primary and secondary schools proceed with STEM instruction based on the most recent curriculum standards and extend the breadth of teaching to cover more students in a given grade. As indicated by our findings (especially Figures 2, 3), nearly all walks of life and most disciplines are linked to STEM. Teachers from numerous disciplines are hence encouraged to join STEM pilot programs (according to Section "Related subjects and class hours," a STEM practice can be related to four or five subjects). Students from the same grade should be welcome to take part in STEM learning based on diverse topics. For instance, students in Grades 1–4 could concentrate on "smart cars," those in Grades 4–8 could learn about "smart alerts," and all others could focus on "sea cleaners." STEM education would therefore be broadened to involve students beyond Grades 4–7 (Section "Teaching objects").

Third, schools should enhance interdisciplinarity to achieve integrated STEM. Interdisciplinarity in education encourages learners to make connections between disciplines (i.e., combining knowledge and skills from two or more subjects) when solving complicated problems or explaining complex phenomena. STEM teachers thus need to master interdisciplinary knowledge; communicate and cooperate in top-level design; and integrate multidisciplinary content in instructional design, course implementation, and self-reflection. In light of the current discipline-based curriculum and the fact that many teachers are unfamiliar with creating instructional materials for integrated STEM, strategies for instructor collaboration across subjects should be carefully planned. Engaging teachers in professional development for curriculum design is critical to developing integrated STEM. This course of action can improve integrated STEM education and rectify insufficient interdisciplinarity. Several provinces have started to explore this prospect, such as by establishing teacher alliances (e.g., Guangdong–Hong Kong–Macao Greater Bay Area STEM Education Alliance) or conducting network-based teaching and research (Hu et al., 2021). Online and offline teacher training has emerged as other strategies. These efforts can routinely bring teachers together to improve interdisciplinarity while presenting opportunities for further collaboration.

Fourth, schools should continue strengthening mathematical applications in learning activities to endow students with a rigorous academic attitude. Most sample cases involved mathematics, albeit to a lesser extent than expected. Simple measurement and data recording dominated math activities in low-scoring cases. In reality, mathematical knowledge encompasses measuring, marking (tagging), calculating, matching, grading, and comparing. These activities should be fully utilized in STEM learning (Küçük-Demir and Düzen, 2022) to better engage students in data analysis. Doing so can also popularize mathematical tools and compel students to seek knowledge more rigorously. Their core literacy will likely be reinforced as a result.

Fifth, schools should focus on archiving documentation (i.e., quantitative and qualitative analyses) to support formative evaluation. This type of evaluation possesses a unique advantage in tracking students' STEM performance (vs. ranking students by level). The availability of evidence is also important. Teachers should retain procedural data as students learn, including answer sheets, design drafts, statistical data, and classroom observation forms. Without this documentation, quantitative or qualitative analysis of teaching effectiveness cannot be objectively obtained. Formative evaluation may be weakened as a result. For instance, Okulu and Oguz-Unver (2021) developed an evaluation form to determine activities' appropriateness with respect to the nature of STEM education, which has four categories (STEM learning environment, activation of students, STEM content and practices, and connecting STEM). The form was found to be useful for evaluating and improving STEM education based on pedagogical approaches such as PjBL and collaborative learning. Moreover, Fitzallen and Watson (2019) stated that statistics provide a firm foundation for bridging STEM disciplines. This perspective recognizes that building on data gathered through classroom activities in STEM settings can potentially support students in honing their statistical literacy. This skill will be advantageous in other social contexts where people encounter such data. These efforts can enhance students' comprehensive qualities and core competencies through formative assessment.

Conclusion

Researchers generally agree about the significance of STEM in K–12 education. Driven by national policies, K–12 STEM education in China has ushered in hundreds of practical cases. Scholars have extensively addressed the effectiveness of applying STEM in K–12 settings. Much less interest has surrounded these practices' curricular similarities. Suggestions for implementing STEM in

primary and secondary schools based on these characteristics are scarce as well. The current study is hoped to enable instructors to better understand STEM implementation, to prompt successful STEM cases, and to promote the localization of STEM education. Common characteristics were discerned using a curriculum perspective, with affiliated suggestions for implementing STEM based on these attributes. We first established analytical metrics (including six primary indicators and 30 secondary indicators) to analyze K–12 STEM cases. Fifty-one typical teaching cases throughout China were next examined based on these metrics. Five characteristics were extracted from the statistical results, complemented by recommendations for promoting STEM education in the country. Findings shed light on STEM implementation, its features, and areas for refinement. These enhancements will help nurture 21st-century talent with scientific inquisitiveness and innovative skills. Future research could apply our suggestions to STEM practices to assess their efficacy and refine the techniques as needed. Including additional cases from throughout China could also unearth more meaningful findings.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by the Academic Committee Office (ACO) of South China Normal University (http://fzghb.scnu.edu.cn/), Guangzhou. Written informed consent from the participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

Author contributions

YZ: conceptualization, methodology, writing—original draft preparation, writing—review and editing, supervision. PL: writing—original draft preparation and writing—review and editing. XY, YG, XQ, and XJ: software, data curation, and formal analysis. XL: writing—original draft preparation. TZ: conceptualization, methodology, writing—review and editing, supervision. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by the Key Project of China Educational Technology Association's 14th Five-Year Plan under grant no. G021 and Jinan University Shenzhen Campus Funding Program under grant no. JNSZQH2105.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Banilower, E. R., Smith, P. S., Malzahn, K. A., Plumley, C. L., Gordon, E. M., and Hayes, M. L. (2018). Report of the 2018 NSSME+. Available at: http://horizon-research.com/NSSME/wp-content/uploads/2020/04/Report_of_the_2018_NSSME.pdf

Barry, N. (2014). The ITEEA 6E learning by DeSIGN model. *Technol. Eng. Teach.* 73, 14–19.

Batdi, V., Talan, T., and Semerci, C. (2019). Meta-analytic and meta-thematic analysis of STEM education. *Int. J. Educ. Math. Sci. Technol.* 7, 382–399.

Bolatli, Z., and Korucu, A. T. (2018). Secondary school students' feedback on course processing and collaborative learning with web 2.0 tools-supported STEM activities. *Bartın Univ. J. Fac. Educ.* 7, 456–478. doi: 10.14686/buefad.358488

Bryan, L., and Guzey, S. S. (2020). K-12 STEM education: an overview of perspectives and considerations. *Hellenic J. STEM Educ.* 1, 5–15. doi: 10.51724/hjstemed.v1i1.5

Carter, V. R. (2013). *Defining Characteristics of an Integrated Stem Curriculum in K–12 Education.* University of Arkansas, Fayetteville, Arkansas, United States.

Change the Equation (2014). Design principles for effective STEM philanthropy. Available at: https://stemworks.wested.org/sites/default/files/STEMworks_Design_Principles.pdf

Chen, T., Hu, X., and Gao, X. (2021). The dilemma and countermeasures of STEAM education in elementary and secondary schools in west part of China. *Teach. Educ. Res.* 33, 59–64. doi: 10.13445/j.cnki.t.e.r.2021.02.009

Chen, K., Yang, Y., and Guo, J. (2021). Content analysis of STEM instructional design from the perspective of middle school science subjects. *Chem. Teach.*, 16–23. doi: 10.3969/j.issn.1005-6629.2021.01.003

Clarke, D. J. (2015). "Comparative research in mathematics education: Boundary crossing and boundary creation," in *Paper Presented at the the 39th Conference of the International Group for the Psychology of Mathematics Education*, Hobart, Australia: PME.

Committee on Highly Successful Schools or Programs in K-12 STEM Education, N. R. C (2011). *Successful K-12 STEM Education: Identifying Effective Approaches in Science, Technology, Engineering, and Mathematics.* DC, Washington: National Academies Press,

Dan, Y., Zhang, X., and Wang, L. (2017). The teaching case design of STEM based on the environment of combining museum and school: water resources project. *Int. J. Eng. Educ.* 33, 1117–1123.

Durovic, G. (2020). "Continuous summative assessment sessions as a motivational tool for STEM students: a case study," in *Paper Presented at the 2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO).*

English, L. D. (2017). Advancing elementary and middle school STEM education. *Int. J. Sci. Math. Educ.* 15, 5–24. doi: 10.1007/s10763-017-9802-x

English, T. (2019). Lessons in STEM success from around the world. Available at: https://www.base11.com/lessons-stem-success/

Falloon, G., Hatzigianni, M., Bower, M., Forbes, A., and Stevenson, M. (2020). Understanding K-12 STEM education: a framework for developing STEM literacy. *J. Sci. Educ. Technol.* 29, 369–385. doi: 10.1007/s10956-020-09823-x

Falloon, G., Stevenson, M., Beswick, K., Fraser, S., and Geiger, V. (2021). Building STEM in schools. *Educ. Technol. Soc.* 24, 110–122.

Fitzallen, N., and Watson, J. (2019). "Using the practice of statistics to design students' experiences in STEM education," in *Harnessing the Transformative Power of Education*, eds Shelley, B., te Riele, K., Brown, N., Crellin, T., (Leiden, The Netherlands: Brill), 74–99.

Freeman, B., Marginson, S., and Tytler, R. (2019). "An international view of STEM education," in *STEM Education 2.0* (Leiden: Brill), 350–363.

Gardner, M., and Tillotson, J. W. (2019). Interpreting integrated STEM: sustaining pedagogical innovation within a public middle school context. *Int. J. Sci. Math. Educ.* 17, 1283–1300. doi: 10.1007/s10763-018-9927-6

Geesa, R. L., Rose, M. A., Stith, K. M., Lowery, K., and Caniglia, J. (2021). "Equity, diversity, and inclusion within integrative STEM education," in *Leadership in Integrative STEM Education: Collaborative Strategies for Facilitating an Experiential and Student-Centered Culture*, 51.

Green, R. A. (2014). The Delphi technique in educational research. *SAGE Open* 4, 215824401452977–215824401452978. doi: 10.1177/2158244014529773

Guzey, S. S., Moore, T. J., and Harwell, M. (2016). Building up STEM: an analysis of teacher-developed engineering design-based STEM integration curricular materials. *J. Pre-College Eng. Educ. Res. (J-PEER)* 6:2. doi: 10.7771/2157-9288.1129

Hansen, M. (2014). Characteristics of schools successful in STEM: evidence from two states' longitudinal data. *J. Educ. Res.* 107, 374–391. doi: 10.1080/00220671.2013.823364

Hartini, S., Mariani, I., and Sulaeman, N. (2020). Developing of students worksheets through STEM approach to train critical thinking skills. *J. Phys. Conf. Ser.* 1567:042029. doi: 10.1088/1742-6596/1567/4/042029

Holmlund, T. D., Lesseig, K., and Slavit, D. (2018). Making sense of "STEM education," in K-12 contexts. *Int. J. STEM Educ.* 5, 32–18. doi: 10.1186/s40594-018-0127-2

Honey, M., Pearson, G., and Schweingruber, H. (2014). *STEM Integration in K-12 Education: Status, Prospects, and an Agenda for Research,* DC, Washington; National Academies Press.

Hu, Y., Peng, X., and Wu, B. (2021). Research on the online training mode of STEM teachers' instructional design ability. *Chin. J. Distance Educ.* 8, 31–39. doi: 10.13541/j.cnki.chinade.2021.08.004

Jiang, F., and Cai, R. (2017). Content analysis on the assessment design of STEM education at home and abroad. *China Audio-Visual Educ.* 6, 59–66. doi: 10.3969/j.issn.1006-9860.2017.06.009

Kang, N.-H. (2019). A review of the effect of integrated STEM or STEAM (science, technology, engineering, arts, and mathematics) education in South Korea. *Asia-Pacific Sci. Educ.* 5, 1–22. doi: 10.1186/s41029-019-0034-y

Kartini, F., Widodo, A., and Winarno, N. (2021). STEM project-based learning on student's STEM literacy: the case of teaching earth layer and disaster. *J. Phys. Conf. Ser.* 1806:012221. doi: 10.1088/1742-6596/1806/1/012221

Khotimah, R. P., Adnan, M., Ahmad, C. N. C., and Murtiyasa, B. (2021). Science, mathematics, engineering, and mathematics (STEM) education in Indonesia: a literature review. *J. Phys. Conf. Ser.* 1776:012028. doi: 10.1088/1742-6596/1776/1/012028

Kim, A. Y., Sinatra, G. M., and Seyranian, V. (2018). Developing a STEM identity among young women: a social identity perspective. *Rev. Educ. Res.* 88, 589–625. doi: 10.3102/0034654318779957

Küçük-Demir, B., and Düzen, Ü. (2022). Investigation of the effects of mathematics-centered stem activities on students' creative thinking skills and student opinions. *J. Gifted Educ. Creativity* 9, 25–42.

Lakanukan, S., Thanyaphongphat, J., and Areeprayolkij, W. (2021). "The design process of STEM learning activities for problem-solving on the PM 2.5 mask: the case of primary School in Thailand," in *Paper Presented at the 29th International Conference on Computers in Education*.

Li, K., and Li, Y. (2019). 5EX design model for interdisciplinary learning activity in STEM education. *e-Educ. Res.* 40, 5–13. doi: 10.13811/j.cnki.eer.2019.04.001

Li, W., and Xu, X. (2018). Cultivation of innovative ability in the integration of art and STEM practice: inspiration from the eight STEAM Education cases in the US. *Primary Secondary Schooling Abroad*, 9–17. doi: 10.3969/j.issn.1007-8495.2018.12.002

Lin, C. (2017). The research of core competencies and values for students in China. *Stud. Psychol. Behav.* 15, 145–154.

Lou, S.-J., Liang, C.-P., and Chung, C.-C. (2017). Effectiveness of combining STEM activities and PBL: a case study of the design of fuel-efficient vehicles. *Int. J. Eng. Educ.* 33, 1763–1775.

Łukaszczyk, Z., and Grebski, M. (2020). Comparative analysis of the curriculum at science and technology (STEM) high schools in Poland and the United States. *Organizacja i Zarządzanie: kwartalnik naukowy.* 2, 75–85. doi: 10.29119/1899-6116.2020.50.6

Lv, W. (2020). *Case study of STEM curriculum in American elementary schools: Taking Lego robot teaching in grades 4–5 as an example. (Master)*. Yunnan Yunnan Normal University.

Lynch, S. J., Burton, E. P., Behrend, T., House, A., Ford, M., Spillane, N., et al. (2018). Understanding inclusive STEM high schools as opportunity structures for underrepresented students: critical components. *J. Res. Sci. Teach.* 55, 712–748. doi: 10.1002/tea.21437

Management Association, I. R (Ed.) (2018). *K-12 STEM Education: Breakthroughs in Research and Practice (2 volumes)*. Hershey, PA: IGI Global.

Maryland State Department of Education (2014). Addressing the standards… what does it look like in practice? Available at: https://news.maryland.gov/msde/wp-content/uploads/sites/12/2015/01/PartnershipE-LitNGSSCCSSSTEMFinal.pdf

Means, B., Wang, H., Wei, X., Lynch, S., Peters, V., Young, V., et al. (2017). Expanding STEM opportunities through inclusive STEM-focused high schools. *Sci. Educ.* 101, 681–715. doi: 10.1002/sce.21281

Miao, M. (2018). *A Case Study of American STEM Curriculum and its Enlightenment. (Master)*. Jiangsu Jiangsu Normal University.

Ministry of Education of the People's Republic of China (2022). Notice of the Ministry of Education on Issuing the Curriculum Plan and Curriculum Standards for Compulsory Education (2022). Available at: http://www.moe.gov.cn/srcsite/A26/s8001/202204/t20220420_619921.html

National Institute of Education Sciences (2017). White paper on STEM education in China. Available at: https://www.doc88.com/p-86416995822344.html

National Research Council (2009). *Engineering in K-12 Education: Understanding the Status and Improving the Prospects,* DC, Washington: National Academies Press.

Nite, S. B., Capraro, M. M., Capraro, R. M., and Bicer, A. (2017). Explicating the characteristics of STEM teaching and learning: a metasynthesis. *J. STEM Teach. Educ.* 52:6. doi: 10.30707/JSTE52.1Nite

Office of Educational Technology (2019). Innovation spotlights: nine dimensions for supporting powerful STEM learning with technology. Available at: https://tech.ed.gov/files/2019/10/stem-innovation-spotlights-research-synthesis.pdf

Okulu, H. Z., and Oguz-Unver, A. (2021). The development and evaluation of a tool to determine the characteristics of STEM activities. *Eur. J. STEM Educ.* 6:6. doi: 10.20897/ejsteme/10894

Qiao, D. (2020). *Analysis Framework Construction and Case Studies of STEM Curriculum From the Perspective of Education for Sus-tainable Development. (Master)*. East China Normal University, Shanghai.

Rowe, G., and Wright, G. (1999). The Delphi technique as a forecasting tool: issues and analysis. *Int. J. Forecast.* 15, 353–375. doi: 10.1016/S0169-2070(99)00018-7

Scott, C. E. (2009). *A Comparative Case Study of the Characteristics of Science, Technology, Engineering, and Mathematics (STEM) Focused High Schools,* George Mason University, Fairfax, Virginia.

Shaw, R. L. (2018). *Using Project-Based Learning to Cultivate 21st Century Skills in STEM Education. (PhD)*. Lamar University, Beaumont, TX.

Sirajudin, N., and Suratno, J. (2021). Developing creativity through STEM education. *J. Phys. Conf. Ser.* 1806:012211. doi: 10.1088/1742-6596/1806/1/012211

Stemworks at Wested (2017). Design Principles Rubric. Available at: https://www.sdstemecosystem.org/sites/default/files/2019-11/STEMworks_Design_Principles_Rubric_0.pdf

The U.S. Department of Education, A. I. f. R (2016). STEM 2026: a vision for innovation in STEM education. Available at: https://oese.ed.gov/files/2016/09/AIR-STEM2026_Report_2016.pdf

Wan, Z. H., Jiang, Y., and Zhan, Y. (2021). STEM education in early childhood: a review of empirical studies. *Early Educ. Dev.* 32, 940–962. doi: 10.1080/10409289.2020.1814986

Wang, B. (2022). The development history, characteristics and prospect of STEM education in China. *Comput. Knowl. Technol.* 18, 159–163. doi: 10.14004/j.cnki.ckt.2022.0522

Wang, J., Shan, J., and Zheng, Y. (2021). Comparative study on STEM curriculum cases between China and the United States: knowledge field, activity design, type and evaluation. *Mod. Distance Educ.* 3, 12–20. doi: 10.13927/j.cnki.yuan.20210512.006

Wang, N., Tang, Q., and Zhang, R. (2019). Research on the evaluation mechanism of American STEM education projects and its enlightenment. *Mod. Educ. Technol.* 29, 108–114. doi: 10.3969/j.issn.1009-8097.2019.09.016

Xu, Y., Wang, X., Wang, X., and Huang, L. (2021). Investigation and research on the status quo of STEM education in the western region. *Sci. Consulting* 11, 180–182.

Yang, H. (2020). Examining the potential of rural education in China — exploration on the localization of STEAM education based on farming culture. *Front. Educ. Res.* 3, 74–80. doi: 10.25236/FER.2020.031017

Yin, Y. (2019). *Research on STEM Education Curriculum in American Primary and Secondary Schools*. Shanghai Normal University, Shanghai.

YouthInsight for the Department of Industry, S., Energy and Resources of Australian Government (2021). 2021-20 STEM influencer report–teachers & career advisors. Available at: https://www.industry.gov.au/sites/default/files/2021-04/stem-influencer-report-teachers-and-career-advisors-2021-20.pdf

Yu, S., and Hu, X. (2015). STEM education and its model for interdisciplinary integration. *Open Educ. Res.* 21, 13–22. doi: 10.13966/j.cnki.kfjyyj.2015.04.002

Zhao, C., Zhao, W., and Jiang, Z. (2018). The design of 5E inquiry teaching model for STEM education. *Mod. Educ. Technol.* 28, 106–112. doi: 10.3969/j.issn.1009-8097.2018.03.016

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 08 November 2022
DOI 10.3389/fpsyg.2022.996338

Check for updates

OPEN ACCESS

EDITED BY
Dimitrios Stamovlasis,
Aristotle University of Thessaloniki, Greece

REVIEWED BY
Şermin Metin,
Hasan Kalyoncu University,
Turkey
Lyubka Aleksieva,
Sofia University,
Bulgaria

*CORRESPONDENCE
Ali İbrahim Can Gözüm
a_ibrahimcan@hotmail.com

SPECIALTY SECTION
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

RECEIVED 17 July 2022
ACCEPTED 18 October 2022
PUBLISHED 08 November 2022

CITATION
Gözüm AİC, Papadakis S and Kalogiannakis M (2022) Preschool teachers' STEM pedagogical content knowledge: A comparative study of teachers in Greece and Turkey.
Front. Psychol. 13:996338.
doi: 10.3389/fpsyg.2022.996338

COPYRIGHT
© 2022 Gözüm, Papadakis and Kalogiannakis. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Preschool teachers' STEM pedagogical content knowledge: A comparative study of teachers in Greece and Turkey

Ali İbrahim Can Gözüm[1]*, Stamatios Papadakis[2] and Michail Kalogiannakis[2]

[1]Dede Korkut Faculty of Education, Kafkas University, Kars, Turkey, [2]Faculty of Education, The University of Crete, Crete, Greece

This study compares the STEM Pedagogical Content Knowledge of Greek and Turkish preschool teachers. The present research is a comparative descriptive study that aims to determine the STEM Pedagogical Content Knowledge of preschool teachers from Greece and Turkey. A descriptive survey model, a method used in quantitative research, was used as this study's primary research method. The STEM Pedagogical Content Knowledge Scale (STEMPCK) was used in this study. Six hundred sixty-nine preschool teachers - 104 Greek and 565 Turkish teachers - participated in this study. The STEMPCK Scale's construct validity and reliability were tested using this study's data set, which was found to be both valid and reliable. No significant difference was found between the STEMPCK scores of Greek and Turkish preschool teachers. The significant differentiation of STEMPCK scores based on whether the teachers had received any STEM training is discussed in light of the relevant literature. This study determines and compares STEMPCK among preschool teachers from disparate countries such as Greece and Turkey and is expected to contribute to the literature.

KEYWORDS

STEMPCK, preschool teachers, Greece, Turkey, STEM

Introduction

The skills of the age of individuals should have been known as 21st-century skills. Science, technology, engineering, and mathematics (STEM) education stands out as an essential practice in developing 21st-century skills (Marsono et al., 2019). By combining science, engineering, technology, and mathematics disciplines, STEM education enables individuals to identify real-life problems, develop alternative and practical solutions, and offer creative, original solutions which are crucial for the 21st century (Cooper and Heaverlo, 2013; Moore et al., 2014).

STEM education is included in the educational practices of many countries from early childhood through to higher education. STEM education practices are critical in early

childhood (Moomaw and Davis, 2010). According to Eshach and Fried (2005), implementing STEM activities in early childhood significantly affects children's attitudes toward science. Children who did not practice STEM activities in their early years lose interest in science later (Allen, 2016) and may not choose STEM-related fields in their academic careers (Brophy et al., 2008). The concept of providing STEM education to children in the fields of science, technology, engineering, and mathematics (Sullivan and Bers, 2016) supports children's cognitive, psychomotor, social, and affective development (Torres-Crospe et al., 2014; Clements and Sarama, 2016) by assisting their reasoning skills and thinking (Gonzalez and Freyer, 2014; Mercan and Kandır, 2022).

Children discuss the present world's critical problems in their early years as part of STEM education. For example, children's scientific thinking and problem-solving skills are supported by examining health problems such as energy, environmental pollution, and epidemics (Bybee, 2010). What matters here is to choose a qualified STEM program (MacFarlane, 2016) and for the teacher to apply that program to have the pedagogic knowledge and proficiency required for STEM (Margot and Kettler, 2019). Park et al. (2017) said in their study that preschool teachers believe the STEM approach should be adopted in the early years but emphasized that teachers cannot implement STEM activities in educational settings and should be supported. Teachers should be taught how to use effective teaching methods to implement STEM activities, and their STEM knowledge and skills should be backed by pedagogic field experience (Lichtenberger and George-Jackson, 2013). Related studies (Reimers et al., 2015; McDonald, 2016; Mercan and Kandır, 2022) revealed that supporting both the content of STEM disciplines and the development of teachers' pedagogical knowledge positively affected in-class STEM practices and supported the children's development. Preschool teachers' content and pedagogical knowledge for STEM disciplines must be supported (Kennedy et al., 2008).

The pedagogical knowledge of preschool teachers emerges in their qualified and effective use of teaching practices while applying STEM activities to children (Hudson et al., 2015). Content knowledge of teachers for STEM activities forms the basis of scientific understanding to be gained by children in STEM activities. However, the critical point to be considered in content knowledge and pedagogy knowledge for STEM education is teacher training. Including STEM content and pedagogical knowledge in the programs where teachers receive pre-service training is a critical element of STEM education (Ball et al., 2008). In the context of teacher education programs, content and pedagogical knowledge adequacy related to STEM are discussed (Kaya and Elster, 2019). Thus, in-service training programs are implemented to support teachers' STEM content knowledge and pedagogical knowledge. Many countries have established policies to promote STEM education and teachers' professional development in STEM (Johnson, 2012). Countries such as Canada, the United States, and Australia have reported an increased need for teachers trained in STEM so that STEM education can be given effectively and proficiently in classrooms (Stohlmann et al., 2014).

There is no specific standard or content agreed upon in the teacher training policies of countries for STEM education. Therefore, it is thought that countries should focus on the difficulties and problems they experience in the STEM education process to meet their expert teacher needs. Implementing STEM activities pedagogically in the learning environment becomes difficult because teachers do not have sufficient knowledge and experience about STEM components, especially engineering, technology and design (Chai et al., 2019; Faikhamta et al., 2020). According to the results of the research conducted by Ültay and Ültay (2020), the STEM performance of teachers who lack knowledge, skills and pedagogical experiences about STEM activities decreases. Accordingly, it is natural for teachers with low STEM knowledge, skills and pedagogical experience to have low STEM performance in the learning process. However, the point that needs to be pointed out is the evaluation of teachers' in-service training to support their STEM knowledge and skills. As a result of the related research (Yıldırım and Türk, 2018;Karademir-Coşkun et al., 2020; Papadakis et al., 2021a; Nikolopoulou, 2022a), it is seen that teachers feel inadequate in their knowledge, skills and pedagogical experiences in STEM education.

Furthermore, they could not improve themselves in STEM education since the training they receive in pre-service training is insufficient, and they do not receive in-service training (El-Deghaidy and Mansour, 2015; Karademir-Coşkun et al., 2020). Research results show that teachers' pedagogical content knowledge (PCK) for STEM education should be supported (Karademir-Coşkun et al., 2020; Papadakis et al., 2021a; Nikolopoulou, 2022a). Based on these findings, teachers must determine and develop their STEM-related pedagogical competencies through in-service training.

There is a certain number of studies examining the STEMPCK skills of teachers and prospective teachers in standard literature (Wang and Fan, 2018; Chai et al., 2019, 2020; Yıldırım and Şahin-Topalcengiz, 2019; Rahman et al., 2022). However, no studies compare the STEMPCK skills of preschool teachers working in different countries, limited to the researchers' literature review. Nevertheless, it is expected that teachers working in different countries will significantly contribute to the literature by examining their STEMPCK skills because of in-service training for STEM. Making an international comparison of the qualifications of the education received by the teachers in their own countries as a result of STEM education will first reveal the quality of the education received. In addition, comparisons between countries can contribute to preparing STEM teacher training programs. Examining subjects with scientific foundations such as STEM in teacher education in different cultures can reveal countries' scientific understanding and pedagogical approaches. Ultimately, it provides significant findings for future research for the countries being compared. The standard or different features that will be determined from the STEMPCK comparison of the countries' teachers can contribute to the teacher education

paradigm for STEM education. What should the content of STEM education be given to teachers? What pedagogical practices should teachers do for STEM education? For the answers to these questions to be universal, it is necessary to conduct many studies in different countries. However, it is essential to examine the STEMPCK status of teachers in the countries compared and how teachers receive STEM education are affected by STEMPCK. Indeed, it is necessary to determine the existing situation to predict the validity of international exams and to lay the foundation for the basics of the in-service training programs to be prepared for teachers. In this context, The Program for International Student Assessment (PISA) 2018 exam results and the Organization for Economic Co-operation and Development (OECD) countries' average scores can be a criterion for comparing STEMPCK teachers of countries with STEM education. PISA Exam Results for Turkey and Greece has given in Table 1 to see the reading, mathematics, and average science results of both countries:

It can be seen that Greece scored 457 in reading comprehension, 451 in mathematics, and 452 in science, while Turkey scored 466 in reading comprehension, 454 in mathematics, and 468 in science. The results show that Greece and Turkey remained below the average of OECD countries in reading, mathematics, and science (OECD, 2019). The fact that both Greece and Turkey practice STEM education in early childhood and fall below the OECD countries' average may be an essential measure of STEM education.

The 2018 PISA test results of Singapore, Hong Kong, China, Finland, Britain, the United States, Germany, and Austria are higher than the average score of OECD countries, and Greece and Turkey are below the average score of OECD countries (OECD, 2019). While there are significant indicators that science education in early childhood education affects the overall results of PISA scores (especially science and mathematics), it can be stated that STEM education also contributes to the success of PISA in these countries. (Havu-Nuutinen et al., 2022). Although STEM education is practised in Greece and Turkey, the fact that it is below the OECD average reveals that countries should be evaluated in terms of having a teacher training policy supported by STEM training (Papadakis et al., 2018, 2019, 2021a; Dorouka et al., 2020; Gözüm and Kandır, 2021; Yıldırım, 2021; Gözüm, 2022; Mercan and Kandır, 2022). What matters here is the STEM pedagogical content knowledge (STEMPCK) of preschool teachers, for it is they who foster attitudes and interest in children who receive STEM education in their early years (Yıldırım and Şahin-Topalcengiz, 2019). This being the case, it is expected that a comparison of the STEMPCK of Greek and Turkish preschool teachers will contribute to the literature and the countries' teacher training policies concerning STEM education. Therefore, this study aims to compare the STEMPCK scores of Greek and Turkish preschool teachers and examine whether there is a significant difference in the STEMPCK scores of teachers based on the STEM training variable.

TABLE 1 2018 PISA exam results for Turkey and Greece*.

Country	Reading average	Mathematics average	Science average
Turkey	466	454	468
Greece	457	451	452
OECD countries	487	489	489

Organisation for Economic Cooperation and Development (OECD, 2019).

Literature review

STEM education in early childhood

The preschool years are the critical period for starting education if STEM literacy continues throughout life (Jipson et al., 2014). Early childhood is critical for children's brain and neuron development. Children's experiences in the first 8 years of life shape their lives in the years to come. Therefore, it is now acknowledged that starting the learning experience with STEM education in early childhood will yield positive outcomes in the future (Moomaw and Davis, 2010; Torres-Crospe et al., 2014; Allen, 2016). Teachers motivate children in early childhood to compare STEM education activities with real-life problems and produce solutions. Looking for solutions to their problems helps develop children's sense of curiosity while supporting their research and inquiry skills (John et al., 2018; Tank et al., 2018). Children test themselves on how to apply their experiences in newly encountered situations by learning new information through research. Developing various projects with STEM applications in early childhood supports children's cognitive, affective and behavioural skills. Children's skills backed by STEM education cover all the 21st-century skills. It is an adequate area for developing problem-solving, cooperation, creativity, communication, and critical thinking (Israsena et al., 2016; Ludwig et al., 2016; Perignat and Katz-Buonincontro, 2019).

Supporting children's STEM skills in early childhood classes can be increased by teaching in play settings prepared with a child-centred approach. Early childhood STEM education also lets children artistically develop their skills, self-expression, technology literacy, and engineering skills by playing games in science, mathematics, engineering, and technology (Van Hoorn et al., 2011). Science and the nature of science are mirrored in STEM education. Since early childhood, STEM education has supported children's interest in and attitude towards all areas of science.

When children's attitudes towards and interest in science are fostered in their early years, this is reflected positively in their academic achievements in later years. Therefore, children should be given STEM education early (Kershaw et al., 2009). Lamb et al. (2015) integrated STEM education into the syllabuses of kindergarten, second-and fifth-grade children and examined how this affected their cognitive and affective development. The STEM-integrated syllabus was applied from kindergarten to fifth grade.

Their study reported that the children's self-efficacy, interest, and knowledge increased significantly. Teachers play a critical role in integrating STEM education into the syllabuses applied to children. It is essential to determine how teachers integrate STEM education into the syllabus (Chalmers et al., 2017). When teachers are being given STEM training, they need to be given information about teaching practices and the materials required for teacher STEM proficiency and to practice integrated STEM education (Stohlmann et al., 2012). Countries are expected to support teachers through their education policies so they can successfully integrate STEM education into the syllabuses applied to children. Accordingly, if qualified teachers conduct STEM activities with children starting in their early years, their knowledge, skills, and academic achievement will increase, and countries will obtain the trained workforce they require (Quigley and Herro, 2016). From this point of view, the current research results, in which Turkish and Greek teachers' views on early childhood STEM education are collected, will be examined, and the situations of the countries where the research is conducted will be described.

In the research conducted by Nikolopoulou (2022a), the opinions of Greek teachers about STEM activities for children between the ages of 4–7 were examined. Greek teachers stated that implementing STEM activities supported children's knowledge, skills and interests in learning. Teachers said they considered children's needs and motivations, cognitive development levels, and learning outcomes while preparing STEM activities. They also mentioned that the main difficulties with STEM activities were teacher training, lack of infrastructure, limited time to implement the activities, and ensuring children's interest and active participation. Papadakis et al. (2021a) profiled the attitudes of Greek and pre-service preschool teachers towards using educational robotics in STEM education. They found no significant difference between preschool and pre-service teachers' attitudes towards using educational robotics. Based on the findings of that study, they discussed the quality of the training to be given to teachers in integrating educational robotics into classrooms. They further recommended revising pre-service teachers' syllabi to consider new STEM and educational robotics technologies. Yıldırım (2021) asked Turkish preschool teachers their opinions on preparing STEM activities and reported that when preschool teachers carry out STEM activities, they use different methods and techniques depending on the STEM content. Teachers had problems planning lessons for STEM education due to their lack of content knowledge. Teachers' STEM education practices were also found to support their professional competencies. The study recommends revising professional development and pre-service teacher training programs to support teachers' STEM competencies and content knowledge.

Research in both countries shows that teachers know the importance of STEM education in early childhood. However, they do not consider themselves sufficient in STEM subjects and applications (Ültay and Ültay, 2020; Yıldırım, 2021; Papadakis et al., 2021a; Nikolopoulou, 2022a). According to the research about STEM education (Ültay and Ültay, 2020; Yıldırım, 2021; Papadakis et al., 2021b; Nikolopoulou, 2022a), PISA 2018 results, Turkey and Greece have the same group of problems with STEM education. Given these results, the STEMPCK of Greek and Turkish preschool teachers who support children's STEM skills in early childhood should be compared. The first research question in this study is: "*Does STEMPCK of preschool teachers differ between the two countries?*" This question clearly can solve many problems about STEM-oriented teacher education. For example, the results of the preparation, implementation and evaluation of STEM-oriented teacher education programs developed in one country may contribute to the other. In addition, solution suggestions determined in one country to support STEMPCK of teachers can be implemented in another country and contribute to the use of time and finance. This research question will help countries develop STEMPCK policies for teachers who support children's STEM skills in early childhood. Next, preschool teachers' STEMPCK is explained in light of the literature.

STEMPCK theoretical framework

Although there is no consensus on pedagogical content knowledge (PCK) in the literature, this model developed by Shulman (1986) is generally accepted. According to Shulman (1986), PCK is pedagogical, contextual, and content knowledge developed by teachers to support children's learning. Teachers create their PCKs using their content knowledge and teaching methods when teaching specific content to children. According to Grossman (1990), PCK consists of content, pedagogical, and contextual knowledge. Teachers use content, pedagogical, and contextual knowledge together in the teaching process to form the PCK for teaching a specific topic. In the Technology Pedagogical Content Knowledge (TPACK) model proposed by Mishra and Koehler (2006), PCK is a combination of pedagogical knowledge (PK) and content knowledge (CK). They argue that PCK contains extensive information on how to organize teaching methods and content-specific characteristics appropriate to a particular context for teaching. PCK creates a link between the pedagogy practised by the teacher, the program, and assessment in the children's learning process. Teachers create ideas for different thinking foundations by using alternative teaching methods to teach specific content. By using PCK effectively, teachers increase children's prior knowledge and awareness when they have to develop different solutions to the same problems (Mishra and Koehler, 2006).

When PCK for STEM education is examined, the definition of STEMPCK made by Saxton et al. (2014) consists of three elements. The first element of STEMPCK is teachers' knowledge of considering what they know about STEM content. The second element is teachers' knowledge of guiding children in the STEM teaching process. The third element is the knowledge of integrating technology into the learning environment to improve teachers' teaching of STEM content. However, for STEM content to be used effectively in the learning environment, teachers are expected to

develop a deep pedagogical understanding of STEM education and the content of STEM education. It is not easy to integrate the disciplines of science, technology, engineering, and mathematics in STEM content and to prepare content for children in line with STEM philosophy (Beswick and Fraser, 2019; Margot and Kettler, 2019). In addition to such difficulties as integrating the disciplines due to the nature of STEM content, the changes occurring today in information and technologies make it difficult for teachers to adapt PCK to the learning process. In this case, teachers' competencies in using information, communication, and technologies (ICT) are expected to improve (Beswick and Fraser, 2019; Penprase, 2020; Silva et al., 2020; Gözüm, 2022). Teachers are expected to have advanced knowledge and skills concerning 21st-century skills and understanding, which are closely related to the nature of STEM education and are among the learning outcomes. Research results show that teachers with improved knowledge and skills regarding 21st-century skills use it more effectively in the PCK learning process for STEM education (Dede, 2010; Howland et al., 2012; Voogt and Roblin, 2012; Ertuğrul Akyol, 2020). For teachers to provide adequate STEM education to children, the STEMPCK theoretical framework consists of interacting with the nature of STEM content and different combinations of PCK elements. Yıldırım and Şahin-Topalcengiz (2019) completed a literature review of the STEM theoretical framework and argued that it consisted of STEM content information (Moore and Smith, 2014; Srikoom et al., 2017), STEM integration information (Bryan et al., 2016; Türk et al., 2018; Beswick and Fraser, 2019; Margot and Kettler, 2019), pedagogical information (Shulman, 1986; Grossman, 1990; Mishra and Koehler, 2006; Yusof et al., 2012; Saxton et al., 2014), 21st-century skills information (Dede, 2010; Howland et al., 2012; Voogt and Roblin, 2012; Ertuğrul Akyol, 2020) and context knowledge (Shulman, 1986; Barnett and Hodson, 2001). Based on this, the elements of the STEMPCK theoretical framework will be discussed under the title of Preschool Teachers' STEMPCK.

Preschool teachers' STEMPCK

While examining the preschool teachers' STEMPCK, this paper will explain STEM content knowledge, STEM integration knowledge, pedagogical knowledge, 21stCentury skill knowledge, and context knowledge within the theoretical framework of STEMPCK.

Teachers must know about STEM content if they are to provide early childhood children with practical and qualified STEM education. Teachers need to have a deep knowledge of STEM content areas and the ability to combine knowledge with experience (Whitebook and Ryan, 2011). Çorlu et al. (2015) emphasized that teachers who do not know about STEM content and education will not be able to acquire children's STEM learning outcomes. Although there is no adoption content in STEM content (Holmlund et al., 2018), it is argued that the integrated delivery of STEM disciplines will be more beneficial for the development of children (Bybee, 2013). However, preschool teachers are expected to be competent in STEM content so they can integrate STEM disciplines and present them to children when teaching them. The study by Moore and Smith (2014) found that teachers used science and mathematics more than technology and engineering when providing STEM education. This research result means that teachers prefer disciplines with STEM content where they feel competent. This result can be considered an essential deficiency in the integrated delivery of STEM education. Knowledge of integration is as necessary as the content of STEM education. The preschool teacher is expected to integrate the subject content pedagogically so that he or she can implement STEM activities in the learning environment (Ostler, 2012). This requires that STEM and PCK combine. A preschool teacher with a good knowledge of STEM components is expected to use his pedagogical knowledge to integrate STEM into early childhood classes. Pedagogical knowledge, which the teacher is expected to have, covers planning activities, implementing them, and evaluating them afterwards. Teachers should have advanced classroom management skills and teaching methods and techniques and be well versed in children's learning psychology (Shulman, 1986; Briscoe and Peters, 1997).

The National Research Council [NRC] (2014) has described the characteristics of STEM education. According to National Research Council [NCR] (1994) , the 21st-century skills of children receiving STEM education are expected to improve. The STEM and PCK knowledge of educators is expected to increase. Equipped with STEM content knowledge and Pedagogical Field Knowledge, the teacher understands the needs of children and aims to furnish them with the skills to possess 21st-century skills by considering the children's development. The 21st-century skills that teachers possess when implementing STEM activities matter. Teachers are expected to be role models for the 21st-century skills they aim to teach children. These 21st-century skills are considered life and career skills. They are expressed as global awareness, information and media literacy, leadership, responsibility, communication, efficiency, technology literacy, creativity, problem-solving, and critical thinking behaviours (Kennedy and Odell, 2014). Teachers should integrate content by creating context according to the characteristics of their region, the children's backgrounds, and the region where the schools are found when planning STEM activities (Barnett and Hodson, 2001; Harris and Hofer, 2011; Gkontelos et al., 2022). Sharapan (2012) suggestions for preschool teachers' context knowledge are as follows: While teachers create a context for STEM, they should select the events, phenomena, and objects children meet in their immediate environment. STEM activities should be built on the events that take place in daily life. Teachers should consider children's interests and needs when creating context. Meaningful contexts for STEM are crucial as they support children's learning. Paint, toys, Lego, parks, etc. can create meaningful contexts for children in STEM education. According to Allen (2016), contexts for STEM activities in early childhood should include concrete experiences. Johnson (2016) says that the contexts created in STEM activities should support children's basic scientific process skills

such as asking questions, guessing, observing, and experimenting. Preschool teachers' knowledge of creating context is expected to affect STEM practices positively. STEMPCK combines teachers' STEM knowledge, pedagogical knowledge, context knowledge, and 21st-century skills.

When teachers know about content in STEM practices, this makes them more confident in planning and implementing activities (Bers et al., 2013; Eng Tek et al., 2016). Teachers enable multidimensional learning by bringing together different disciplines in STEM content knowledge and forming connections between these disciplines (Smith and Karr-Kidwell, 2000). When preschool teachers experience uncertainty about teaching STEM content, this can cause them to feel anxiety when implementing activities, leading to a reduction in teacher confidence concerning conducting STEM activities and a corresponding drop in the quality and effectiveness of the STEM education activity (Hedlin and Gunnarsson, 2014; Cohrssen and Page, 2016). For teachers to implement STEM activities effectively and proficiently, unique teaching methods and pedagogical skills should be supported by in-service and pre-service training (Atiles et al., 2013; Bers et al., 2013). To support the STEMPCK competencies of teachers who plan and implement STEM activities, teachers should have hands-on practice, and the training they receive should not be theoretical only (Bers et al., 2013; Alexander et al., 2014). When teachers' STEM knowledge and skills are supported, their beliefs, attitudes, and feelings towards STEM practices can improve and make STEM activities more effective for children (Hedlin and Gunnarsson, 2014; Aldemir and Kermani, 2017; Park et al., 2017).

The study conducted by Koyunlu-Ünlü and Dere (2019) found that when preschool teachers received STEM training, this positively affected their STEM awareness. The study by Chanunan (2021) found that STEM PCK-based training positively affects the STEMPCK of pre-service science and mathematics teachers. The results of this research shed light on the training of teachers equipped with the knowledge and skills to teach STEM. Faikhamta et al. (2020) developed a pedagogical content knowledge-based STEM professional development program and applied it to science teachers. The research found that the implemented program positively supported teachers' STEM knowledge and practices and developed awareness about STEM disciplines. This being the case, the STEMPCK of preschool teachers in Greece and Turkey who received and did not receive STEM training should be examined. Teachers who received STEM training are expected to have high STEMPCK. This is thought to be necessary for countries' teacher training policies. Therefore, this study's second research question is, *"How do preschool teachers' training influence their STEMPCK?"* It is expected that teachers will be supported by STEM training, that children will be provided with sufficient support and infrastructure for STEM education, and that Greece and Turkey will develop economically and technologically. How this situation interacts with the STEM education variable by country is expected to be similar for all the world countries (*Australia*, *Indian* and *Malaysian,* etc.) (Thomas and Watters, 2015). Although the countries where teachers supported by STEM training live vary, support for their STEMPCK competencies should not cause a significant difference. In this case, the study's research question is *"What is the relation between the teachers' country of origin and their training in STEMPCK scores?"*

Materials and methods

Research model

The primary purpose of this study is to compare the STEMPCK of Greek and Turkish preschool teachers. The research was conducted using the descriptive survey method, a quantitative research approach. The study aimed to describe the STEMPCK status of Greek and Turkish preschool teachers by using the descriptive survey method. The descriptive survey model is used to learn individuals' attitudes, opinions, beliefs, and demographic characteristics in educational sciences (Johnson and Christensen, 2014).

Participants

The study participants were 104 Greek and 565 willing Turkish preschool teachers selected using simple random sampling in line with the quantitative survey method. With simple random sampling, each participant has an equal chance of participating in the study. As the participants participated in the study independently of each other, the probability of them representing the universe is high (Büyüköztürk et al., 2011). Table 2 shows the participants' demographic details by country.

Table 2 shows that a total of 104 preschool teachers from Greece participated in the study, of whom 80.8% ($n = 84$) were female and 19.2% ($n = 20$) were male; 43.3% of the teachers held a bachelor's degree ($n = 84$), 56.7% held a master's degree or doctorate. While 34.6% ($n = 36$) of the Greek preschool teachers who participated in the study had received STEM training, 65.4% ($n = 68$) had not. The average age of the Greek teachers participating in the study was 35.09 years. Furthermore, 19.2% ($n = 20$) of the Greek preschool teachers participating in the study taught children aged 36–48 months, 11.5% ($n = 12$) taught children aged 49–60 months, and 69.2% ($n = 72$) taught children aged 61–72 months.

A total of 565 preschool teachers from Turkey participated in the study, of whom 87.8% ($N = 496$) were female, 12.2% ($N = 69$) were male, and 70.4% of the teachers held a bachelor's degree ($N = 84$), and 29.6% held master's degree or doctorate. While 37.2% ($n = 210$) of the Turkish preschool teachers participating in the study had received STEM training, 62.8% ($n = 355$) had not. The Turkish teachers' average age in the study was 33.29 years. Of the Turkish preschool teachers participating in the study 10.4% taught children aged 36–48 months ($n = 59$), 28.1% taught children aged 49–60 months ($n = 159$), and 61.4% taught children aged 61–72 months ($n = 347$).

TABLE 2 Participants' demographic details.

Country		Gender		Education level		Have you received any form of STEM training?		Age						
		Female	Male	Bachelor's degree	Master or doctorate	Yes	No	20–25	26–30	31–35	36–40	41–45	46 – +	Total
Greek	*n*	84	20	45	59	36	68	7	13	26	48	10	–	104
	%	80.8	19.2	43.3	56.7	34.6	65.4	6.7	12.5	25.0	46.2	9.6	–	100.0
Turkey	*n*	496	69	398	167	210	355	96	119	112	172	45	21	565
	%	87.8	12.2	70.4	29.6	37.2	62.8	17.0	21.1	19.8	30.4	8.0	3.7	100.0
Total	*n*	580	89	443	226	246	423	103	132	138	220	55	21	669
	%	86.7	13.3	66.2	33.8	36.8	63.2	15.4%	19.7	20.6	32.9	8.2	3.1	100.0

Data collection tool

The data collection tool used in this study was the STEM Pedagogical Content Knowledge Scale (STEMPCK Scale) developed by Yıldırım and Şahin-Topalcengiz (2019). When they developed their scale, they performed exploratory factor analysis on the data of 443 pre-service teachers and confirmatory factor analysis on the data of 212 pre-service teachers. The construct validity studies for the STEMPCK Scale found six factors, namely, 21st-Century Skills, Pedagogical Knowledge, Mathematics, Science, Engineering, and Technology. The fit index values resulting from the confirmatory factor analysis of the scale revealed a good fit: (CFI = 0.93, NFI = 0.93, GFI = 0.99, AGFI = 0.97, IFI = 0.93, SRMR = 0.043, RMSEA = 0.034). The internal consistency coefficient of the total STEMPCK Scale was found to have a Cronbach's Alpha value of 0.95. The internal consistency coefficients of the sub-factors of the STEMPCK Scale ranged between 0.78 and 0.90. These results showed that the STEMPCK Scale was valid and reliable data collection tool to measure pre-service teachers' STEM pedagogical content knowledge.

Validity and reliability study of the data collection tool

The participants in this study were Turkish and Greek preschool teachers, yet the data collection tool had been developed using pre-service preschool teachers. As the data collection tool's target group had changed, this study's data set was used to perform the validity and reliability analyses. Construct validity was examined by subjecting the Greek and Turkish preschool teachers' data sets to separate CFA. The number of participants ($N_{Greece} = 104$; $N_{Turkey} = 565$) varied by country. Fit index values are affected by the number of participants. Therefore the CFA analysis fit index values to be applied to the data sets for the Greek and Turkish teachers were worked out differently as the data set for the Greek participants is $N < 250$, the Comparative Fit Index (CFI) value, which is less affected by the sample, was examined first. The Normed Fit Index (NFI) and Incremental Fit Index (IFI) values were examined together with the CFI value within the scope of the comparative model fit index. The IFI value matters in that it is calculated by considering the sample size and the complexity of the model. The Goodness of Fit Index (GFI) value, which tests the model regardless of the sample size, is also considered. The Adjustment Goodness of Fit Index (AGFI) value should be examined to decide the fit value adjusted according to the degree of freedom of GFI. It is recommended that the Root Mean Square Error of Approximation (RMSEA) value is preferred less in samples where $N < 250$. Like the RMSEA value, the Standardized Root Mean Square Residual (SRMR) value also does not show a good fit value in small samples. The chi-square value (X^2) is susceptible to sampling. At the same time, the X^2 value is significant in samples where $N < 250$ ($p > 0.05$) is insignificant in large samples ($p < 0.05$). Therefore, it is argued that the value of X^2 divided by the degree of freedom (*df*) or (X^2/df) will yield better results to evaluate the model's overall goodness of fit (Gürbüz, 2019). Therefore, (X^2/df), the CFI, NFI, GFI, AGFI, and IFI values were considered in the CFA analysis applied to the Greek data set. By contrast, the Turkish data set considered (X^2/df), the CFI, NFI, GFI, AGFI, IFI, SRMR, and RMSEA values because the sample size was sufficient. The same fit index values were examined for the entire data set. Table 3 shows the fit index values (fiv) resulting from the CFA analysis by country.

On examination of Table 3, the X^2 value for the fit index value of the Greek participants' data set is small and significant ($p > 0.05$), and it is thought that this is because the sample was small ($N < 250$). Examination of the data set for the Turkish and total participants shows the X^2 value to be large and not significant ($p < 0.05$). This lack of significance may be because the Turkish and comprehensive data set was ($N > 250$). The X^2/sd values were examined because fit values for data sets are affected by the number of participants. The X^2/df value for the Greek data set was 1.08. Being <3, it showed a good fit. The X^2/df value for the Turkish data set was 3.94, and 4.67 for the total participants. X^2/df values between 3 and 5 show an acceptable fit (Munro, 2005; Şimsek, 2007; Hooper et al., 2008). When the fit index values of the Greek

TABLE 3 Fit index values as a result of CFA analysis applied to the data set.

Country	X^2	*df*	X^2/df	CFI	NFI	GFI	AGFI	IFI	SRMR	RMSEA
Greek	1587.54	1,469	1.080	0.91	0.92	0.91	0.92	0.92	–	–
Turkey	5793.65	1,469	3.944	0.93	0.90	0.92	0.94	0.93	0.057	0.065
Data	6862.42	1,469	4.671	0.94	0.91	0.92	0.93	0.93	0.060	0.068
Acceptable Fiv		$3 < X^2/df < 5$		$90 \leq$ CFI, NFI, GFI, AGFI, IFI < 0.95					$0.05 \leq$ SRMR, RMSEA ≤ 0.08	
Good Fiv		$X^2/df < 3$		$0.95 \leq$ CFI, NFI, GFI, AGFI, IFI ≤ 0.99					SRMR, RMSEA ≤ 0.05	

TABLE 4 Cronbach's alpha values for STEMPCK scale and its component factors.

STEMPCK scale and factor	Greek	Turkey	Total data
STEMPCK scale	0.85	0.88	0.90
Pedagogical knowledge	0.82	0.84	0.87
Science knowledge	0.79	0.83	0.85
Technology knowledge	0.78	0.82	0.86
Engineering knowledge	0.81	0.85	0.87
Mathematical knowledge	0.77	0.79	0.84
21st-century skills knowledge	0.80	0.83	0.91

and Turkish participants are examined, the CFI, NFI, GFI, AGFI, and IFI values range between 0.90 and 0.95, thus showing acceptable fit values. On examination of the fit index values of the data set for the Greek and Turkish participants in total, they were found to be acceptable (Bentler, 1980; Bentler and Bonett, 1980; Baumgartner and Homburg, 1996; Marsh et al., 2006). SRMR and RMSEA values range between.05 and.08, showing an acceptable level (Browne and Cudeck, 1993). Therefore, the STEMPCK Scale was valid at an acceptable level.

Table 4 shows the reliability values of the STEMPCK Scale and its factors based on the data set of the participants in this study.

Table 4 shows that when the Cronbach's Alpha (α) values of the data set of Greek and Turkish participants are examined, the Cronbach's Alpha (α) value of STEMPCK Scale and Its Factors is more significant than 0.70. The Cronbach's Alpha (α) values for the STEMPCK Scale and its Factors for the Greek and Turkish participants' total data set a range between.84 and.91. According to George and Mallery (2003), a Cronbach's Alpha value of $0.7 \leq \alpha < 0.9$ shows that the scale is reliable in terms of internal consistency. Therefore, the STEMPCK Scale and Its Factors used in the study can be considered reliable.

Data collection

The researchers converted the data collection tools into Google Forms and had them published in Greek and Turkish. The researchers then distributed the Google Forms to the participants using social media apps such as Facebook and WhatsApp. The institutions where the participants worked did not assist in distributing the Google Forms. The Google Form includes a consent form for the participants to complete saying they are voluntarily participating in the study. It also includes an ethics declaration, saying that the participant's data will not be shared. The researchers shared their email addresses and contact information on Google Forms so the participants could obtain information about the research and ask questions. After the participants approved the consent form saying they wanted to participate voluntarily in the study, they shared personal information and fill in the data collection tool's fields. All the scale items on the Google Form had the mandatory box selected to avoid data loss and were completed accordingly. The Google Form also included an open-ended question for those teachers who did not want to have to fill in the scale items and those who wanted to express their opinions about the comprehensibility of the scale items. The participants voiced no negative opinions about the mandatory items or their comprehensibility. This may indicate that the participants filled in the Google Form without social desirability bias.

All procedures performed in studies involving human participants were following the ethical standards of by Kafkas University Ethics Committee in Turkey - Türkiye research committee (document no: E.30529/24–05.2022) and by the University of Crete ethics committees in Greece (document no: 606/18-5-2022) and also the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Data analysis

This study's data were analyzed for two different reasons. The purpose of the analyses was to determine the validity and reliability of the data collection tool. The purpose of the second analysis was to determine the findings for the research problems. Validity and reliability analyses were performed by dividing the data for the Greek and Turkish participants into two sets. CFA and internal consistency coefficients were calculated for both data sets. Validity and reliability analyses were made for the data set, which was the sum of the data of the Greek and Turkish participants. AMOS and SPSS computer software programs were used for these statistical operations. The mean, standard deviation, percentage, and frequencies were calculated as required by descriptive

analysis. A MANOVA test was performed for the research problem to examine whether the dependent variables had a significant difference over the independent variable. Tukey, a *post hoc* test, was used to determine the direction in cases where significant differences were found. An independent t-test was conducted to see if there was a significant difference based on whether the Turkish and Greek participants received STEM training in their respective countries.

Data analysis assumptions

The assumptions of the CFA and MONOVA tests were examined when the data were analyzed. The Greek, Turkish, and complete data sets were analyzed separately when examining the DFA analysis assumptions. The data for the Greek and Turkish participants were analyzed as a single data set when conducting the MANOVA test.

Before conducting Confirmatory Factor Analysis, multiple normality values were examined in the Greek and Turkish participants' data set. The CR values for the Greek and Turkish participants in the data set were below 10, and the kurtosis values for the data sets ranged between −3 and + 3. The multiple normality values for the participants' data set show normal distribution. It can therefore be argued that the multiple normality assumption is met in the CFA analysis (Gürbüz and Şahin, 2018).

When conducting the MANOVA test, Box's M test values were examined to test whether the variance and covariance matrices were equal. It was understood that the variance and covariance matrices of the dependent variables of *country* and *teachers receiving STEM training* were equal (Box's $M = 88.304$, $F = 1.075$, $sd_1 = 28$, $sd_2 = 18447.287$, $p = 0.374$). Levene's test results were examined to determine whether the variance distributions of independent variables were homogeneous. The Levene test results showed that the scale and the factors were homogeneous as there was no significant difference from $p > 0.05$. Wilks' Lambda values were examined for the multivariate test when performing MANOVA analyses.

Findings

This section gives the findings related to the research questions. Table 5 gives the findings to the research question, "*Does STEMPCK of preschool teachers differ between the two countries?*"

Since multiple Anovas have been made in Tables 6 a type I error is possible. Therefore, Bonferroni correction is used to check type 1 error significant difference. Since Bonferroni correction was used in this study, each ANOVA was evaluated at a significance level (p) of.017.

"*Does STEMPCK of preschool teachers differ between the two countries?*" When the findings of this research question are examined, Table 5 shows ($\lambda = 0.985$, $F_{(6.660)} = 1.678$, $p = 0.124$, $p > 0.01$) showing no significant difference in STEMPCK scores between countries.

TABLE 5 Descriptive statistic and multivariate test.

Scale / Component	Country	Q*	N	Mean	SD
Pedagogical knowledge	Total	Yes	246	53,45	6,41
		No	423	51,62	5,81
		Total	669	52,29	6,10
Science knowledge	Total	Yes	246	33,14	4,60
		No	423	29,28	5,80
		Total	669	30,70	5,70
Technology knowledge	Total	Yes	246	28,31	5,16
		No	423	25,32	4,96
		Total	669	26,42	5,23
Engineering knowledge	Total	Yes	246	27,30	5,27
		No	423	24,28	4,59
		Total	669	25,39	5,06
Mathematical knowledge	Total	Yes	246	34,07	5,12
		No	423	30,90	5,86
		Total	669	32,06	5,80
21st century skills knowledge	Total	Yes	246	18,67	1,88
		No	423	17,77	2,25
		Total	669	18,10	2,16
STEMPCK	Total	Yes	246	194,95	20,74
		No	423	179,20	20,09
		Total	669	184,99	21,69

Multivariate test

Effect	λ	F	Hypothesis df	Error df	Sig.	η_p^2
Intercept	0.018	6102.288	6.000	660.000	0.000	0.982
Country	0.985	1.678	6.000	660.000	0.124	0.115
STEM training	0.919	9.724	6.000	660.000	0.000	0.181
Country* STEM training	0.991	.974	6.000	660.000	0.442*	0.009

*Have you received any form of STEM training?

"How do preschool teachers' training influence their STEMPCK?" When the findings of the research question are examined, the multivariate test results in Table 5 show ($\lambda = 0.919$, $F_{(6.660)} = 9.724$, $p = 0.000$, $p < 0.01$), indicating a significant difference in teachers' STEMPCK scores based on whether they received STEM training or not. When the total STEMPCK score and STEMPCK factors are examined, a significant difference is found in the Pedagogical knowledge factor ($F_{(1.665)} = 3.229$, $p = 0.073$, $p > 0.01$), science knowledge ($F_{(1.665)} = 43.855$, $p = 0.000$, $p < 0.01$), technology knowledge ($F_{(1.665)} = 28.181$, $p = 0.000$, $p < 0.01$), engineering knowledge ($F_{(1.665)} = 27.559$, $p = 0.000$, $p < 0.01$), mathematical knowledge ($F_{(1.665)} = 22.762$, $p = 0.000$, $p < 0.01$), 21st-Century Skills knowledge ($F_{(1.665)} = 21.313$, $p = 0.000$, $p < 0.01$) and STEMPCK total score ($F_{(1.665)} = 44.346$, $p = 0.000$, $p < 0.01$).

"What is the relation between the teachers' country of origin and their training in terms of STEMPCK scores?" When the findings of the research question are examined, the multivariate test results in

TABLE 6 Multidirectional analysis of variance values and pairwise comparisons.

Tests of between-subjects effects

Corrected Model	Dependent variable	Sum of squares	*df*	Mean Square	*F*	Sig.	η_p^2
	Pedagogical knowledge	617.862	3	205.954	5.650	0.001	0.025
	Science knowledge	2410.044	3	803.348	27.619	0.000	0.111
	Technology knowledge	1445.485	3	481.828	18.977	0.000	0.079
	Engineering knowledge	1418.467	3	472.822	19.972	0.000	0.083
	Mathematical knowledge	1569.295	3	523.098	16.583	0.000	0.070
	21st-century skills knowledge	134.909	3	44.970	9.986	0.000	0.043
	STEMPCK	39223.572	3	13074.524	31.602	0.000	0.125
Intercept	Pedagogical knowledge	871610.382	1	871610.382	23909.836	0.000	0.973
	Science knowledge	304730.901	1	304730.901	10476.483	0.000	0.940
	Technology knowledge	225074.817	1	225074.817	8864.833	0.000	0.930
	Engineering knowledge	211683.467	1	211683.467	8941.520	0.000	0.931
	Mathematical knowledge	335863.933	1	335863.933	10647.077	0.000	0.941
	21st-century skills knowledge	107254.320	1	107254.320	23817.931	0.000	0.973
	STEMPCK	11070067.491	1	11070067.491	26757.427	0.000	0.976
Country	Pedagogical knowledge	64.365	1	64.365	1.766	0.184	0.003
	Science knowledge	76.319	1	76.319	2.624	0.106	0.004
	Technology knowledge	53.517	1	53.517	2.108	0.147	0.003
	Engineering knowledge	2.081	1	2.081	0.088	0.767	0.000
	Mathematical knowledge	3.168	1	3.168	0.100	0.751	0.000
	21st-century skills knowledge	5.630	1	5.630	1.250	0.264	0.002
	STEMPCK	621.205	1	621.205	1.502	221	0.002
STEM Training	Pedagogical knowledge	117.720	1	117.720	3.229	0.073	0.005
	Science knowledge	1275.617	1	1275.617	43.855	0.000	0.062
	Technology knowledge	715.516	1	715.516	28.181	0.000	0.041
	Engineering knowledge	652.426	1	652.426	27.559	0.000	0.040
	Mathematical knowledge	718.045	1	718.045	22.762	0.000	0.033
	21st-century skills knowledge	95.974	1	95.974	21.313	0.000	0.031
	STEMPCK	18346.808	1	18346.808	44.346	0.000	0.063
Country * STEM	Pedagogical knowledge	60.798	1	60.798	1.668	0.197	0.003
Training	Science knowledge	4.043	1	4.043	0.139	0.709	0.000
	Technology knowledge	0.066	1	0.066	0.003	0.959	0.000
	Engineering knowledge	3.993	1	3.993	0.169	0.681	0.000
	Mathematical knowledge	4.777	1	4.777	0.151	0.697	0.000
	21st-century skills knowledge	6.184	1	6.184	1.373	0.242	0.002
	STEMPCK	52.228	1	52.228	0.126	0.722	0.000
Error	Pedagogical knowledge	24241.944	665	36.454			
	Science knowledge	19342.946	665	29.087			
	Technology knowledge	16884.103	665	25.390			
	Engineering knowledge	15743.354	665	23.674			
	Mathematical knowledge	20977.542	665	31.545			
	21st-century skills knowledge	2994.556	665	4.503			
	STEMPCK	275123.427	665	413.719			

(Continued)

TABLE 6 (Continued)

Tests of between-subjects effects

Total	Pedagogical knowledge	1854591.000	669	Corrected total	Pedagogical knowledge	24859.806	668
	Science knowledge	652506.000	669		Science knowledge	21752.990	668
	Technology knowledge	485515.000	669		Technology knowledge	18329.587	668
	Engineering knowledge	448591.000	669		Engineering knowledge	17161.821	668
	Mathematical knowledge	710550.000	669		Mathematical knowledge	22546.837	668
	21st-century skills knowledge	222449.000	669		21st-century skills knowledge	3129.465	668
	STEMPCK	23210502.000	669		STEMPCK	314346.99	668
Pairwise comparisons							
	Dependent variable	(I)	(J)		Mean difference (I-J)	Std. error	Sig.
	Pedagogical knowledge	Yes	No		1.214	0.675	0.073
	Science knowledge	Yes	No		3.996*	0.603	0.000
	Technology knowledge	Yes	No		2.993*	0.564	0.000
	Engineering knowledge	Yes	No		2.858*	0.544	0.000
	Mathematical knowledge	Yes	No		2.998*	0.628	0.000
	21st-century skills knowledge	Yes	No		1.096*	0.237	0.000
	STEMPCK	Yes	No		15.153 *	2.276	0.000

(I)/(J) = Have you received any form of STEM training?

Table 5 show ($\lambda = 0.991$, $F_{(6.660)} = 0.974$, $p = 0.442$, $p > 0.01$), indicating no significant difference in the teachers' STEMPCK scores.

A significant difference is found between the teachers' STEMPCK scores based on whether they received STEM training or not (Table 6). The direction of significant difference is examined in Table 6.

When the direction of significant difference in Table 6 is examined, a significant difference is found in the science knowledge ($p < 0.01$), technology knowledge ($p < 0.01$), engineering knowledge ($p < 0.01$), mathematical knowledge ($p < 0.01$), 21st-century skills knowledge ($p < 0.01$) factors and the STEMPCK total score ($p < 0.01$) in favor of teachers who received STEM training.

The figures below were used to examine by country if there was a significant difference between the STEMPCK scores based on whether or not the teachers had received STEM training by looking at the estimated marginal means values. Also examined was whether the Turkish and Greek teachers receiving STEM training or not made a significant difference in their respective countries. The STEMPCK and sub-factors are examined in Table 7.

Table 7 shows a significant difference in the pedagogical knowledge scores of Greek preschool teachers based on whether they received STEM training or not ($t_{102} = 3.308$; $p = 0.000$; $p < 0.050$). There is a significant difference in Turkish preschool teachers' pedagogical knowledge scores based on whether they received STEM training or not ($t_{563} = 3.807$; $p = 0.000$, $p < 0.050$). When the direction of the significant difference was examined, it was found to favour the pedagogical knowledge of preschool teachers who received STEM training.

There is a significant difference in Greek preschool teachers' science knowledge scores based on whether they received STEM training or not ($t_{102} = 5.134$; $p = 0.000$, $p < 0.050$). There is a significant difference in Turkish preschool teachers' science knowledge scores based on whether they received STEM training or not ($t_{563} = 3.807$; $p = 0.000$, $p < 0.050$). When the direction of the significant difference was examined, it was found to favour the science knowledge of preschool teachers who received STEM training.

There is a significant difference in the technology knowledge scores of Greek preschool teachers based on whether they received STEM training or not ($t_{102} = 3.279$; $p = 0.001$; $p < 0.050$). There is a significant difference in Turkish preschool teachers' technology knowledge scores based on whether they received STEM training or not ($t_{563} = 3.807$; $p = 0.000$, $p < 0.050$). When the direction of the significant difference is examined, it is found to favour the technical knowledge of preschool teachers who received STEM training.

There is a significant difference in the engineering knowledge scores of Greek preschool teachers based on whether they received STEM training or not ($t_{102} = 3.279$; $p = 0.001$; $p < 0.050$). There is a

10.3389/fpsyg.2022.996338

TABLE 7 Independent *t*-test result of STEM education status by country.

	Country	Q*	*N*	Mean	*sd*	*df*	*t*	*p*
Pedagogical knowledge	Greece	Yes	36	52,94	5,23	102	3.308	0.000
		No	68	50,60	3,82			
		Total	104	51,41	4,33			
	Turkey	Yes	210	53,71	6,57	563	3.807	0.000
		No	355	51,62	6,12			
		Total	565	52,40	6,36			
Science knowledge	Greece	Yes	36	32,50	3,10	102	5.134	0.000
		No	68	28,27	4,38			
		Total	104	29,74	4,45			
	Turkey	Yes	210	33,25	4,81	563	7.720	0.000
		No	355	29,48	6,03			
		Total	565	30,88	5,89			
Technology knowledge	Greece	Yes	36	27,63	5,09	102	3.279	0.001
		No	68	24,61	4,10			
		Total	104	25,66	4,67			
	Turkey	Yes	210	28,42	5,17	563	6.630	0.000
		No	355	25,46	5,11			
		Total	565	26,56	5,32			
Engineering knowledge	Greece	Yes	36	26,97	4,87	102	3.306	0.001
		No	68	24,33	3,21			
		Total	104	25,25	4,04			
	Turkey	Yes	210	27,35	5,35	563	7.042	0.000
		No	355	24,27	4,82			
		Total	565	25,42	5,23			
Mathematical knowledge	Greece	Yes	36	33,69	4,95	102	2.768	0.007
		No	68	30,94	4,75			
		Total	104	31,89	4,97			
	Turkey	Yes	210	34,13	5,16	563	6.479	0.000
		No	355	30,89	6,06			
		Total	565	32,10	5,95			
21st Century Skills knowledge	Greece	Yes	36	19,13	1,53	102	2.963	0.004
		No	68	17,76	2,54			
		Total	104	18,24	2,33			
	Turkey	Yes	210	18,59	1,92	563	4.477	0.000
		No	355	17,77	2,19			
		Total	565	18,08	2,13			
STEMPCK	Greece	Yes	36	191,88	12,18	102	5.439	0.004
		No	68	177,54	13,10			
		Total	104	182,50	14,46			
	Turkey	Yes	210	195,48	21,85	563	8.558	0.000
		No	355	179,52	21,16			
		Total	565	185,45	22,75			

*Have you received any form of STEM training?

significant difference in Turkish preschool teachers' engineering knowledge scores based on whether they received STEM training or not ($t_{563} = 3.807$; $p = 0.000$, $p < 0.050$). When the direction of the significant difference is examined, it is found to favour the engineering knowledge of preschool teachers who received STEM training.

There is a significant difference in the mathematical knowledge scores of Greek preschool teachers based on whether they received STEM training or not ($t_{102} = 3.279$; $p = 0.007$; $p < 0.050$). There is a significant difference in Turkish preschool teachers' mathematical knowledge scores based on whether they received STEM training or not ($t_{563} = 3.807$; $p = 0.000$, $p < 0.050$).

When the direction of the significant difference is examined, it is found to favour the mathematical knowledge of preschool teachers who received STEM training.

There is a significant difference in 21st-century skills knowledge scores of Greek preschool teachers based on whether they received STEM training or not ($t_{102}=2.963$; $p=0.004$; $p<0.050$). There is a significant difference in 21st-century skills knowledge scores of Turkish preschool teachers based on whether they received STEM training or not ($t_{563}=4.477$; $p=0.000$, $p<0.050$). When the direction of the significant difference was examined, it was found to favour the 21st-century skills knowledge of preschool teachers who received STEM training.

There is a significant difference in the STEM PCK scores of Greek preschool teachers based on whether they received STEM training or not ($t_{102}=5.439$; $p=0.004$; $p<0.050$). There is a significant difference in Turkish preschool teachers' STEMPCK scores based on whether they received STEM training or not ($t_{563}=8.558$; $p=0.000$, $p<0.050$). When the direction of the significant difference was examined, it was found to favour the STEMPCK scores of Greek and Turkish preschool teachers who had received STEM training.

Discussion and conclusions

The primary purpose of this study is to compare the STEMPCK of Greek and Turkish preschool teachers. The study data were examined to determine whether there was not a significant difference between Greek and Turkish preschool teachers in STEMPCK and its components (Pedagogical knowledge, science knowledge, technology knowledge, engineering knowledge, mathematical knowledge, and 21st-century skills knowledge; Table 5). The results show that Greek and Turkish teachers have similar levels of pedagogical content knowledge concerning STEM education. According to the 2018 PISA exam results, Turkey and Greece scored below the OECD average for reading, science, and mathematics, and their respective average scores were similar. The interpretation of the results reveals that there is no significant difference between Greek and Turkish preschool teachers' STEMPCK scores and that this is consistent with the 2018 PISA results (OECD, 2019). However, the fact that Turkey and Greece's 2018 PISA results are below OECD countries indicates that children's STEM knowledge and skills are not adequately supported. However, getting quality STEM education early can enable children to influence their academic success in the following years. At this point, it is expected that their teachers will have developed STEMPCK so that children can receive quality STEM education. According to the results of the research (Ültay and Ültay, 2020; Yıldırım, 2021; Papadakis et al., 2021a; Nikolopoulou, 2022a) conducted with Greek and Turkish preschool teachers, it has been determined that there are PCK deficiencies in STEM education.

According to the results of the study by Ling et al. (2020), in which they detected teachers' PCK shortcomings concerning STEM education, the teachers were found to have PCK shortcomings regarding STEM. Given the results of this research, teachers should receive STEM training to address their PCK shortcomings. Among those influential variables in identifying STEM needs and addressing them, one key variable is teachers receiving STEM training. To this end, the study examined *"How do preschool teachers' training influence their STEMPCK?"* The participating teachers receiving STEM training significantly differed in the total STEMPCK score (Table 6). Both Greek and Turkish preschool teachers receiving STEM training resulted in a significant differentiation in STEMPCK scores. Another result of the research is that there is no significant difference in the interaction between the country variable and the STEM training variable. When the STEM training rates of the teachers participating in the study were examined, it was seen that 34.6% of the Greek participants ($n=36$) and 37.2% of the Turkish participants ($n=210$) received STEM training. Given these results, a critical topic of discussion is how practical STEM activities will be in the classrooms of those Greek and Turkish teachers who have not received STEM training. Teachers who do not receive STEM education will likely experience difficulties planning, implementing and evaluating STEM activities. As a result of supporting Greek and Turkish teachers with STEMPCK, they can make an adequate STEM education. It is recommended that teachers take STEM education, according to the literature.

In their qualitative research, Weng et al. (2020) held semi-structured interviews with teachers. They concluded that teachers lacked STEM knowledge and PCK for STEM education. They recommended that teachers attend STEM training courses to increase their STEMPCK. The study conducted by Shernoff et al. (2017) identified professional needs concerning STEM based on the opinions of teachers and administrators. The teachers said that they considered themselves inadequate in terms of STEMPCK. They stated that they did not receive enough STEM training before service or in-service and did not know enough about planning, implementing, and assessing activities. Preschool teachers in Greece and Turkey should be given support in the form of STEM training before service and in-service. Though there are science, mathematics and technology courses related to STEM education, there are no specific courses in the integrated STEM content when the content of the preschool teacher training program in Turkey is examined. The same issue is also observed in Greece. Although preschool teachers in Greece have classes on STEM, they do not take classes for integrated STEM education.

Nonetheless, the most basic STEMPCK of teachers should be given in the pre-service period. Current developments and new approaches to STEM should be supported by training during the service process, and STEMPCK developments should be provided. Furthermore, the STEMPCK deficiencies of teachers are so much higher than in the pre-service period that it may be necessary to make intense efforts to eliminate them during the service period. Hence, it is not a coincidence that these countries are below OECD, according to PISA results. However, it should be stated that having a STEM education does not make the STEM teaching

process efficient because STEM education is a complex education process. PCK support of teachers alone is also not enough for STEM education. At the same time, it has been determined that teachers need training for STEM content knowledge. Teachers' lack of STEM content and PCK knowledge may reduce their tendency to avoid STEM activities or to engage children in STEM activities. When the relevant literature is examined,

The study conducted by Hsu et al., 2011 examined primary school teachers' perceptions of engineering, design, and technology concerning STEM. They determined that while teachers had strong beliefs in engineering, design, and technology, they did not feel competent enough in knowledge or educational practices. Training is vital if teachers are to feel competent in STEM content knowledge. Teachers can practice STEM education effectively by receiving training and increasing their self-efficacy beliefs (Rahman et al., 2017; Ling et al., 2020). The STEM content knowledge of Greek and Turkish teachers who receive and do not receive STEM training differs significantly from those who receive STEM training (see Table 6). According to Bandura (1986), teachers need the first-hand experience to increase their self-efficacy. Therefore preschool teachers' STEM training background is critically important. The study by Yılmaz (2019) determined that pre-service teachers with STEM experience were interested in STEM education and tended to engage in more STEM activities than those without STEM experience. The STEM education of teachers from Turkey and Greece should be beyond just the cognitive acquisition of STEM content and pedagogy. STEM education supports children's learning outcomes by influencing teachers' self-efficacy in organizing and conducting STEM activities. Teachers' doing STEM activities has an effect that improves their self-efficacy in this stage (Nathan et al., 2011). Teachers' STEM self-efficacy affects not only their current STEM education experiences but also their past direct experiences with STEM readiness. Other studies point out preschool teachers receive less training in the STEM education discipline than primary and secondary education teachers (Aldemir and Kermani, 2017). This literature information is valid for both Greek and Turkish preschool teachers. Therefore, preschool teachers need to acquire the necessary content information in the pre-service period to do STEM activities in their educational processes. Mathematics and science course hours of pre-service teachers in pre-service programs play an influential role in their STEM education. The lack of teachers' science and mathematics courses may be the main reason teachers have difficulties providing quality STEM education (King et al., 2013) because teachers should have the content knowledge to make STEM education effective. In this study, the significant differentiation in the STEM content knowledge of the teachers who received STEM education is in line with the literature results. Wan et al., (2021) argue that preschool teachers must be given STEM training during teacher training because preschool teachers have little experience in carrying out STEM activities involving multidisciplinary issues. Considering that Greek and Turkish preschool teachers have similar STEMPCK, STEM training should be given to preschool teachers of both countries as a matter of urgency, and their experiences should be supported. Papadakis et al. (2021a) and Yıldırım (2021) recommend that in-service training programs for STEM be provided for both Greek and Turkish preschool teachers and pre-service teachers and that STEM be included in teacher training programs. Another significant result of this study is that STEMPCK and its components differ significantly in favour of Greek and Turkish teachers who received STEM training (Table 7).

Papadakis et al. (2021a) and Yıldırım (2021) state that although it is helpful to provide teachers with STEM training, this is not enough to integrate STEM into classroom practices. The development of STEM content and pedagogy content knowledge of Turkish and Greek preschool teachers makes classroom STEM activities more understandable and suitable for child development (Williams, 2016). Therefore, education programs that include STEM content and pedagogical knowledge should be prepared for preschool teachers who are intended to provide STEM education (Schuster et al., 2012). It is understood that teachers who receive STEM education apply more qualified STEM activities to children. However, the teacher does not explain all aspects of STEM education which influence children's academic success and learning in STEM education (Fore et al., 2015). Training that will make teachers experienced in STEM activities is considered critical in the tendency to do STEM education because the STEMPCKs of teachers who receive STEM education differ significantly. This situation has essential effects on the reflection of STEM activities in practice. Thus, the results of this research are in line with the literature.

Wang et al. (2011) cited the lack of a STEM education program specifically tailored for teachers and children as the reason for not implementing qualified STEM activities in the classroom. Considering that there is no vocational training program specially prepared to train Greek and Turkish pre-service preschool teacher, it can be thought that Greek and Turkish teachers face similar difficulties in STEM activities. It is believed that implementing programs that support preschool teachers' knowledge, skills, and experiences in STEM, not just in Greece and Turkey but also those teaching early childhood classes in other countries too, will significantly affect how those countries develop (Cunningham and Higgins, 2015). Greece and Turkey can be considered representative examples here.

Moore et al. (2014) argue that the STEM training programs to be prepared for Greek and Turkish preschool teachers should be prepared in line with the contextual content for the conditions of the country where the teachers are located and the locations of the schools where they work (urban, rural and suburban). In-service training programs to be developed for different learning environments can meet specific pedagogical needs. It should not be forgotten that preschool teachers should plan and implement STEM activities according to the needs of the children. This shows that teachers' pedagogical knowledge concerning STEM practices according to the cultures of the different countries where children live is just as crucial as preschool teachers' STEM knowledge and experience. For example, according to Sullivan and Bers (2018),

children aged 3–6 years do not all possess the same programming skills for use in STEM activities. The cultural effects of countries on children's programming skills are observable. The average programming skill scores of children aged 3–6 years studying in Singapore kindergartens may match the average programming skill scores of first- and second-grade children in America. Cultural differences should therefore be considered when developing STEM and vocational education programs for Greek and Turkish preschool teachers and their children. The following should be considered when designing educational programs: teachers and pre-service teachers should gain practical experience in authentic learning environments where STEM education is done for children and develop their pedagogical knowledge to plan STEM content activities based on children's development levels and learning needs (Brenneman et al., 2019).

Teacher training and in-service programs to be prepared in support of Greek and Turkish preschool teachers' STEMPCK should be current, consistent, and developed in line with studies and practices in specific developmentally appropriate cultures (Papadakis et al., 2018, 2019, 2021a,b; Dorouka et al., 2020; Gkontelos et al., 2021; Gözüm and Kandır, 2021; Yıldırım, 2021; Gözüm, 2022; Mercan and Kandır, 2022). In addition, Greek and Turkish preschool teachers should be provided with resources prepared by experts in STEM education, and blogs, social media, and video sharing platforms should be created for more accurate and effective practices (Early Childhood STEM Working Group, 2017; Gözüm, 2021).

STEM education relates to 21st-century skills because teachers provide up-to-date information and support some skills. In this research, teachers' knowledge of 21st-century skills was also examined within the scope of STEMCPK. The present study found that Greek and Turkish preschool teachers receive STEM training, positively affecting their 21st-century skills. It is expected that the 21st-century skills of Turkish and Greek teachers are developed to provide qualified STEM education. STEM education programs that will include 21st-century skills should be developed starting from the pre-service period to support the 21st-century skills of Turkish and Greek teachers and to develop their practice attitudes towards STEM education. For teachers to be effective in STEM activities, they are expected to be technology literate, which is crucial in the digital age. The study by Çetin and Kahyaoğlu (2018) investigated how the application of STEM activities affected pre-service teachers' attitudes towards STEM education and 21st-century skills. The study concluded that pre-service teachers' attitudes towards STEM education and 21st-century skills improved in classes where STEM activities were applied. In the preschool years, when teachers are role models for children, the development of 21st-century skills of teachers can enable more effective implementation of STEM practices in the classroom. Therefore, the fact that Turkish and Greek preschool teachers are equipped with 21st-century skills enables them to acquire cognitive and affective skills that will support PCK. The mixed research study by Ertuğrul Akyol (2020) examined the effect of STEM activities on pre-service teachers' 21st-century skills. It concluded that STEM activities positively affected pre-service teachers' problem-solving, computational, critical, and creative thinking skills. The qualitative results showed that pre-service teachers with 21 t century skills effectively used simple materials in robotic and coding-based STEM activities. Teachers' role is essential in using digital applications in STEM activities in classroom applications (Papadakis et al., 2021b; Nikolopoulou, 2022b).

Teachers with advanced 21st-century skills are expected to be able to adapt technology to STEM activities. Moreover, Turkish and Greek preschool teachers can cope with many factors that prevent them from including educational robotics (Papadakis et al., 2021b) or mobile applications (Nikolopoulou, 2021) in classroom STEM activities due to their advanced 21st-century skills. Support for teachers' 21st- century skills will help them to plan and implement quality STEM activities and assess them. One of the goals of STEM education is for children to produce creative solutions to daily life problems because of technology-based learning and STEM activities supported by technology (Dede, 2010; Howland et al., 2012). Furthermore, increased teacher knowledge of 21st-century skills will support the development of children's skills such as problem-solving, collaborative learning, creative thinking, and self-learning through STEM activities (Dede, 2010; Voogt and Roblin, 2012; Gkontelos et al., 2022). In this context, teachers' knowledge of 21st-century skills from a STEM standpoint must be supported if they are continuously open to innovation and integrate technology and other variables into learning environments. This is why 21st-century skills should be included in the training programs that support STEMPCK.

Consequently, it was determined that Turkish and Greek teachers were similar to STEMPCK, and STEMPCK positively supported teachers who received STEM education in both countries. Another critical result of the research is that the country variable does not affect STEM education. In this context, a joint emergency action plan can be prepared to support STEMPCK of Turkish and Greek preschool teachers. The researchers believe that the Results and Discussion part of this study will help shape the content and structure of the vocational education planned for preschool teachers.

Recommendations

Preschool teachers in Greece and Turkey have similar countries and plan teacher training programs and in-service training activities together, considering cultural factors. Furthermore, they can prepare projects jointly with preschool STEMPCK. Greek and Turkish preschool teachers can take advantage of being neighbourly teachers from a country with an above-average PISA score to investigate the STEMPCK profiles of Greek and Turkish preschool teachers and prepare urgent training programs. Teachers must examine countries' early childhood education policies and invest in technology-supported applications to provide adequate education using STEMPCK.

Limitations and future research

One fundamental limitation of the study is the number of teachers participating in the study from Greece and Turkey. The results of this study are limited to the participating teachers. Although the number of participants is thought to be good when the ratio of preschool teachers working in their countries is considered, future studies can be planned with more participants. Nevertheless, despite these limitations, it is expected that comparing the STEMPCK of Greek and Turkish teachers will contribute to future research. Future research could benefit from mixed research models combining qualitative and quantitative research methods to add more detail to the results of this study.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by all procedures performed in studies involving human participants were following the ethical standards of by Kafkas University Ethics Committee in Turkey - Türkiye research committee (document no: E.30529/24-05.2022) and by the University of Crete ethics committees in Greece (document no: 606/18-5-2022) and also the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

AG: writing, literature review, methodology, design, data collecting, data analysis, data validation, conclusion and discussion. SP: literature review, methodology, data collecting and editing. MK: literature review, design, and data collecting. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Aldemir, J., and Kermani, H. (2017). Integrated STEM curriculum: improving educational outcomes for head-start children. *Early Child Dev. Care* 187, 1694–1706. doi: 10.1080/03004430.2016.1185102

Alexander, C., Knezek, G., Christensen, R., Tyler-Wood, T., and Bull, G. (2014). The impact of project-based learning on pre-service teachers' technology attitudes and skills. *J Comput. Math. Sci. Teach.* 33, 257–282.

Allen, A. (2016). Don't fear STEM-you already teach it. *Sch. Age Sch. Exch.* 56–59.

Atiles, J. T., Jones, J. L., and Anderson, J. A. (2013). More than a read-aloud: preparing and inspiring early childhood teachers to develop our future scientists. *Teach. Educ. Pract.* 26, 285–299.

Ball, D. L., Thames, M. H., and Phelps, G. (2008). Content knowledge for teaching: what makes it special? *J. Teach. Educ.* 59, 389–407.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, N. J: Prentice-Hall.

Barnett, J., and Hodson, D. (2001). Pedagogical context knowledge: towards a fuller understanding of what good science teachers know. *Sci. Teach. Educ.* 85, 426–453. doi: 10.1002/sce.1017

Baumgartner, H., and Homburg, C. (1996). Applications of structural equation modelling in marketing and consumer research: a review. *Int. J. Res. Mark.* 13, 139–161.

Bentler, P. M. (1980). Multivariate analysis with latent variables: causal modelling. *Annu. Rev. Psychol.* 31, 419–456.

Bentler, P. M., and Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychol. Bull.* 88, 588–606.

Bers, M. U., Seddighin, S., and Sullivan, A. (2013). Ready for robotics: bringing together the T and E of STEM in early childhood teacher education. *J. Technol. Teach. Educ.* 21, 355–377.

Beswick, K., and Fraser, S. (2019). Developing mathematics teachers' 21stcentury competence for teaching in STEM contexts. *ZDM* 51, 955–965.

Brenneman, K., Lange, A., and Nayfeld, I. (2019). Integrating STEM into preschool education; designing a professional development model in diverse settings. *Early Childhood Educ. J.* 47, 15–28.

Briscoe, C., and Peters, J. (1997). Teacher collaboration across and within schools: supporting individual change in elementary science teaching. *Sci. Educ.* 81, 51–65.

Brophy, S., Klein, S., Portsmore, M., and Rogers, C. (2008). Advancing engineering education in P-12 classrooms. *J. Eng. Educ.* 97, 369–387. doi: 10.1002/j.2168-9830.2008.tb00985.x

Browne, M. W., and Cudeck, R. (1993). "Alternative ways of assessing model fit" in *Testing Structural Equation Models*. eds. K. A. Bollen and J. S. Long (Beverly Hills, CA: Sage), 136–162.

Bryan, L. A., Moore, T. J., Johnson, C. C., and Roehrig, G. H. (2016). "Integrated STEM education" in *STEM Road Map: A Framework for Integrated STEM Education*. eds. C. C. Johnson, E. E. Peters-Burton and T. J. Moore (New York: Routledge).

Büyüköztürk, Ş., Kılıç Çakmak, E., Akgün, Ö.E., and Karadeniz, Ş. ve Demirel, F. (2011). *Bilimsel Araştırma Yöntemleri*. Ankara: Pegem Yayıncılık.

Bybee, R. W. (2010). What is STEM education? *Science* 329:996.

Bybee, R. W. (2013). *The Case of STEM Education: Challenges and Opportunities*. Arlington, VA: NSTA Press.

Çetin, A., and Kahyaoğlu, M. (2018). STEM temelli etkinliklerin fen bilgisi öğretmen adaylarının fen, matematik, mühendislik ve teknoloji ile 21. yüzyıl becerilerine yönelik tutumlarına etkisi. *Ekev Akademi Dergisi* 75, 15–28.

Chai, C. S., Jong, M., Yin, H. B., Chen, M., and Zhou, W. (2019). Validating and modelling 'teachers' technological pedagogical content knowledge for integrative science, technology, engineering and mathematics education. *J. Educ. Technol. Soc.* 22, 61–73.

Chai, C. S., Rahmawati, Y., and Jong, M. S.-Y. (2020). Indonesian science, mathematics, and engineering preservice 'teachers' experiences in STEM-TPACK design-based learning. *Sustainability* 12:9050. doi: 10.3390/su12219050

Chalmers, C., Carter, M. L., Cooper, T., and Nason, R. (2017). Implementing 'big ideas' to advance the teaching and learning of science, technology, engineering, and mathematics (STEM). *Int. J. Sci. Math. Educ.* 15, 25–43.

Chanunan, S. (2021). Enhancıng preservıce STEM teachers' STEM PCK and teachıng self-efficacy through stem pck-based course wıth the uses of experıentıal learnıng coupled wıth worked example ınstructıonal prıncıples. *J. Educ. Naresuan Unv.* 23, 45–73.

Clements, D. H., and Sarama, J. (2016). Math, science, and technology in the early grades. *Futur. Child.* 26, 75–90.

Cohrssen, C., and Page, J. (2016). Articulating a rights-based argument for mathematics teaching and learning in early childhood education. *Australas. J. Early Childhood* 41, 104–108.

Cooper, R., and Heaverlo, C. (2013). Problem solving and creativity and design: what influence do they have on girls' interest in STEM subject areas? *Am. J. Eng. Educ.* 4, 27–38.

Çorlu, M. S., Capraro, R. M., and Çorlu, M. A. (2015). Investigating the mental readiness of pre-service teachers for ıntegrated teaching. *Int. Online J. Educ. Sci.* 7, 17–28.

Cunningham, C. M., and Higgins, M. (2015). Engineering for everyone. *Educ. Leadersh.* 72, 42–47.

Dede, C. (2010). "Comparing frameworks for 21st century skills" in *21st Century Skills: Rethinking How Students Learn*. eds. J. Bellanca and R. Brandt (Bloomington, IN: Solution Tree Press), 51–75.

Dorouka, P., Papadakis, S., and Kalogiannakis, M. (2020). Tablets and apps for promoting robotics, mathematics, STEM education, and literacy in early childhood education. *Int. J. Mobile Learn. Organ.* 14, 255–274.

Early Childhood STEM Working Group. (2017). *Early STEM Matters: Providing High-Quality STEM Experiences for all Young Learners. Policy Report*. Chicago, IL: Erikson Institute.

El-Deghaidy, H., and Mansour, N. (2015). Science 'teachers' perceptions of STEM education: possibilities and challenges. *Int. J. Learn. Teach.* 1, 51–54. doi: 10.18178/ijlt.1.1.51-54

Eng Tek, O. N. G., Ayob, A., Ibrahim, M. N., Adnan, M., Shariff, J., and Ishak, N. (2016). The effectiveness of in-service training of early childhood teachers on STEM integration through project-based inquiry learning (PIL). *J. Turkish Sci. Educ.* 13, 44–58. doi: 10.12973/tused.10170a

Ertuğrul Akyol, B. (2020). *STEM Etkinliklerinin fen Bilgisi Öğretmen Adaylarının bilgi İşlemsel, Eleştirel, Yaratıcı Düşünme ve Problem Çözme Becerilerine Etkisi*. (Yayımlanmamış Doktora Tezi). Erciyes Üniversitesi, Eğitim Bilimleri Enstitüsü, Kayseri.

Eshach, H., and Fried, M. N. (2005). Should science be taught in early childhood? *J. Sci. Educ. Technol.* 14, 315–336.

Faikhamta, C., Lertdechapat, K., and Prasoblarb, T. (2020). The impact of a PCK-based professional development program on science teachers' ability to teaching STEM. *J. Sci. Math. Educ. Southeast Asia* 43, 1–22.

Fore, G. A., Feldhaus, C. R., Sorge, B. H., Agarwal, M., and Varahramyan, K. (2015). Learning at the nano-level: accounting for complexity in the internalization of secondary STEM teacher professional development. *Teach. Teach. Educ.* 51, 101–112.

George, D., and Mallery, P. (2003). *SPSS for Windows Step by Step: A Simple Guide and Reference. 11.0 Update. 4th Edn.* Boston: Allyn & Bacon.

Gkontelos, A., Vaiopoulou, J., and Stamovlasis, D. (2021). Teachers' irrational belief scale: psychometric properties of the Greek version and measurement invariance across genders. *Behav. Sci.* 11:160.

Gkontelos, A., Vaiopoulou, J., and Stamovlasis, D. (2022). Teachers' innovative work behavior scale: psychometric properties of the Greek version and measurement invariance across genders. *Soc. Sci.* 11:306.

Gonzalez, M., and Freyer, C. (2014). A collaborative initiative: STEM and universally designed curriculum for at-risk preschoolers. *Natl Teach. Educ. J.* 7, 21–29.

Gözüm, A. İ. C. (2021). Blog Design in the Scope of Community Service in the Pandemic Process: An Investigation of Pre-Service Pre-School Teachers' Use of Information and Communication Technologies. 8th International Conference on Instructional Technology and Teacher Education, Trabzon University, Trabzon. Available at: https://ittes.org.tr/dosyalar/files/ittes2021_ozetler_v4_1.pdf

Gözüm, A. İ. C. (2022). "Digital games for STEM in early childhood education: active co-playing parental mediation and educational content examination" in *STEM, Robotics, Mobile Apps in Early Childhood and Primary Education. Lecture Notes in Educational Technology*. eds. S. Papadakis and M. Kalogiannakis (Singapore: Springer).

Gözüm, A. İ. C., and Kandır, A. (2021). Digital games pre-schoolers play: parental mediation and examination of educational content. *Educ. Inf. Technol.* 26, 3293–3326.

Grossman, P. L. (1990). *The Making of a Teacher: Teacher Knowledge and Teacher Education.* New York: Teachers College Press.

Gürbüz, S. (2019). *AMOS Ile Yapısal Eşitlik Modellemesi*. Ankara: Seçkin Yayıncılık.

Gürbüz, S., and Şahin, F. (2018). Sosyal Bilimlerde Araştırma Yöntemleri Felsefe-Yöntem-Analiz. Seçkin Yayıncılık.

Harris, J. B., and Hofer, M. J. (2011). Technological pedagogical content knowledge (TPACK) in action: a descriptive study of secondary teachers' curriculum-based, technology-related instructional planning. *J. Res. Technol. Educ.* 43, 211–229. doi: 10.1080/15391523.2011.10782570

Havu-Nuutinen, S., Kewalramani, S., Veresov, N., Pöntinen, S., and Kontkanen, S. (2022). Understanding early childhood science education: comparative analysis of Australian and Finnish curricula. *Res. Sci. Educ.* 52, 1093–1108.

Hedlin, M., and Gunnarsson, G. (2014). Preschool student teachers, technology, and gender: positive expectations despite mixed experiences from their school days. *Early Child Dev. Care* 184, 1948–1959.

Holmlund, T. D., Lesseig, K., and Slavit, D. (2018). Making sense of "STEM education" in K 12 contexts. *Int. J. STEM Educ.* 5, 1–18.

Hooper, D., Coughlan, J., and Mullen, M. R. (2008). Structural equation modelling: guidelines for determining model fit. *Electron. J. Bus. Res. Methods* 6, 53–60.

Howland, J., Jonassen, D., and Marra, R. (2012). *Meaningful Learning with Technology. 4th* Edn, NJ: Pearson.

Hsu, M. C., Purzer, S., and Cardella, M. E. (2011). Elementary teachers' views about teaching design, engineering, and technology. *J. Pre Coll. Eng. Educ. Res.* 1:5. doi: 10.5703/1288284314653

Hudson, P., English, L., Dawes, L., King, D., and Baker, S. (2015). Exploring links between pedagogical knowledge practices and student outcomes in STEM education for primary schools. *Aust. J. Teach. Educ.* 40, 134–151. doi: 10.3316/informit.277020514111390

Israsena, V., Dejakaisaya, P., and Santhanakul, P. (2016). Introducing STEM to early childhood education. *Thammasat Int. J. Sci. Technol.* 4, 22–26.

Jipson, J. L., Callanan, M. A., Schultz, G., and Hurst, A. (2014). "Scientists not sponges: STEM interest and inquiry in early childhood" in *Ensuring STEM Literacy*. eds. J. G. Manning, J. B. Jensen, M. K. Hemenway and M. G. Gibbs (San Francisco: Astronomical Society of the Pacific), 149–156.

John, M. S., Sibuma, B., Wunnava, S., Anggoro, F., and Dubosarsky, M. (2018). An iterative participatory approach to developing an early childhood problem-based STEM curriculum. *Eur. J. STEM Educ.* 3, 1–12.

Johnson, C. C. (2012). Implementation of STEM education policy: challenges, progress, and lessons learned. *Sch. Sci. Math.* 112, 45–55.

Johnson, B., and Christensen, L. (2014). *Educational Research: Quantitative, Qualitative, and Mixed Approaches*, Lose Angeles: SAGE Publications.

Karademir-Coşkun, T., Alakurt, T., and Yılmaz, B. (2020). STEM education from the perspective of ınformatıon technologies teachers. *Abant İzzet Baysal Univ. J. Fac. Educ.* 20, 820–836. doi: 10.17240/aibuefd.2020-536856

Kaya, V. H., and Elster, D. (2019). Environmental science, technology, engineering, and mathematics pedagogical content knowledge: teacher's professional development as environmental science, technology, engineering, and mathematics literate individuals in the light of experts' opinions. *Sci. Educ. Int.* 30, 11–20.

Kennedy, M. M., Ahn, S., and Choi, J. (2008). The value added by teacher education. In M. Cochran-Smith, S. Feiman-Nemser and J. McIntyre (with K. E. Demers) (Eds.), *Handbook of Research on Teacher Education: Enduring Issues in Changing Contexts* (3rd Edn.). New York: Routledge, pp. 1249–1273.

Kennedy, T. J., and Odell, M. R. L. (2014). Engaging students in STEM education. *Sci. Educ. Int.* 25, 246–258.

Kershaw, P., Anderson, L., and Warburton, B. (2009). *15 by 15 A Comprehensive Policy Framework for Early Human Capital Investment*. Vancouver, BC: Human Early Learning Partnership.

King, C. M., Lancaster, P. E., Defrance, N. L., Melin, J. A., and Cleveland, R. (2013). Curricular and pedagogical imperatives for fostering effective stem teacher preparation. *Teach. Educ. Pract.* 26, 706–721.

Koyunlu-Ünlü, Z., and Dere, Z. (2019). Okul öncesi öğretmen adaylarının FeTeMM farkındalıklarının değerlendirilmesi. *Erzincan Üniversitesi Eğitim Fakültesi Dergisi* 21, 44–55.

Lamb, R. L., Akmal, T. T., and Petrie, K. (2015). Development of a cognition-priming model describing learning in a STEM classroom. *J. Res. Sci. Teach.* 52, 410–437.

Lichtenberger, E., and George-Jackson, C. (2013). Predicting high school students' interest in majoring in a STEM field: insight into high school students' postsecondary plans. *J. Career Tech. Educ.* 28, 19–38. doi: 10.21061/jcte.v28i1.571

Ling, L. S., Pang, V., and Lajium, D. (2020). A case study of teachers' pedagogical content knowledge in the implementation of integrated STEM education. *J. Pendidikan Sains Dan Matematik Malaysia* 10, 49–64.

Ludwig, M., Marklein, M. B., and Song, M. (2016). Arts integration: a promising approach to improving early learning. *Am. Inst. Res.* 5, 1–14.

MacFarlane, B. (2016). "Infrastructure of comprehensive STEM programming for advanced learners" in *STEM education for high-ability learners designing and implementing programming*. ed. B. MacFarlane (Waco, TX:: Prufrock Press), 139–160.

Margot, K. C., and Kettler, T. (2019). Teachers' perception of STEM integration and education: a sySTEMatic literature review. *Int. J. STEM Educ.* 6, 1–16. doi: 10.1186/s40594-018-0151-2

Marsh, H. W., Hau, K. T., Artelt, C., Baumert, J., and Peschar, J. L. (2006). OECD's brief self-report measure of educational psychology's most useful affective constructs: cross-cultural, psychometric comparisons across 25 countries. *Int. J. Test.* 6, 311–360.

Marsono, M., Khasanah, F., and Yoto, Y. (2019). "Integrating STEM (science technology engineering and mathematics)" in *Education on Advancing Vocational Student's Creative Thinking Skills. 2nd International Conference on Vocational Education and Training (ICOVET 2018)* (Atlantis Press), 170–173.

McDonald, C. V. (2016). STEM education: a review of the contribution of the disciplines of science, technology, engineering, and mathematics. *Sci. Educ. Int.* 27, 530–569.

Mercan, Z., and Kandır, A. (2022). The effect of the early STEAM education program on the visual-spatial reasoning skills of children: research from Turkey. *Education* 3-13, 1–31.

Mishra, P., and Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108, 1017–1054.

Moomaw, S., and Davis, J. A. (2010). STEM comes to preschool. *Young Child.* 65, 12–18.

Moore, T. J., Glancy, A. W., Tank, K. M., Kersten, J. A., Smith, K. A., and Stohlmann, M. S. (2014). A framework for quality K-12 engineering education: Research and development. *J. Precoll. Eng. Educ. Res.* 4, 1–13.

Moore, T. J., and Smith, K. A. (2014). Advancing the state of the art of STEM integration. *J. STEM Educ.* 15, 5–10.

Munro, B. H. (2005). *Statistical Methods for Health Care Research*. Philadelphia: Lippincott Williams & Wilkins.

Nathan, M. J., Atwood, A. K., Prevost, A., Phelps, L. A., and Tran, N. A. (2011). How professional development in project Lead the way changes high school STEM teachers' beliefs about engineering education. *J. Precoll. Eng. Educ. Res.* 1, 15–29.

National Research Council [NCR]. (1994). *National Science Education Standards*. Washington: National Academy Press.

National Research Council [NRC]. (2014). *STEM Integration in K-12 Education: Status, Prospects, and an Agenda for Research*. Washington, DC: National Academies Press.

Nikolopoulou, K. (2021). Mobile devices in early childhood education: teachers' views on benefits and barriers. *Educ. Inf. Technol.* 26, 3279–3292.

Nikolopoulou, K. (2022a). STEM activities for children aged 4–7 years: teachers' practices and views. *Int. J. Early Years Educ.*, 1–16. doi: 10.1080/09669760.2022.2128994

Nikolopoulou, K. (2022b). "Digital Technology in Early STEM education: exploring its supportive role" in *STEM, Robotics, Mobile Apps in Early Childhood and Primary Education. Lecture Notes in Educational Technology*. eds. S. Papadakis and M. Kalogiannakis (Singapore: Springer), 103–115.

OECD (2019). *PISA 2018 Results (Volume I): What Students Know and Can Do*, PISA, OECD Publishing, Paris.

Ostler, E. (2012). 21st century STEM education: a tactical model for long-range success. *Int. J. Appl. Sci. Technol.* 2, 28–33.

Papadakis, S., Kalogiannakis, M., and Zaranis, N. (2018). The effectiveness of computer and tablet-assisted intervention in early childhood students' understanding of numbers. An empirical study conducted in Greece. *Educ. Inf. Technol.* 23, 1849–1871.

Papadakis, S., Kalogiannakis, M., and Zaranis, N. (2021b). Teaching mathematics with mobile devices and the realistic mathematical education (RME) approach in kindergarten. *Adv. Mobile Learn. Educ. Res.* 1, 5–18.

Papadakis, S., Vaiopoulou, J., Sifaki, E., Stamovlasis, D., and Kalogiannakis, M. (2021a). Attitudes towards the use of educational robotics: exploring pre-service and in-service early childhood teacher profiles. *Educ. Sci.* 11:204.

Papadakis, S., Zaranis, N., and Kalogiannakis, M. (2019). Parental involvement and attitudes towards young Greek children's mobile usage. *Int. J. Child Comput. Interact.* 22:100144

Park, M.-H., Dimitrov, D. M., Patterson, L. G., and Park, D.-Y. (2017). Early childhood teachers' beliefs about readiness for teaching science, technology, engineering, and mathematics. *J. Early Child. Res.* 15, 275–291. doi: 10.1177/1476718X15614040

Penprase, B. E. (2020). *STEM Education for the 21st Century*, Switzerland: Springer.

Perignat, E., and Katz-Buonincontro, J. (2019). STEAM in practice and research: an integrative literature review. *Think. Skills Creat.* 31, 31–43.

Quigley, C. F., and Herro, D. (2016). 'Finding the joy in the unknown': implementation of STEAM teaching practices in middle school science and math classrooms. *J. Sci. Educ. Technol.* 25, 410–426.

Rahman, S. M., Chacko, S. M., and Kapila, V. (2017). Building Trust in Robots in Robotics-Focused STEM Education under TPACK Framework in Middle Schools. In the 2017 ASEE Annual Conference & Exposition. (25–28).

Rahman, N. A., Rosli, R., Rambely, A. S., Siregar, N. C., Capraro, M. M., and Capraro, R. M. (2022). Secondary school teachers' perceptions of STEM pedagogical content knowledge. *J. Math. Educ.* 13, 119–134.

Reimers, J. E., Farmer, C. L., and Klein-Gardner, S. S. (2015). An introduction to the standards for preparation and professional development for teachers of engineering. *J. Precoll. Eng. Educ. Res.* 5, 40–60.

Saxton, E., Burns, R., Holveck, S., Kelley, S., Prince, D., Rigelman, N., et al. (2014). A common measurement system for K-12 STEM education: adopting an educational evaluation methodology that elevates theoretical foundations and systems thinking. *Stud. Educ. Eval.* 40, 18–40.

Schuster, D., Buckwalter, J., Marrs, K., Pritchett, S., Sebens, J., and Hiatt, B. (2012). Aligning university-based teacher preparation and new STEM teacher support. *Sci. Educ.* 21, 39–44.

Sharapan, H. (2012). From STEM to STEAM: how early childhood educators can apply Fred Rogers' approach. *YC Young Child.* 67, 36–40.

Shernoff, D. J., Sinha, S., Bressler, D. M., and Ginsburg, L. (2017). Assessing teacher education and professional development needs for the implementation of integrated approaches to STEM education. *Int. J. STEM Educ.* 4, 1–16.

Shulman, L. S. (1986). Those who understand: knowledge growth in teaching. *Educ. Res.* 15, 4–14. doi: 10.3102/0013189X015002004

Silva, R., Bernardo, C. D. P., Watanabe, C. Y. V., Silva, R. M. P. D., and Neto, J. M. D. S. (2020). Contributions of the internet of things in education as a support tool in the educational management decision-making process. *Int. J. Innov. Learn.* 27, 175–196.

Şimsek, O. F. (2007). *Yapisal Esitlik Modellemesine Giris: Temel Ilkeler ve LISREL Uygulamalari*. Ankara: Ekinoks.

Smith, J., and Karr-Kidwell, P. (2000). The Interdisciplinary Curriculum: A Literary Review and a Manual for Administrators and Teachers. Retrieved from ERIC Database. (ED443172). https://files.eric.ed.gov/fulltext/ED443172.pdf

Srikoom, W., Hanuscin, D., and Faikhamta, C. (2017). Perceptions of in-service teachers toward teaching STEM in Thailand. *Asia Pac. Forum Sci. Lear. Teach.* 18, 1–23.

Stohlmann, M., Moore, T. J., and Roehrig, G. H. (2012). Considerations for teaching integrated STEM education. *J. Precoll. Eng. Educ. Res.* 2:4.

Stohlmann, M. S., Roehrig, G. H., and Moore, T. J. (2014). "The need for STEM teacher education development" in *STEM education: How to train 21st century teachers* (Nova Science Publishers, Inc), 17–32.

Sullivan, A., and Bers, M. U. (2016). Robotics in the early childhood classroom: learning outcomes from an 8-week robotics curriculum in pre-kindergarten through second grade. *Int. J. Technol. Des. Educ.* 26, 3–20. doi: 10.1007/s10798-015-9304-5

Sullivan, A., and Bers, M. U. (2018). Dancing robots: integrating art, music, and robotics in Singapore's early childhood centres. *Int. J. Technol. Des. Educ.* 28, 325–346.

Tank, K. M., Rynearson, A. M., and Moore, T. J. (2018). Examining student and teacher talk within engineering Design in Kindergarten. *Eur. J. STEM Educ.* 3:10.

Thomas, B., and Watters, J. J. (2015). Perspectives on Australian, Indian, and Malaysian approaches to STEM education. *Int. J. Educ. Dev.* 45, 42–45.

Torres-Crospe, M. N., Kraatz, K., and Pallansch, L. (2014). From fearing STEM to playing with it: the natural integration of STEM into the preschool classroom. *SRATE J.* 23, 8–16.

Türk, N., Kalaycı, N., and Yamak, H. (2018). New trends in higher education in the globalizing world: stem in teacher education. *Univ. J. Educ. Res.* 6, 1286–1304.

Ültay, N., and Ültay, E. (2020). A comparative investigation of the views of preschool teachers and teacher candidates about STEM. *J. Sci. Learn.* 3, 67–78.

Van Hoorn, J., Monighan, P., Scales, B., and Rodriguez, K. (2011). *Play at the Centre of the Curriculum. 5th Edn.* Boston: Pearson.

Voogt, J., and Roblin, N. P. (2012). A comparative analysis of international frameworks for 21st century competences: implications for national curriculum policies. *J. Curric. Stud.* 44, 299–321.

Wang, Z., and Fan, R. (2018). Analysis of TPCK knowledge structure of STEM teachers in middle school. *J. Guangxi Teach. Educ. Univ.* 54, 105–111. doi: 10.16088/j.issn.1001-6597.2018.02.015

Wang, H. H., Moore, T. J., Roehrig, G. H., and Park, M. S. (2011). STEM integration: teacher perceptions and practice. *J. Precoll. Eng. Educ. Res.* 1, 1–13. doi: 10.5703/1288284314636

Weng, X., Jong, M. S. Y., and Chıu, T. K. F. (2020). "Implementation challenges of STEM education: from teachers' perspective," in *Proceedings of the 28th International Conference on Computers in Education (ICCE 2020), Vol. 1,* 683–685.

Whitebook, M., and Ryan, S. (2011). Degrees in Context: Asking the Right Questions about Preparing Skilled and Effective Teachers of Young Children. Available at: http://www.nieer.org/resources/policybriefs/23.pdf

Williams, R. (2016). STEM Professional Development: What's Going on from the Presenters' and Participants' Perspectives? (Order No. 10103407). Available from ProQuest Central. (1785367212). Available at: https://www.proquest.com/docview/1785367212?parentSessionId=SXjJneaUGK9HJoDixcebiwFAD%2FnbMru%2Fh%2FLcdUijwqI%3D

Yıldırım, B. (2021). Preschool STEM activities: preschool teachers' preparation and views. *Early Childhood Educ. J.* 49, 149–162.

Yıldırım, B., and Şahin-Topalcengiz, E. (2019). STEM pedagogical content knowledge scale (STEMPCK): a validity and reliability study. *J. STEM Teach. Educ.* 53, 1–20.

Yıldırım, B., and Türk, C. (2018). Pre-service primary school 'teachers' views about STEM education: an applied study. *Trakya J. Educ.* 8, 195–213. doi: 10.24315/trkefd.310112

Yılmaz, K. G. (2019). *Fen bilgisi öğretmen adaylarının fen, teknoloji, mühendislik ve matematik (STEM) alanlarına yönelik ilgi düzeyleri.* (Yayımlanmamış Yüksek Lisans Tezi), Bartın Üniversitesi, Eğitim Bilimleri Enstitüsü Bartın.

Yusof, Y. M., Zakaria, E., and Maat, S. M. (2012). Teachers' general pedagogical content knowledge (PCK) and content knowledge of algebra. *Soc. Sci.* 7, 668–672.

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 12 December 2022
DOI 10.3389/fpsyg.2022.1082659

Check for updates

OPEN ACCESS

EDITED BY
Stamatios Papadakis,
University of Crete, Greece

REVIEWED BY
Özgen Korkmaz,
Amasya University, Turkey
Janika Leoste,
Tallinn University, Estonia

*CORRESPONDENCE
Laila El-Hamamsy
laila.elhamamsy@epfl.ch

SPECIALTY SECTION
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

RECEIVED 28 October 2022
ACCEPTED 15 November 2022
PUBLISHED 12 December 2022

CITATION
El-Hamamsy L, Zapata-Cáceres M, Marcelino P, Bruno B, Dehler Zufferey J, Martín-Barroso E and Román-González M (2022) Comparing the psychometric properties of two primary school Computational Thinking (CT) assessments for grades 3 and 4: The Beginners' CT test (BCTt) and the competent CT test (cCTt). *Front. Psychol.* 13:1082659. doi: 10.3389/fpsyg.2022.1082659

COPYRIGHT
© 2022 El-Hamamsy, Zapata-Cáceres, Marcelino, Bruno, Dehler Zufferey, Martín-Barroso and Román-González. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Comparing the psychometric properties of two primary school Computational Thinking (CT) assessments for grades 3 and 4: The Beginners' CT test (BCTt) and the competent CT test (cCTt)

Laila El-Hamamsy[1,2]*, María Zapata-Cáceres[3], Pedro Marcelino[4], Barbara Bruno[5], Jessica Dehler Zufferey[2], Estefanía Martín-Barroso[3] and Marcos Román-González[6]

[1]MOBOTS Group, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, [2]LEARN - Center for Learning Sciences, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, [3]Laboratory of Information Technologies in Education, Rey Juan Carlos University, Madrid, Spain, [4]Computational Thinking Department, TreeTree2 (T2), Lisbon, Portugal, [5]CHILI Laboratory, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, [6]Faculty of Education, Universidad Nacional de Educación a Distancia (UNED), Madrid, Spain

Introduction: With the increasing amount of research around Computational Thinking (CT) and endeavors introducing CT into curricula worldwide, assessing CT at all levels of formal education is of utmost importance to ensure that CT-related learning objectives are met. This has contributed to a progressive increase in the number of validated and reliable CT assessments for K-12, including primary school. Researchers and practitioners are thus required to choose among multiple instruments, often overlapping in their age validity.

Methods: In this study, we compare the psychometric properties of two of these instruments: the Beginners' CT test (BCTt), developed for grades 1–6, and the competent CT test (cCTt), validated for grades 3–4. Classical Test Theory and Item Response Theory (IRT) were employed on data acquired from 575 students in grades 3–4 to compare the properties of the two instruments and refine the limits of their validity.

Results: The findings (i) establish the detailed psychometric properties of the BCTt in grades 3–4 for the first time, and (ii) through a comparison with students from the same country, indicate that the cCTt should be preferred for grades 3–4 as the cCTt is able to discriminate between students of low and medium ability. Conversely, while the BCTt, which is easier, shows a ceiling effect, it is better suited to discriminate between students in the low ability range. For these grades, the BCTt can thus be employed as a screening mechanism to identify low ability students.

Discussion: In addition to providing recomendations for use of these instruments, the findings highlight the importance of comparing the psychometric properties of existing assessments,

so that researchers and practitioners, including teachers and policy makers involved in digital education curricular reforms, may take informed decisions when selecting assessments.

KEYWORDS

Computational Thinking, assessment, primary school, validation, developmental appropriateness, psychometrics

1. Introduction and related work

Computational Thinking (CT) is more and more often considered to be an essential twenty-first century skill (Li et al., 2020), that is as important as reading, writing, and arithmetic (Wing, 2006) and must be taught at a young age. Despite the lack of consensus regarding the definition of CT, CT is traditionally defined by Wing (2006) as "an approach to solving problems, designing systems, and understanding human behavior that draws on concepts fundamental to computing" which was later reformulated by Aho (2012) as "the thought processes involved in formulating problems so their solutions can be represented as computational steps and algorithms." As such CT has often been associated with Computer Science (CS), although many researchers consider CT to be transversal (Mannila et al., 2014; Weintrop, 2016; Denning and Tedre, 2021; Weintrop et al., 2021b), and not exclusively related to CS or mathematics (Li et al., 2020). This has lead to a "tremendous growth in curricula, learning environments, and innovations around CT education" (Weintrop et al., 2021b). To be successful, these initiatives rely on the constructive alignment between the learning objectives, teaching and learning activities, and *assessments* (Biggs, 1996). Developing and implementing effective CT interventions thus requires expanding the portfolio of developmentally appropriate instruments to assess CT at all levels of formal education, for use by researchers and educators alike (Weintrop et al., 2021a).

Developing CT assessments requires having better insight into what composes this competence, with a competence referring to "the proven ability to use knowledge, skills, and personal, social, and/or methodological abilities, in work or study situations and in professional and personal development" (European Union, 2006). As such, Brennan and Resnick (2012) proposed an operational definition of CT by decomposing CT into three dimensions. The first is CT-concepts, i.e., "the concepts designers engage with as they program, such as iteration, parallelism," (Brennan and Resnick, 2012), which thus includes sequences, loops, if-else statements and so forth at the primary school level. These elements can be adequately assessed through diagnostic and summative tools (Román-González et al., 2019). The second is CT-practices i.e., "the practices designers develop as they engage with the concepts, such as debugging projects or remixing others' (Brennan and Resnick, 2012), which thus requires understanding the thought processes involved in resolving CT problems. These may include elements of abstraction, decomposition, evaluation, and so forth and can be adequately assessed through formative-iterative tools and data-mining tools (Román-González et al., 2019). The third is CT-perspectives, i.e., "the perspectives designers form about the world around them and about themselves" (Brennan and Resnick, 2012), and therefore their perception of CT which can be adequately evaluated through perception and attitude scales and vocabulary assessments (Román-González et al., 2019).

Despite the increase in research around CT in the past two decades, and the various means of assessing CT identified by Tang et al. (2020) [i.e., "traditional test(s) composed of selected- or constructed response questions, portfolio assessment(s), interviews, and surveys"], few validated and reliable instruments exist for CT, and even less at the primary school level (Román-González et al., 2019; Basu et al., 2020; Zapata-Cáceres et al., 2020; Clarke-Midura et al., 2021). This limitation was highlighted by Tang et al. (2020) in their recent meta review on CT assessments: out of 96 studies, only 45% provided reliability evidence and just 18% provided validity evidence. This mirrors the findings of Bakala et al. (2021) who, in their literature review on the effects of robots on preschool children's CT, found that most studies employed *ad-hoc* evaluations, typically neither standardized nor validated. Bakala et al. (2021) attributed this to the fact that only two recent valid and reliable tests for that age group existed at the time of their review [the TechCheck by Relkin et al., 2020; Relkin and Bers, 2021 and the Beginners' CT test (BCTt) by Zapata-Cáceres et al., 2020] and recommended that researchers aim to employ them in future studies. To further limit the available choices, many existing assessments are strongly tied to specific CS frameworks (Rowe et al., 2021) [e.g., Dr., Scratch (Moreno-León and Robles, 2015) or the Fairy assessments (Werner et al., 2012)]. As stated by Relkin and Bers (2021) and Rowe et al. (2021), being strongly tied to specific frameworks means that the instrument risks conflating with programming abilities. This contributes to a lack of generalizability and thus limits the range of applications of such instruments (Tikva and Tambouris, 2021), which for example should be avoided in the context of pre-post test experimental designs. It is essential to provide researchers and practitioners

(e.g., teachers and policy makers involved in digital education curricular reforms) the means to assess CT:

1. at all levels of education
2. independently from specific studies or programming environments
3. in a valid and reliable way to ensure that there is sufficient "evidence and theory [to] support the interpretations of test scores entailed by proposed uses of tests" (Clarke-Midura et al., 2021)
4. with an instrument which can easily be administered.

Without these, it is not possible to ensure that CT-related learning objectives are met, whether in individual interventions or in the context of large scale CS and/or CT curricular reform initiatives (El-Hamamsy et al., 2021a,b).

Unfortunately, while an increasing number of instruments have been recently developed, several do not meet these criteria (Hubwieser and Mühling, 2014; Bellettini et al., 2015; Gane et al., 2021; Parker et al., 2021). For example, the Bebras challenge is sometimes used to assess CT skills, but has undergone limited psychometric validation (Hubwieser and Mühling, 2014; Bellettini et al., 2015). Gane et al. (2021)'s assessment require manual grading and multiple annotators, thus limiting the test's scalability and its usability by other researchers and practitioners. Parker et al. (2021) assessment which is based on a combination of block-based and Bebras-style questions, has been piloted with just 57 fourth graders. Finally Chen et al. (2017)'s assessment for 5th graders appears highly dependent on the robotics programming context, includes open questions and was administered to just 37 students, thus including the limitations of all the aforementioned assessments, in addition to limiting its use in other CT-related contexts.

Instruments meeting the aforementioned criteria, and having undergone a psychometric validation and reliability assessment process at the level of primary school (see section 2.2), include the TechCheck for lower primary school (grades 1–2, ages 6–8, Relkin et al., 2020), the TechCheck-K, which is an adaptation of the former for kindergarden (ages 4–6, Relkin and Bers, 2021), the BCTt for grades 1–6 (ages 5–10, Zapata-Cáceres et al., 2020), the competent CT test (cCTt) for grades 3–4 (ages 7–9, El-Hamamsy et al., 2022a), the Computational Thinking Assessment for Chinese Elementary Students (CTA-CES) for grades 3–6 (ages 9–12, Li et al., 2021), and Kong and Lai (2022)'s CT-concepts test for grades 3–5. A synthesis of these instruments is provided in Table 1 and shows that these instruments often differ in the underlying definition of CT employed to define the test items which makes it complex to compare them pyschometrically. Furthermore, these instruments are all relatively new and adopt an unplugged approach, using multiple choice questions to assess primary school students' CT abilities. Furthermore, there is an overlap in their target age ranges. It is thus important for researchers and practitioners to not only identify instruments that best assess the learning objectives of their interventions, but also to understand the limits of validity of these instruments to make informed decisions for their own studies. Such instruments are unfortunately not often compared against one another to determine which may be more appropriate for a given age range. To the best of our knowledge, only the TechCheck and TechCheck-K were compared to establish whether the TechCheck-K would be an adequate instrument for kindergarden students (Relkin and Bers, 2021), with the TechCheck being more appropriate for first and second graders.

In this paper, we are interested in the overlap between the BCTt and the cCTt for students in grades 3 and 4 as these two instruments overlap in their targets, and are from the same "family" of CT tests, and thus cover the same concepts. Therefore, the BCTt and cCTt cannot be considered complementary within a system of assessments, and thus require choosing between them. It is therefore essential to establish their limits of validity for the considered age group to provide recommendations to help researchers make an informed decision when selecting CT-assessments in accordance with their study requirements. Indeed, while the BCTt was initially developed as an instrument looking to cover all of primary school, the validation procedure appeared to indicate that the BCTt was too easy for students in upper primary school (Zapata-Cáceres et al., 2020). As the cCTt was derived from the BCTt to adapt the instrument in terms of format and content to improve its validity for students in grades 3 and 4 (7–9 year old students), the present study therefore investigates how the BCTt and cCTt complement each other in assessing CT in grades 3 and 4, to propose recommendations for their use for these grades. More specifically, we look to answer the following research questions:

1. *How do the psychometric properties of the BCTt and the cCTt compare for students in grades 3–4 (7–9 years old)?*
2. *How does the psychometric comparison inform us about how the instruments should be used in grades 3–4 (7–9 years old)?*

2. Methodology

2.1. The BCTt, cCTt, and their validation

The BCTt and the cCTt are two 25-item multiple choice CT assessments[1] of progressive difficulty, targeting the CT-concepts posed by Brennan and Resnick (2012) in their decomposition of CT into concepts, practices, and perspectives. More specifically, the two tests evaluate notions of sequences, simple loops (only one instruction is repeated), complex loops

1 Please note that the full BCTt is available upon request to the co-authors of Zapata-Cáceres et al. (2020), and the cCTt items are presented in El-Hamamsy et al. (2022a) with an editable version available upon request to the co-authors of the article.

TABLE 1 Synthesis of validated and scalable primary school unplugged CT assessments and corresponding validation processes adapted from El-Hamamsy et al. (2022a).

Test	Format	Target age group	CT definition	Validation process	Sample	Validity established for
TechCheck (Relkin et al., 2020) and TechCheck-K (Relkin and Bers, 2021)	15 item MCQ	1st and 2nd graders (6–9 year old students) and kindergarden (5–6 year old students)	Algorithms, Modularity, Design Process, Debugging, Control Structures, Hardware/Software	Expert validation, psychometric analysis (Classical Test Theory and Item Response Theory), convergent validation with the TACTIC-KIBO	768 5–9 year old students participating in a robotics coding curriculum and 89 kindergarden students without coding experience	Full sample
Beginner's CT test (Zapata-Cáceres et al., 2020; Zapata-Cáceres and Fanchamps, 2021)	25 item MCQ	Primary school (5–12 year old students) and Kindergarden (4–5 years old students)	Computational concepts, practices, perspectives (Brennan and Resnick, 2012)	Expert validation, and psychometric analysis (Classical Test Theory)	299 primary school students from grades 1 to 6 and 5 kindergarden students	4–7 year old students
The competent CT test (cCTt) (El-Hamamsy et al., 2022a)	25 item MCQ	Primary school (7–9 year old students)	Computational concepts, practices, perspectives (Brennan and Resnick, 2012)	Expert validation and psychometric analysis (Classical Test Theory, Item Response Theory), Confirmatory Factor Analysis	1,519 primary school students from grades 3 to 4	Full sample
CT Assessment for Chinese Elementary Students (CTA-CES, Li et al., 2021)	25 item MCQ	Grades 3–6 (ages 9–12)	Abstraction, algorithmic thinking, decomposition, evaluation, pattern recognition, generalization (Selby and Woollard, 2013)	Expert validation, Classical Test Theory, Item Response Theory, Construct validity by comparing two groups of students, criterion validity through correlations with reasoning, spatial ability, and verbal ability	280 grade 3–6 students	Full sample
Kong and Lai (2022)'s CT-concepts test	14 item MCQ	Grades 3–5 (ages 8–10)	Sequences, conditionals, repetition (Brennan and Resnick, 2012)	Item Response Theory	13,670 grade 3 to 5 students	Full sample

10.3389/fpsyg.2022.1082659

(two or more instructions are repeated), conditionals and while statements (see the distribution of items in Table 2), with the factor structure pertaining to these concepts having been validated through Confirmatory Factor Analysis by El-Hamamsy et al. (2022a). The BCTt was derived from the CTt (Román-González et al., 2017, 2018, 2019), with changes in terms of format and content to adapt it to primary school. In a similar spirit, the cCTt made alterations to the format and content of the BCTt to more specifically target students in grades 3 and 4 (El-Hamamsy et al., 2022a). Both instruments, like their predecessor the CTt, employ grid-type and canvas-type questions (see Figure 1) and employ the same type of tasks. The individual questions differ (see Table 2) as the cCTt (i) favors questions on 4 × 4 grids, (ii) replaces BCTt questions of low difficulty with questions related to complex concepts (e.g., while statements), (iii) alters the disposition of objects on the grids, and responses, with respect to the BCTt equivalents.

Both the BCTt and cCTt instruments were validated by starting with an evaluation by experts and making adjustments based on their suggestions, prior to administration to students in the target age groups. The BCTt, which was designed for grades 1–6, was administered to 200 students in that age group (Zapata-Cáceres et al., 2020). The authors found that the test had good reliablity with Cronbach's $\alpha = 0.824$. The results indicated that the students improved as they got older, and started to exhibit a ceiling effect in grades 3–4[2]. The results indicated that the differences were significant between all grades, excepted those in grades 4–6 who already exhibit a ceiling effect (Zapata-Cáceres et al., 2020). These results indicate that students begin to exhibit a ceiling effect either in grade 3 or grade 4. The cCTt, which was designed for grades 3 and 4, was administered to 1,519 students in that age group and analyzed through Classical Test Theory and Item Response Theory (El-Hamamsy et al., 2022a). The results indicated that the grade 4 students scored significantly better than the grade 3 students (out of 25 pts, the one-way ANOVA indicates that $p < 0.001$, $\Delta_{grades} = +2.9$ pts, Cohen's $d = 0.57$, $\mu_3 = 12.62 \pm 5.18$, $n = 711$; $\mu_4 = 15.49 \pm 4.96$, $n = 749$). The Classical Test Theory results indicated that the test had good reliability with Cronbach's $\alpha = 0.85$, levels of discrimination, and a wide range of question difficulties. Item Response Theory was employed to support these findings and indicated that the test was better suited at evaluating and discrimination between students with low and medium abilities.

2 In the original validation of the BCTt by Zapata-Cáceres et al. (2020) they obtained average scores of $\mu_{grade1} = 16.52 \pm 3.31$, $\mu_{grade2} = 16.78 \pm 2.49$, $\mu_{grade4} = 21.57 \pm 3.04$, $\mu_{grade5} = 21.84 \pm 2.61$, $\mu_{grade6} = 21.72 \pm 2.62$ out of 25. Please note that their sample did not include grade 3 students.

2.2. Psychometric analysis

The objective of this study is to compare the psychometric properties of the BCTt and cCTt for students in grades 3 and 4. Classical Test Theory and Item Response Theory are two complementary (De Champlain, 2010; Awopeju and Afolabi, 2016) approaches typically employed to analyse the validity and reliability of scales and assessments. The Classical Test Theory and Item Response Theory (IRT) analyses are conducted in R (version 4.2.1, R Core Team, 2019) using the following packages: lavaan (version 0.6-11, Rosseel, 2012), CTT (version 2.3.3, Willse, 2018), psych (version 2.1.3, Revelle, 2021), mirt (version 1.36.1, Chalmers, 2012), and subscore (version 3.3, Dai et al., 2022).

2.2.1. Classical test theory

Classical Test Theory "comprises a set of principles that allow us to determine how successful our proxy indicators are at estimating the unobservable variables of interest" (DeVellis, 2006). Classical test theory focuses on test scores (Hambleton and Jones, 1993) and computes:

- Reliability of the scale using Cronbach's α measurement of internal consistency of scales (Bland and Altman, 1997). In the context of assessments, $0.7 < \alpha < 0.9$ is considered high and $0.5 < \alpha < 0.7$ is considered moderate (Hinton et al., 2014; Taherdoost, 2016). The drop alpha is computed per question as it indicates of the reliability of the test without said question, and thus whether the internal consistency of the test improves without it.
- Item difficulty index, i.e., the proportion of correct responses. *Please note that this means that a question with a high difficulty index is an easy question.* Determining whether questions are too easy or too difficult is often based on arbitrary thresholds which vary around what are considered to be ideal item difficulties. Indeed, some researchers have posited that item difficulties should vary between 0.4 and 0.6 as these are claimed to have maximum discrimination indices (Vincent and Shanmugam, 2020). As such, thresholds employed in the literature have varied around these values, with items being classified as difficult for a range of thresholds between 0.1 and 0.3, and items being classified as easy for a range of thresholds varying between 0.7 and 0.9.

  In this study, to remain coherent with the first cCTt validation in grades 3–4, we consider that questions with a difficulty index above 0.85 are too easy, while those with a difficulty index below 0.25 are too hard and could be revised.
- Point biserial correlation, or item discrimination. This is a measure of discrimination between the high ability examinees and low ability examinees. A point-biserial

TABLE 2 Comparison between the BCTt and the cCTt in terms of question concepts and question types (Table taken from El-Hamamsy et al., 2022a).

	BCTt				cCTt			
Blocks	**Grid** (3 × 3)	**Grid** (4 × 4)	**Canvas**	**Total**	**Grid** (3 × 3)	**Grid** (4 × 4)	**Canvas**	**Total**
Sequences	3	1	2	6	1	1	2	4
Simple loops	3	2	0	5	0	4	0	4
Complex loops	0	5	2	7	0	5	2	7
Conditional statements	1	3	0	4	1	3	0	4
While statements	1	2	0	3	1	3	0	4
Combinations	0	0	0	0	0	2	0	2
Total	8	13	4	25	3	18	4	25

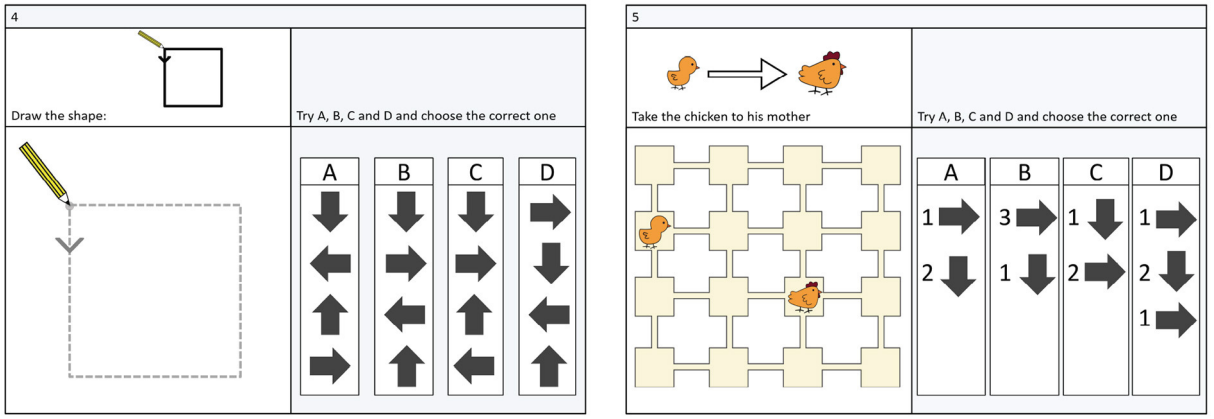


FIGURE 1
The two main question formats of the BCTt and cCTt: grid **(Left)** and canvas **(Right)** (Figure taken from El-Hamamsy et al., 2022a).

correlation above 0.15 is recommended, with good items generally having point biserial correlations above 0.25 (Varma, 2006). In this article, we consider a threshold of 0.2, which is commonly employed in the field (Chae et al., 2019).

Unfortunately, Classical Test Theory suffers from several limitations, including that the analysis is sample-dependent (Hambleton and Jones, 1993). As such, analyzing an instrument from the lens of Classical Test Theory on two different populations may not yield consistent results. The literature thus recommends employing Item Response Theory to complement the results of Classical Test Theory.

2.2.2. Item Response Theory (IRT)

According to Hambleton and Jones (1993), (i) IRT *is sample independent* so scores describing examinee proficiency are not dependent on the test difficulty, (ii) test items can be matched to ability levels, and (iii) the test models do not require strict parallel tests to assess reliability. This is because IRT models the link between a students' latent ability and their probability of correctly answering a question. Indeed, by evaluating the tests' questions with respect to latent ability:

- The results are more likely to be sample independent, and therefore more likely to generalize beyond a specific sample of learners (Xie et al., 2019), thus providing consistency between two different populations.
- Item Response Theory is more adapted to compare multiple assessments through the latent ability scale (Jabrayilov et al., 2016; Dai et al., 2020), and thus including cases where different populations have taken the tests. Comparing two assessments can indeed be done in cases where the instruments measure the same latent traits (Xie et al., 2019), which we believe is possible in the present case because both instruments measure the same CT-concepts, using the same symbols. This can be

verified through Confirmatory Factor Analysis, as done by Kong and Lai (2022).

Item Response Theory models estimate the probability of a person of a given ability (measured in standard deviations from the mean) answering each question correctly. This is visualized through a logistic Item Characteristic Curve (ICC) for each question. As Figure 2A shows, an item's difficulty (b_i) is the x-value (θ) where the ICC reaches a $y = 0.5$ probability of answering correctly, and represents the number of standard deviations from the mean the question difficulty is. Items to the left of the graph are considered easier while items on the right are considered harder. According to De Ayala and Little (2022), "typical item and person locations fall within -3 to +3", with easy items having scores below -2, average items having scores between -2 and +2 and hard items having scores above +2.

Several IRT models exist for binary response data, however given the low sample size (Sahin and Anil, 2017), we focus on one parameter logistic (1-PL) and 2-PL models. While 1-PL models consider that only difficulty varies across items, 2-PL models also take into account that some questions can discriminate more or less well between students of different ability, and thus exhibit varying ICC slopes. In the example in Figure 2B, blue and red items are of equal difficulty b_i ($y = 0.5$ crossing) and relatively similar discrimination a_i, while items green and purple are of equal difficulty and varying discrimination. As the blue item is steeper, it has a higher discrimination than the black and green items. According to De Ayala and Little (2022), reasonably good discrimination values range from approximately 0.8–2.5. Indeed, questions with steeper ICC slopes are better suited at discriminating between students at a given ability, while questions with lower discrimination power have more gentle slopes.

Items that discriminate better (steeper ICC slopes) thus provide more information about the ability level at which students are likely to start answering correctly, which results in higher bell shaped Item Information Curves, or IICs. The bell shaped curves in Figure 2C represent the amount of information I_i provided for each of the test's items according to the student's ability θ. These IICs vary in both maximum value (dependent on the item's discriminability, i.e., the ICC slope), and the x-value at which they reach it (the item's difficulty). Here, the blue and red curves, as well as the green and purple curves, have the same difficulty (they both reach their maximum around x = -2 and x = 0, respectively), but are of different discriminability: the blue item discriminates more than the red, the red more than the green and the green more than the purple (steeper ICC slope, and higher maximum IIC value).

Taking into account the different test items and the amount of information provided by each question, one can obtain the resulting Test Information Function (TIF) and Standard Error of Measurements (SEM). In Figure 2D, the TIF (blue) is the sum of the instrument's IICs from Figures 2B,C, while the SEM is the square root of the variance. The TIF shows that the instrument displays maximum information around -2 and provides more information in the low-medium ability range than in the high ability range. The SEM (red) is at its lowest where the test provides the most information (maximum of the TIF) and at its highest where the test provides the least information (minimum of the TIF).

Please note that prior to applying IRT, it is recommended to verify whether the data meets the unidimensionality criteria. If the unidimensionality criteria is not met, the higher the misspecification, then the higher the impact on the estimated parameters, and in particular on the discriminatoin parameter (with little impact on the difficulty parameter, Kahraman, 2013; Rajlic, 2019). The unidimensionality criteria can be verified through Confirmatory Factor Analysis (CFA) as done by Kong and Lai (2022) for instance. As the input data is binary (with a score of 0 or 1 per question), the CFA analysis is conducted using an estimator which is adapted to non-normal data and employs diagonally weighted least squared and robust estimators to estimate the model parameters (Schweizer et al., 2015; Rosseel, 2020).

When analyzing the results of IRT, as in the case of Confirmatory Factor Analysis, and other similar statistical approaches, multiple fit indices should be considered to establish the goodness of fit of the model. Model fit indices include the following metrics:

- The chi-square χ^2 statistic which should have $p_{\chi^2} > 0.05$. However, the larger the sample, the larger the χ^2 statistic, and the lower the p-value (Prudon, 2015; Alavi et al., 2020). The literature therefore suggests employing the ratio between the χ^2 statistic and the degrees of freedom with a cutoff at $\chi^2/df \leq 3$ (Kyriazos, 2018). At the individual item level for IRT models, Orlando and Thissen's signed χ^2 statistic ($S-\chi^2$) is recommended, with a ratio of $\chi^2/df \leq 5$ being acceptable (Wheaton et al., 1977; Kong and Lai, 2022) and a ratio below 3 being considered good.
- The root mean square error of approximation or RMSEA which should be < 0.06 for good fit and < 0.08 for acceptable fit (Hu and Bentler, 1999; Chen et al., 2008; Xia and Yang, 2019).
- The standardized root mean square residual or SRMR (Hu and Bentler, 1999; Xia and Yang, 2019) which should be < 0.08.
- The comparative fit index (CFI) and Tucker Lewis index (TLI) with values > 0.95 indicating a good fit, and acceptable values being > 0.90 (Kong and Lai, 2022).

Finally, more specifically to IRT, are

- Yen (1984)'s Q3 statistic to measure local independence which requires that none of the pairs of item residuals have a high correlation to ensure that local independence is not

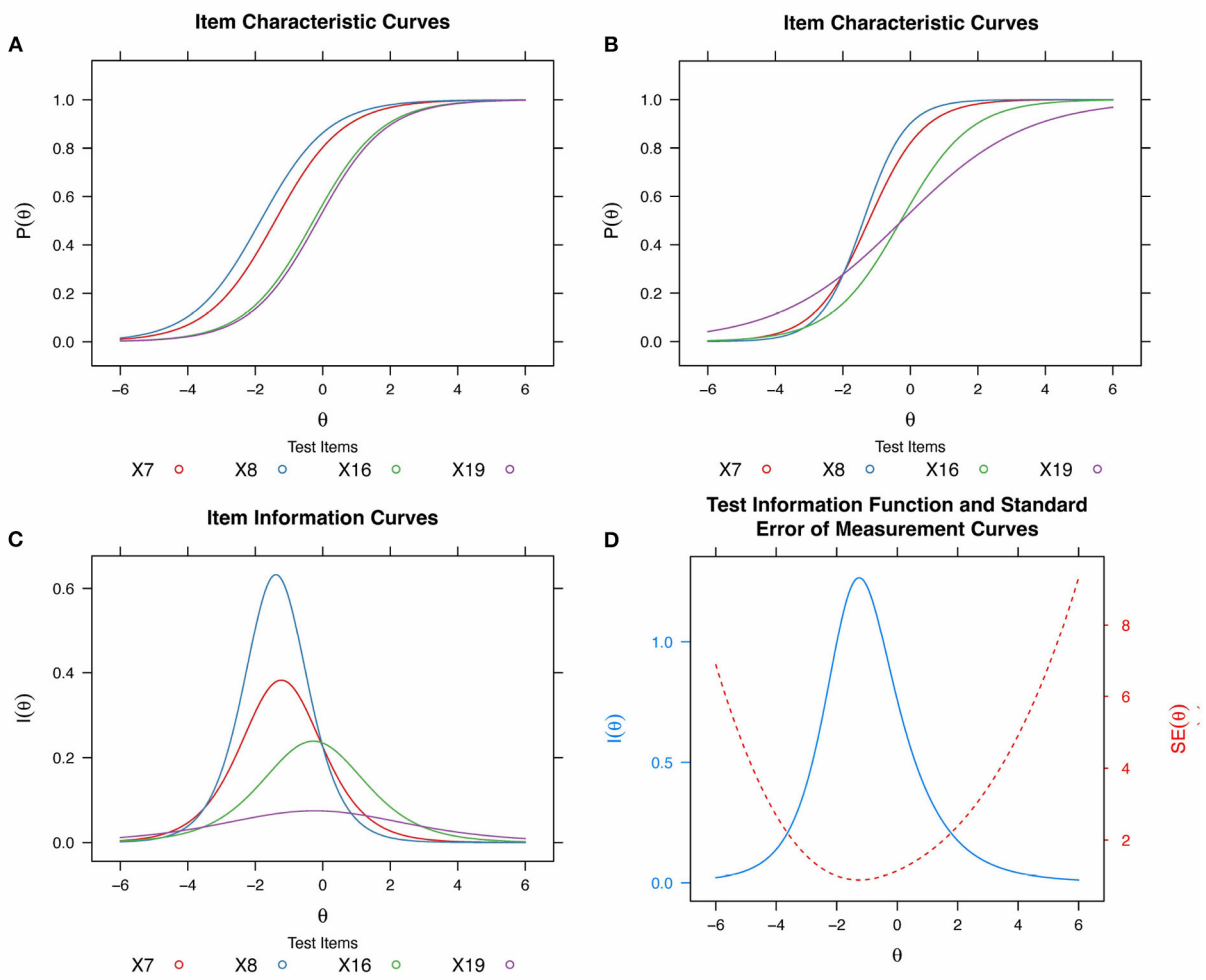


FIGURE 2
IRT Theory plots. **(A)** Item Characteristic Curves for four items of equal discrimination (slope) and varying difficulty (using a 1-PL model on the cCTt test data). **(B)** Item Characteristic Curves (ICC) for four items (blue, red, green, purple) of varying difficulty and discrimination (using a 2-PL model on cCTt test data). **(C)** Item Information Curves (IICs) for the items in **(B)**. **(D)** Test Information Function (TIF, in blue) for the four items from Panels **(B)** and **(C)** (IIC, in black), and the standard error of measurement (SEM, in red).

violated for the given model type. Critical values for the Q3 statistic are often arbitrary (Christensen et al., 2017) (e.g., 0.2 Christensen et al., 2017; Kong and Lai, 2022 or 0.3 Marais, 2012). As in our case the sample size is small (around 200 for the cCTt and 300 for the BCTt), and the number of items is high, the threshold of 0.3 is chosen as a critical value as the Q3 statistic is expected to be higher here than in cases with large samples and low number of items (Christensen et al., 2017). Similarly, as the number of items is high, the critical values are also expected to be higher (Christensen et al., 2017). As such, we consider the 0.3 threshold for the present study.

The Q3 statistic is computed once the model with the best fit has been selected.

- The M_2 statistics by Maydeu-Olivares and Joe "which have been found to be effective in evaluating the goodness of fit of IRT models" (Kong and Lai, 2022).
- The IRT reliability for each ability θ which is "closely related to test information and standard error, as it concerns the measurement precision and can be calculated with the equation $r = 1 - SEM(\theta^2)$" (Kong and Lai, 2022) where SEM represents the SEM for each ability.
- Wainer and Thissen (2001)'s marginal reliability metric (r_{xx}) which "denotes the ratio of the true score variance to the total variance, expressed with respect to the estimated latent abilities" (Andersson and Xin, 2018).

TABLE 3 Participants.

Test	Gender	Number of participants per grade			
		Grade 3	Grade 4	Undisclosed	Total
BCTt	Female	80	82	5	167
	Male	78	61	6	145
	Undisclosed			62	62
	Total	158	143	73	374
cCTt	Female	36	68		104
	Male	38	59		97
	Total	74	127		201

2.3. Participants and data collection

To compare the instruments, we used data collected by researchers and practitioners using the BCTt and cCTt in a study looking to evaluate the impact of a CT intervention conducted in public schools in Portugal. The recruitment for the intervention was done in three stages. First a call was sent out to schools and teachers to ask whether they were interested in participating in the CT intervention which included a pre-post test assessment using either the BCTt (in spring 2020) or the cCTt (in spring 2021). Secondly, teachers who were interested were briefed about the intervention and the assessments before agreeing or not to participate with their classrooms. Thirdly, consent forms were sent out to the parents of the concerned students.

The administration of both instruments was done in the classrooms following the protocol established for the BCTt, and its adaptation for the cCTt. In order to compare the instruments and avoid biases from the interventions themselves (whose goals and outcomes are outside the scope of this article), we only consider the results of the pre-tests administered to 575 students prior to the interventions (El-Hamamsy et al., 2022b).[3] More specifically, we analyse the results of the BCTt pre-test administered in March 2020 to 374 students in grades 3–4, and the results of the cCTt pre-test administered in April 2021 to 201 other students in grades 3–4 (see Table 3). All participants were enrolled in the same school districts in Portugal and did not have any prior experience with the CT-concepts measured with the instruments, as this is not part of the national curriculum. Please note that while the populations are not identical, they are considered to be comparable, and a comparison of both instruments is possible through the lens of IRT which is sample agnostic (see section 2.2.2) and complements the results of Classical Test Theory which may be subject to sample dependency. Comparing the properties of the instruments on two distinct samples also helps avoid the testing-effect, i.e., having students' performance improve on the second instrument because the questions employ the same modalities as the first instrument, and are therefore familiar and easier due to practice, rather than being due to a difference between the instruments (Knapp, 2016).

3 The data is available on Zenodo (El-Hamamsy et al., 2022b).

3. Results

3.1. Score distribution

The distribution of scores obtained in the two tests (both out of a maximum of 25 points) is shown in Figure 3. The Shapiro-Wilk test of normality indicates that the distribution of the cCTt is normal ($p > 0.05$, fails to reject H_0) and that the distribution of BCTt is not ($p < 0.0001$, rejects H_0). This is due to a ceiling effect, which is apparent for the BCTt (skew $= -1.23$, kurtosis $= 1.98$), but is not present in the case of the cCTt (skew $= -0.07$, kurtosis $= -0.13$).[4] Neither instrument shows significant differences in scores between genders [one-way ANOVA $F_{BCTt}(1) = 0.19$, $p_{BCTt} = 0.67$; one-way ANOVA $F_{cCTt}(1) = 0.03$, $p_{cCTt} = 0.86$].

Where the BCTt is concerned, students in grade 4 ($\mu_4 = 20.62 \pm 3.66$) perform significantly better than students in grade 3 ($\mu_3 = 19.18 \pm 4.16$). Indeed, the one-way ANOVA indicates that the difference between grades is significant [$F_{(1)} = 10.18$, $p = 0.0016$, $\Delta\mu = 1.44$ out of 25] with a medium-small effect size (Cohen's $d = 0.37$[5] Lakens, 2013). This would appear to confirm the progression between grades on the BCTt observed in the original BCTt validation.

Where the cCTt is concerned, no significant differences exist between grades [one-way ANOVA $F_{(1)} = 1.63$, $p = 0.2$]. The lack of distinction between grades in this sample is related to the fact that the grade 3 students are performing well on the

4 Skew (i.e., the asymmetry of a distribution) and kurtosis (i.e., the location of the peak of a distribution) of a normal distribution are close to 0 (Kim, 2013).

5 Cohen's *d* effect size is a quantitative measure of the magnitude of the observed difference. It is a standardized measure of the difference between the two means which is calculated by dividing the difference of the means by the standard deviation. Cohen suggested that 0.2 is a small effect size, 0.5 a medium effect size, and 0.8 a large effect size (Lakens, 2013).

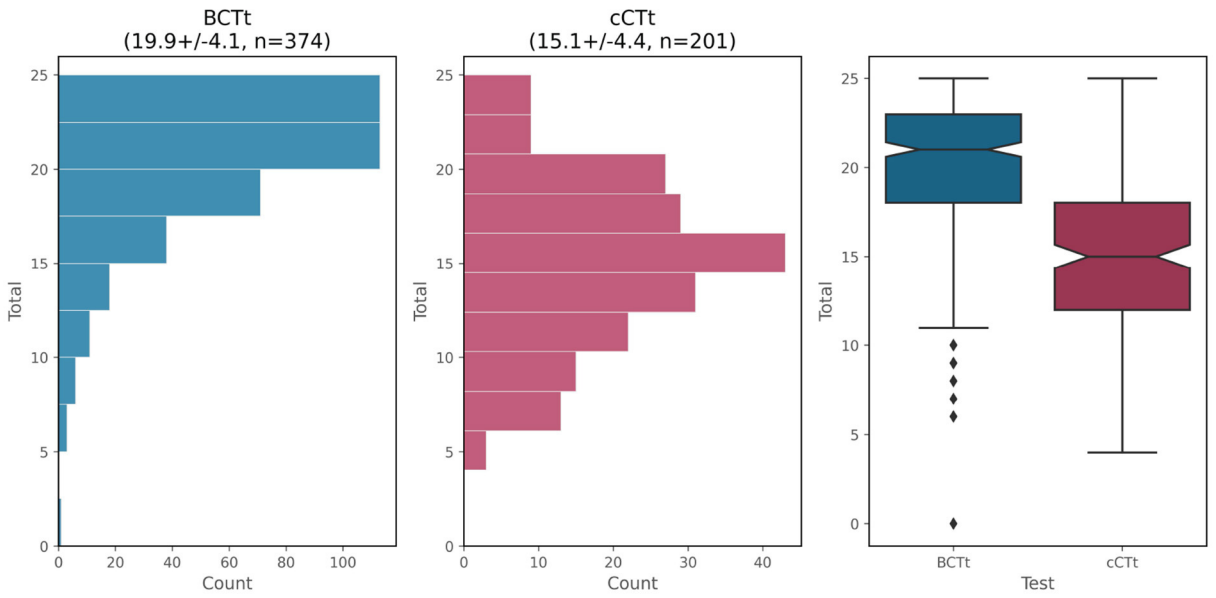


FIGURE 3
Score distribution for the BCTt and cCTt. The histogram and boxplots show the ceiling effect of the BCTt while the cCTt exhibits a normal distribution centered around 15/25 (i.e., 60%).

test ($\mu = 14.64 \pm 3.75$ out of 25), and specifically as well as the grade 4 students ($\mu = 15.45 \pm 4.68$). Indeed, in the first study validating the cCTt, the grade 3 students scored an average of $\mu = 12.62 \pm 5.18$ ($n = 711$) and the grade 4 students $\mu = 15.49 \pm 4.96$ ($n = 749$) out of 25.

3.2. Classical Test Theory

Cronbach's α (Bland and Altman, 1997) measurement of internal consistency of scales was used as an indicator of the instruments' reliability. According to the thresholds of Hinton et al. (2014) and Taherdoost (2016), both instruments exhibit high reliability ($\alpha_{BCTt} = 0.82 > 0.7$, $\alpha_{cCTt} = 0.78 > 0.7$). Nonetheless, the individual item difficulties (i.e., the proportion of correct answers) and point biserial correlations (i.e., the difference between the high scorers and the low scorers of the sample population) provide useful insights into the developmental appropriateness of the instruments, by indicating which items could be revised to improve the validity of the instruments for the target populations.

Figure 4 shows that both instruments present questions of decreasing difficulty (i.e., that are harder). The BCTt counts 13 questions which are above the maximum difficulty index threshold (i.e., are too easy) for the target age group, as opposed to 5 for the cCTt (including the 3 that were too easy in the original cCTt validation). The cCTt also exhibits two questions which are too hard (the same ones as in the original cCTt validation), which is not the case of the BCTt. Indeed, as Figure 4 shows, the BCTt covers a smaller range of item difficulties (BCTt difficulty indices min = 0.97, max = 0.49, range = 0.48; cCTt difficulty indices min = 0.96, max = 0.18, range = 0.79), lacking items in the lower half of the difficulty index range.

In terms of point-biserial correlation (see Figure 4), questions that could be revised for students in grades 3–4 are those below the 0.2 threshold. The metric indicates that only one item could be revised for the BCTt (question 24), while four items of the cCTt could be revised (questions 2, 17, 22, and 24). Interestingly, most of these questions were among the most difficult ones for the students.

Table 4 reports the Classical Test Theory analysis results for all questions in the two tests. Accounting for both difficulty indices and point biserial correlation, the number of questions that could be revised for students in grades 3 and 4 are higher for the BCTt ($n = 14$) than the cCTt ($n = 8$), as can be seen in Table 4.

3.3. Item Response Theory (IRT)

3.3.1. Verifying the unidimensionality to compare instruments through Confirmatory Factor Analysis

One criteria required to compare instruments through IRT is that the data measure the same latent trait. We thus employed Confirmatory Factor Analysis (CFA) as done by Kong and Lai

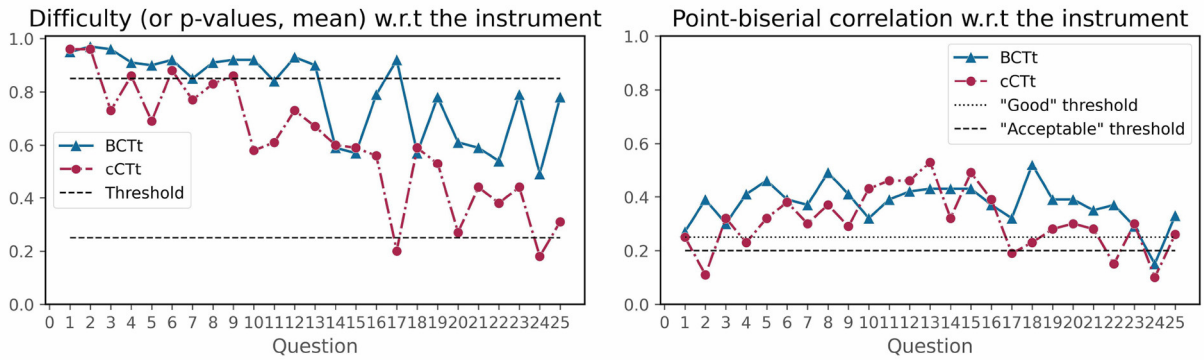


FIGURE 4
Classical Test Theory—Item Difficulty indices (i.e., the proportion of correct responses) on the left, and Point-Biserial Correlation on the right. Items with difficulty indices above the 0.85 threshold are considered too easy while items with difficulty indices below the 0.25 threshold are considered too difficult. Items with a point-biserial correlation above the 0.2 threshold are considered acceptable while those above 0.25 are considered good.

TABLE 4 Full BCTt (Cronbach's $\alpha_{BCTt} = 0.82$) and cCTt (Cronbach's $\alpha_{cCTt} = 0.78$) Classical Test Theory Analysis.

	BCTt						cCTt				
Q	Difficulty index	std	PBC	Drop alpha	Revision	Q	Difficulty index	std	PBC	Drop alpha	Revision
1	**0.95**	0.22	0.27	0.82	x	1	**0.96**	0.19	0.25	0.78	x
2	**0.97**	**0.16**	0.39	0.82	x	2	**0.96**	0.19	**0.11**	0.78	x
3	**0.96**	0.2	0.3	0.82	x	3	0.73	0.44	0.32	0.77	
4	**0.91**	0.29	0.41	0.82	x	4	**0.86**	0.35	0.23	0.78	x
5	**0.9**	0.3	0.46	0.81	x	5	0.69	0.46	0.32	0.77	
6	**0.92**	0.27	0.39	0.82	x	6	**0.88**	0.32	0.38	0.77	x
7	**0.85**	0.35	0.37	0.82	x	7	0.77	0.42	0.3	0.77	
8	**0.91**	0.29	0.49	0.81	x	8	0.83	0.38	0.37	0.77	
9	**0.92**	0.27	0.41	0.82	x	9	**0.86**	0.35	0.29	0.78	x
10	**0.92**	0.27	0.32	0.82	x	10	0.58	0.49	0.43	0.77	
11	0.84	0.37	0.39	0.82		11	0.61	0.49	0.46	0.76	
12	**0.93**	0.25	0.42	0.82	x	12	0.73	0.45	0.46	0.77	
13	**0.9**	0.3	0.43	0.81	x	13	0.67	0.47	0.53	0.76	
14	0.59	0.49	0.43	0.81		14	0.6	0.49	0.32	0.77	
15	0.57	0.5	0.43	0.81		15	0.59	0.49	0.49	0.76	
16	0.79	0.41	0.37	0.82		16	0.56	0.5	0.39	0.77	
17	**0.92**	0.27	0.32	0.82	x	17	**0.2**	0.4	**0.19**	0.78	x
18	0.57	0.5	0.52	0.81		18	0.59	0.49	0.23	0.78	
19	0.78	0.42	0.39	0.82		19	0.53	0.5	0.28	0.78	
20	0.61	0.49	0.39	0.82		20	0.27	0.45	0.3	0.77	
21	0.59	0.49	0.35	0.82		21	0.44	0.5	0.28	0.78	
22	0.54	0.5	0.37	0.82		22	0.38	0.49	**0.15**	0.78	x
23	0.79	0.41	0.29	0.82		23	0.44	0.5	0.3	0.77	
24	0.49	0.5	**0.15**	0.83	x	24	**0.18**	0.38	**0.1**	0.78	x
25	0.78	0.41	0.33	0.82		25	0.31	0.46	0.26	0.78	

Q, question; Difficulty index, proportion of correct responses; std, standard deviation; PBC, Point-Biserial Correlation. Items that are too easy (i.e., $\mu > 0.85$), too difficult (i.e., $\mu < 0.25$), or with a low point-biserial correlation (< 0.2) are marked in bold as elements which could be revised.

(2022), with a Diagonally Weighted Least Squares estimator to account for the binary inputs (see Table 5 for the fit indices). The Kaiser, Meyer, Olkin (KMO) measure of sampling adequacy indicates that the data is appropriate for factor analysis in both cases. Bartlett's test of sphericity also suggests that there is sufficient significant correlation in the data for factor analysis. For the full instruments (with 25 items) the model fit indices are also adequate in terms of the χ^2 criteria statistic, the CFI and TLI indices for both instruments. The RMSEA is below 0.6 in both cases. Finally, the SRMR is considered acceptable for the cCTt and just shy of the limit for the BCTt ($SRMR_{BCTt} = 0.084$). The modification indices for the BCTt-CFA indicate high correlations between 3 items from the BCTt (Q14, Q15, and Q18) which address the notions of complex loops. Removing item 15 from the factor analysis improves the model fit and meets the threshold requirements for the different fit indices (see Table 5). Furthermore, we exclude items with low CFA factor loadings (< 0.2) from the IRT analysis. Please note that all remaining items have significant factor loadings and that the excluded items correspond to questions which have low point biserial correlations (namely Q24 in the BCTt, and Q2, Q17, Q22, and Q24 in the cCTt). The corresponding fit indices for the final 1 factor CFA are provided in Table 5. With these adjustments, a 1 factor structure appears suitable for both instruments (when excluding Q15 and Q24 from the BCTt, and Q2, Q17, Q22, and Q24 from the cCTt).

TABLE 5 Confirmatory factor analysis fit indices for unidimensionality.

	Modification	KMO	Bartlett's test of sphericity	χ^2	χ^2/df	CFI	TLI	RMSEA	SRMR
BCTt (25 items)		0.84	$\chi^2(300) = 2026$, $p < 0.001$	$\chi^2(275) = 464$, $p < 0.001$	True	0.92	0.913	$0.43, 90\%ci = [0.036 - 0.05]$	0.084
BCTt (24 items)	Removing Q15 (high correlations with Q14 % Q18)	0.84	$\chi^2(276) = 1825$, $p < 0.001$	$\chi^2(252) = 357$, $p < 0.001$	True	0.945	0.940	$0.033, 90\%ci = [0.025 - 0.041]$	0.076
BCTt (23 items)	Removing Q24 (low factor loading)	0.84	$\chi^2(253) = 1779$, $p < .001$	$\chi^2(230) = 322$, $p < 0.001$	True	0.951	0.946	$0.033, 90\%ci = [0.024 - 0.041]$	0.074
cCTt (25 items)		0.75	$\chi^2(300) = 877$, $p < 0.001$	$\chi^2(275) = 350$, $p = 0.001$	True	0.935	0.929	$0.037, 90\%ci = [0.024 - 0.049]$	0.077
cCTt (21 items)	Removing Q2, Q17, Q22, Q24 (low factor loading)	0.77	$\chi^2(210) = 761$, $p < .001$	$\chi^2(189) = 216$, $p < 0.089$	True	0.975	0.972	$0.027, 90\%ci = [0.000 - 0.043]$	0.071

3.3.2. Comparing the instruments

As indicated previously, we only consider the 1-PL and 2-PL models in our study due to the low sample sizes which prevent us from finding stable solutions in the case of the 3-PL model and prevent us from converging in the case of the 4-PL model (see global model fit indices for the 1-PL and 2-PL models in Table 6). For both the BCTt and the cCTt, the 2-PL model was selected as an ANOVA indicated that the 2-PL model improved the fit significantly compared to the 1-PL model in both cases [$\chi^2_{BCTt}(22) = 62.92$, $p_{BCTt} < 0.0001$, $\chi^2_{cCTt}(20) = 79.84$, $p_{cCTt} < 0.0001$]. Individual item discrimination, difficulties, and fit indices are provided for the 2-PL models in Table 7. The results indicate that the $\chi^2/df < 3$ criterion is achieved for all items, and that all but three items have RMSEA just shy of the 0.6 threshold (considering that the rounded values would be equal to 0.6 these can be considered acceptable, Ockey and Choi, 2015). We then verify the local independence using Yen (1984)'s Q3 statistic and find that it is below the 0.3 threshold for all pairs of items in the BCTt and in the cCTt.

The results of the IRT analyses are shown in Figures 5A–D. While the Item Characteristic curves (Figure 5A) appear to indicate that the BCTt questions have higher "discrimination power" than the cCTt questions, this difference is not significant [one-way ANOVA $F_{(1)} = 3.11$, $p = 0.085$, see Figure 6]. This means that both tests are as good at discriminating between

TABLE 6 IRT model parameter fit indices for 1-PL and 2-PL models with the BCTt and cCTt.

		M_2	*df*	*p*	RMSEA	ci RMSEA 5%	ci RMSEA 95%	SRMR	TLI	CFI
BCTt (23 items)	1-PL	514	253	0.000	0.053	0.046	0.059	0.098	0.929	0.929
	2-PL	415	230	0.000	0.046	0.039	0.053	0.068	0.945	0.950
cCTt (21 items)	1-PL	392	210	0.000	0.067	0.056	0.077	0.102	0.849	0.849
	2-PL	294	189	0.000	0.053	0.041	0.065	0.075	0.903	0.913

students, however where they discriminate best differs[6]. The Item Information Curves (Figure 5B) shows that the BCTt questions provide most information in the low ability range, while the Item Information is more distributed along the low-medium range for the cCTt. The resulting TIFs (Figure 5C) therefore confirm that the BCTt is better at discriminating between students with low ability, while the cCTt is better at discriminating between low-medium abilities. As such, the IRT findings support that the cCTt overall fits grade 3–4 individuals and it decently works all along the ability range.

3.4. Limitations

As in all studies, the study presents certain limitations. Aside the inherent limitations pertaining to the specific use of Classical Test Theory and Item Response Theory which are well documented in the literature, the following elements are specific to the current study.

The instruments were tested on two populations from different schools, one year apart, and may thus differ in their CT abilities. While the students in the same grades should be expected to have the same level of CT-skills, this may not be the case. However, certain elements help mitigate this risk and counter the limitation: the schools are in the same country and district and thus follow the same mandatory curriculum (which does not include CS or CT), the measurements took place at the same time of the academic year, and we employed IRT as it tends to be sample agnostic.

The relatively small sample sizes prevented us from testing more complex models, such as 3-PL and 4-PL models. Indeed, larger sample sizes, in particular for the cCTt (n = 200), would have likely improved the model fit and reliability of the item difficulty and discrimination indices. These indices should only be considered as indicative of where the test provides more information, also since the IRT analysis was conducted on a subset of the items to meet the unidimensionality criteria. However, please note that the IRT analysis was also conducted with the full subset of items (although not presented in the article) and lead to the same conclusions. Such an analysis is possible as the violation of the unidimentionality criteria leads to "an overestimation of the discrimination parameter, (ii) with little impact on the difficulty estimation" (Kahraman, 2013; Rajlic, 2019), with "the impact on the estimated parameters [being] smaller the closer we are to the unidimensionality criteria" (Kahraman, 2013; Rajlic, 2019). Given the small samples and the fact that the IRT parameters were estimated on a subset of the items, it would be best to avoid using the IRT parameter estimates of the present study, in particular for the cCTt, to estimate the students' abilities on the latent ability scale.

4. Recommendations for the use of the BCTt and the cCTt

Considering (i) the present BCTt-cCTt comparison, (ii) the results of the BCTt validation conducted by Zapata-Cáceres et al. (2020) over grades 1–6, and (iii) the cCTt validation conducted by El-Hamamsy et al. (2022a) over grades 3–4, we propose the following recommendations with respect to these two instruments for grades 3–4:

- The cCTt should be preferred for grades 3–4 as it differentiates better between students in this age group and ability level, in addition to discriminating moderately well along the entire ability range. The cCTt is thus better suited to evaluate the efficacy of the intervention itself, in a pre-post-test design.
- The BCTt could be employed for low-ability students in grades 3–4, depending on the assessors' prior knowledge of the context and the students being assessed given the good discriminability the BCTt offers in grades 3–4 for low ability students.
- The BCTt could be employed as a screening mechanism to identify low-ability students which could prove useful for practitioners prior to an intervention, e.g., to ensure

6 The "discrimination power" of the instrument which relates to how high the discrimination is over all the questions of the assessment and is provided by slope of the ICCs, maximum values of the IICs. This is related to where the assessment, and thus the individual questions, discriminate best (which is provided by the $y = 0.5$ crossing of the ICCs, or the peak of the IICs).

TABLE 7 BCTt and cCTt item parameters and fit indices.

Item	Dscr	Dffc	BCTt S-χ^2	BCTt df S-χ^2	BCTt RMSEA S-χ^2	BCTt p S-χ^2	BCTt S-χ^2/df	Item	Dscr	Dffc	cCTt S-χ^2	cCTt df S-χ^2	cCTt RMSEA S-χ^2	cCTt p S-χ^2	cCTt S-χ^2/df
Q1	1.29	-2.8	11.5	8	0.034	0.175	1.44	Q1	2	-2.38	3.3	2	0.058	0.191	1.66
Q2	2.79	-2.28	2.3	2	0.019	0.323	1.13	Q3	0.95	-1.27	13.7	11	0.035	0.251	1.24
Q3	1.73	-2.47	3.2	6	0.0	0.786	0.53	Q4	0.76	-2.6	15.2	10	0.052	0.124	1.52
Q4	1.82	-1.85	13.8	11	0.026	0.245	1.25	Q5	0.95	-1	11.6	11	0.017	0.392	1.06
Q5	2.41	-1.6	20.3	9	0.058	0.016	2.26	Q6	2.01	-1.56	9	7	0.038	0.254	1.28
Q6	1.66	-2.03	13.1	11	0.022	0.29	1.19	Q7	1.02	-1.4	11	11	0.0	0.447	1
Q7	1.38	-1.67	15.8	12	0.029	0.2	1.32	Q8	1.46	-1.45	9.5	9	0.016	0.396	1.05
Q8	2.56	-1.6	11.8	9	0.029	0.224	1.31	Q9	1.47	-1.67	16.2	9	0.064	0.062	1.8
Q9	2.16	-1.82	8.6	10	0.0	0.574	0.86	Q10	1.52	-0.33	14.1	10	0.046	0.167	1.41
Q10	1.41	-2.26	18.2	11	0.042	0.077	1.66	Q11	1.76	-0.4	5.3	10	0.0	0.871	0.53
Q11	1.42	-1.55	14.6	13	0.018	0.334	1.12	Q12	1.98	-0.82	14.3	8	0.063	0.075	1.79
Q12	2.02	-1.95	9.9	10	0.0	0.451	0.99	Q13	2.75	-0.53	7.9	7	0.026	0.338	1.13
Q13	1.93	-1.73	13.3	11	0.024	0.275	1.21	Q14	0.98	-0.49	17.1	10	0.061	0.071	1.71
Q14	1.31	-0.36	14.7	10	0.035	0.145	1.47	Q15	2.35	-0.3	13	8	0.057	0.112	1.62
Q16	1.25	-1.37	19.3	13	0.036	0.113	1.49	Q16	1.09	-0.27	5.7	10	0.0	0.837	0.57
Q17	1.43	-2.22	11.8	11	0.014	0.383	1.07	Q18	0.53	-0.75	16.3	12	0.043	0.178	1.36
Q18	1.77	-0.26	8.6	8	0.014	0.379	1.07	Q19	0.62	-0.22	9.2	12	0.0	0.687	0.77
Q19	1.21	-1.3	11.4	11	0.01	0.408	1.04	Q20	0.65	1.66	12	10	0.032	0.284	1.2
Q20	1.13	-0.5	9.2	11	0.0	0.607	0.83	Q21	0.49	0.53	13.4	12	0.024	0.343	1.11
Q21	0.85	-0.52	7.8	11	0.0	0.733	0.71	Q23	0.63	0.38	9.5	11	0.0	0.577	0.86
Q22	1.08	-0.18	17	10	0.043	0.073	1.7	Q25	0.5	1.66	11.3	11	0.011	0.421	1.02
Q23	0.86	-1.74	18	13	0.032	0.157	1.39								
Q25	0.92	-1.63	14.2	13	0.016	0.357	1.1								

Dscr, Discrimination; Dffc, Difficulty.

10.3389/fpsyg.2022.1082659

that the intervention is well-tailored to the abilities of the students and ensure that nobody is "left behind."

5. Discussion and conclusion

The BCTt and the cCTt are two instruments that expand the portfolio of validated CT assessments, in particular, at the level of primary education. These instruments overlap in their target age ranges, notably in grades 3–4, and had not yet been compared psychometrically for those age groups. This study thus looked to establish the limits of validity of the two instruments by providing a detailed comparison of their psychometric properties on data acquired from 575 students (374 doing the BCTt and 201 doing the cCTt). Indeed, as:

1. The BCTt and the cCTt were validated in different countries, and thus potentially different contexts
2. There were only $n = 52$ grade 4 students in the BCTt validation, and $n = 0$ grade 3 students, with limited psychometric analyses conducted for the BCTt in those grades specifically.

The present study looked to conduct a detailed psychometric analysis of the BCTt in grades 3–4 (which was not yet conducted)

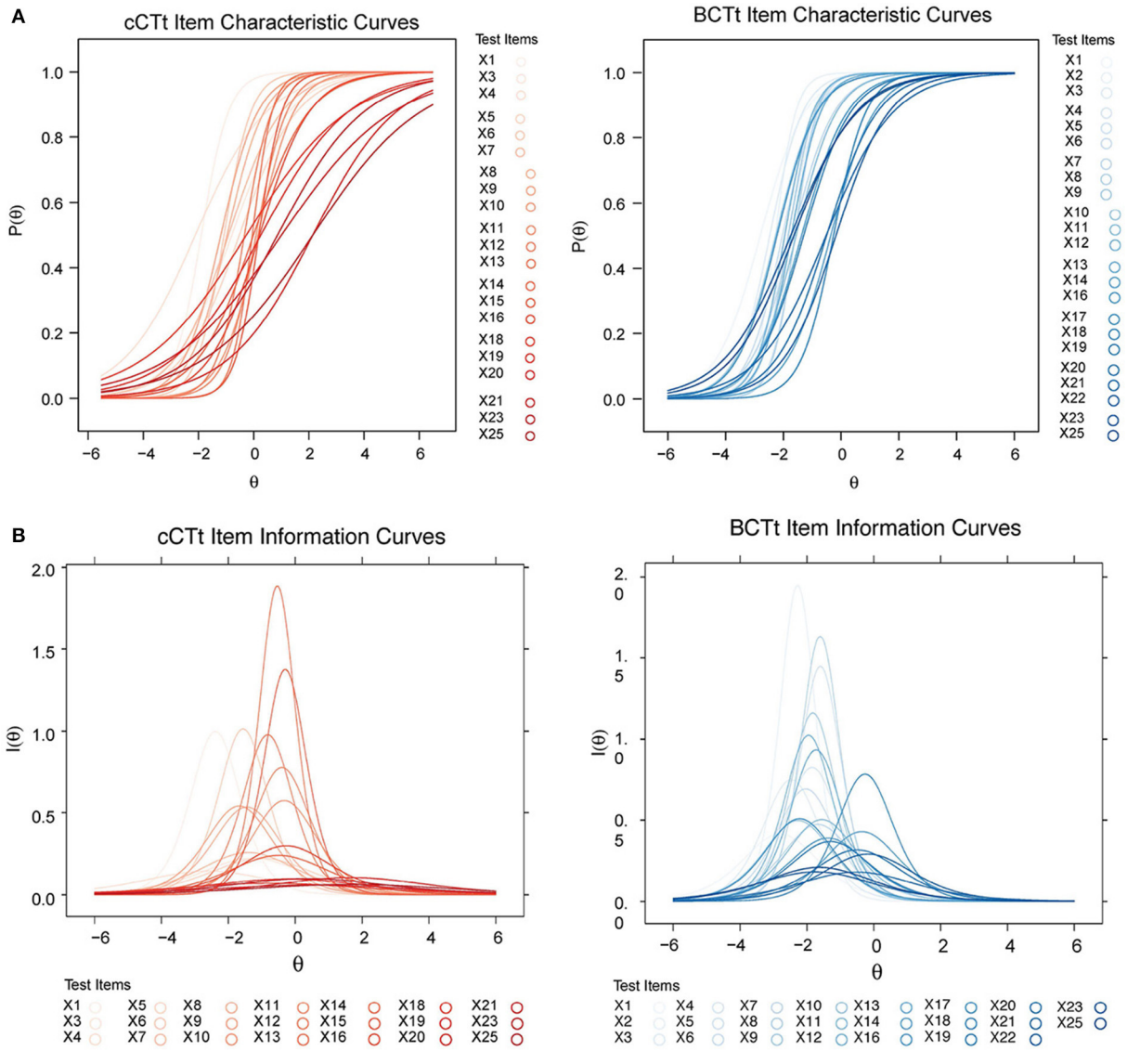


FIGURE 5
(Continued)

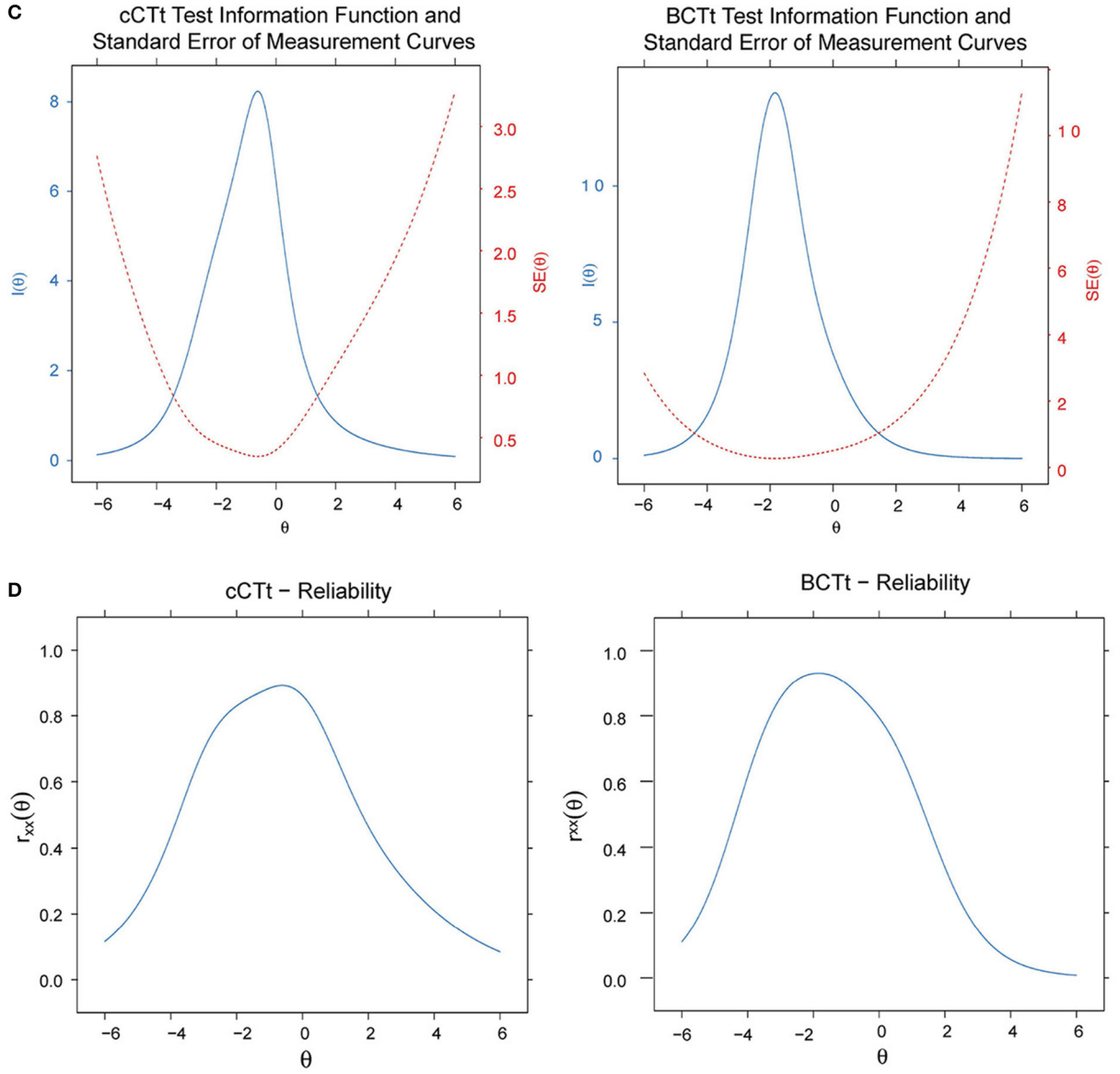


FIGURE 5 (Continued)
Item Response Theory curves for the BCTt and the cCTt. **(A)** Item Characteristic Curves (ICC). The figure shows that the items have varying difficulties and discrimination (slopes), with BCTt items showing higher discriminability in the low ability range and cCTt items showing higher discriminability in the low and medium ability ranges. **(B)** Item Response Theory Item Information Curves (IIC). Items in both instruments provide varying amount of information at different ability levels. Similarly to the ICC curves in Panel **(A)**, the information of the BCTt is mainly in the low ability range, while the information of the cCTt is in the low and medium ability ranges. Item Response Theory curves for the BCTt and the cCTt. **(C)** Test information function (TIF). The TIF being the sum of each instruments' Item Information Curves [see Panel **(B)**], the results confirm prior observations: the BCTt provides most of its information in the low ability range while the cCTt provides most information in the low and medium ability ranges. **(D)** Reliability at different ability levels. The figures show that both instruments have low reliability in the high ability range. The BCTt reliability peak is shifted toward the lower ability range while the cCTt reliability peak is toward the medium ability range. Please note that the marginal reliability r_{xx} for the BCTt is $r_{xx}(BCTt) = 0.75$, and for the cCTt $r_{xx}(cCTt) = 0.80$.

and compare the validity of the two instruments on a large and comparable pool of grade 3–4 students from a third, and single, country.

The findings from the psychometric analyses of the two instruments help re-establish their validity in grades 3 and 4 with both a new population and with students from a new country (here n = 575in Portugal, while the cCTt was validated with n = 1,519 grade 3–4 students in Switzerland, El-Hamamsy et al., 2022a, and the BCTt with n = 299 grade 1–6 students in Spain, Zapata-Cáceres et al., 2020). Where the cCTt is concerned, while there were no differences between students in grades 3–4 in the present sample, the general

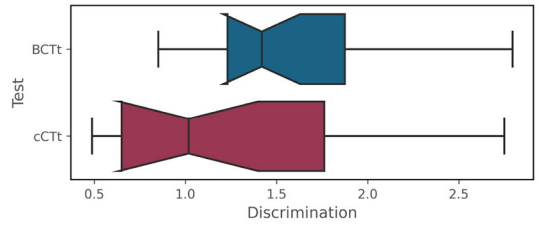


FIGURE 6
Item Response Theory BCTt–cCTt item discrimination comparison [one-way ANOVA $F_{(1)} = 3.11$, $p = 0.085$].

conclusions drawn from the Classical Test Theory analysis and overall IRT are coherent with those obtained by El-Hamamsy et al. (2022a). Where the BCTt is concerned, the results confirm the ceiling effect observed in grade 4 in the original study (Zapata-Cáceres et al., 2020) and extend it to students in grade 3 who were not part of the initial pool of students who were administered the BCTt. The psychometric comparison indicates that *the cCTt should be preferred for students in grades 3 and 4*, as students already have a good assimilation of basic CT concepts pertaining to sequences and loops. Therefore, students in grades 3–4 perform too well on the easier BCTt (which employs smaller 3 × 3 grids), giving rise to a ceiling effect. The *BCTt should instead be preferred if the objective is to discriminate between students with low abilities in grades 3 and 4*.

The findings are consistent with other studies that found that simple loops are already mastered in early primary school (Montuori et al., 2022), with very young students (starting 3 years old) already being able to solve algorithmic problems and their results improving with age (Piatti et al., 2022). As CT skills relate to students' numerical, verbal, and non-verbal reasoning abilities (Tsarava et al., 2022), it is likely that the findings align with students' maturation, increase in working memory (which is required to achieve tasks, Cowan, 2016), and executive functions over time. Therefore, as students get older, they should be able to deal with more complex computational concepts (e.g., conditionals and while loops), including those with more complex perceptual configurations (e.g., the 4 × 4 grids), corroborating the differences observed between both instruments. Future work should therefore consider continuing to refine the limits of validity of the instruments. Indeed, refinement studies are common in educational psychology, with similar work having already been undertaken for (i) the original CTt (aimed at 10–16 year old students) to improve it's validity for 16 year old students and above (Guggemos et al., 2022), and (ii) The TechCheck and it's variants to improve the validity for kindergarden students (Relkin et al., 2020; Relkin and Bers, 2021).

Two key takeaways emerge from the present study:

1. The importance of building and validating CT assessments for each specific age: children in the early stages of education undergo rapid cognitive development, so an instrument designed for a specific age range is likely to be too difficult for those immediately younger and too easy for those immediately older.
2. The importance of psychometrically comparing existing, overlapping CT instruments to establish their limits of validity. By providing detailed comparisons, researchers and practitioners may be able to choose the assessment in an informed way, and in accordance with their requirements and objectives.

As numerous researchers have put forward, instruments such as the BCTt and the cCTt should be combined with other forms of assessments in a systems of assessments (Grover et al., 2015; Román-González et al., 2019; Weintrop et al., 2021a) to accurately measure the full range of competencies at play when considering CT (Brennan and Resnick, 2012; Piatti et al., 2022). The systems of assessments could therefore include other instruments which assess CT practices such as the test by Li et al. (2021), employ direct observations of students' thought processes and strategies (Lye and Koh, 2014; Chevalier et al., 2020), or learning analytics and educational data mining techniques (Cock et al., 2021; Nasir et al., 2021; Zapata-Cáceres and Martín-Barroso, 2021). Complementary assessments would not only help gain a more accurate and in-depth picture of student learning but also feed into the learning activity design and intervention process (Chevalier et al., 2022). For completeness, the system of assessments should also include instruments that measure CT perspectives (e.g., such as those developed for high school, Yagci, 2019 and undergraduates, Korkmaz et al., 2017).

Provided that validation is a multi-step process that requires "collect[ing] multiple sources of evidence to support the proposed interpretation and use of assessment result[s] [and] multiple methodologies, sources of data, and types of analysis" (Gane et al., 2021), it is important to note that the BCTt and cCTt may still undergo further validation by including evidence of criterion validity. This can be achieved through several means. The first is comparing with other existing validated assessments. For instance, Relkin et al. (2020) compared the TechCheck with the TACTIC-KIBO, while (Li et al., 2021) went one step further and correlated the CTA-CES with reasoning, spatial abilities, and verbal abilities. The second is establishing the test's predictive validity, for example by establishing whether the instrument can predict academic performance and coding achievement as done by Román-González et al. (2018). The third is determining the instruments' concurrent validity, that is to say seeing whether the instrument is able to distinguish between two groups that differ, for instance novices and experts, or according

to students expressed digital proficiency as done by Li et al. (2021).

Data availability statement

The data presented in this study can be found on Zenodo (El-Hamamsy et al., 2022b).

Ethics statement

The studies involving human participants were reviewed and approved by Comité de Ética de la Investigación de la Universidad Rey Juan Carlos. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

LEH, MZC, PM, BB, and EMB: conceptualization. LEH, MZC, and PM: methodology. MRG: validation. LEH: formal analysis, writing—original draft and preparation, and visualization. PM: investigation. PM and LEH: data curation. LEH, MZC, PM, JD, BB, EMB, and MRG: writing–review and editing. BB and EMB: supervision. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by (i) the NCCR Robotics, a National Centre of Competence in Research, funded by the Swiss National Science Foundation (grant number 51NF40_185543), (ii) the Madrid Regional Government through the project e-Madrid-CM (P2018/TCS-4307) which is co-financed by the Structural Funds (FSE and FEDER), and (iii) the Fundação Calouste Gulbenkian (Gulbenkian Programme for Knowledge).

Acknowledgments

We would like to thank Professor Jean-Philippe Antonietti for answering our questions pertaining to Confirmatory Factor Analysis and Item Response Theory.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Aho, A. V. (2012). Computation and computational thinking. *Comput. J.* 55, 832–835. doi: 10.1093/comjnl/bxs074

Alavi, M., Visentin, D. C., Thapa, D. K., Hunt, G. E., Watson, R., and Cleary, M. (2020). Chi-square for model fit in confirmatory factor analysis. *J. Adv. Nurs.* 76, 2209–2211. doi: 10.1111/jan.14399

Andersson, B., and Xin, T. (2018). Large sample confidence intervals for item response theory reliability coefficients. *Educ. Psychol. Meas.* 78, 32–45. doi: 10.1177/0013164417713570

Awopeju, O. A., and Afolabi, E. R. I. (2016). Comparative analysis of classical test theory and item response theory based item parameter estimates of senior school certificate mathematics examination. *Eur. Sci. J.* 12:263. doi: 10.19044/esj.2016.v12n28p263

Bakala, E., Gerosa, A., Hourcade, J. P., and Tejera, G. (2021). Preschool children, robots, and computational thinking: a systematic review. *Int. J. Child Comput. Interact.* 29:100337. doi: 10.1016/j.ijcci.2021.100337

Basu, S., Rutstein, D., Xu, Y., and Shear, L. (2020). "A principled approach to designing a computational thinking practices assessment for early grades," in *SIGCSE '20: Proceedings of the 51st ACM Technical Symposium on Computer Science Education* (Portland OR: ACM), 912–918. doi: 10.1145/3328778.3366849

Bellettini, C., Lonati, V., Malchiodi, D., Monga, M., Morpurgo, A., and Torelli, M. (2015). "How challenging are bebras tasks? An IRT analysis based on the performance of Italian students," in *Proceedings of the 2015 ACM Conference on Innovation and Technology in Computer Science Education, ITiCSE '15* (New York, NY: Association for Computing Machinery), 27–32. doi: 10.1145/2729094.2742603

Biggs, J. (1996). Enhancing teaching through constructive alignment. *High. Educ.* 32, 347–364. doi: 10.1007/BF00138871

Bland, J. M., and Altman, D. G. (1997). Statistics notes: Cronbach's alpha. *BMJ* 314:572. doi: 10.1136/bmj.314.7080.572

Brennan, K., and Resnick, M. (2012). "New frameworks for studying and assessing the development of computational thinking," in *Proceedings of the 2012 Annual Meeting of the American Educational Research Association*, Vol. 1 (Vancouver) p. 25.

Chae, Y.-M., Park, S. G., and Park, I. (2019). The relationship between classical item characteristics and item response time on computer-based testing. *Korean J. Med. Educ.* 31, 1–9. doi: 10.3946/kjme.2019.113

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *J. Stat. Softw.* 48, 1–29. doi: 10.18637/jss.v048.i06

Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., and Paxton, P. (2008). An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. *Sociol. Methods Res.* 36, 462–494. doi: 10.1177/0049124108314720

Chen, G., Shen, J., Barth-Cohen, L., Jiang, S., Huang, X., and Eltoukhy, M. (2017). Assessing elementary students? computational thinking in everyday reasoning and robotics programming. *Comput. Educ.* 109, 162–175. doi: 10.1016/j.compedu.2017.03.001

Chevalier, M., Giang, C., El-Hamamsy, L., Bonnet, E., Papaspyros, V., Pellet, J.-P., et al. (2022). The role of feedback and guidance as intervention methods to foster computational thinking in educational robotics learning activities for primary school. *Comput. Educ.* 180:104431. doi: 10.1016/j.compedu.2022.104431

Chevalier, M., Giang, C., Piatti, A., and Mondada, F. (2020). Fostering computational thinking through educational robotics: a model for creative computational problem solving. *Int. J. STEM Educ.* 7:39, 1–18. doi: 10.1186/s40594-020-00238-z

Christensen, K. B., Makransky, G., and Horton, M. (2017). Critical values for Yen's Q_3 : identification of local dependence in the Rasch model using residual correlations. *Appl. Psychol. Meas.* 41, 178–194. doi: 10.1177/0146621616677520

Clarke-Midura, J., Silvis, D., Shumway, J. F., Lee, V. R., and Kozlowski, J. S. (2021). Developing a kindergarden computational thinking assessment using evidence-centered design: the case of algorithmic thinking. *Comput. Sci. Educ.* 31, 117–140. doi: 10.1080/08993408.2021.1877988

Cock, J., Marras, M., Giang, C., and Kaser, T. (2021). "Early prediction of conceptual understanding in interactive simulations," in *Proceedings of The 14th International Conference on Educational Data Mining (EDM21)* (Paris), 161–171.

Cowan, N. (2016). Working memory maturation: can we get at the essence of cognitive growth? *Perspect. Psychol. Sci.* 11, 239–264. doi: 10.1177/1745691615621279

Dai, B., Zhang, W., Wang, Y., and Jian, X. (2020). Comparison of trust assessment scales based on item response theory. *Front. Psychol.* 11:10. doi: 10.3389/fpsyg.2020.00010

Dai, S., Wang, X., and Svetina, D. (2022). *subscore: Computing Subscores in Classical Test Theory and Item Response Theory*. R Package Version 3.3.

De Ayala, R. J., and Little, T. D. (2022). *The Theory and Practice of Item Response Theory, 2nd Edn.* Methodology in the Social Sciences. New York, NY: Guilford Press.

De Champlain, A. F. (2010). A primer on classical test theory and item response theory for assessments in medical education. *Med. Educ.* 44, 109–117. doi: 10.1111/j.1365-2923.2009.03425.x

Denning, P. J., and Tedre, M. (2021). Computational thinking: a disciplinary perspective. *Inform. Educ.* 20, 361–390. doi: 10.15388/infedu.2021.21

DeVellis, R. F. (2006). Classical test theory. *Med. Care* 44, S50–S59. doi: 10.1097/01.mlr.0000245426.10853.30

El-Hamamsy, L., Bruno, B., Chessel-Lazzarotto, F., Chevalier, M., Roy, D., Zufferey, J. D., et al. (2021a). The symbiotic relationship between educational robotics and computer science in formal education. *Educ. Inform. Technol.* 26, 5077–5107. doi: 10.1007/s10639-021-10494-3

El-Hamamsy, L., Chessel-Lazzarotto, F., Bruno, B., Roy, D., Cahlikova, T., Chevalier, M., et al. (2021b). A computer science and robotics integration model for primary school: evaluation of a large-scale in-service K-4 teacher-training program. *Educ. Inform. Technol.* 26, 2445–2475. doi: 10.1007/s10639-020-10355-5

El-Hamamsy, L., Zapata-Cáceres, M., Barroso, E. M., Mondada, F., Zufferey, J. D., et al. (2022a). The competent computational thinking test: Development and validation of an unplugged computational thinking test for upper primary school. *J. Educ. Comput. Res.* 60:07356331221081753. doi: 10.1177/07356331221081753

El-Hamamsy, L., Zapata-Cáceres, M., Marcelino, P., Zufferey, J. D., Bruno, B., Barroso, E. M., et al. (2022b). Dataset for the comparison of two Computational Thinking (CT) test for upper primary school (grades 3-4) : the Beginners' CT test (BCTt) and the competent CT test (cCTt). *Zenodo*. doi: 10.5281/zenodo.5885033

European Union (2006). *Recommendation of the European Parliament and of the Council of 18 December 2006 on Key Competences for Lifelong Learning.* Technical report.

Gane, B. D., Israel, M., Elagha, N., Yan, W., Luo, F., and Pellegrino, J. W. (2021). Design and validation of learning trajectory-based assessments for computational thinking in upper elementary grades. *Comput. Sci. Educ.* 31, 141–168. doi: 10.1080/08993408.2021.1874221

Grover, S., Pea, R., and Cooper, S. (2015). ""Systems of assessments" for deeper learning of computational thinking in k-12," in *Proceedings of the 2015 Annual Meeting of the American Educational Research Association* (Chicago, IL), 15–20.

Guggemos, J., Seufert, S. and Román-González, M. Computational thinking assessment - towards more vivid interpretations. *Tech Know Learn*. (2022). doi: 10.1007/s10758-021-09587-2

Hambleton, R. K., and Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educ. Meas. Issues Pract.* 12, 38–47. doi: 10.1111/j.1745-3992.1993.tb00543.x

Hinton, P., McMurray, I., and Brownlow, C. (2014). *SPSS Explained*. London: Routledge. doi: 10.4324/9781315797298

Hu, L.-T., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equat. Model. Multidiscipl. J.* 6, 1–55. doi: 10.1080/10705519909540118

Hubwieser, P., and Mühling, A. (2014). "Playing PISA with bebras," in *Proceedings of the 9th Workshop in Primary and Secondary Computing Education*, WiPSCE '14 (New York, NY: Association for Computing Machinery), 128–129. doi: 10.1145/2670757.2670759

Jabrayilov, R., Emons, W. H. M., and Sijtsma, K. (2016). Comparison of classical test theory and item response theory in individual change assessment. *Appl. Psychol. Meas.* 40, 559–572. doi: 10.1177/0146621616664046

Kahraman, N. (2013). Unidimensional interpretations for multidimensional test items. *J. Educ. Meas.* 50, 227–246. doi: 10.1111/jedm.12012

Kim, H.-Y. (2013). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restor. Dent. Endod.* 38, 52–54. doi: 10.5395/rde.2013.38.1.52

Knapp, T. R. (2016). Why is the one-group pretest–posttest design still used? *Clin. Nurs. Res.* 25, 467–472. doi: 10.1177/1054773816666280

Kong, S.-C., and Lai, M. (2022). Validating a computational thinking concepts test for primary education using item response theory: an analysis of students? responses. *Comput. Educ.* 187:104562. doi: 10.1016/j.chb.2017.01.005

Korkmaz, z., akir, R., and zden, M. Y. (2017). A validity and reliability study of the computational thinking scales (CTS). *Comput. Hum. Behav.* 72, 558–569.

Kyriazos, T. A. (2018). Applied psychometrics: writing-up a factor analysis construct validation study with examples. *Psychology* 9, 2503–2530. doi: 10.4236/psych.2018.911144

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for *t*-tests and ANOVAs. *Front. Psychol.* 4:863. doi: 10.3389/fpsyg.2013.00863

Li, Y., Schoenfeld, A. H., diSessa, A. A., Graesser, A. C., Benson, L. C., English, L. D., et al. (2020). Computational thinking is more about thinking than computing. *J. STEM Educ. Res.* 3, 1–18. doi: 10.1007/s41979-020-00030-2

Li, Y., Xu, S., and Liu, J. (2021). Development and validation of computational thinking assessment of Chinese elementary school students. *J. Pacific Rim Psychol.* 15:183449092110102. doi: 10.1177/18344909211010240

Lye, S. Y., and Koh, J. H. L. (2014). Review on teaching and learning of computational thinking through programming: what is next for K-12? *Comput. Hum. Behav.* 41, 51–61. doi: 10.1016/j.chb.2014.09.012

Mannila, L., Dagiene, V., Demo, B., Grgurina, N., Mirolo, C., Rolandsson, L., et al. (2014). "Computational thinking in K-9 education," in *ITiCSE-WGR '14: Proceedings of the Working Group Reports of the 2014 on Innovation & Technology in Computer Science Education Conference* (Uppsala: ACM), 1–29, doi: 10.1145/2713609.2713610

Marais, I. (2012). "Local dependence," in *Rasch Models in Health*, eds K. B. Christensen, S. Kreiner, and M. Mesbah (Hoboken, NJ: John Wiley & Sons, Ltd.), 111–130. doi: 10.1002/9781118574454.ch7

Montuori, C., Ronconi, L., Vardanega, T., and Arfé, B. (2022). Exploring gender differences in coding at the beginning of primary school. *Front. Psychol.* 13:887280. doi: 10.3389/fpsyg.2022.887280

Moreno-León, J., and Robles, G. (2015). "Dr. scratch: a web tool to automatically evaluate scratch projects," in *Proceedings of the Workshop in Primary and Secondary Computing Education, WiPSCE '15* (New York, NY: ACM), 132–133. doi: 10.1145/2818314.2818338

Nasir, J., Kothiyal, A., Bruno, B., and Dillenbourg, P. (2021). Many are the ways to learn: identifying multi-modal behavioral profiles of collaborative learning in constructivist activities. *Int. J. Comput.-Support. Collab. Learn.* 16, 485–523. doi: 10.1007/s11412-022-09368-8

Ockey, G. J., and Choi, I. (2015). Structural equation modeling reporting practices for language assessment. *Lang. Assess. Quart. Int. J.* 12, 305–319. doi: 10.1080/15434303.2015.1050101

Parker, M., Kao, Y., Saito-Stehberger, D., Franklin, D., Krause, S., Richardson, D., et al. (2021). "Development and preliminary validation of the assessment of computing for elementary students (ACES)," in *SIGCSE '21: Proceedings of the 52nd ACM Technical Symposium on Computer Science Education* (Virtual Event), 10–16. doi: 10.1145/3408877.3432376

Piatti, A., Adorni, G., El-Hamamsy, L., Negrini, L., Assaf, D., Gambardella, L., et al. (2022). The CT-cube: a framework for the design and the assessment of computational thinking activities. *Comput. Hum. Behav. Rep.* 5:100166. doi: 10.1016/j.chbr.2021.100166

Prudon, P. (2015). Confirmatory factor analysis as a tool in research using questionnaires: a critique. *Comprehens. Psychol.* 4:03.CP.4.10. doi: 10.2466/03.CP.4.10

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Rajlic, G. (2019). *Violations of Unidimensionality and Local Independence in Measures Intended as Unidimensional: Assessing Levels of Violations and the Accuracy in Unidimensional IRT Model Estimates*. Ph.D. thesis, University of British Columbia.

Relkin, E., and Bers, M. (2021). "Techcheck-k: a measure of computational thinking for kindergarden children," in *2021 IEEE Global Engineering Education Conference (EDUCON)* (Vienna), 1696–1702. doi: 10.1109/EDUCON46332.2021.9453926

Relkin, E., de Ruiter, L., and Bers, M. U. (2020). TechCheck: development and validation of an unplugged assessment of computational thinking in early childhood education. *J. Sci. Educ. Technol.* 29, 482–498. doi: 10.1007/s10956-020-09831-x

Revelle, W. (2021). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Evanston, IL: Northwestern University.

Román-González, M., Moreno-León, J., and Robles, G. (2019). "Combining assessment tools for a comprehensive evaluation of computational thinking interventions," in *Computational Thinking Education*, eds S.-C. Kong and H. Abelson (Singapore: Springer), 79–98. doi: 10.1007/978-981-13-6528-7_6

Román-González, M., Pérez-González, J.-C., and Jiménez-Fernández, C. (2017). Which cognitive abilities underlie computational thinking? Criterion validity of the computational thinking test. *Comput. Hum. Behav.* 72, 678–691. doi: 10.1016/j.chb.2016.08.047

Román-González, M., Pérez-González, J.-C., Moreno-León, J., and Robles, G. (2018). Can computational talent be detected? Predictive validity of the computational thinking test. *Int. J. Child Comput. Interact.* 18, 47–58. doi: 10.1016/j.ijcci.2018.06.004

Rosseel, Y. (2012). lavaan: an R package for structural equation modeling. *J. Stat. Softw.* 48, 1–36. doi: 10.18637/jss.v048.i02

Rosseel, Y. (2020). *Structural Equation Modeling with lavaan [PowerPoint Slides]*. Department of Data Analysis, Ghent University.

Rowe, E., Almeda, M. V., Asbell-Clarke, J., Scruggs, R., Baker, R., Bardar, E., et al. (2021). Assessing implicit computational thinking in Zoombinis puzzle gameplay. *Comput. Hum. Behav* 120:106707. doi: 10.1016/j.chb.2021.106707

Sahin, A., and Anil, D. (2017). The effects of test length and sample size on item parameters in item response theory. *Educ. Sci. Theory Pract.* 17, 321–335. doi: 10.12738/estp.2017.1.0270

Schweizer, K., Ren, X., and Wang, T. (2015). "A comparison of confirmatory factor analysis of binary data on the basis of tetrachoric correlations and of probability-based covariances: a simulation study," in *Quantitative Psychology Research, Springer Proceedings in Mathematics & Statistics,* eds R. E. Millsap, D. M. Bolt, L. A. van der Ark, and W.-C. Wang, (Cham: Springer International Publishing), 273–292. doi: 10.1007/978-3-319-07503-7_17

Selby, C., and Woollard, J. (2013). "Computational thinking: the developing definition," in *Special Interest Group on Computer Science Education (SIGCSE) 2014* (Atlanta, GA).

Taherdoost, H. (2016). Validity and reliability of the research instrument; how to test the validation of a questionnaire/survey in a research. *SSRN Electr. J.* 5, 28–36. doi: 10.2139/ssrn.3205040

Tang, X., Yin, Y., Lin, Q., Hadad, R., and Zhai, X. (2020). Assessing computational thinking: a systematic review of empirical studies. *Comput. Educ.* 148:103798. doi: 10.1016/j.compedu.2019.103798

Tikva, C., and Tambouris, E. (2021). Mapping computational thinking through programming in K-12 education: a conceptual model based on a systematic literature review. *Comput. Educ.* 162:104083. doi: 10.1016/j.compedu.2020.104083

Tsarava, K., Moeller, K., Román-González, M., Golle, J., Leifheit, L., Butz, M. V., et al. (2022). A cognitive definition of computational thinking in primary education. *Comput. Educ.* 179:104425. doi: 10.1016/j.compedu.2021.104425

Varma, S. (2006). *Preliminary Item Statistics Using Point-Biserial Correlation and p-Values*. Morgan Hill, CA: Educational Data Systems Inc.

Vincent, W., and Shanmugam, S. K. S. (2020). The role of classical test theory to determine the quality of classroom teaching test items. *Pedagog. J. Pendid.* 9, 5–34. doi: 10.21070/pedagogia.v9i1.123

Wainer, H., and Thissen, D. (2001). "True score theory: The traditional method," in *Test Scoring*, eds D. Thissen and H. Wainer (Mahwah, NJ: Lawrence Erlbaum Associates), 35–84.

Weintrop, D. (2016). Defining computational thinking for mathematics and science classrooms. *J. Sci. Educ. Technol.* 25, 127–147. doi: 10.1007/S10956-015-9581-5

Weintrop, D., Rutstein, D., Bienkowski, M., and McGee, S. (2021a). "Assessment of computational thinking," in *Computational Thinking in Education, 1st Edn.* (New York, NY: Routledge), 90–111. doi: 10.4324/9781003102991-6

Weintrop, D., Wise Rutstein, D., Bienkowski, M., and McGee, S. (2021b). Assessing computational thinking: an overview of the field. *Comput. Sci. Educ.* 31, 113–116. doi: 10.1080/08993408.2021.1918380

Werner, L., Denner, J., Campe, S., and Kawamoto, D. C. (2012). "The fairy performance assessment: measuring computational thinking in middle school," in *SIGCSE '12: The 43rd ACM Technical Symposium on Computer Science Education* (Raleigh, NC), 215–220. doi: 10.1145/2157136.2157200

Wheaton, B., Muthen, B., Alwin, D. F., and Summers, G. F. (1977). Assessing reliability and stability in panel models. *Sociol. Methodol.* 8, 84–136. doi: 10.2307/270754

Willse, J. T. (2018). *CTT: Classical Test Theory Functions*. R Package Version 2.3.3

Wing, J. M. (2006). Computational thinking. *Commun. ACM* 49, 33–35. doi: 10.1145/1118178.1118215

Xia, Y., and Yang, Y. (2019). RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: the story they tell depends on the estimation methods. *Behav. Res.* 51, 409–428. doi: 10.3758/s13428-018-1055-2

Xie, B., Davidson, M. J., Li, M., and Ko, A. J. (2019). "An item response theory evaluation of a language-independent CS1 knowledge assessment," in *Proceedings of the 50th ACM Technical Symposium on Computer Science Education* (Minneapolis MN: ACM), 699–705. doi: 10.1145/3287324.3287370

Yagci, M. (2019). A valid and reliable tool for examining computational thinking skills. *Educ. Inf. Technol.* 24, 929–951. doi: 10.1007/s10639-018-9801-8

Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Appl. Psychol. Meas.* 8, 125–145. doi: 10.1177/014662168400800201

Zapata-Cáceres, M., and Fanchamps, N. (2021). "Using the beginners computational thinking test to measure development on computational concepts among preschoolers," in *Proceedings of the 5th APSCE International Computational Thinking and STEM in Education Conference 2021* (Delft: Asia-Pacific Society for Computers in Education), 32–37.

Zapata-Cáceres, M., and Martín-Barroso, E. (2021). Applying game learning analytics to a voluntary video game: intrinsic motivation, persistence, and rewards in learning to program at an early age. *IEEE Access.* 9, 123588–123602. doi: 10.1109/ACCESS.2021.3110475

Zapata-Cáceres, M., Martín-Barroso, E., and Román-González, M. (2020). "Computational thinking test for beginners: design and content validation," in *2020 IEEE Global Engineering Education Conference (EDUCON)* (Porto), 1905–1914. doi: 10.1109/EDUCON45650.2020.9125368

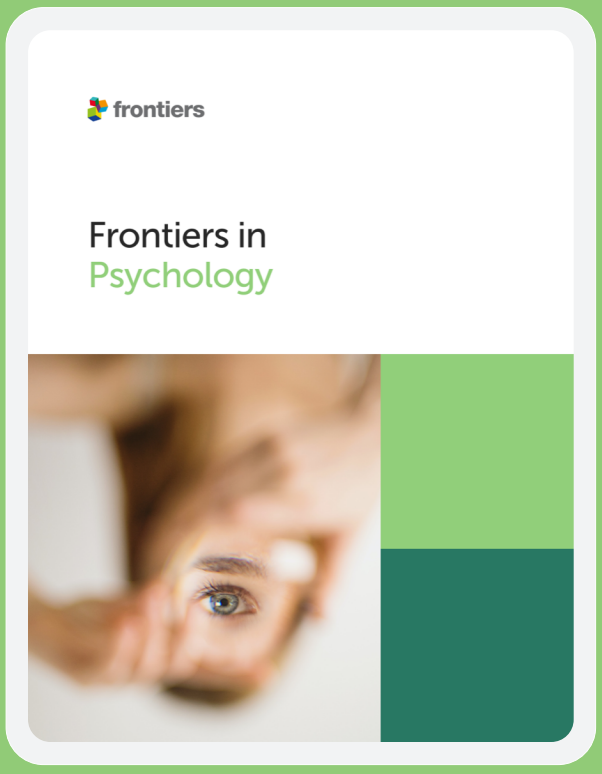
Frontiers in Psychology

Paving the way for a greater understanding of human behavior

The most cited journal in its field, exploring psychological sciences - from clinical research to cognitive science, from imaging studies to human factors, and from animal cognition to social psychology.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact