frontiers in Education

TYPE Original Research
PUBLISHED 04 March 2026
DOI 10.3389/feduc.2026.1759274

OPEN ACCESS

EDITED BY
Guillermo M. Chans,
Institute for the Future of Education, Monterrey Institute of Technology and Higher Education (ITESM), Mexico

REVIEWED BY
Sezer Kanbul,
Near East University, Cyprus
Rusen Meylani,
Dicle University, Türkiye

*CORRESPONDENCE
Johannes Huwer
johannes.huwer@uni-konstanz.de

RECEIVED 02 December 2025
REVISED 14 February 2026
ACCEPTED 18 February 2026
PUBLISHED 04 March 2026

CITATION
Brückner M, Thyssen C and Huwer J (2026) AI for adaptive science teaching: strengthening teacher self-efficacy and perceived usefulness. *Front. Educ.* 11:1759274.
doi: 10.3389/feduc.2026.1759274

AI for adaptive science teaching: strengthening teacher self-efficacy and perceived usefulness

Mathea Brückner[1,2], Christoph Thyssen[3] and Johannes Huwer[1,2]*

[1]Department of Chemistry, Science Education, University of Konstanz, Konstanz, Germany, [2]Department of Chemistry, Science Education, University of Education Thurgau, Kreuzlingen, Switzerland, [3]Department of Biology, Biology Education, University of Education Freiburg, Freiburg, Germany

Integrating Artificial Intelligence (AI) into everyday school practice holds great potential for implementing adaptive teaching. AI-supported tools enable learning processes to be individualized and facilitate a more effective consideration of students' diverse needs. However, to realize these benefits, adequate technical infrastructure, teachers' willingness, and relevant competencies are essential. This pilot study investigates whether a short, targeted intervention can enhance science teachers' Artificial Intelligence Self-Efficacy Expectation (AISEE) and their Perceived Usefulness (PU) of AI in adaptive science teaching. In addition, teachers' conceptual understanding of the adaptive teaching components 'assessment', 'feedback', and 'adaptivity' was examined by asking them to provide descriptive terms for each component. Their responses were analyzed and inductively categorized to gain deeper insights into teachers' understanding of the concepts. The participants were German lower secondary education science teachers in multiplier roles. The results show a significant increase in both PU and AISEE after the intervention and a post-intervention correlation between these two variables. The results underscore the value of hands-on training formats in fostering Self-Efficacy (SE) and PU for AI-supported adaptive science teaching.

KEYWORDS
adaptivity, artificial intelligence, stem education, teacher training, technology acceptance

1 Introduction

In today's classrooms, teachers are increasingly confronted with growing heterogeneity (Lewalter et al., 2023), larger class sizes, and multilingual learning environments. These developments pose major challenges for educational practice, as they require instruction that can flexibly respond to diverse learning needs, interests, and levels of prior knowledge. Adaptive teaching is a promising approach to meeting these demands, as it tailors learning processes to students' individual strengths and needs (Corno, 2008; Decristan et al., 2015; Hardy et al., 2022).

Although adaptive teaching is becoming increasingly important, it remains difficult to implement and systematically capture in everyday school life. This is partly due to the lack of validated instruments for recording adaptive teaching processes and partly due to the inconsistent use and interpretation of the concept itself (Plass and Pawar, 2020; Vaughn et al., 2022). In addition, the high level of complexity and organization demands associated with adaptive teaching pose further challenges, which can hinder its practical application in school settings (Pelánek, 2025). One key to the wider adoption of adaptive teaching may be the use of AI (Celik et al., 2022; Kasneci et al., 2023). AI-based systems offer the potential to automatically track learning progress (Maestrales et al., 2021), create differentiated materials (Celik et al., 2022), and provide personalized feedback (Labadze et al., 2023; Su et al., 2023). All of this would allow teachers to respond more specifically to students' individual needs with less organizational effort. Especially in science education, AI offers great potential, as learners often start with very different levels of prior knowledge in this subject area. Despite this potential, the actual use of AI in everyday school life remains low (Figueroa et al., 2024; Galindo-Domínguez et al., 2024b). Many teachers feel overwhelmed when using AI-based tools or have sufficient knowledge of the didactically appropriate use of such technologies (Ng et al., 2023; Yue et al., 2024). In addition, teachers have legitimate concerns that the uncritical use of AI could lead to a decline in students' critical thinking skills, as the technology may take over certain important cognitive processes (Gerlich, 2025). To address these challenges, professional development and teacher education must play a bigger role. Given current political developments - in particular the training obligations for staff provided in the European Union's AI Act - the discussion about qualification programs for teachers is becoming increasingly relevant (European Parliament, 2024). Targeted programs in teacher training are therefore needed to effectively leverage AI's potential for adaptive teaching.

To better understand how teachers can effectively implement AI-supported adaptive teaching, it is first necessary to clarify the concept of adaptive teaching itself and identify the competencies required for its successful implementation. After all, while adaptive instruction is often cited as a promising approach, it remains unclear to what extent teachers can adequately define and differentiate its key components, such as adaptivity, assessment, and feedback. Against this background, the following theoretical framework outlines the key conceptual foundations relevant to this pilot study.

2 Theoretical background

2.1 Adaptivity

The idea of adapting instruction to learners' individual needs has a long tradition, with early conceptual roots dating back to the 5th century BC (Corno, 2008). Even then, it was recognized that learners have different prerequisites, needs, and approaches to knowledge. These early ideas are also reflected in later pedagogical-psychological concepts. Bandura (1977), for example, provided important theoretical foundations for adaptive teaching with his Social Learning Theory. The focus here is on learning through observation and the importance of self-efficacy expectations - i.e. the conviction that one can carry out one's own actions successfully. High self-efficacy is typically accompanied by increased motivation and greater willingness to work hard. In the school context, this applies not only to students, whose confidence in their own abilities is crucial for learning, but also to teachers, whose self-efficacy strongly influences their instructional practices (Choi and Lee, 2017). Adaptive teaching addresses, among other things, the self-efficacy of students by creating learning environments that take individual requirements into account, provide tailored support, and enable targeted experiences of success. This not only promotes technical skills but also sustainably strengthens learners' confidence in their own abilities. In this context, Vygotskij and Cole (1979) use the theory of the zone of proximal development to describe an understanding of learning that is based on individual development. They emphasize that teachers should recognize students' current competencies and provide targeted support to optimally promote learning. In almost the same period, Glaser (1978) and Snow (1980) also defined adaptive teaching as a means of addressing the individual learning needs of students, both in terms of common and personal learning goals. Building on these theoretical foundations, Corno (2008) understands adaptive teaching as a dynamic process that goes beyond content adaptation. The focus is on the situational response to individual learning needs, combined with the promotion of self-regulatory learning strategies. Corno (2008) thus expands the concept of adaptivity to include a process-oriented perspective in which learners are supported in developing sustainable strategies for problem-solving and independent learning. Adaptivity can be seen on two levels: in micro adaptation, i.e., the immediate reaction to learning processes in the classroom, and in macro adaptation, which includes longer-term decisions planned in advance, such as grouping students or adapting learning materials to different performance levels (Corno, 2008). Both adaptations are made based on diagnostic findings.

Diagnosis in this context refers to the recording of individual learning requirements, interests, and support needs, which form the basis for micro- and macro-level adaptations. Both micro and macro adaptations occur within specific teaching conditions and are closely linked to learners' individual learning prerequisites and the sociocultural context of the learners (Corno, 2008). This perspective is also consistent with Mishra's (2019) notion of contextual knowledge in the TPACK model, which emphasizes that the effective integration of pedagogy, content, and technology is inseparably tied to the concrete educational setting and its contextual factors (Mishra, 2019; Rosenberg and Koehler, 2015). Adaptive teaching is thus understood not only as a methodological approach, but also as a cyclical, pedagogical process that combines individual support and the development of autonomous learning skills (Corno, 2008).

More recent definitions of adaptivity increasingly include technological components (Natriello, 2013; Plass and Pawar, 2020; Shute and Zapata-Rivera, 2012). In this context, Plass and Pawar (2020) define adaptivity as the ability of a system to capture a variety of learning-relevant variables and make appropriate adjustments on this basis in order to optimally support learning processes. Technology-supported adaptive learning systems offer a decisive advantage in this context: they enable continuous adaptation based on the learner's current performance (Natriello, 2013; Shute and Zapata-Rivera, 2012). Such systems can continuously record individual performance levels and then dynamically adapt content, level of difficulty, and teaching methods to optimize the learning process. Especially, the integration of AI has high potential here (Gligorea et al., 2023). Among other things, AI-based systems can analyze strengths and weaknesses, control the learning pace individually and adapt the presentation of content in a targeted manner. This individualized approach can promote engagement, increase motivation, and lead to greater long-term learning retention and improved academic performance (Gligorea et al., 2023; Yang et al., 2022). Using Yang et al. (2022) as an example, the approach of using repeated, adaptive formative assessments could promote intrinsic motivation by supporting the experience of autonomy and the development of competence—two central psychological needs in Ryan and Deci's (2017) Self-Determination Theory. However, if adaptive systems primarily provide directive instructions without any possibility of self-determination, the sense of autonomy may in fact be undermined. Moreover, the need for social connectedness may not be adequately addressed in this approach.

2.2 Assessment

Assessment is a central element of adaptive systems. This refers to the systematic process of recording, evaluating, and interpreting data on learners' knowledge, skills or behavior (Mislevy et al., 2003). The aim is to draw well-founded conclusions about learning progress and skills development. At an individual level, assessment and performance are closely linked, as performance is evaluated based on specific tasks (Willmott, 1978). Assessment can be formative—i.e., accompanying the learning process - and summative - for final evaluation (Plass and Pawar, 2020). Only by continuously collecting this data on knowledge, skills, and behavior can adaptive systems make targeted adjustments that promote learning.

2.3 Feedback

Feedback is often used as a consequence of assessment or as part of an adaptive process or learning scenario. Feedback is informative input provided in response to demonstrated performance or behavior (Hattie and Timperley, 2007). It serves to reduce the discrepancy between the current and intended learning status and is therefore a central instrument for the learning process (Hattie and Timperley, 2007). Effective feedback answers three key questions: "Where am I?" (Feed Back), "Where am I going?" (Feed Up) and "What are the next steps?" (Feed Forward) (Hattie and Timperley, 2007). According to Winne and Butler (1994), feedback can contain information that intervenes in existing knowledge structures in different ways: It can confirm existing knowledge and thus contribute to strengthening security and stability in the learning process. It can also serve to expand knowledge, for example, by introducing new information, pointing out connections, or offering alternative perspectives. Finally, feedback can also initiate a restructuring of knowledge - for example, when it helps to recognize and correct misconceptions. Feedback can be divided into different forms: implicit (e.g., non-verbal signals such as frowning) and explicit (direct verbal or written feedback). It can also be classified according to source (teacher, peers, self-feedback), time aspects, form (verbal, written) or scope (Narciss and Huth, 2002). Hattie and Timperley (2007) also distinguish between four target levels of feedback: task, process, self-regulation and self. Feedback at the task level concerns the accuracy or quality of a result, whereas feedback at the process level addresses strategies and ways of thinking in the learning process. Feedback on self-regulation helps learners to control their own learning, whereas feedback at the self-level - such as praise of the individual - is often considered to be the least effective in terms of learning, as it only contributes to cognitive development to a limited extent. Feedback is used to provide diagnostic findings to learners and is closely related to assessment, as the evaluation of learning performance forms the basis for targeted feedback. Without any kind of prior diagnosis, no targeted feedback could therefore be given.

In summary, assessment forms the basis for both feedback and adaptivity. Without a diagnostic assessment of learning performance (including data on knowledge, skills, and behaviors), neither targeted feedback nor adaptive adjustment of the learning process would be possible. Feedback helps the learner to understand the current learning status and to work specifically on improvement. At the same time, adaptivity aims to design the learning environment so that it (repeatedly) responds optimally to individual needs. Ideally, an adaptive learning scenario is designed as an iterative loop in which learning processes and content are adapted after a continuous diagnosis. The current status and further steps are communicated through appropriate feedback.

2.4 Assessment, feedback, adaptivity and AI in the DiKoLAN framework

With the increasing combination of adaptivity and technological support, teachers are not only required to have pedagogical sensitivity but also need profound knowledge about the didactically reflected use of digital tools - especially if digital systems are to be used to support diagnostic processes or differentiation. Basic digital technologies are already integrated into current didactic frameworks, meaning that technology-unspecific information and communication technology (ICT) skills should be included as an integral part of training in many teacher training courses. For example, TPACK (Koehler et al., 2013) and DigCompEdu (Redecker, 2017) offer initial frameworks for teachers and trainers in dealing with digital technologies in an educational context, in theory also including skills for the use of artificial intelligence. However, a more recent framework, DiKoLAN, 'Digital Competencies for Teaching in Science Education', explicitly addresses not only competencies for science teachers (Kotzebue Kotzebue et al., 2021), but also, in an extension of the framework, the DiKoLAN^{AI}, competencies specifically for a purposeful application of AI in and for science teaching (Huwer et al., 2024). The DiKoLAN framework builds upon the TPACK and DigCompEdu models, adapting and extending them to address the specific demands of science education. DiKoLAN (Kotzebue Kotzebue et al., 2021) delineates the digital competencies of pre-service teachers into seven core domains that are particularly pertinent to teaching in the natural sciences. Within DiKoLAN PLUS (Meier et al., 2025), a distinction is made between the three more subject-specific competence areas 'data acquisition' (DAQ), 'data processing' (DAP) and 'simulation and modeling' (SIM) as well as the four more general competence areas 'documentation' (DOC), 'presentation' (PRE), 'communication/collaboration' (COM), 'information search and evaluation' (ISE) and the newly added area 'assessment, feedback, adaptivity' (AFA) (Meier et al., 2025). Finally, the AI-specific interpretation of DiKoLAN PLUS, the competency framework DiKoLAN^{AI} describes the competencies required to implement adaptive science teaching with AI in the area of 'assessment, feedback and adaptivity'. In addition to the skills listed in DiKoLAN PLUS that are required for the use of AI in the classroom, the individual willingness of teachers also plays a central role. This is where the Technology Acceptance Model (TAM) comes in (Davis, 1986).

2.5 The technology acceptance model (TAM)

According to the Technology Acceptance Model (TAM) (Davis, 1986; Venkatesh and Bala, 2008; Venkatesh and Davis, 2000), a teachers' intention to use a technology such as AI in the classroom is shaped by several key factors. The TAM, originally developed by Davis (1986), was subsequently extended by Venkatesh and Davis (2000) as TAM2 and later also by Venkatesh and Bala (2008) as TAM3, incorporating additional explanatory variables (Figure 1). The

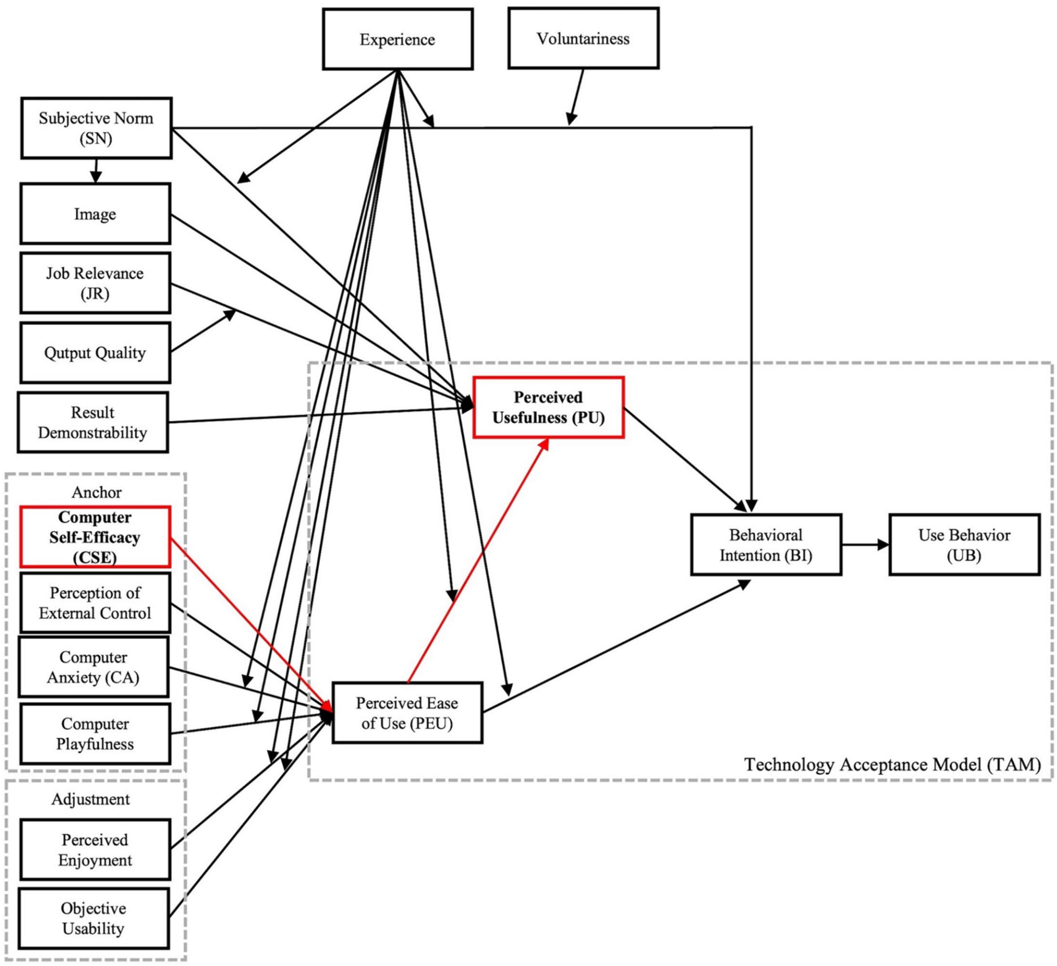

FIGURE 1
The technology acceptance model 3 (based on Venkatesh and Bala, 2008, p. 280) with marked correlations, which were examined in more detail in this pilot study.

Technology Acceptance Model 1 (TAM1) (Davis, 1986) describes the central factors that influence the use and acceptance of technologies. Two of the important factors that can be found in all three TAMs are the Perceived Usefulness (PU) and the Perceived Ease of Use (PEU) of the technology. Venkatesh and Davis (2000) subsequently established further variables mostly to explain the PU such as Subjective Norm (SN) and Job Relevance (JR) in the TAM2. Finally, in an extended version of the TAM according to Venkatesh and Bala (2008), further variables are considered which influence the PEU, including Computer Self-Efficacy (CSE) and Computer Anxiety (CA). Self-Efficacy (SE) describes a person's conviction that they can successfully master a certain task (Bandura, 1978). In their work, Venkatesh and Bala (2008) refer to CSE, i.e., a person's personal belief about the extent to which they are able to use a particular system independently and successfully. In terms of teachers, this means confidence in their own ability to use digital technologies, including AI-supported tools, effectively in the classroom. This expectation significantly influences the PEU, i.e., the subjectively perceived ease of using a technology (Venkatesh and Davis, 1996). People with high SE often perceive digital tools as more intuitive and easier to use (Venkatesh and Davis, 2000). In this context, Holden and Rada (2011) investigated both Technology Self-Efficacy (TSE) and CSE and found that teachers' TSE has a greater influence on the TAM than their CSE. TSE is similar to CSE, but refers specifically to confidence in one's ability to accomplish specific tasks with a particular technology named by the user - as opposed to general computer skills (Holden and Rada, 2011).

At the same time, the PU plays a crucial role. The term 'Perceived Usefulness' refers to the perceived usefulness of a technology for one's own work (Venkatesh and Bala, 2008). According to the TAM, it is positively influenced by the PEU, which means that if teachers perceive a technology as easy to understand, their perception of its actual usefulness for teaching also increases. Both the PU and the PEU in turn directly influence the Behavioral Intention (BI), i.e., the intention to use a technology in the future. The higher the PU, the more likely it is that teachers will integrate the technology into their lessons.

As PEU and PU remain key predictors of the intention to use AI (Ma and Lei, 2024; Nja et al., 2023; Zhang et al., 2023), examining these relationships within the context of science teacher education and the increasing integration of AI into schools and teaching is becoming relevant. AI-based systems offer diverse potential for tailoring learning content to students' individual needs; however, their effective use requires a high degree of acceptance and willingness not only from teachers but also from the students themselves. Although several studies on teachers' acceptance of AI technology are already available (Ma and Lei, 2024; Velli and Zafiropoulos, 2024; Zhang et al., 2023), there is a lack of studies that specifically analyze the factors influencing the acceptance of AI for adaptive use in the classroom. Given the diverse potential that AI offers in this area, such a study appears urgently needed. In addition to the central predictors (PU and PEU) for BI, it is also important to consider the influencing factors of these core variables (PU, PEU) – including self-efficacy (SE) – as well as other potential influencing factors. In summary, a more in-depth examination appears urgently necessary, as AI applications have already become an integral part of many students' daily lives—and thus of everyday school life—often emerging through students' independent use rather than guided by teachers.

3 Rationale

As previously outlined, the integration of AI in schools presents a significant opportunity to advance adaptive teaching. Beyond technical infrastructure, the successful implementation depends crucially on teachers' acceptance of adaptive AI-based systems and competencies to utilize AI-based tools effectively.

In the context of examining adaptive teaching, it is useful to explore teachers' conceptual understanding of adaptivity and related constructs. Moreover, as previously mentioned, PU (Perceived Usefulness) and SE (Self-Efficacy) are considered key factors influencing teachers' BI (Behavioral Intention) to integrate AI into their teaching. Given the specific focus of the present pilot study on AI-supported technologies, Computer Self-Efficacy (CSE) is hereafter conceptualized as Artificial Intelligence Self-Efficacy Expectation (AISEE). Against this background, the question arises as to whether the technology acceptance can be increased through targeted interventions. Furthermore, it is of interest to examine the relationship between key influencing factors (PU and AISEE) in the context of AI-driven design of adaptive science education. From this, the following research questions arise:

> *RQ1*: How do teachers describe the terms 'assessment', 'feedback', and 'adaptivity' in the school context?
>
> *RQ2*: To what extent does a targeted intervention change teachers' Perceived Usefulness of AI for adaptive science teaching?
>
> *RQ3*: To what extent does a targeted intervention influence teachers' AI Self-Efficacy Expectation in using AI for adaptive science teaching?
>
> *RQ4*: Is there a relationship between teachers' AI Self-Efficacy Expectation and their Perceived Usefulness of AI in the context of adaptive science teaching?
>
> *RQ5*: Are there any further relationships between the variables age, years of service, gender, Perceived Usefulness and AI Self-Efficacy Expectation in using AI for adaptive science teaching?

Against this background, the present pilot study investigates the levels of AISEE (AI Self-Efficacy Expectation) and PU (Perceived Usefulness) among science teachers with regard to selected competencies in using AI for adaptive teaching—both before and after a targeted intervention. Furthermore, the study explores which terms teachers primarily associate with the concepts of 'assessment', 'feedback', and 'adaptivity'.

It can be assumed that a targeted intervention may increase both PU and AISEE (Galindo-Domínguez et al., 2024a; Täschner et al., 2025). Based on this assumption, it is expected that teachers' PU of AI for adaptive teaching will increase after the intervention. Furthermore, teachers' AISEE regarding the use of AI for adaptive teaching is also expected to improve as a result of the intervention.

4 Methods

The pilot study was conducted in the context of a two-day professional development event for science teachers. The study reports on two distinct data collections. The first one focuses on exploring terms teachers primarily associate with the concepts of 'assessment', 'feedback', and 'adaptivity' (Section 2.1). This data collection was conducted during a lecture session about 'Competencies for Teachers in Times of Transformation' offered to all participants of the two-day professional development program. The second employed a single-group pre-post design was embedded in a one-hour workshop, which participants could voluntarily choose from a range of available workshops (Section 4.2). Participants were surveyed in a pre-post design regarding the workshop to their AISEE (AI Self-Efficacy Expectation) and their PU (Perceived Usefulness) of AI for adaptive teaching. Below you will find a description of the procedure and participants involved in the word clouds (Section 4.1), followed by more detailed information about the survey conducted in the workshop (Section 4.2).

4.1 Collection of describing terms for 'assessment', 'feedback' and 'adaptivity' in the word cloud

The word cloud data collection relates to research question 1 (RQ1) and, although it took place within the same main event, it was conducted independently of the workshop in terms of both time and location.

4.1.1 Participants

The survey was conducted among teachers who attended a lecture as part of a professional development conference. These teachers also serve as official multipliers in the state of Baden-Württemberg, Germany, in the STEM field for lower secondary education (grades 5–10). Accordingly, all participants teach at least one STEM subject. A total of 61 teachers attended the lecture.

4.1.2 Measurement

During the lecture, participants were asked to name three terms each, to describe the concepts of 'assessment', 'feedback', and 'adaptivity' (e.g., 'Which three terms would you use to describe 'assessment'?').

To collect the responses anonymously, the tool 'AnswerGarden' (Creative Heroes AnswerGarden, 2025) was used. This tool not only generates a word cloud but also includes the possibility to download a CSV file that includes the recorded terms and their frequencies, enabling a precise analysis. Teachers were asked to provide three terms for each word cloud. However, due to system requirements, it was also possible to enter only one or two terms or to leave the word cloud completely blank.

4.1.3 Data analysis

To evaluate the word clouds, the teachers' terms relating to the description of the terms 'assessment', 'feedback', and 'adaptivity' were grouped into inductive categories using a consensus coding procedure (Kuckartz, 2010). We exported all entries from AnswerGarden as a CSV file and processed them in four steps. First, we removed non-meaningful inputs (e.g., "?," "no idea," empty strings) and entries that were truncated to non-identifiable fragments due to the character limit. Second, we standardized spelling variants and obvious typos and harmonized German/English duplicates (e.g., "Rückmeldung/feedback") while retaining the original meaning. Third, we split multi-word entries into single terms when they contained clearly separable concepts (e.g., "individualization and differentiation") but retained multi-word phrases as one unit when they represented a single concept (e.g., "formative assessment"). Fourth, we removed exact duplicates within a single participant's submission; duplicates across participants were retained and counted as repeated mentions. This pipeline resulted in final datasets of $n = 63$ (assessment), $n = 80$ (feedback), and $n = 71$ (adaptivity) valid terms for coding.

For coding, two researchers independently assigned an initial subset of terms to inductively developed categories using a shared codebook. Interrater agreement on this subset was quantified (percent agreement and Cohen's κ). Cohen's Kappa indicated a substantial agreement between the two raters for every term (adaptivity: $\kappa = 0.96$, based on 34 items, assessment: $\kappa = 0.89$, based on 43 items, feedback: $\kappa = 0.90$, based on 48 items). Disagreements and ambiguous cases were then discussed and resolved to establish the final category system and a consensual coding of the full dataset (see Supplementary material of this paper).

4.2 Workshop

The data collected in the workshop serve to answer research questions 2 to 4 (RQ 2–4). It should be noted that the collection of word clouds is described in Section 4.1 took place independently of the workshop, both spatially and temporally. However, the participant group overlapped, as both events were part of the same larger event.

4.2.1 Workshop content

The workshop was divided into a theoretical input section and a practical application phase. At the beginning, participants received a brief overview of where AI already plays a role in the educational context or could play a role in the future (input-phase = 10 min). The discussion highlighted the potential of AI for designing adaptive teaching and why its use is becoming increasingly relevant for teachers, students, and the education system. Emphasis was placed on the connection to the natural sciences: Since AI has long been part of scientific research, it not only serves as a methodological tool but also as an authentic subject matter in STEM education.

As following, a brief introduction was provided on the practical use of AI for a) lesson planning and b) the preparation of teaching materials. Teachers were introduced to useful prompting strategies to effectively interact with AI-based systems (prompting strategies = 10 min). Four common prompting strategies (Zero-Shot-, Few-Shot-, Chain-Of-Thought- and Ask-Before-Answer-Prompting) were selected for the intervention, some of which teachers may already have been using unconsciously. However, the presentation of the respective advantages and disadvantages makes it clear that the wording of the prompts has a significant influence on the output of the AI. Afterwards, they had the opportunity to develop an AI-supported teaching element using one of our prepared examples, either the personalized creation of a tutor bot with poe.com (Quora Poe, 2017) or the use of an LLM of their choice to generate critical statements for a discussion or even an application of their own choosing that they wished to try in their teaching at some point. Specifically, the teachers were tasked with selecting a possible application for AI and trying it out in relation to a teaching topic of their choice, ideally one that they were currently covering with their students (hands-on = 30 min). An interactive mind map was available throughout, allowing participants to review all the topics covered. It also included practical tips, such as how to create true-false tasks or develop a personalized chatbot for the classroom. Finally, the teachers were able to share their products and experiences and discuss their use in an exchange (discussion = 10 min).

4.2.2 Participants

The present pilot study involves science teachers who participated in the described professional development program. These teachers act as multipliers in Baden-Württemberg (Germany) in the STEM field of secondary education (grades 5–10). Multiplier teachers are specially trained educators who share their knowledge and skills with colleagues to effectively spread and implement new pedagogical or technological approaches in schools. In the workshop $n = 39$ science teachers participated (Figure 2). Of these, $n = 35$ participants agreed to the data protection provisions for the pre-survey and $n = 18$ for the post-survey. Questionnaires with incomplete information on demographic data or the subjects taught were not included. This resulted in $n = 30$ pre-surveys and $n = 16$ post-surveys. Only questionnaires that were completed in both the pre- and post-survey were included in the evaluation. The scales for PU (Perceived Usefulness) and AI Self-Efficacy Expectation (AISEE) were analyzed separately. For the PU scale, a complete data set of $n = 16$ participants in the pre- and post-surveys was obtained, of which $n = 13$ teachers also completed the AISEE survey in full. The following describes the demographic characteristics of the sample ($n = 16$) valid for the evaluation of PU. The sample consists of $n = 16$ teachers, who provide complete data for both time points. This sample includes $n = 8$ female and $n = 8$ male teachers. The ages of the participants range from 29 to 57 years, with an average age of 45, with a 95% confidence interval from 40.72 to 49.28. Regarding years of service, the average is 18.19 years. As teachers in Germany generally teach at least two subjects, with a possible combination of two or more sciences, the following subject distribution emerges within the sample: $n = 7$ teach biology, $n = 9$ teach chemistry, $n = 5$ teach physics, and $n = 11$ teach science and technology or comparable science-oriented subjects.

4.2.3 Measurement of AI self-efficacy expectation and perceived usefulness

The survey took place immediately before and after the workshop. All participants in the pilot study gave their informed consent in

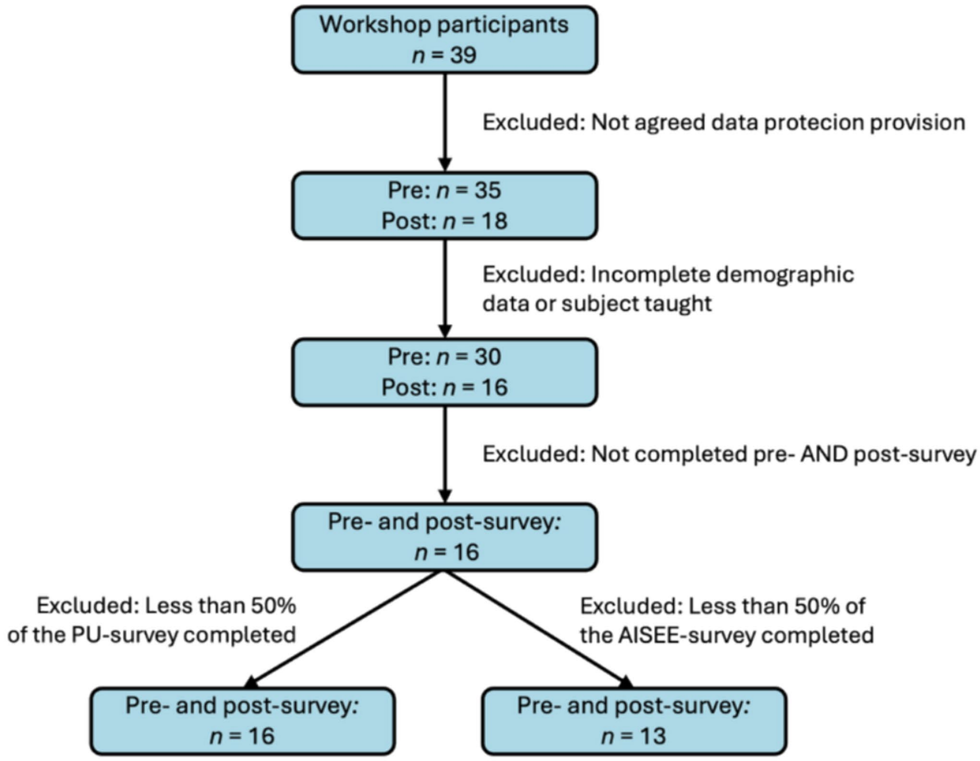

FIGURE 2
Description of the sample based on the analyzed questionnaires.

accordance with applicable data protection regulations prior to the survey. A three-part questionnaire was used as a measurement instrument. In the first section, teachers were asked to generate an individual code to allow for anonymous matching of the pre- and post-tests. This section also included questions on demographic information.

The second part of the questionnaire was based on an excerpt from the TAM by Davis (1986). The PU (Perceived Usefulness) scale was developed by Holden and Rada (2011) and administered using a seven-point Likert scale. This scale was adapted to assess the PU of artificial intelligence in the context of adaptive teaching and translated into German (see Supplementary material 1).

The final section of the questionnaire focused on assessing AISEE (AI Self-Efficacy Expectation) related to the area 'assessment, feedback, adaptivity' (AFA) competencies addressed in the training, based on the DiKoLAN^{AI} competency framework (Huwer et al., 2024; Huwer et al., 2025). The instrument was based on existing and validated questionnaires from the DiKoLAN framework (Kotzebue et al., 2021; Meier et al., 2024). As the DiKoLAN^{AI} follows the same structure, this approach can be transferred accordingly. This procedure was confirmed through an expert interview with the *Working Group Digital Core Competencies*. In this case, the measures were collected using a ten-point Likert scale, which was chosen based on findings indicating that a ten-point format is particularly suitable when the scale range is freely selectable (Preston and Colman, 2000). Henne et al. (2022) further demonstrated that the instrument reliably measures the effects of interventions on the main learning objectives; therefore, for reasons of test economy, the survey focused on these objectives and the corresponding competencies. Specifically, the test included questions about DiKoLAN^{AI} (Huwer et al., 2024) self-efficacy expectations for the following competencies: KI. AFA. T. N1 + 2, KI. AFA. T. B2, KI. AFA. M. N1, KI. AFA. M. B1 + 3, KI. AFA. U. N2. One example would be a self-assessment in the following competency (KI. AFA. M. B3): 'I can describe strategies for using AI-based feedback, tutoring, and assessment/diagnostic technologies in teaching and learning processes'. The complete questionnaire can be found in Supplementary material 1. The datasets analyzed for this study are openly available on Zenodo[1] (Brückner, 2025).

4.2.4 Data analysis

If fewer than 50% of the items in a scale were answered, the corresponding scale score was treated as missing (NA). The data were assessed for normality using the Shapiro–Wilk test and Q-Q plots. For the paired-sample analysis, normality was specifically evaluated for the difference scores between pre- and post-test measurements, as these are the relevant values for paired t-tests, for PU (Perceived Usefulness) and AISEE (AI Self-Efficacy Expectation). We used one-tailed tests for the pre–post comparisons because directional hypotheses (Time 2 > Time 1) were specified *a priori* based on the workshop's objective and the study's assumptions stated in the Rationale section. This decision was made before data analysis to reflect the pre-specified direction of change rather than being chosen *post-hoc*. We chose a one-tailed inferential framework (rather than the more conservative two-tailed alternative) because the pre–post hypotheses were strictly directional and derived from the intervention logic: the workshop was designed to increase PU and AISEE through hands-on engagement, and effects in the opposite direction were not theoretically expected

1 doi: 10.5281/zenodo.16911860

in this pilot context. We acknowledge that one-tailed testing is less conservative; therefore, to support conventional interpretation and demonstrate robustness, we additionally report the corresponding two-sided 95% confidence intervals (and two-tailed p-values) alongside the primary one-tailed results. None of the Shapiro–Wilk tests were significant, indicating no statistically significant deviations from normality. A potential outlier was identified in the Q-Q plot for PU; sensitivity analyses were conducted excluding this case, and the reported results remained statistically significant. Given the acceptable approximation to normality, parametric tests were used as the primary analyses to address the research questions. In addition, Cronbach's alpha was calculated to assess the reliability of the PU (Perceived Usefulness) scale by Holden and Rada (2011) and for the AISEE (AI Self-Efficacy Expectation) scale for both the pre- and post-tests. According to the PU scale, the pre-test yielded a Cronbach's alpha of $\alpha = 0.879$, while the post-test produced a value of $\alpha = 0.899$, indicating high reliability of the scale. The Cronbach's alpha for the AISEE scale was $\alpha = 0.913$ for the pre-test and $\alpha = 0.971$ for the post-test, indicating that some items may have been redundant for participants.

A paired t-test was conducted to compare teachers' AISEE (AI Self-Efficacy Expectation) regarding AI for adaptive teaching and their PU before and after the intervention. Furthermore, a Pearson correlation was calculated to examine associations among age, years of service, AISEE, and PU (RQ4–5). In addition, a point-biserial correlation was used to analyze the relationship between gender and PU (RQ5). A significance level of $\alpha = 0.05$ was applied for all statistical analyses throughout the study. Due to the small sample size, the test power was also specified for each calculation. As nonparametric sensitivity analyses, pre–post changes in PU and AISEE were additionally tested using Wilcoxon signed-rank tests. Associations between PU and AISEE were additionally examined using Spearman's rank correlation (ρ). The correlation between gender and PU was also tested using the Mann–Whitney U test. Results were compared with the parametric analyses to evaluate robustness in this small-sample, Likert-derived dataset. Results.

4.3 Word clouds

To address RQ1, teachers were asked to provide three descriptive words for each of the terms 'assessment', 'feedback', and 'adaptivity'. For the terms 'assessment' and 'adaptivity', there were cases where participants did not provide a description and instead entered terms like 'no idea' or '?'. These responses were assigned to the category no description. Table 1 lists all the categories formed for the terms, which are explained below with examples.

4.3.1 Assessment

A total of $n = 63$ raw entries for 'assessment' were provided, all of which could be integrated into the analysis. The categorization resulted in seven categories: *data evaluation*, *diagnostic (process and format)*, *target of assessment* (educational), *response*, *collaboration*, *no description*, and *other*. A detailed classification of the terms into categories can be found in the Supplementary material 2. The largest category was *data evaluation* ($n = 17$), which included terms such as 'evaluation' (Beurteilung) and 'judgment' (Bewertung). This was followed by *diagnostic (process and format)* ($n = 16$), with examples like 'survey' (Erhebung), 'data collection' (Datensammlung), and 'competence test' (Kompetenzüberprüfung). The categories *target of assessment* [e.g., 'subject knowledge' (Fachwissen), 'competences' (Kompetenzen)) and *(educational) response* (e.g., 'support' (Unterstützung) and 'exchange' (Austausch)] both had $n = 10$ entries. The *(educational) response* category also included feedback-related terms such as 'feedback' ($n = 2$), or the German translation for feedback, 'Rückmeldung' ($n = 3$).

TABLE 1 Overview of all categories related to the terms.

Assessment	Feedback	Adaptivity
Data evaluation ($n = 17$)	Definition ($n = 26$)	Adaptation ($n = 46$)
Diagnostic (process and format) ($n = 16$)	Goal of feedback ($n = 16$)	Variability ($n = 7$)
Target of assessment ($n = 10$)	Type of feedback ($n = 12$)	Implementation ($n = 4$)
(educational) response ($n = 10$)	(interpersonal) mode ($n = 5$)	Effect ($n = 3$)
Collaboration ($n = 3$)	Feedback medium ($n = 5$)	Diagnosis ($n = 2$)
No description ($n = 3$)	Way of communication ($n = 3$)	Tool ($n = 2$)
Other ($n = 4$)	Prerequisite ($n = 3$)	No description ($n = 2$)
	Current situation ($n = 2$)	Other ($n = 5$)
	Other ($n = 8$)	

4.3.2 Feedback

A total of $n = 80$ raw entries for 'feedback' were provided, all of which could be integrated into the analysis. The categorization of responses resulted in nine categories: *definition*, *goal of feedback*, *types of feedback*, *feedback medium*, *(interpersonal) mode*, *way of communication, prerequisite*, *current situation*, and *other*. The largest category was *definition* ($n = 26$), which was primarily represented by the German translation 'Rückmeldung' (feedback), with $n = 23$ entries. This was followed by the category *goal of feedback* ($n = 16$), including examples such as 'improvement' (Verbesserung), 'development' (Entwicklung), and 'support' (Förderung). The categories *feedback medium* (e.g., 'tips'(Tipps), 'grades' (Noten)) and *(interpersonal) mode* [e.g., 'appreciation'(Wertschätzung), 'respect' (Respekt)] were each represented with $n = 5$. In the category *prerequisite*, terms related to diagnosis were also included, such as 'evaluation' (Auswertung) ($n = 2$). A detailed assignment of terms to categories can be found in the Supplementary material 3.

4.3.3 Adaptivity

A total of $n = 73$ raw entries for 'adaptivity' were collected. Two entries were deemed invalid (e.g., unidentifiable words or '.') and were therefore excluded. This resulted in a final sample of $n = 71$ valid entries. The categorization resulted in eight categories: *adaptation*, *variability*, *implementation*, *effect*, *diagnosis*, *tool*, *no description*, and *other*. A total of $n = 71$ terms were included. The largest category is *adaptation* ($n = 46$). The most frequently mentioned term within this category was 'adaptation' (Anpassung) ($n = 19$), which also serves as the category's defining term, followed by terms like 'individualization' (Individualisierung) ($n = 13$) and 'differentiation' (Differenzierung) ($n = 3$). The next-largest category is *variability* ($n = 7$), with examples such as 'flexibility' (Flexibilität), 'malleability' (Formbarkeit), and 'changeable' (veränderbar). The category *implementation* (e.g., 'self-directed' (selbstbestimmt), 'social format' (Sozialform)) and the category *effect* (e.g., 'support' (Förderung), 'appropriate learning' (angemessenes Lernen) was each represented with $n = 4$ and $n = 3$ terms, respectively. The *diagnosis* category was represented by only $n = 2$ entries, including 'self-evaluation' (Selbstevaluation) and 'comparison' (Abgleich). The category *tools* is represented in the same size ($n = 2$). A detailed classification of the terms into categories is provided in Supplementary material 4.

4.4 Workshop

4.4.1 Correlation matrix

Table 2 presents the bivariate relationships among the key variables examined in the workshop questionnaire results. Significant positive correlations were found between gender and PU (Perceived Usefulness) in the pre-test. Furthermore, a significant relationship was observed both before and after the intervention between AISEE (AI Self-Efficacy Expectation) and PU. The pre- and post-test values of these two variables also show a significant correlation.

4.4.2 Increase of perceived usefulness after the intervention (RQ2)

The following section examines whether and to what degree the targeted intervention influenced teachers' PU (Perceived Usefulness) of AI for adaptive science teaching. A paired-samples t-test showed significantly higher values for the PU of AI use for adaptive teaching after the workshop [$t(15) = 2.65$, $^{**}p = 0.009$, $d = 0.66$]; the one-tailed 95% confidence interval for the mean difference (Time 1- Time 2) had an upper bound of −0.22. This means that after the training ($M = 5.15$, $SD = 1.03$), participants rated the usefulness significantly higher than before ($M = 4.50$, $SD = 1.07$) (Figure 3). The median also increased from 4.40 (pre) to 5.17 (post). The effect size of $d = 0.66$ corresponds to a medium effect (Cohen, 1992). The statistical power ($1 - \beta$) of the conducted analyses was 0.81, corresponding to a β error of 0.19. A two-sided Wilcoxon signed-rank test confirmed significant pre–post increases in PU ($V = 14$, $^{*}p = 0.030$, $r = 0.54$), with a 95% confidence interval from −1.42 to −0.08.

4.4.3 Increase of AI self-efficacy expectation after the intervention (RQ3)

The results address the question of whether the targeted intervention influenced teachers' AISEE (AI Self-Efficacy Expectation) in the context of adaptive science teaching. The paired-samples t-test of the AISEE scores shows a significant increase after the professional development intervention [$t(12) = 3.15$, $^{**}p = 0.004$, $d = 0.87$] (Figure 4); the one-tailed 95% confidence interval for the mean difference (Time 1- Time 2) had an upper bound of −0.52. The mean score increased from $M = 4.94$ ($SD = 1.77$) in the pre-test to $M = 6.13$ ($SD = 1.62$) in the post-test. Similarly, the median increased significantly from 4.33 to 6.57. The effect size of $d = 0.87$ indicates a strong effect according to Cohen (1992). The statistical power (1 − β) of the conducted analyses was 0.90, corresponding to a β error of 0.10. A two-sided Wilcoxon signed-rank test confirmed significant pre–post increases in AISEE ($V = 10.5$, $^{*}p = 0.015$, $r = 0.67$), with a 95% confidence interval from −2.0 to −0.43.

4.4.4 Relationship between AI self-efficacy expectation and perceived usefulness (RQ4)

A Pearson correlation was conducted to examine the relationship between AISEE (AI Self-Efficacy Expectation) and PU (Perceived Usefulness) scores before and after the workshop. A moderate positive correlation ($r = 0.54$, $p = 0.055$, $n = 13$) between AISEE (AI Self-Efficacy Expectation) and PU (Perceived Usefulness), with a 95% confidence interval from −0.01 to 0.84 was observed in the pre-test with a low statistical power (1 − β) of 0.51, corresponding to a β error of 0.49. After the intervention (in the post-test), a significant correlation between AISEE and PU was found ($r = 0.93$, $p < 0.001$, $n = 13$) (Figure 5), with a 95% confidence interval from 0.79 to 0.98. The statistical power (1 − β) of the conducted analyses was high with >0.99, corresponding to a β error of <0.01. A Spearman correlation corroborated the post-test association between AISEE and PU ($p < 0.001$, $\rho = 0.76$, $n = 13$), with a 95% confidence interval from 0.55 to 0.89. Figure 4 shows a scatter plot with regression line illustrating the relationship between AISEE and PU.

4.4.5 Further results (RQ5)

Due to the notable pattern observed in the correlation matrix (Table 2), a Pearson correlation was conducted to examine the relationship between gender and PU (Perceived Usefulness) score. The

TABLE 2 Correlation matrix.

Variable	(1)	(2)	(3)[1]	(4)	(5)	(6)	(7)
(1) Age							
(2) Years of service	0.95**						
(3) Gender[1]	−0.16	−0.29					
(4) PU (Pre)	−0.21	−0.21	0.60*				
(5) PU (Post)	−0.01	0.11	0.33	0.56*			
(6) AISEE (Pre)	0.08	0.21	0.20	0.54	0.68*		
(7) AISEE (Post)	−0.23	−0.09	0.22	0.48	0.93**	0.77**	

$^{*}p < 0.05$; $^{**}p < 0.01$; [1]coded according to 0 = female and 1 = male.

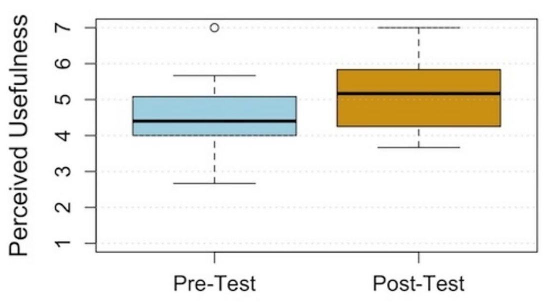

FIGURE 3
Boxplot of perceived usefulness scale scores (pre-post comparison).

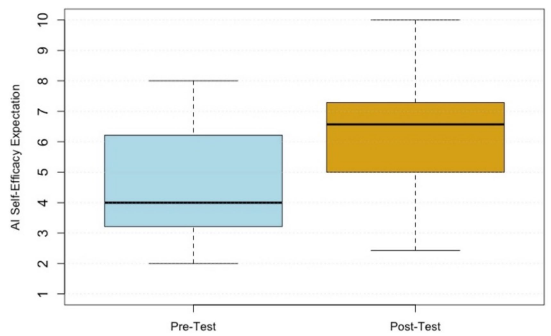

FIGURE 4
Boxplot of AI self-efficacy expectation scale scores (pre-post comparison).

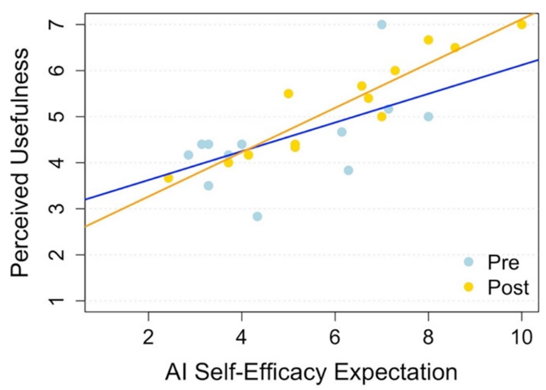

FIGURE 5
Scatter plot and regression of the scale values of AI self-efficacy expectation to perceived usefulness.

analysis revealed a significant relationship between gender and PU (Perceived Usefulness) before the intervention ($r = 0.60$, $^*p = 0.015$, $n = 16$) (statistical power = 0.73), with a 95% confidence interval ranging from 0.14 to 0.84. After the intervention this correlation was no longer significant ($r = 0.33$, $p = 0.218$, $n = 16$) (statistical power = 0.24), with a 95% confidence interval from −0.20 to 0.71. Given the multiple comparisons involved, this analysis should be considered exploratory. The results were confirmed using a non-parametric Mann–Whitney U test, which yielded a similar pattern of findings (Pre: $W = 10$, $^*p = 0.023$, $r = 0.58$; Post: $W = 22$, $p = 0.318$, $r = 0.26$), with a 95% confidence interval for pre from −2.33 to −0.20 and for post from −2.0 to −0.67. Overall, the reported results indicate that men had significantly higher PU scores than women prior to the intervention.

5 Discussion

The majority of multipliers surveyed in the workshop also took part in the subsequent survey on word clouds, which was supplemented by additional (multiplier) teachers. Because many participants served as multiplier teachers, it is plausible that workshop materials and experiences may be shared with colleagues within their professional networks. However, the present pilot data do not allow conclusions about dissemination, transfer, or impact beyond the immediate workshop context.

To outline teachers' understanding of the terms (RQ1), the results of the word clouds were analyzed. The analysis of the word clouds for the terms 'assessment', 'feedback' and 'adaptivity' provides information about which aspects teachers emphasize in these concepts. In the survey, the number of terms per word cloud varied. While teachers were asked to provide three terms, the system also allowed one, two, or no entries. Additionally, the intended procedure was for terms to be submitted individually, but in some cases, not all three terms were entered because some teachers did not understand this. The input field's word limit also caused some terms to be truncated or not submitted when multiple terms were entered simultaneously. The categorization of the terms revealed that teachers primarily associate 'assessment' with the categories of data evaluation and diagnostic (process and format) (cf. Table 1). For data evaluation, the most frequently mentioned terms were 'evaluation' (Beurteilung) ($n = 7$), followed by 'estimation' (Einschätzung) ($n = 3$) and 'judgment' (Bewertung) ($n = 2$). This shows that assessment is primarily understood as the outcome of a performance evaluation. Terms from the categories diagnostic (process and format) and target of assessment were also frequently mentioned, which indicates that teachers also highlight that assessment is interpreted as an instrument for collecting data about knowledge and skills. The mentioned terms in the category (educational) response ($n = 10$), like 'support' (Unterstützung) 'assistance' (Hilfestellung) or 'exchange (information)' (Austausch), show that teachers do not see the assessment process exclusively for the purpose of summative diagnosis, but rather as an opportunity to react to the findings from the (formative) assessment. In contrast, current studies revealed, that in everyday school life (formative) assessment is still rarely used in many places, or not used with sufficient intensity (Baran-Łucarz, 2019; Johnson et al., 2019). However, whether the teachers surveyed prefer to use assessment for formative or summative diagnosis would need to be asked more specifically.

Turning to the term 'feedback', it was not surprising that the German equivalent 'Rückmeldung' was by far the most frequently mentioned response ($n = 23$), as it is the most immediate and intuitive association when prompted with this term. A few teachers explicitly linked feedback to more assessment-linked terms within the category prerequisite, suggesting a reasoned understanding that effective feedback should always relate to a concrete target, which, in practice, typically presupposes some form of assessment. The categories goal of feedback [represented by, e.g., 'improvement' (Verbesserung), 'support' (Förderung)] and feedback medium [represented by, e.g., 'tips' (Tipps), 'conversation' (Gespräch)] make it clear that feedback is primarily seen as a supportive measure for the further development of learners. At the same time, however, some terms were also mentioned that are more associated with summative aspects, such as 'grades' (Noten), 'conclusive' (abschließend) or 'summative' (summativ). The teachers also implied statements such as 'coaching' or 'participation' for the description of 'feedback', which aligns with Haughney et al. (2020) view that effective feedback involves actively engaging learners and providing them with agency after it is delivered. The term 'appreciation' (Wertschätzung) as well as terms such as 'respect' (Respekt) and 'empathy' (Empathie) were also mentioned by some teachers, making it clear that feedback as a process has not only a cognitive but also an emotional dimension in the perception of teachers and, as in other studies, is perceived as a means of shaping relationships between teachers and students (Carless and Winstone, 2023).

Given that there is no universally accepted definition of adaptivity even within the scientific community (Parsons et al., 2018; Plass and Pawar, 2020), it is to be expected that a uniform understanding is also absent among teachers. Consequently, the analysis of the descriptive terms in the 'adaptivity' word cloud represents a useful initial step in capturing teachers' conceptualizations of adaptivity. The high number of mentions in the category adaptation ($n = 46$) in the word cloud suggests that teachers clearly recognize the core of the concept, namely the adaptation of instruction and responsiveness to students´ individual needs. Grant and Basye (2014) also show that teachers associate adaptivity with terms such as individualization and differentiation. These terms were also mentioned by participants in this pilot study, with 'individualization' (Individualisierung) ($n = 13$) and 'differentiation' (Differenzierung) ($n = 3$). The high number of mentions of the term individualization confirms that teachers' understanding of this term also includes its proximity to consideration of the individual level, which is strongly emphasized in many common definitions (Corno, 2008; Plass and Pawar, 2020). As adaptivity is closely linked to technological systems in many definitions (Natriello, 2013; Plass and Pawar, 2020; Shute and Zapata-Rivera, 2012), it is noticeable that only a few participants explicitly mentioned this connection by naming concrete tools. Interestingly, the tools mentioned, such as 'Bettermarks' and 'Duolingo', both rely on AI to enable adaptive learning experiences. Similarly, the term 'adaptivity' was rarely used in diagnostic contexts within the adaptive process. This could indicate that teachers understand adaptivity primarily as a punctual or situation-dependent pedagogical measure, while the continuously diagnostically controlled (in form of formative assessment) and iterative character of adaptive processes has so far received less attention. The results indicate that teachers are generally able to categorize key terms but do not primarily focus on conceptual terms or technological connections in initial descriptions, possibly because these are less prevalent in everyday school life or were not cognitively activated in the survey context.

Given the small sample size, the data were analyzed using both parametric and nonparametric methods. All nonparametric tests support the same conclusions as the parametric analyses. For RQ2–4, the PU (Perceived Usefulness) and the AISEE (AI Self-Efficacy Expectation) of AI for adaptive science teaching among teachers were recorded during a workshop. Although the workshop was not designed to provide a comprehensive insight into all possible uses of AI in adaptive science teaching due to the limited time available, significant increases in both AISEE and PU were observed (RQ2 + 3). The results provide initial evidence of short-term, self-reported change following brief engagement with the topic, within this pilot workshop context. Other studies also confirm the effectiveness of compact workshop formats (O'Meara and Faulkner, 2022; Zenni and Turner, 2021). However, it should be noted that the transfer of training content to the school use depends to a large extent on the intensity of its application (Lipowsky, 2010; Vigerske, 2017): the more intensively teachers put the content into practice, the more likely it is that teachers attempt to apply the content in their professional context and potentially share it with colleagues. Gaining practical experience is also relevant, as it can strengthen participants' self-efficacy (Bray-Clark and Bates, 2003). Even though attempts were made to create opportunities for practical experimentation within the limited time available, the brevity of the intervention could nevertheless have a negative effect on the sustainability of implementation and the long-term transfer of the content taught. Furthermore, Vigerske (2017) shows that the self-efficacy is a key factor influencing the decision to transfer training content: teachers who are confident that they can implement the content taught at school are more likely to make a positive transfer decision. Vigerske (2017) also emphasizes the importance of colleagues, which means that a joint implementation by the colleagues significantly increases the quality of transfer. The role of multiplier teachers, who actively bring continuing education content into the collegium, is particularly important here. As the intervention was primarily attended by multiplier teachers and an increase in self-efficacy expectations was observed, this may indicate potentially favorable conditions for transfer, which should be examined in future studies using follow-up measures and classroom-level indicators.

A further finding emerged from the analysis of the relationship between the AISEE (AI Self-Efficacy Expectation) and the PU (Perceived Usefulness) of AI for adaptive teaching (RQ4): A moderate positive correlation ($r = 0.54$) between AISEE and PU was observed before the intervention, but it was not statistically significant, so the relationship should be interpreted with caution. After the intervention, the correlation between AISEE and PU remains significant, suggesting that after the workshop, participants with higher AISEE also tend to report higher PU. Given the small pilot sample and shared-method measurement, this association should be interpreted cautiously and does not imply a directional or causal relationship. The TAM (Technology Acceptance Model) provides a suitable theoretical frame of reference for appropriately classifying the results (Davis, 1986; Venkatesh and Bala, 2008). According to the TAM, self-efficacy (in this case the AISEE) do not directly influence PU, but rather PEU (Perceived Ease of Use). However, in this pilot study only PU was assessed, as it has a strong impact on BI (Behavioral Intention) (Teo, 2009; Venkatesh and Bala, 2008; Zuo et al., 2025), while PEU typically affects BI indirectly through PU; additionally, PEU could not be measured uniformly because participants used different AI-based systems. The significant post-intervention correlation between AISEE and PU suggests that higher levels of AISEE may be related to a higher PU, which may reflect co-variation in self-reports within this specific context. As shown in other studies where PU is described as a strong or even the strongest predictor of BI (Behavioral Intention) (Teo, 2009; Venkatesh and Bala, 2008; Zuo et al., 2025), the observed association between AISEE and PU, together with the increases in these measures, might tentatively indicate a possible influence on BI (Behavioral Intention). However, a more extensive empirical investigation would be required to test the structural relationships among these variables, going beyond the scope of the present pilot study.

Overall, the intervention can be characterized as hands-on, practical experience. In TAM-related research, hands-on experience is discussed as a factor that may relate to perceptions such as perceived usefulness (PU). Accordingly, future teacher education formats could be designed to provide structured opportunities for practical engagement and reflection, and should evaluate whether such experiences are associated with sustained changes and classroom transfer using follow-up assessments and behavioral/performance indicators. The present pilot study is consistent with this line of reasoning but cannot determine whether the brief intervention influenced actual classroom implementation, as no follow-up assessment of AI use was conducted.

Furthermore, it became apparent that there was a gender difference prior to the intervention (RQ5). Male participants gave significantly higher assessments of the PU of AI for adaptive teaching than female participants. This value converged after the intervention. The findings indicate that the observed gender difference in PU was smaller in the post-test than in the pre-test. Because this is a single-group, short-term observation, it remains unclear whether this pattern reflects a workshop-related change, regression to the mean, or other contextual factors. However, it remains unclear whether this convergence is only a short-term effect or whether it will also be observed in the long term. The decline in gender differences could be related either to the free choice of AI applications, which allowed participants to engage with tools that suited their own teaching contexts, and the wide range of possible uses demonstrated. Additionally, the open discussion format at the end of the workshop may have provided a valuable space for shared reflection and the exchange of both critical and encouraging experiences, potentially leveling initial uncertainties or biases. Other studies have also identified gender-specific differences in attitudes toward artificial intelligence (Aldasoro et al., 2024; Zhang et al., 2023). It would therefore be desirable for future studies to also look at the extent to which certain event formats can help to reduce gender-specific differences.

6 Implications

The present pilot study provides preliminary evidence of short-term, self-reported change in teachers' attitudes and beliefs related to using AI for adaptive teaching following a brief, practice-oriented workshop. Given the single-group pre–post design and reliance on self-report, these results should not be interpreted as evidence of causal impact or effectiveness beyond this setting. Instead, they suggest that AI-related teacher education formats that include structured hands-on phases may be a promising direction for supporting teachers' perceived usefulness and (AI) self-efficacy.

In this sense, future teacher education curricula could consider integrating recurring opportunities for practical experience with AI across phases of teacher training, while evaluating whether such experiences are associated with sustained changes and classroom transfer (Vigerske, 2017; Vogelsang et al., 2019). Similarly, low-threshold, experience-oriented formats that allow teachers to develop and reflect on their own skills in dealing with AI may help to build experience-based confidence, which prior research discusses as relevant for (AI) self-efficacy (Bray-Clark and Bates, 2003; Gale et al., 2021; Vogelsang et al., 2019; Wray et al., 2022).

Moreover, teacher education and school-based professional learning should provide space for collegial exchange—not only to enable multiplier teachers to share workshop content, but also to share experiences (Siciliano, 2016), clarify misunderstandings, and explicitly address ethical considerations and potential risks. While the present study cannot address long-term outcomes, such structures could be explored in future work as potential supports for sustained competence development, self-efficacy, and acceptance, which may facilitate AI integration in everyday school practice. Overall, the present findings should be read as preliminary and context-specific, reflecting short-term self-reported change under a single-group pre–post design rather than causal evidence of effectiveness.

7 Limitations

One limitation of the pilot study is the absence of follow-up or a subsequent survey of teachers, which precludes assessing the sustainability of the skills taught in the intervention. At the same time, the positive changes in PU (Perceived Usefulness) and AISEE (AI Self-Efficacy Expectation) show that even short interventions can be effective. For this reason, this limitation should be considered in the context of the study's objectives. Secondly, the teachers participated in the training voluntarily, suggesting that their expectations regarding AISEE and PU may already be higher at baseline than in a representative sample of teachers, as they selected the workshop topic independently. Thirdly, the sample size was comparatively small, which limits the generalizability of the findings. An *a priori* calculation of the sample size for more meaningful analyses would not have been useful in this case, as the number of participants for this single data collection date could not be determined in advance. It is fortunate that teachers could be surveyed at all, as such events are rare. For this reason, the present study was conducted as a pilot study. In addition to the statistical analyses, the test statistic was specified in each case to ensure the significance of the results was transparent. Additionally, we must consider potential general conference effects, social desirability, novelty or Hawthorne effects, participants' awareness of the workshop's purpose, and measurement sensitization from completing the pre-test. Moreover, in the survey in which teachers were asked to provide three terms each for a word cloud on the topics of 'assessment', 'feedback', and 'adaptivity', the intention was that the terms be submitted individually. However, in some cases, not all three terms were entered because some teachers did not understand this procedure. In addition, the input field's word limit meant that, when attempting to enter multiple terms at once, individual terms were truncated or not submitted. Furthermore, the study was limited to two central constructs of the TAM (AISEE and PU), which does not allow for a comprehensive modeling of the acceptance process. It should also be noted that established framework models such as TAM can reach their limits in the context of AI, as they do not adequately reflect the specific ethical, educational, institutional, and technical complexities of AI applications. Future studies should therefore also include scales for other factors influencing the TAM in the questionnaire to obtain a more comprehensive picture of AI acceptance in the school context. Finally, it should be noted that no comparative values for the AISEE in conjunction with the DiKoLAN^{AI} competencies (Huwer et al., 2024) have been available to date, making it impossible to compare the values collected with other empirical findings.

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 in the article/Supplementary material.

Ethics statement

Ethical approval was not required for the studies involving humans because all participants were teachers at the Bundesland Baden-Württemberg. They took part voluntarily and with informed consent. Pseudonymization of the participants was ensured during the study. Due to all these measures in the conduct of the study, an audit by an Ethics Committee was waived. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

MB: Conceptualization, Data curation, Writing – review & editing, Methodology, Writing – original draft, Investigation, Visualization, Formal analysis. CT: Supervision, Writing – review & editing, Project administration, Validation. JH: Validation, Resources, Supervision, Funding acquisition, Project administration, Writing – review & editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This research was funded by the Federal Ministry of Education and Research (BMFTR), project "MINT-ProNeD" (grant number 01JA23M02K).

Acknowledgments

We would like to thank all participating teachers for their time and valuable insights, which made this study possible. Special thanks go to the Center for School Quality and Teacher Education (Zentrum für Schulqualität und Lehrerbildung, ZSL) for its support in organizing the workshops, and to the Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF) for its financial support.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that Generative AI was used in the creation of this manuscript. For translation purposes, the AI-based tool DeepL was used. ChatGPT (versions GPT-4 and GPT-5-mini, OpenAI) was used to support the stylistic refinement and phrasing of selected sentences. All conceptual thinking, data analysis, interpretation, and argumentation were entirely conducted by the author, who assumes full responsibility for the intellectual content of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

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.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feduc.2026.1759274/full#supplementary-material

References

Aldasoro, I., Armantier, O., Doerr, S., Gambacorta, L., and Oliviero, T. (2024). The gen AI gender gap. *Econ. Lett.* 241:111814. doi: 10.1016/j.econlet.2024.111814

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (1978). Self-efficacy: toward a unifying theory of behavioral change. *Adv. Behav. Res. Ther.* 1, 139–161. doi: 10.1016/0146-6402(78)90002-4

Baran-Łucarz, M. (2019). Formative assessment in the English as a foreign language classroom in secondary schools in Poland. Report on a mixed-method study. *J. Educ. Cult. Soc.* 10:327. doi: 10.15503/jecs20192.309.327

Bray-Clark, N., and Bates, R. (2003). Self-efficacy beliefs and teacher effectiveness: implications for professional development. *Prof. Educ.* 26, 13–22.

Brückner, M. (2025). Survey of science teachers on AI self-efficacy expectations, perceived usefulness, and expectations of colleagues' AI skills for adaptive teaching. *Zenodo*. doi: 10.5281/zenodo.16911860

Carless, D., and Winstone, N. (2023). Teacher feedback literacy and its interplay with student feedback literacy. *Teach. High. Educ.* 28, 150–163. doi: 10.1080/13562517.2020.1782372

Celik, I., Dindar, M., Muukkonen, H., and Järvelä, S. (2022). The promises and challenges of artificial intelligence for teachers: a systematic review of research. *TechTrends* 66, 616–630. doi: 10.1007/s11528-022-00715-y

Choi, E., and Lee, J. (2017). EFL teachers' self-efficacy and teaching practices. *ELT J.* 72, 175–186. doi: 10.1093/elt/ccx046

Cohen, J. (1992). A power primer. *Psychol. Bull.* 112, 155–159. doi: 10.1037/0033-2909.112.1.155

Corno, L. (2008). On teaching adaptively. *Educ. Psychol.* 43, 161–173. doi: 10.1080/00461520802178466

Creative Heroes AnswerGarden (2025). Available online at: answergarden.ch (accessed August 20, 2025).

Davis, F. (1986). *A technology acceptance model for empirically testing new end-user information systems: Theory and results*. Cambridge, MA: Massachusetts Institute of Technology.

Decristan, J., Hondrich, A., Büttner, G., Hertel, S., Klieme, E., Kunter, M., et al. (2015). Impact of additional guidance in science education on primary students' conceptual understanding. *J. Educ. Res.* 108:957. doi: 10.1080/00220671.2014.899957

European Parliament (2024). *Artificial Intelligence Act*. Strasbourg: The European Parlament and the Council of the European Union.

Figueroa, E., Batista, E., Palau, R., Unciti, O., Ferre, M., and Martínez-Ballesté, A. (2024). The use of artificial intelligence techniques in smart classrooms is in its infancy. *IEEE Access* 12, 125179–125193. doi: 10.1109/ACCESS.2024.3454372

Gale, J., Alemdar, M., Cappelli, C., and Morris, D. (2021). A mixed methods study of self-efficacy, the sources of self-efficacy, and teaching experience. *Front. Educ.* 6:750599. doi: 10.3389/feduc.2021.750599

Galindo-Domínguez, H., Delgado, N., Campo, L., and Losada, D. (2024a). Relationship between teachers' digital competence and attitudes towards artificial intelligence in education. *Int. J. Educ. Res.* 126:102381. doi: 10.1016/j.ijer.2024.102381

Galindo-Domínguez, H., Delgado, N., Losada, D., and Etxabe, J.-M. (2024b). An analysis of the use of artificial intelligence in education in Spain: the in-service teacher's perspective. *J. Digit. Learn. Teach. Educ.* 40, 41–56. doi: 10.1080/21532974.2023.2284726

Gerlich, M. (2025). AI tools in society: impacts on cognitive offloading and the future of critical thinking. *Societies* 15:6. doi: 10.3390/soc15010006

Glaser, R. (1978). Adaptive education: individual diversity and learning. *J. Sch. Psychol.* 16, 284–285. doi: 10.1016/0022-4405(78)90014-6

Gligorea, I., Cioca, M., Oancea, R., Gorski, A.-T., Gorski, H., and Tudorache, P. (2023). Adaptive learning using artificial intelligence in e-learning: a literature review. *Educ. Sci.* 13:1216. doi: 10.3390/educsci13121216

Grant, P., and Basye, D. (2014). *Personalized learning: A guide for engaging students with technology*. Eugene: International Society for Technology.

Hardy, I., Meschede, N., and Mannel, S. (2022). Measuring adaptive teaching in classroom discourse: effects on student learning in elementary science education. *Front. Educ.* 7:1041316. doi: 10.3389/feduc.2022.1041316

Hattie, J., and Timperley, H. (2007). The power of feedback. *Rev. Educ. Res.* 77, 81–112. doi: 10.3102/003465430298487

Haughney, K., Wakeman, S., and Hart, L. (2020). Quality of feedback in higher education: a review of literature. *Educ. Sci.* 10:60. doi: 10.3390/educsci10030060

Henne, A., Möhrke, P., Thoms, L.-J., and Huwer, J. (2022). Implementing digital competencies in university science education seminars following the DiKoLAN framework. *Educ. Sci.* 12:356. doi: 10.3390/educsci12050356

Holden, H., and Rada, R. (2011). Understanding the influence of perceived usability and technology self-efficacy on teachers' technology acceptance. *J. Res. Technol. Educ.* 43, 343–367. doi: 10.1080/15391523.2011.10782576

Huwer, J., Becker-Genschow, S., Thyssen, C., Thoms, L.-J., Kotzebue, L. V., Finger, A., et al. (2024). *Kompetenzen für den Unterricht mit und über Künstliche Intelligenz*. Münster: Waxmann.

Huwer, J., Thyssen, C., Becker-Genschow, S., von Kotzebue, L., Finger, A., Kremser, E., et al. (2025). Competencies for teaching with and about artificial intelligence in the natural sciences—DiKoLAN AI. *Comput. Educ. Open* 9:100303. doi: 10.1016/j.caeo.2025.100303

Johnson, C. C., Sondergeld, T. A., and Walton, J. B. (2019). A study of the implementation of formative assessment in three large urban districts. *Am. Educ. Res. J.* 56, 2408–2438. doi: 10.3102/0002831219842347

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learn. Individ. Differ.* 103:102274. doi: 10.1016/j.lindif.2023.102274

Koehler, M. J., Mishra, P., and Cain, W. (2013). What is technological pedagogical content knowledge (TPACK)? *J. Educ.* 193, 13–19. doi: 10.1177/002205741319300303

Kotzebue, L.v., Meier, M., Finger, A., Kremser, E., Huwer, J., Thoms, L.-J., et al. (2021). The framework DiKoLAN (digital competencies for teaching in science education) as basis for the self-assessment tool DiKoLAN-grid. *Educ. Sci.* 11:775. doi: 10.3390/educsci11120775

Kuckartz, U. (2010). "Die Kategorien und das Codieren von Texten" in *Einführung in die computergestützte Analyse qualitativer Daten*. ed. U. Kuckartz (Wiesbaden: VS Verlag für Sozialwissenschaften), 57–71.

Labadze, L., Grigolia, M., and Machaidze, L. (2023). Role of AI chatbots in education: systematic literature review. *Int. J. Educ. Technol. High. Educ.* 20:56. doi: 10.1186/s41239-023-00426-1

Lewalter, D., Diedrich, J., Goldhammer, F., Köller, O., and Reiss, K. (2023). *PISA 2022: Analyse der Bildungsergebnisse in Deutschland*. Münster: Waxmann.

Lipowsky, F. (2010). "Lernen im Beruf – Empirische Befunde zur Wirksamkeit von Lehrerfortbildung" in *Lehrerinnen und Lehrer lernen – Konzepte und Befunde zur Lehrerfortbildung*. eds. F. Müller, A. Eichenberger, M. Lüders and J. Mayr (Münster: Waxmann), 51–72.

Ma, S., and Lei, L. (2024). The factors influencing teacher education students' willingness to adopt artificial intelligence technology for information-based teaching. *Asia Pac. J. Educ.* 44, 94–111. doi: 10.1080/02188791.2024.2305155

Maestrales, S., Zhai, X., Touitou, I., Baker, Q., Schneider, B., and Krajcik, J. (2021). Using machine learning to score multi-dimensional assessments of chemistry and physics. *J. Sci. Educ. Technol.* 30, 239–254. doi: 10.1007/s10956-020-09895-9

Meier, M., Thyssen, C., Becker-Genschow, S., Bruckermann, T., Finger, A., Huwer, J., et al. (2025). "Orientierungsrahmen Digitale Kompetenzen für das Lehramt in den Naturwissenschaften – DiKoLAN PLUS" in *Digitale Kompetenzen für das Lehramt in den Naturwissenschaften - DiKoLAN PLUS*. eds. C. Thyssen, M. Meier and S. Becker-Genschow (Hamburg: Joachim Herz Stiftung).

Meier, M., Thyssen, C., Kremser, E., Thoms, L.-J., Becker-Genschow, S., Bruckermann, T., et al. (2024). *DiKoLAN-Grid – Kompetenzraster zur Selbsteinschätzung der naturwissenschaftlichen digitalen Lehrkompetenz*. Geneva: Zenodo.

Mishra, P. (2019). Considering contextual knowledge: the TPACK diagram gets an upgrade. *J. Digit. Learn. Teach. Educ.* 35, 76–78. doi: 10.1080/21532974.2019.1588611

Mislevy, R., Steinberg, L., and Almond, R. (2003). Focus article: on the structure of educational assessments. *Meas. Interdiscip. Res. Perspect.* 1, 3–62. doi: 10.1207/S15366359MEA0101_02

Narciss, S., and Huth, K. (2002). *How to design informative tutoring feedback for multimedia learning. Instrucional Design for Mulitmedia Learning*. Cham: Springer.

Natriello, G. (2013). *Adaptive educational technologies: Tools for learning and for learning about learning*. Washington, DC: National Academy of Education.

Ng, D. T. K., Leung, J. K. L., Su, J., Ng, R. C. W., and Chu, S. K. W. (2023). Teachers' AI digital competencies and twenty-first century skills in the post-pandemic world. *Educ. Technol. Res. Dev.* 71, 137–161. doi: 10.1007/s11423-023-10203-6

Nja, C. O., Idiege, K. J., Uwe, U. E., Meremikwu, A. N., Ekon, E. E., Erim, C. M., et al. (2023). Adoption of artificial intelligence in science teaching: from the vantage point of the African science teachers. *Smart Learn. Environ.* 10:42. doi: 10.1186/s40561-023-00261-x

O'Meara, N., and Faulkner, F. (2022). Professional development for out-of-field post-primary teachers of mathematics: an analysis of the impact of mathematics specific pedagogy training. *Ir. Educ. Stud.* 41, 389–408. doi: 10.1080/03323315.2021.1899026

Parsons, S. A., Vaughn, M., Scales, R. Q., Gallagher, M. A., Parsons, A. W., Davis, S. G., et al. (2018). Teachers' instructional adaptations: a research synthesis. *Rev. Educ. Res.* 88, 205–242. doi: 10.3102/0034654317743198

Pelánek, R. (2025). Adaptive learning is hard: challenges, nuances, and trade-offs in modeling. *Int. J. Artif. Intell. Educ.* 35, 304–329. doi: 10.1007/s40593-024-00400-6

Plass, J., and Pawar, S. (2020). Toward a taxonomy of adaptivity for learning. *J. Res. Technol. Educ.* 52, 275–300. doi: 10.1080/15391523.2020.1719943

Preston, C. C., and Colman, A. M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychol.* 104, 1–15. doi: 10.1016/S0001-6918(99)00050-5

Quora Poe (2017). Available online at: poe.com

Redecker, C. (2017). *European framework for the digital competence of educators: DigCompEdu*. Luxembourg: Publications Office of the European Union.

Rosenberg, J. M., and Koehler, M. J. (2015). Context and technological pedagogical content knowledge (TPACK): a systematic review. *J. Res. Technol. Educ.* 47, 186–210. doi: 10.1080/15391523.2015.1052663

Ryan, R. M., and Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. New York, NY: Guilford Publications.

Shute, V., and Zapata-Rivera, D. (2012). Adaptive educational systems. *Adaptive Technol. Training Educ.* 22, 7–27. doi: 10.1017/CBO9781139049580.004

Siciliano, M. D. (2016). It's the quality not the quantity of ties that matters: social networks and self-efficacy beliefs. *Am. Educ. Res. J.* 53, 227–262. doi: 10.3102/0002831216629207

Snow, R. E. (1980). Aptitude, learner control, and adaptive instruction. *Educ. Psychol.* 15, 151–158. doi: 10.1080/00461528009529223

Su, Y., Lin, Y., and Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assess. Writing* 57:100752. doi: 10.1016/j.asw.2023.100752

Täschner, J., Dicke, T., Reinhold, S., and Holzberger, D. (2025). "Yes, I can!" a systematic review and meta-analysis of intervention studies promoting teacher self-efficacy. *Rev. Educ. Res.* 95, 3–52. doi: 10.3102/00346543231221499

Teo, T. (2009). Modelling technology acceptance in education: a study of pre-service teachers. *Comput. Educ.* 52, 302–312. doi: 10.1016/j.compedu.2008.08.006

Vaughn, M., Parsons, S. A., and Gallagher, M. A. (2022). Challenging scripted curricula with adaptive teaching. *Educ. Res.* 51, 186–196. doi: 10.3102/0013189x211065752

Velli, K., and Zafiropoulos, K. (2024). Factors that affect the acceptance of educational AI tools by Greek teachers—a structural equation modelling study. *Eur. J. Inv. Health Psychol. Educ.* 14, 2560–2579. doi: 10.3390/ejihpe14090169

Venkatesh, V., and Bala, H. (2008). Technology acceptance model 3 and a research agenda on interventions. *Decis. Sci.* 39, 273–315. doi: 10.1111/j.1540-5915.2008.00192.x

Venkatesh, V., and Davis, F. D. (1996). A model of the antecedents of perceived ease of use: development and test. *Decis. Sci.* 27, 451–481. doi: 10.1111/j.1540-5915.1996.tb00860.x

Venkatesh, V., and Davis, F. D. (2000). A theoretical extension of the technology acceptance model: four longitudinal field studies. *Manag. Sci.* 46, 186–204. doi: 10.1287/mnsc.46.2.186.11926

Vigerske, S. (2017). Transfer von Lehrerfortbildungsinhalten in die Praxis. Wiesbaden: Springer Fachmedien Wiesbaden.

Vogelsang, C., Finger, A., Laumann, D., and Thyssen, C. (2019). Vorerfahrungen, Einstellungen und motivationale Orientierungen als mögliche Einflussfaktoren auf den Einsatz digitaler Werkzeuge im naturwissenschaftlichen Unterricht. *Z. Didakt. Naturwiss.* 25, 115–129. doi: 10.1007/s40573-019-00095-6

Vygotskij, L. S., and Cole, M. (1979). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard Univ. Press.

Willmott, A. (1978). Assessment and performance. *Oxf. Rev. Educ.* 4, 51–64. doi: 10.1080/0305498780040105

Winne, P. H., and Butler, D. L. (1994). *"Student cognition in learning from teaching," in international encyclopedia of education*. Oxford: Pergamont, 5738–5745.

Wray, E., Sharma, U., and Subban, P. (2022). Factors influencing teacher self-efficacy for inclusive education: a systematic literature review. *Teach. Teach. Educ.* 117:103800. doi: 10.1016/j.tate.2022.103800

Yang, A. C. M., Flanagan, B., and Ogata, H. (2022). Adaptive formative assessment system based on computerized adaptive testing and the learning memory cycle for personalized learning. *Comput. Educ. Artif. Int.* 3:100104. doi: 10.1016/j.caeai.2022.100104

Yue, M., Jong, M. S.-Y., and Ng, D. T. K. (2024). Understanding K–12 teachers' technological pedagogical content knowledge readiness and attitudes toward artificial intelligence education. *Educ. Inf. Technol.* 29, 19505–19536. doi: 10.1007/s10639-024-12621-2

Zenni, E. A., and Turner, T. L. (2021). Planning and presenting workshops that work: a faculty development workshop. *MedEdPORTAL* 17:11158. doi: 10.15766/mep_2374-8265.11158

Zhang, C., Schießl, J., Plößl, L., Hofmann, F., and Gläser-Zikuda, M. (2023). Acceptance of artificial intelligence among pre-service teachers: a multigroup analysis. *Int. J. Educ. Technol. High. Educ.* 20:49. doi: 10.1186/s41239-023-00420-7

Zuo, Z., Luo, Y., Yan, S., and Jiang, L. (2025). From perception to practice: artificial intelligence as a pathway to enhancing digital literacy in higher education teaching. *Systems* 13:664. doi: 10.3390/systems13080664

Glossary

AFA - Assessment, Feedback and Adaptivity

AI - Artificial Intelligence

AISEE - Artificial Intelligence Self-Efficacy Expectation

BI - Behavioral Intention

CA - Computer Anxiety

COM - Communication/Collaboration

CSE - Computer Self-Efficacy

DAP - Data Processing

DAQ - Data Acquisition

DigCompEdu - Digital Competence Framework for Educators

DiKoLAN - Digital Competencies for Teaching in Science Education

DOC - Documentation

ICT - Information and Communication Technology

ISE - Information Search and Evaluation

LLM - Large Language Model

PEU - Perceived Ease of Use

PRE - Presentation

PU - Perceived Usefulness

RQ - Research Question

SE - Self-Efficacy

SIM - Simulation and Modeling

SN - Subjective Norm

STEM - Science, Technology, Engineering and Mathematics

TAM - Technology Acceptance Model

TPACK - Technological Pedagogical Content Knowledge

TSE - Technology Self-Efficacy