



OPEN ACCESS

EDITED BY
Paolino Di Felice,
University of L'Aquila, Italy

REVIEWED BY
Saulius Gudas,
Vilnius University, Lithuania
Amir Mashmool,
University of Bremen, Germany

*CORRESPONDENCE
Niklas Retzlaff
✉ niklas.retzlaff.dba@edu.triagon-academy.com

RECEIVED 10 May 2025
ACCEPTED 06 October 2025
PUBLISHED 27 October 2025

CITATION
Retzlaff N and Spörrle M (2025) Measuring
agility in software development teams:
development and initial validation of the Agile
Team Practice Inventory for Software
Development (ATPI-SD).
Front. Comput. Sci. 7:1626456.
doi: 10.3389/fcomp.2025.1626456

COPYRIGHT
© 2025 Retzlaff and Spörrle. 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.

Measuring agility in software development teams: development and initial validation of the Agile Team Practice Inventory for Software Development (ATPI-SD)

Niklas Retzlaff^{1,2*} and Matthias Spörrle³

¹Department of Business Administration, Triagon Academy, Marsa, Malta, ²cosinex GmbH, Bochum, Germany, ³Seeburg Castle University (SCU), Seekirchen, Austria

Introduction: Agile methodologies are ubiquitous in software development, yet their measurement remains challenging due to a lack of validated instruments. This paper details the development and initial validation of the Agile Team Practice Inventory for Software Development (ATPI-SD), a new questionnaire measuring team-level agility based on core agile values and practices.

Methods: Starting from a comprehensive literature review (258 items) and expert consultations ($n = 7$), five dimensions were initially identified, leading to 67 generated items. Expert feedback refined this to 37 items across 4 dimensions, which were tested in Study 1 ($n = 199$). Further analysis resulted in a final 20-item scale with four dimensions: Customer Involvement (CI), Team Collaboration (TC), Iterative and Incremental Development Processes (IIDP), and Continuous Development Process Improvement (CDPI).

Results: Data from our study ($n = 237$) showed good internal consistency for the total scale ($\alpha = 0.89$) and subscales (ranging from 0.69 to 0.84). Confirmatory Factor Analysis indicated a moderate-to-acceptable model fit (e.g., CFI = 0.88, TLI = 0.86). Moderate convergent validity was supported by a significant correlation with a single-item self-rating of team agility ($p = 0.404$, $p < 0.001$).

Discussion: While suggesting potential for refinement, the ATPI-SD provides a systematically developed and initially validated instrument for researchers and practitioners assessing agility in software development teams.

KEYWORDS

agile software development, measuring agility, scale development, questionnaire, agility assessment, team agility, instrument validation, software engineering

1 Introduction

Agile methodologies have fundamentally reshaped the landscape of software development over the past two decades. Originating from the Agile Manifesto (Beck et al., 2001), principles emphasizing flexibility, customer collaboration, iterative progress, and rapid response to change have become cornerstones for countless organizations seeking to enhance productivity, product quality, market responsiveness, and employee engagement (Dybå and Dingsøyr, 2008; Kalenda et al., 2018; Tripp et al., 2016). The successful implementation of these principles hinges critically on the ability of software development teams to internalize and enact agile values and practices effectively. Consequently, understanding and assessing the degree of agility present within these

teams is of paramount importance for both academic research and practical organizational development.

Despite the widespread adoption and recognized importance of team agility, its accurate measurement remains a significant challenge. The existing landscape of assessment tools is fragmented and often lacks rigorous scientific validation (Chronis and Gren, 2016; Gren et al., 2017; Maneva et al., 2017; Yürüm et al., 2018). Many instruments have emerged from practical or consulting contexts and, while potentially useful, have not undergone systematic psychometric evaluation regarding their reliability and validity. Even instruments developed within academic settings (Leffingwell, 2007; Looks et al., 2021; Qumer and Henderson-Sellers, 2008; Sidky et al., 2007; So and Scholl, 2009; Soundararajan, 2013) often suffer from limited empirical testing or validation across diverse contexts, lack widespread adoption, or, in some cases, have yielded conflicting results or failed validation attempts in subsequent empirical studies (Chronis and Gren, 2016; Gren et al., 2017). This situation hinders the ability to reliably compare agility levels across teams or studies, track improvements over time, or confidently investigate the relationship between team agility and crucial outcomes like performance, innovation, or job satisfaction (Chronis and Gren, 2016; Tripp et al., 2016).

This situation highlights a clear research gap: there is a need for a theoretically grounded, practically applicable, and empirically validated instrument to measure agility specifically at the team level within the context of software development. Such an instrument should capture the core dimensions inherent in agile principles and practices, be concise enough for use in survey research and organizational diagnostics, and adhere to established psychometric standards.

This paper addresses this gap by detailing the systematic development and initial psychometric validation of the Agile Team Practice Inventory for Software Development (ATPI-SD), a novel questionnaire designed to assess team agility in software development environments. Drawing upon a comprehensive literature review and expert consultations, the ATPI-SD aims to provide a multi-dimensional yet integrated measure of agile practices. Its development followed a rigorous multi-stage process, including item generation based on established literature and agile principles, expert review for content validity and relevance, Study 1 for initial refinement, and Study 2 involving a substantial sample of software development professionals.

The contribution of this paper is twofold. First, it provides a transparent account of the development process, addressing the often-lamented gap between practical agility assessment needs and academic rigor. Second, it presents the ATPI-SD itself, along with initial evidence of its reliability and validity, offering researchers and practitioners a potentially valuable tool. For researchers, the ATPI-SD can serve as a standardized measure to investigate the antecedents and consequences of team agility. For practitioners, it offers a diagnostic instrument to assess current agile practices, identify areas for improvement, and benchmark progress in agile transformations.

The remainder of this paper is structured as follows: Section 2 outlines the theoretical foundations of agility and reviews existing measurement approaches. Section 3 describes the systematic, multi-phase process undertaken to develop the ATPI-SD items

and dimensions. Section 4 presents the methodology and results of the main validation study, including reliability analyses and confirmatory factor analysis based on data from 237 software professionals. Section 5 discusses the findings, interprets the psychometric properties of the ATPI-SD, acknowledges limitations, and outlines implications for research and practice. Finally, Section 6 concludes with a summary and directions for future research concerning the instrument's further validation and application.

2 Theoretical background: defining and measuring agility

Before embarking on the development of a new measurement instrument, it is crucial to establish a clear conceptual understanding of agility within the software development context and to review how it has been operationalized and measured previously. This section first defines team agility by drawing on foundational principles and literature, then identifies key dimensions that constitute the construct, and finally reviews existing measurement approaches, highlighting their limitations and justifying the need for the new instrument developed in this study.

2.1 Defining agility in software development teams

The term “agility” in software development originates largely from the Agile Manifesto (Beck et al., 2001), which outlined four core values and twelve principles. These emphasize individuals and interactions over processes and tools, working software over comprehensive documentation, customer collaboration over contract negotiation, and responding to change over following a plan. While various specific methodologies like Scrum, Kanban, Extreme Programming (XP), and others implement these principles differently, they share a common philosophical core centered on adaptability, iterative progress, and human-centric collaboration (Highsmith and Highsmith, 2002; Moniruzzaman and Hossain, 2013).

Traditional software development models, such as the Waterfall model, often rely on linear, sequential phases with extensive upfront planning and documentation (Royce, 1970). In contrast, agile approaches embrace uncertainty and changing requirements, utilizing short development cycles (iterations or sprints) to deliver functional software incrementally and gather frequent feedback (Leffingwell, 2007), a key differentiator from more rigid, plan-driven models particularly challenged by dynamic environments (Jin, 2024; Rahman, 2024).

For the purpose of measurement, particularly within organizational research, agility is increasingly conceptualized not merely as adherence to a specific named methodology (e.g., Scrum), but fundamentally as a set of capabilities and enacted practices manifested by the development team. This perspective emphasizes aspects like team design, collaborative processes, adaptability, empowerment, and the effective application of agile taskwork and teamwork behaviors, rather than simply

following prescribed rituals (Junker et al., 2021; Rathor et al., 2023; Uraon et al., 2023). Therefore, for the purpose of this study, *team agility* is conceptualized as the consistent application of practices and behaviors within a software development team that enable it to respond effectively to changing requirements, deliver value incrementally, foster close collaboration, and continuously improve its processes. This definition deliberately focuses on the self-reported enactment of agile processes rather than on objective output metrics, which can be influenced by numerous external factors. This process-oriented view aligns with recent conceptualizations of team effectiveness that highlight specific work practices and adaptive capacity (Sathe and Panse, 2022; Strode et al., 2022; Uraon et al., 2023).

2.2 Dimensions of software development team agility

To operationalize the definition of team agility for measurement in a way that addresses the fragmentation and validation gaps in existing tools (as discussed in Section 1), the broad concept needs to be broken down into distinct, yet interrelated, dimensions grounded in core agile principles. Based on the foundational Agile Manifesto (Beck et al., 2001) and an initial review of theoretical literature and established agility assessment frameworks, five core dimensions were first identified as potentially crucial for capturing team agility in software development: Customer Involvement, Software Testing, Team Collaboration, Iterative and Incremental Development Processes, and Continuous Development Process Improvement. This multi-dimensional approach was chosen deliberately to move beyond simplistic or methodology-specific assessments toward a more nuanced understanding of agile practice adoption.

However, as detailed in the instrument development process (Section 3.2), the “Software Testing” dimension was subsequently excluded following expert review. The consensus was that while vital for software quality, specific testing techniques might be less indicative of overall team-level agile “practices” compared to the other dimensions, or perhaps better captured implicitly within them. This exclusion aligns with findings in recent literature where testing practices in agile contexts are often integrated within broader dimensions such as technical capability, quality assurance practices, or iterative development workflows, rather than consistently treated as a standalone dimension of “team” agility (Berlas, 2024; Chronis and Gren, 2016; Latif et al., 2017).

Consequently, the Agile Team Practice Inventory for Software Development (ATPI-SD) developed in this research focuses on measuring the four remaining dimensions. These were selected for their direct theoretical linkage to the Agile Manifesto’s principles and values, their emphasis on observable team-level behaviors and practices (making them suitable for measurement), and their demonstrated relevance to successful team functioning in contemporary software development. While distinct for analytical purposes, these dimensions are understood to be synergistic aspects of holistic team agility (Rathor et al., 2023; Strode et al., 2022). The ATPI-SD aims to assess the extent to which teams enact practices reflecting these dimensions:

Customer involvement (CI): This dimension captures the team’s enactment of practices ensuring active participation and collaboration with the customer or end-user throughout the development lifecycle. It operationalizes the Manifesto’s value of “Customer collaboration over contract negotiation” and the principle of satisfying the customer through early and continuous delivery (Beck et al., 2001). Regularly involving the customer ensures the product meets actual user needs, facilitates rapid feedback loops essential for adaptation, enhances alignment with expectations, and is crucial for delivering value and achieving project success in agile environments (Bhatta and Thite, 2018; Rahy and Bass, 2021; Silva-Martinez, 2023).

Team collaboration (TC): This dimension assesses the observable practices related to the quality and intensity of interaction, communication, mutual support, and shared responsibility within the development team and with closely related stakeholders. It directly reflects the Manifesto’s prioritization of “Individuals and interactions over processes and tools” and the principle of daily collaboration between business people and developers (Beck et al., 2001). Strong collaborative practices foster flexibility, enhance transparency, enable effective problem-solving and shared decision-making, and are demonstrably linked to improved team dynamics, performance, and stakeholder satisfaction in agile settings (Avila et al., 2022; Dugbartey and Kehinde, 2025; Rathor et al., 2023; Shah, 2024).

Iterative and incremental development processes (DP): This dimension measures the team’s adherence to core agile mechanics—specifically, working in short, time-boxed cycles (iterations) to produce functional parts of the system (increments). It embodies principles like “Deliver working software frequently” and “Welcome changing requirements” (Beck et al., 2001). Consistent application of iterative and incremental practices allows teams to adapt effectively to evolving requirements, incorporate feedback systematically, manage risks proactively, and demonstrably leads to enhanced project performance, predictability, and stakeholder satisfaction (Esang et al., 2024; Mahadik et al., 2022; Rathor et al., 2023; Shah, 2024).

Continuous development process improvement (PI): This dimension evaluates the team’s engagement in practices aimed at ongoing reflection, learning, and adaptation of its own processes and workflows. It operationalizes the Manifesto’s principle that “At regular intervals, the team reflects on how to become more effective, then tunes and adjusts its behavior accordingly” (Beck et al., 2001). Regularly engaging in continuous improvement practices (e.g., retrospectives, seeking feedback, experimenting with processes) fosters adaptability, enhances efficiency and product quality, cultivates a learning orientation, and contributes significantly to sustained team effectiveness and successful project outcomes (Dugbartey and Kehinde, 2025; Goyal, 2023; Maharao, 2024; Moraga-Díaz and Piñango, 2023).

By focusing on these four empirically supported and theoretically grounded dimensions, the ATPI-SD seeks to provide a robust framework for assessing team agility. This structure directly addresses the need identified earlier for a multi-dimensional, principle-based instrument focused on observable team practices, thereby offering a more comprehensive and valid approach compared to overly simplistic or narrowly focused existing measures.

2.3 Existing approaches to measuring agility and their limitations

While the importance of team agility in software development is widely acknowledged, its reliable and valid measurement presents considerable challenges, a situation highlighted consistently in recent literature reviews surveying the state of agility assessment (Cunha et al., 2024; Menezes et al., 2024; Ojha, 2023; Rathor et al., 2023). Existing approaches vary significantly in their origin, focus, format, and, critically, their level of empirical validation. A notable category includes tools originating from practitioners or consultants, such as various agile maturity models. Often designed for specific diagnostic or coaching purposes, these tools frequently lack published, rigorous evidence of psychometric soundness. Indeed, when subjected to empirical scrutiny, they often exhibit significant limitations concerning core psychometric properties like construct validity (i.e., measuring the intended concept), predictive validity (i.e., predicting relevant outcomes), and reliability (i.e., measurement consistency) (Chronis and Gren, 2016; Gunsberg et al., 2018; Henriques and Tanner, 2017). Furthermore, their common reliance on self-assessment introduces potential response biases (Thiyagarajan et al., 2024), and their developmental transparency is often limited, hindering systematic evaluation and comparison (Özcan Top and Demirsors, 2019; Yürüm et al., 2018).

Within the academic sphere, several questionnaires and frameworks have been proposed (Leffingwell, 2007; Looks et al., 2021; Qumer and Henderson-Sellers, 2008; Sidky et al., 2007; So and Scholl, 2009; Soundararajan, 2013), often offering a more systematic and theoretically grounded foundation. However, recent critiques and validation studies reveal that these academic instruments also face significant limitations that hinder their widespread utility and the comparability of research findings. A primary concern is the often limited or questionable psychometric validation. Many instruments lack comprehensive evidence for construct, discriminant, or predictive validity, or demonstrated reliability across diverse samples (Chronis and Gren, 2016; Cunha et al., 2024). Moreover, empirical validation attempts for some scales have yielded disappointing or conflicting results, casting doubt on their psychometric robustness and practical applicability (Gren et al., 2015). Compounding this is the considerable heterogeneity among existing scales regarding their unit of analysis, dimensionality, and specific practices measured. This heterogeneity, coupled with variability in how core constructs are defined and operationalized (Dutra et al., 2022; Ojha, 2023), makes rigorous comparison across studies extremely difficult. Furthermore, many instruments lack adaptability to diverse organizational contexts (Njanka et al., 2021) or adequate consideration of cultural nuances that influence agile practices (Hukkelberg and Berntzen, 2019). Limitations also arise from the scope and focus of existing instruments. Some are narrowly tied to specific methodologies like Scrum, limiting their applicability to teams using other or hybrid approaches (Menezes et al., 2024). Others face practicality challenges due to length or complexity. Conversely, some tools, including those embedded in tracking software, are criticized for being overly simplistic, relying heavily on quantitative metrics while neglecting important qualitative

aspects, non-technical dimensions like team culture or leadership (Făgărăsan et al., 2022; Zapata et al., 2021), or the underlying processes driving outcomes. The static nature of some tools, failing to evolve with agile practices, also poses a challenge (Ganesh and Narayanan, 2019).

This critical review of existing measurement approaches underscores a persistent gap, frequently highlighted in recent research (Cunha et al., 2024; Menezes et al., 2024; Rathor et al., 2023): the lack of a widely accepted, comprehensively validated, and practically feasible instrument for measuring team-level agility based on core agile principles applicable across different methodologies and contexts. The development of the Agile Team Practice Inventory for Software Development (ATPI-SD), described next, directly aims to address these identified shortcomings by providing such an instrument.

2.4 Rationale for a new instrument: the agile team practice inventory for software development

The limitations inherent in existing approaches to measuring team agility, particularly concerning validation, scope, practicality, and theoretical grounding as detailed previously, underscore a clear need for a new, improved instrument. The Agile Team Practice Inventory for Software Development (ATPI-SD) was developed specifically to address these identified shortcomings. Critically, the design principles guiding the ATPI-SD's creation align directly with explicit calls from recent literature urging the development of more effective team agility measures to fill the current void (Cunha et al., 2024; Dingsoeyr et al., 2019; Menezes et al., 2024; Rathor et al., 2023). Therefore, the ATPI-SD was deliberately designed with several key characteristics, each responding to both the identified gaps and the expressed needs within the research community. Firstly, it maintains a distinct Team-Level Focus, assessing practices and capabilities as perceived collectively by the team. This specifically addresses the need for measures centered on team dynamics and performance, rather than individual or purely organizational assessments, a requirement emphasized in recent calls for new metrics (Arumugam and Vaidyanathan, 2022; Mashmool et al., 2021; Rathor et al., 2023). Secondly, the ATPI-SD is Principle-Based, focusing on core agile practices derived from the Agile Manifesto, thereby enhancing its applicability across diverse methodologies and directly answering the call for measures grounded in fundamental agile philosophy rather than rigid adherence to specific frameworks (Dingsoeyr et al., 2019; Gren, 2022; Mashmool et al., 2021; Rathor et al., 2023). Thirdly, its Multi-Dimensional structure, based on the theoretically derived and empirically supported dimensions outlined in Section 2.2, aims to capture the multifaceted nature of agility, addressing critiques of overly simplistic measures and aligning with recommendations for metrics that better reflect the complexity of team interactions (Arumugam and Vaidyanathan, 2022; Hukkelberg and Berntzen, 2019; Rathor et al., 2023). Fourthly, a central objective was to ensure the ATPI-SD is Psychometrically Sound. Its development process, detailed in

Section 3, incorporated systematic steps including expert review, pilot testing, and confirmatory factor analysis, directly tackling the widespread problem of inadequate validation in existing tools and responding to the urgent call for rigorously validated instruments (Dingsoeyr et al., 2019; Mashmool et al., 2021; Rathor et al., 2023). Finally, the instrument was designed for Practical Applicability and Adaptability, aiming for a reasonable completion time suitable for research and organizational diagnostics, while the principle-based approach enhances its potential relevance across different contexts, addressing the need for context-appropriate measures (Dingsoeyr et al., 2019; Rathor et al., 2023). The dimensions identified previously served as the conceptual blueprint, and the subsequent development phases sought to realize these design principles while meeting robust psychometric standards. This systematic approach intends to provide the more robust foundation for future empirical research on agility that the literature indicates is needed.

3 Instrument development process: crafting the agile team practice inventory for software development

The development and validation of measurement scales are fundamental undertakings in social and behavioral research, yet the process can be complex and resource-intensive (Boateng et al., 2018). Recognizing this, the development of the Agile Team Practice Inventory for Software Development (ATPI-SD) adhered to established best practices, following a systematic, multi-phase process designed to ensure content validity, relevance, and preliminary psychometric soundness. As recommended by frameworks for rigorous scale development (Boateng et al., 2018), this approach integrated theoretical insights from the literature with practical expertise from the field, moving progressively from a broad initial item pool to a refined set of items suitable for empirical testing. It encompassed key phases including initial item generation and content validity assessment, followed by scale construction through expert review and item reduction, and preliminary scale evaluation via Study 1 (the overall process is depicted in Figure 1). Each phase is detailed below.

3.1 Item generation and dimension definition

The initial phase focused on establishing a strong theoretical and empirical foundation for the instrument by grounding it in the existing body of knowledge on agile software development. To achieve this, a comprehensive literature review was conducted, encompassing academic publications, foundational texts like the Agile Manifesto (Beck et al., 2001), established agility frameworks, prior measurement attempts (both academic and practitioner-focused, see Table 1 for examples), and empirical studies investigating agile practices.

This review informed the generation of an extensive initial pool of 258 potential items. These items were derived directly from, or adapted based on, descriptions of agile practices, principles, and values found in the literature, including scales developed by

researchers such as Leffingwell (2007); Looks et al. (2021); Sidky et al. (2007); So and Scholl (2009); Soundararajan (2013). The primary objective during this stage was to capture a broad spectrum of behaviors and characteristics potentially reflecting team agility. A comprehensive list of these initial items was maintained internally during the development process.

Concurrently, guided by the theoretical framework developed from the literature review (as outlined in Section 2), five initial dimensions were defined *a priori*: Customer Involvement, Software Testing, Team Collaboration, Iterative and Incremental Development Processes, and Continuous Development Process Improvement.

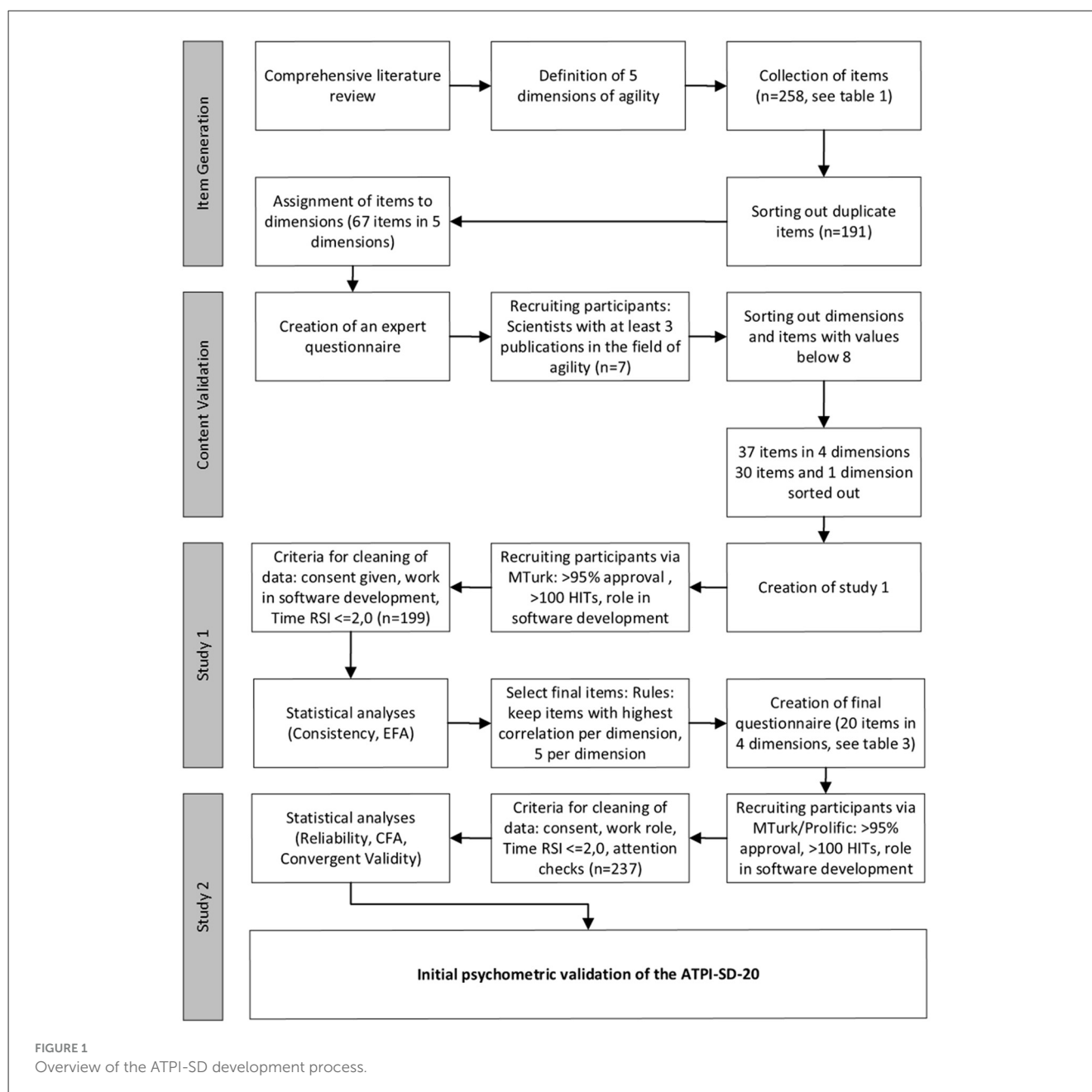
Each of the 258 generated items was subsequently allocated to one of these five dimensions based on its primary conceptual content. An initial content refinement process followed, during which items were eliminated if they were considered clearly redundant, ambiguous, overly specific to a single methodology (conflicting with the goal of broader applicability), or irrelevant to the core construct of team agility. This refinement process yielded a reduced pool of 67 items distributed across the five dimensions: Customer Involvement (number of items, $n = 8$), Software Testing ($n = 7$), Team Collaboration ($n = 20$), Iterative and Incremental Development Processes ($n = 24$), and Continuous Development Process Improvement ($n = 8$). This set of 67 items formed the basis for the subsequent expert review phase aimed at establishing content validity.

3.2 Expert review for content validation

To ensure the content validity and practical relevance of the initial 67 items and the proposed dimensional structure, an expert review process was conducted, a crucial step in scale development (Boateng et al., 2018). Seven experts in the field of agile software development were recruited for this phase. Selection criteria required experts to be active researchers with recent publications on agile topics, ensuring they possessed both deep theoretical knowledge and familiarity with current industry practices.

The experts were provided with the list of 67 items, categorized according to the five proposed dimensions (Customer Involvement, Software Testing, Team Collaboration, Iterative and Incremental Development Processes, Continuous Development Process Improvement), along with clear operational definitions for each dimension. They were tasked with rating the relevance of each dimension and each individual item for measuring team agility using a scale from 0 (“not relevant”) to 10 (“extremely relevant”). Furthermore, experts were invited to provide qualitative feedback regarding item clarity, wording appropriateness, and the suitability of the overall dimensional structure.

The quantitative relevance ratings were subsequently averaged across the seven experts for both the dimensions and individual items. Following recommended practices in scale development to retain only highly relevant items, a stringent relevance threshold of 8.0 was applied for item and dimension retention. Analysis of the dimensional relevance revealed average scores of 8.4 for Customer Involvement, 7.4 for Software Testing, 9.4 for Team Collaboration, 9.3 for Iterative/Incremental Processes, and 9.4



for Continuous Improvement. Notably, the “Software Testing” dimension fell below the 8.0 threshold. Furthermore, none of the individual items initially allocated to this dimension achieved the relevance threshold. This quantitative finding, combined with qualitative expert feedback suggesting that specific testing practices might be less indicative of overall team-level agility or better captured implicitly within other dimensions (such as Iterative Processes or Continuous Improvement), led to the decision to remove the “Software Testing” dimension entirely, along with its 7 associated items.

For the remaining four dimensions, individual items which scored below the 8.0 relevance threshold were also eliminated. This rigorous expert review and subsequent reduction process resulted in a substantially refined pool of 37 items distributed across the four retained dimensions: Customer Involvement

($n = 8$ items), Team Collaboration ($n = 13$ items), Iterative and Incremental Development Processes ($n = 11$ items), and Continuous Development Process Improvement ($n = 5$ items). This critical step significantly enhanced the content validity of the emerging instrument by ensuring that only items deemed highly relevant by domain experts were carried forward to the next phase of empirical testing.

3.3 Study 1 and final item selection

Following the expert review, the refined 37-item version of the questionnaire was subjected to an initial empirical evaluation in Study 1. The primary goals of this study were to gather

TABLE 1 Overview of selected academically developed methods for measuring agility.

Author(s)	Year	Scale type	Items	Publication type
Looks et al.	2021	7-point Likert	36	Conference
Sidky et al.	2007	5-point Likert	29	Journal
Qumer and Henderson-Sellers	2008	Yes/No	4	Journal
So and Scholl	2009	7-point Likert	46	Conference
Leffingwell	2007	7-point Likert	67	Book
Soundararajan	2013	5-point Likert	75	Dissertation

data on item performance, assess preliminary scale reliability and factor structure, and further refine the instrument for clarity and brevity.

Data for Study 1 were collected from a sample of 199 participants recruited via online platforms, primarily Amazon Mechanical Turk (MTurk). The target population consisted of individuals currently working in software development roles. The sample was composed of 146 male (73.4%) and 53 female (26.6%) participants. A significant portion of the sample held developer roles ($n = 86$, 43.2%), followed by project managers ($n = 49$, 24.6%), team leaders ($n = 22$, 11.1%), and business analysts ($n = 20$, 10.1%), ensuring that the initial item testing was conducted with a relevant professional audience. Geographically, the sample was heavily concentrated in the United States ($n = 180$, 90.5%). Participants responded to the 37 items using a 7-point Likert scale, ranging from 1 (“Almost Never True”) to 7 (“Almost Always True”). Standard data quality assurance measures were employed, including the use of attention check items and filtering based on response time analysis. Specifically, responses with a Time Relative Speed Index (Time RSI) lower than 2.0 were excluded, as such values may indicate overly rapid completion potentially compromising data validity ([Leiner, 2019](#)).

Initial analysis focused on the internal consistency of the 37-item scale and its four dimensions using Cronbach’s Alpha. The results, presented in [Table 2](#), indicated good to excellent reliability for the dimensions based on the 37-item pool, providing preliminary evidence for the coherence of the item set.

To further validate the questionnaire structure beyond internal consistency and to guide item reduction, an Exploratory Factor Analysis (EFA) based on principal axis analysis was performed. Given potential deviations from normality common in survey data (which were confirmed using Mardia’s tests for multivariate skewness and kurtosis in this dataset), preliminary z-standardization of the data was performed to unify scaling, although this does not alter the underlying distribution shape. Consequently, an estimation method robust to violations of the normality assumption was selected. Maximum Likelihood (ML) EFA was chosen, as ML estimation is considered relatively robust in such cases. The analysis was based on the correlation matrix, and an oblique Promax rotation was applied, as the underlying dimensions of agility were theoretically

TABLE 2 Cronbach’s Alpha for the 37-item version in Study 1.

Dimension (initial 37 items)	Cronbach’s Alpha
Customer involvement (8 items)	0.85
Team collaboration (13 items)	0.82
Iterative and incremental development processes (11 items)	0.87
Continuous development process improvement (5 items)	0.83
Total scale (37 items)	0.95

expected to be correlated. Parallel analysis was used alongside eigenvalue criteria to help determine the number of factors to retain.

The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was excellent ($KMO = 0.94$), and Bartlett’s test of sphericity ($\chi^2(190) = 2455.192$, $p < 0.001$), both indicated that the data were suitable for factor analysis. The EFA results suggested a multi-dimensional structure broadly consistent with the intended four dimensions, although some items exhibited cross-loadings or lower-than-ideal primary loadings.

Based on these EFA findings (considering factor loadings, cross-loadings, and item uniqueness values), alongside item-total correlations and analyses of how deleting specific items would impact subscale reliability (Cronbach’s Alpha if item deleted), the item pool was further reduced. The objective was to achieve a more parsimonious scale with strong psychometric properties and clear factor interpretability, while maintaining adequate content coverage across the four dimensions. A key decision during this refinement stage was to retain an equal number of items (five) per dimension. This standardization was chosen to enhance the comparability between the subscale scores and mitigate potential biases in analysis that can arise from unequal scale lengths ([Frisbie and Brandenburg, 1979](#); [Schriesheim et al., 1989](#)). Furthermore, balancing the number of items per dimension contributes to a more consistent respondent experience, potentially reducing survey fatigue and dropout rates in online administration ([Toepoel et al., 2009](#)), and facilitates the standardized application of the instrument across different studies and contexts ([Moran et al., 2001](#)). While shorter scales can increase the risk of measurement error compared to longer ones ([Kruyen et al., 2014](#)), retaining five well-performing items per dimension was deemed an optimal compromise between psychometric robustness and practical usability. The five items selected for each dimension were those demonstrating the best combination of high primary factor loadings, low cross-loadings, strong item-total correlations, and positive contribution to subscale reliability.

This analysis in Study 1 and systematic item reduction process led to the final 20-item version of the ATPI-SD (referred to as ATPI-SD-20), consisting of four subscales with five items each: Customer Involvement (CI), Team Collaboration (TC), Iterative and Incremental Development Processes (DP), and Continuous Development Process Improvement (PI). The final items are listed in [Table 3](#). This version balances breadth of content coverage with psychometric considerations and practical brevity, forming the instrument evaluated in Study 2, which is described in the subsequent section.

TABLE 3 Final 20 items of the Agile Team Practice Inventory for Software Development (ATPI-SD-20).

ID	Item wording	Item inspired by
Dimension: Customer Involvement (CI)		
CI1	Each iteration concludes with the delivery of a working product to the customer.	(Looks et al., 2021)
CI2	Customers play a central role in establishing the priorities of our project stories.	(Soundararajan, 2013)
CI3	Customer-requested changes are well-accommodated in our development process.	(Soundararajan, 2013)
CI4	Our team can easily reach out to customers without bureaucratic delays.	(So and Scholl, 2009)
CI5	It is acceptable for team members to express disagreement with customers and suggest alternative ideas.	(Sidky et al., 2007)
Dimension: Team Collaboration (TC)		
TC1	Team members effectively determine their own workload.	(Soundararajan, 2013)
TC2	Team members take full ownership of their assigned work items.	(Soundararajan, 2013)
TC3	Team members reliably fulfill their work responsibilities.	(Soundararajan, 2013)
TC4	Our team manages day-to-day activities autonomously with minimal management intervention.	(Soundararajan, 2013)
TC5	Our working environment promotes collaborative efforts.	(Soundararajan, 2013)
Dimension: Iterative and Incremental Development Processes (DP)		
DP1	Story prioritization is effectively managed in our backlog.	(Soundararajan, 2013)
DP2	We keep a well-maintained product backlog.	(Soundararajan, 2013)
DP3	Our code is implemented in short, manageable iterations.	(So and Scholl, 2009)
DP4	Our iteration-end software consistently meets the production code quality requirements.	(So and Scholl, 2009)
IIDP5	The progress of our project is primarily gauged by the working software produced.	(So and Scholl, 2009)
Dimension: Continuous Development Process Improvement (PI)		
PI1	Changing requirements are viewed as value additions, not extra burdens.	(Looks et al., 2021)
PI2	Regular retrospectives lead to actionable strategies for project improvement.	(Looks et al., 2021)
PI3	All team members are proactive in the continuous improvement of the project.	(Looks et al., 2021)
PI4	Insights from retrospectives result in tangible improvements.	(Looks et al., 2021)
PI5	Experimental improvements are explored throughout the project's duration.	(Looks et al., 2021)

Item IDs (e.g., CI1, TC1) were added for clarity in referencing the final items. "Item inspired by" indicates the study from which the initial item concept was drawn or adapted. Responses were given on a 7-point scale ranging from 1 ("never true") to 7 ("always true").

4 Study 2: methodology and results

Following the development and initial refinement in Study 1, the final 20-item Agile Team Practice Inventory (ATPI-SD-20) was administered to a larger, independent sample in Study 2. This main validation study aimed to rigorously assess the psychometric properties of the newly developed instrument. This section outlines the methodology employed and presents the empirical results concerning the ATPI-SD-20's reliability, factor structure, and initial convergent validity.

4.1 Methodology

A quantitative, cross-sectional survey design was utilized for Study 2. Data were collected online via electronic questionnaires at a single point in time.

Participants were recruited primarily through established crowdsourcing platforms, namely Amazon Mechanical Turk (MTurk) and Prolific, which are commonly used for accessing diverse participant pools in social science research. Recruitment efforts were supplemented by postings on relevant professional social media platforms (e.g., LinkedIn) to enhance reach within the target population of software development professionals. Eligibility criteria required participants to be at least 18 years old and currently working as part of a software development team. To ensure data quality, standard procedures were implemented, including the use of attention check items embedded within the survey and response time analysis. Participants with exceptionally fast completion times [specifically, a Time Relative Speed Index (Time RSI) below 2.0, suggesting potentially non-attentive responding (Leiner, 2019)] were excluded from the final analysis. This process resulted in a final sample size of $N = 237$ valid responses.

As intended for a study focused on software development agility, the sample was primarily composed of professionals in core technical and leadership roles. The majority identified as Developers (54.0%), complemented by significant representation from Managers (19.0%) and QA/Testers (17.3%). This composition ensures that the collected data reflects the perspectives of those directly involved in enacting and overseeing agile practices. The participants predominantly came from medium-sized organizations (51–500 employees, 44.7%) and possessed considerable professional experience, with over half of the sample (50.2%) having between 3 and 9 years in the field. The sample was also geographically diverse, with the largest contingents from the United States (35.9%) and India (16.9%), representing major hubs of the global software industry. This international distribution provides a broad perspective, which strengthens the external validity of our findings. Participation was entirely voluntary, and informed consent outlining the study's purpose, procedures, confidentiality, and voluntary nature was obtained electronically from all participants prior to their engagement with the survey. Data collection took place between July and August 2024.

The survey instrument consisted of three sections, collecting data on the following measures:

Agile team practice inventory for software development (ATPI-SD-20)

The primary measure was the newly developed 20-item ATPI-SD, finalized during the preceding development phase (the items are listed in Table 3). The ATPI-SD-20 comprises four distinct subscales, each containing five items: Customer Involvement (CI), Team Collaboration (TC), Iterative and Incremental Development Processes (DP), and Continuous Development Process Improvement (PI). Participants were asked to rate the extent to which each statement accurately described their current software development team’s practices. Responses were captured using a 7-point Likert scale, anchored at 1 (“Almost Never True”) and 7 (“Almost Always True”). Higher scores on the items, subscales, and the total scale reflect a perception of more frequent or intense application of agile practices within the team.

Self-rated team agility

To provide an initial assessment of convergent validity, a single-item measure was included. Participants were asked to provide a global rating of their team’s overall agility (“How agile do you rate your team?”) on a 7-point scale, ranging from 1 (“Fully Traditional”/“Not agile at all”) to 7 (“Fully Agile”).

Demographic and professional information

Participants provided information on their professional role, work experience, company size, and geography. This data was used to characterize the sample and ensure its relevance to the research question.

Data analysis procedures were planned to evaluate the psychometric properties of the ATPI-SD-20. Descriptive statistics, including means (M) and standard deviations (SD), were calculated for the ATPI-SD total score and each of its four subscales to characterize the sample’s responses. Internal consistency reliability was assessed for the total scale and each subscale using Cronbach’s Alpha coefficients. To evaluate the underlying factor structure of the instrument, a Confirmatory Factor Analysis (CFA) was performed using JASP (Version 0.19.1). The CFA tested the hypothesized four-factor model (CI, TC, DP, PI), specifying that each factor was indicated by its corresponding five items. Given the ordinal nature of the Likert-scale response data and potential deviations from multivariate normality (as observed in Study 1), Maximum Likelihood (ML) estimation, which offers some robustness, was employed. Model fit was evaluated using a range of established fit indices, including the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Normed Fit Index (NFI), Parsimony Normed Fit Index (PNFI), Relative Fit Index (RFI), Incremental Fit Index (IFI), and Relative Noncentrality Index (RNI), following recommendations for comprehensive model fit assessment (Hair et al., 2010; Hu and Bentler, 1999). Standardized factor loadings for each item on its designated factor and the correlations between the latent factors were examined to further interpret the structural model. Finally, convergent validity was assessed by calculating the Spearman rank correlation coefficient (ρ) between the ATPI-SD total score and the single-item self-rated team agility measure. Spearman correlation was chosen due to the ordinal nature of the single-item measure and its robustness to potential non-normality in the score distributions.

4.2 Results

This section presents the results of the data analysis conducted to evaluate the psychometric properties of the ATPI-SD-20 based on the sample of $N = 237$ software development professionals.

First, descriptive statistics were calculated for the ATPI-SD total score and its four subscales using mean scores on the 1-to-7 scale, as recommended for clearer interpretation. As shown in Table 4, participants, on average, rated their teams’ agility as moderately high, with mean scores consistently above the scale’s midpoint. The mean total agility score across all 20 items was $M = 5.32$ ($SD = 0.73$). The subscale means ranged from $M = 5.17$ for Continuous Development Process Improvement (PI) to a high of $M = 5.48$ for Team Collaboration (TC), indicating a generally positive perception of agile practices within the sampled teams.

Next, the internal consistency reliability of the scales was assessed using Cronbach’s Alpha. The results, presented in Table 5, indicated excellent internal consistency for the total 20-item ATPI-SD scale ($\alpha = 0.89$). The reliability coefficients for the four subscales were acceptable to good: Customer Involvement ($\alpha = 0.69$), Team Collaboration ($\alpha = 0.70$), Iterative and Incremental Development Processes ($\alpha = 0.78$), and Continuous Development Process Improvement ($\alpha = 0.84$). These findings demonstrate good overall scale cohesion and suggest that the items within each subscale measure their respective underlying construct consistently, meeting commonly accepted thresholds for research instruments (Nunnally and Bernstein, 1994).

To evaluate the instrument’s factor structure, a Confirmatory Factor Analysis (CFA) was conducted using Maximum Likelihood estimation. This analysis tested the hypothesized four-factor structure of the ATPI-SD-20, where Customer Involvement, Team Collaboration, Iterative and Incremental Development Processes,

TABLE 4 Descriptive statistics for ATPI-SD scales ($N = 237$).

Scale	Min	Max	M	SD
ATPI-SD total score (Mean)	3.70	6.95	5.32	0.73
Customer involvement (CI)	1.20	7.00	5.21	0.90
Team collaboration (TC)	3.40	7.00	5.48	0.82
Iterative and incremental dev. proc. (DP)	2.80	7.00	5.42	0.88
Continuous dev. process improvement (PI)	2.40	7.00	5.17	1.05

TABLE 5 Internal consistency reliability (Cronbach’s Alpha) for ATPI-SD scales.

Scale	Items	Cronbach’s Alpha
Agility total score (ATPI-SD-20)	20	0.89
Customer involvement (CI)	5	0.69
Team collaboration (TC)	5	0.70
Iterative and incremental dev. proc. (DP)	5	0.78
Continuous dev. process improvement (PI)	5	0.84

TABLE 6 Confirmatory factor analysis fit indices for the four-factor ATPI-SD model.

Index	Model value	Recommended threshold
Comparative fit index (CFI)	0.88	> 0.90
Tucker-Lewis index (TLI)	0.86	> 0.90
Normed fit index (NFI)	0.80	> 0.90
Incremental fit index (IFI)	0.88	> 0.90
Relative noncentrality index (RNI)	0.88	> 0.90
Parsimony normed fit index (PNFI)	0.69	> 0.50 (indicates parsimony)

and Continuous Development Process Improvement were specified as distinct but correlated latent factors, each indicated by its five respective items. The assessment of model fit, based on several established indices (see Table 6), suggested a moderate to acceptable representation of the observed data by the proposed four-factor model. Specifically, the Comparative Fit Index (CFI = 0.88) and Tucker-Lewis Index (TLI = 0.86) fell slightly below the conventional threshold of 0.90 for good fit but were within ranges often considered acceptable for newly developed scales in social science research (Hair et al., 2010; Hu and Bentler, 1999). Other indices, such as the Normed Fit Index (NFI = 0.80) and the Incremental Fit Index (IFI = 0.88), corroborated this assessment, while the Parsimony Normed Fit Index (PNFI = 0.69) indicated moderate parsimony. Overall, these fit indices suggest that the hypothesized four-factor structure provides a reasonable approximation of the relationships in the data, acknowledging potential room for model improvement.

Further examination within the CFA focused on the standardized factor loadings. Most items loaded substantially and significantly ($p < 0.001$) onto their intended latent factors, with values typically ranging from moderate to high (0.43 to 0.96). This indicates that the items generally serve as reasonably good indicators of their respective dimensions. Notably, the weakest loading was observed for item CI5 (“It is acceptable for team members to express disagreement with customers...”), suggesting it might be a comparatively less precise indicator of Customer Involvement than the other items in that subscale. The analysis also estimated the correlations between the four latent factors. All factors were found to be significantly and positively correlated with each other ($p < 0.001$), with correlation coefficients ranging from moderate to strong (ρ between 0.54 for CI ↔ TC, and 0.79 for DP ↔ PI). These inter-factor correlations support the conceptualization of the ATPI-SD dimensions as distinct yet related facets of the broader construct of team agility.

Finally, to assess initial convergent validity, the relationship between the ATPI-SD total score and the single-item self-rating of overall team agility was examined using Spearman’s rank correlation. A statistically significant, positive correlation was found ($\rho = 0.40$, $p < 0.001$). This moderate correlation indicates that teams scoring higher on the specific practices measured by the ATPI-SD-20 also tend to be perceived as more agile overall by their members, providing support for the convergent validity of the new instrument.

4.3 Summary of validation results

The results from Study 2 provide initial support for the psychometric properties of the 20-item ATPI-SD. The scale demonstrated excellent overall internal consistency and acceptable to good reliability for its four subscales. The confirmatory factor analysis suggested that the proposed four-factor structure (Customer Involvement, Team Collaboration, Iterative/Incremental Processes, Continuous Improvement) offers a moderate to acceptable fit to the empirical data, with most items showing strong loadings on their intended factors and factors being appropriately correlated. Furthermore, the significant positive correlation with a global self-rating of team agility supports the convergent validity of the ATPI-SD. While the model fit is not perfect and some items show weaker loadings or higher residuals, the overall findings suggest the ATPI-SD is a promising instrument for measuring team agility practices.

5 Discussion

This Study addressed the identified gap in rigorously validated instruments for measuring team agility within software development by detailing the systematic development and initial psychometric validation of the Agile Team Practice Inventory for Software Development (ATPI-SD-20). This 20-item, four-dimensional questionnaire aims to capture core agile practices at the team level. The preceding section presented the validation results based on data from 237 software development professionals. This section discusses these findings, interprets the psychometric properties of the ATPI-SD-20, acknowledges the study’s strengths and limitations, outlines implications for research and practice, and suggests directions for future research.

5.1 Interpretation of psychometric findings

Study 2 yielded encouraging initial results regarding the ATPI-SD-20’s psychometric quality, while also illuminating areas for potential future refinement. Regarding reliability, the instrument demonstrated strong internal consistency. The overall 20-item ATPI-SD scale achieved excellent reliability ($\alpha = 0.89$), indicating that the items collectively form a cohesive measure of the target construct. The four subscales also showed acceptable to good internal consistency, with Cronbach’s Alpha coefficients ranging from 0.69 (Customer Involvement) to 0.84 (Continuous Development Process Improvement). These values meet or exceed commonly accepted thresholds for reliability in organizational research (Nunnally and Bernstein, 1994), suggesting the items within each dimension consistently measure their intended underlying facet of agility. While acceptable, the slightly lower alpha for the Customer Involvement subscale might suggest this dimension captures a slightly more heterogeneous set of practices compared to the others, potentially warranting closer examination of its items in future research.

Structurally, the Confirmatory Factor Analysis (CFA) provided initial support for the hypothesized four-factor model representing Customer Involvement, Team Collaboration,

Iterative and Incremental Development Processes, and Continuous Development Process Improvement. Key model fit indices, such as the CFI (0.88) and TLI (0.86), indicated a moderate to acceptable fit of the model to the observed data. While these values do not meet the most stringent criteria for “excellent” fit sometimes proposed, they fall within a range often considered reasonable and informative in social science research, particularly for newly developed instruments assessing complex, multifaceted constructs like team agility (Hair et al., 2010). This suggests that the theoretically derived four-dimensional model offers a plausible representation of the structure underlying the ATPI-SD-20 items. Further supporting this, the latent factors were significantly and positively correlated (ranging from 0.54 to 0.79), consistent with the conceptualization of these dimensions as distinct but related aspects of overall team agility. Examination of the standardized factor loadings showed that most items loaded significantly and substantially onto their intended factors, confirming their relevance. However, the moderate overall model fit, along with specific findings like the weaker loading of item CI5, suggests that the model could potentially be improved through future refinements, such as revising specific items or exploring minor structural modifications based on further empirical data.

Initial evidence supporting the convergent validity of the ATPI-SD-20 emerged from its relationship with a global self-rating of team agility. A statistically significant, moderate positive correlation ($\rho = 0.40, p < 0.001$) was found between the ATPI-SD total score and participants' single-item global rating of their team's agility. This finding indicates that teams perceived as more frequently enacting the specific agile practices measured by the ATPI-SD-20 are also generally viewed as more agile overall by their members. The moderate magnitude of this correlation is sensible, given that the ATPI-SD-20 assesses specific behavioral practices across four dimensions, whereas the single item captures a broader, potentially more subjective, overall impression. Such divergence is commonly observed when comparing multi-item, multi-dimensional scales with single-item global assessments.

In summary, the ATPI-SD-20 demonstrates promising initial psychometric characteristics. It appears capable of reliably measuring four distinct but interconnected dimensions of agile team practices, and its scores show a meaningful relationship with global perceptions of team agility.

5.2 Strengths and contributions

This research contributes several strengths to the field. First, the development of the ATPI-SD followed a systematic and transparent multi-phase process, integrating theoretical grounding from the agile literature, input from domain experts for content validity, and empirical testing via pilot and main studies. This structured approach directly addresses a common criticism regarding the lack of rigorous development and validation in many existing agility assessment tools (Chronis and Gren, 2016; Yürüm et al., 2018). Second, by focusing on core agile principles rather than prescribing adherence to a specific named methodology (e.g., Scrum), the ATPI-SD potentially offers broader applicability across diverse teams employing various agile, hybrid, or customized approaches.

Third, its multi-dimensional structure (CI, TC, DP, PI) allows for a more nuanced assessment of team agility, moving beyond a single overall score to enable the identification of specific areas of strength or weakness in a team's practices. Fourth, this study provides crucial initial empirical evidence regarding the ATPI-SD's reliability and construct validity, establishing a psychometric foundation often missing for practitioner-derived assessment tools. Finally, the resulting 20-item instrument offers a balance between comprehensive coverage and practical brevity, making it feasible for use in research surveys and organizational diagnostics without imposing an excessive burden on respondents.

5.3 Limitations and considerations

Despite our promising initial findings, this study has several limitations that can be framed as threats to different forms of validity. These also highlight important avenues for future research.

A primary concern relates to conclusion validity, which addresses the statistical robustness of our findings. The CFA results indicated a moderate-to-acceptable model fit rather than an excellent one. This suggests that while the proposed four-factor structure is plausible, the item structure may not perfectly represent the underlying reality due to the inherent complexity and overlap of agility dimensions. The slightly lower reliability of the Customer Involvement subscale also highlights it as an area for potential future improvement, potentially through refinement of items like CI5 which showed weaker properties.

The validation evidence also faces challenges to its construct validity, meaning whether we are truly measuring the intended concept. The validation evidence is initial; further research is necessary to establish discriminant validity (demonstrating that ATPI scores are distinct from related constructs like general team cohesion) and predictive validity (assessing if ATPI scores predict meaningful outcomes like performance or innovation). Furthermore, data were collected via self-report measures. A key consideration is that this self-reported assessment of agile processes is distinct from objective, indicator-based measures of agile outputs. While attention checks were employed, self-reports remain susceptible to biases such as social desirability. Future research could significantly strengthen construct validity by triangulating ATPI scores with behavioral data from agile software development tools like Jira or Azure DevOps. Correlating self-reported practices (e.g., “Story prioritization is effectively managed”) with objective metrics (e.g., backlog refinement frequency, cycle time) would provide a more holistic and robust picture of team agility.

The study's internal validity is constrained by its cross-sectional design, which precludes causal inferences. We can only report associations, not cause-and-effect relationships. Test-retest reliability, to examine the stability of scores over time, also needs investigation to further strengthen claims about the measure's consistency.

Finally, there are significant threats to external validity (generalizability). Although efforts were made to recruit a relevant sample, its specific demographic profile (e.g., concentration in the USA and India) may limit the generalizability of findings.

Cultural nuances could profoundly affect how agile practices are perceived and reported, which directly impacts the instrument's content validity in different contexts. For instance, in cultures that highly value structure and predictability, practices associated with high agility might be negatively connoted as “planless” or chaotic. Conversely, in cultures that prize adaptability, the same concepts may carry a strongly positive connotation. This potential for culturally-dependent interpretation underscores the critical need for cross-cultural validation before the ATPI is applied globally. Moreover, our study did not account for other important boundary conditions that could shape how agility is perceived and enacted. Contextual variables such as the specific agile methodology employed (e.g., Scrum, Kanban), organizational culture, team size, work experience, or project complexity were not included in the analysis. Future research should investigate the impact of these factors to provide a more nuanced understanding of the ATPI's performance across diverse environments.

5.4 Implications for research and practice

The development and initial validation of the ATPI-SD-20 hold implications for both the research community and practitioners involved in agile software development. For researchers, the ATPI-SD offers a standardized, theory-grounded instrument with demonstrated initial reliability and construct validity, enabling empirical investigation into the antecedents of team agility, such as organizational culture or leadership styles, and examination of its consequences, including team performance, innovation, or job satisfaction. It also facilitates comparative studies assessing agility levels across different types of teams, organizations, or methodology implementations, and can serve as a foundation for developing adapted or translated versions following appropriate validation. For practitioners, organizations and agile teams can leverage the ATPI-SD as a practical diagnostic tool to assess their current state of agile practice adoption across the four key dimensions (CI, TC, DP, PI). This allows them to identify specific areas where practices are strong versus those needing improvement, thereby guiding targeted interventions like coaching or training. Furthermore, the ATPI-SD can be used to benchmark agility profiles against reference points or track progress longitudinally during agile transformations, and to facilitate data-informed team discussions and retrospectives focused on enhancing specific agile practices beyond purely anecdotal assessments.

5.5 Directions for future research

Building upon this initial validation work, several key avenues for future research emerge. A primary focus should be on comprehensive psychometric validation, conducting studies explicitly designed to assess discriminant validity (e.g., differentiating ATPI-SD scores from related team constructs), predictive validity (e.g., linking ATPI-SD scores to objective project outcomes), and test-retest reliability over appropriate time intervals. Furthermore, investigating measurement invariance is crucial to determine if the ATPI-SD functions equivalently

across important subgroups, such as different roles, industries, or cultural contexts, thereby ensuring fair comparisons. Attention should also be directed toward item refinement and model optimization; utilizing advanced psychometric techniques like Item Response Theory (IRT) on larger datasets could help diagnose item performance more precisely and guide potential revisions, particularly for the Customer Involvement subscale, aiming to enhance both subscale homogeneity and overall model fit. To bolster confidence in generalizability, cross-validation of the ATPI-SD-20 with diverse and larger samples from varied geographical, industrial, organizational, and cultural backgrounds is necessary. Longitudinal applications represent another valuable direction, enabling researchers to track the evolution of team agility over time, assess the impact of specific interventions, and evaluate the instrument's sensitivity to change. Finally, comparative analyses are warranted, directly contrasting the performance and utility of the ATPI-SD with other existing agility assessment instruments and correlating its scores with more objective indicators of team process or output where available.

6 Conclusion

The pervasive adoption of agile methodologies in software development underscores the critical need for reliable and valid tools to assess team agility. Existing measurement approaches often lack sufficient empirical validation or practical applicability, hindering both rigorous research and effective organizational practice. This paper addressed this gap by detailing the systematic development and initial psychometric validation of the Agile Team Practice Inventory for Software Development (ATPI-SD), a new 20-item, four-dimensional questionnaire designed to measure core agile practices at the team level.

Grounded in agile principles and refined through expert review and empirical testing in Study 1, the ATPI-SD aims to provide a balanced measure applicable across various agile contexts. Study 2, the main validation study involving 237 software development professionals, provided promising initial support for the instrument's psychometric properties. The ATPI-SD demonstrated excellent overall internal consistency, with its four subscales (Customer Involvement, Team Collaboration, Iterative and Incremental Development Processes, Continuous Development Process Improvement) showing acceptable to good reliability. Confirmatory Factor Analysis indicated a moderate but acceptable fit for the hypothesized four-factor structure, with most items strongly representing their intended dimensions. Furthermore, a significant positive correlation with a global self-assessment of team agility provided initial evidence for convergent validity.

While acknowledging the need for further validation—particularly concerning discriminant and predictive validity, refinement of specific items, and testing across more diverse samples and cultural contexts—the ATPI-SD represents a significant step forward. It offers researchers a standardized, theory-based tool to investigate the complex dynamics of team agility and its relationship with various organizational outcomes. For practitioners, the ATPI-SD provides a concise, actionable diagnostic instrument to assess current practices, guide

improvement efforts, and foster data-driven discussions within agile teams and organizations.

Ultimately, the development of the ATPI-SD contributes to the ongoing effort to bridge the gap between the practical realities of agile software development and the requirements of rigorous scientific measurement. By providing a more robust means of assessing team agility, this work aims to facilitate a deeper understanding and more effective implementation of agile principles, ultimately contributing to the success of software development endeavors. Future research focused on the continued refinement and validation of the ATPI-SD will be essential to fully realize its potential as a valuable tool for the agile community.

Data availability statement

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

Ethics statement

Ethical approval was not required for the studies involving humans because ethical review and approval were waived for this study due to the nature of the research involving anonymized survey data on non-sensitive topics from non-vulnerable adult populations, which falls outside the requirements for formal review according to institutional guidelines. Informed consent was obtained from all subjects involved in the study. Participants were presented with a consent form detailing the study purpose, procedures, risks, benefits, confidentiality measures, and voluntary nature of participation before beginning the survey, and indicated their consent by clicking an agreement button.

Author contributions

NR: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing – original draft, Writing

– review & editing. MS: Conceptualization, Formal analysis, Methodology, Supervision, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Conflict of interest

NR was employed at cosinex GmbH.

The remaining author declares that the research 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) declare that Gen AI was used in the creation of this manuscript. To assist with language refinement, grammar and style checking, and to explore alternative phrasing and structural suggestions for improving clarity and flow. It was also utilized for adapting the manuscript to specific journal template requirements (LaTeX). The core scientific content, data analysis, interpretations, and conclusions were solely developed by the human authors.

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.

References

- Arumugam, C., and Vaidyanathan, S. (2022). Agile team measurement to review the performance in global software development," in *Research Anthology on Agile Software, Software Development, and Testing* (IGI Global Scientific Publishing), 2015–2025. doi: 10.4018/978-1-6684-3702-5.ch096
- Avila, D. T., Petegem, W. V., and Snoeck, M. (2022). Improving teamwork in agile software engineering education: the ASEST+ framework. *IEEE Trans. Educ.* 65, 18–29. doi: 10.1109/TE.2021.3084095
- Beck, K., Beedle, M., van Bennekum, A., Cockburn, A., Fowler, M., Grenning, J., et al. (2001). Manifest für Agile Softwareentwicklung. *Abgerufen am 20:2024*.
- Berlas, M. F. (2024). *Software Metrics in Agile Software Development: A Review Report*. Authorea Preprints. doi: 10.36227/techrxiv.171084962.20068546/v1
- Bhatta, N., and Thite, M. (2018). "Agile approach to E-HRM project management," in *e-HRM* (Routledge), 57–72. doi: 10.4324/9781315172729-4
- Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quinonez, H. R., and Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: a primer. *Front. Public Health* 6:149. doi: 10.3389/fpubh.2018.00149
- Chronis, K., and Gren, L. (2016). "Agility measurements mismatch: a validation study on three agile team assessments in software engineering," in *Agile Processes, in Software Engineering, and Extreme Programming. Lecture Notes in Business Information Processing*, eds. H. Sharp, and T. Hall (Cham: Springer International Publishing), 16–27. doi: 10.1007/978-3-319-33515-5_2
- Cunha, F., Perkusich, M., Guimarães, E., Santos, R. R., Rique, T., Albuquerque, D., et al. (2024). An insight into the capabilities of professionals and teams in agile software development: an update of the systematic literature review. *J. Commun. Softw. Syst.* 20, 99–112. doi: 10.24138/jcomss-2023-0172

- Dingsoeyr, T., Falessi, D., and Power, K. (2019). Agile development at scale: the next frontier. *IEEE Softw.* 36, 30–38. doi: 10.1109/MS.2018.2884884
- Dugbartey, A. N., and Kehinde, O. (2025). Optimizing project delivery through agile methodologies: balancing speed, collaboration and stakeholder engagement. *World J. Adv. Res. Rev.* 25, 1237–1257. doi: 10.30574/wjarr.2025.25.1.0193
- Dutra, E., Lima, P., Cerdeiral, C., Djiir, B., and Santos, G. (2022). TACT: an instrument to assess the organizational climate of agile teams - a preliminary study. *J. Softw. Eng. Res. Dev.* 10, 1–11. doi: 10.5753/jserd.2021.1973
- Dybå, T., and Dingsøyr, T. (2008). Empirical studies of agile software development: a systematic review. *Inf. Softw. Technol.* 50, 833–859. doi: 10.1016/j.infsof.2008.01.006
- Esang, M. O., Johnson, E. A., Attai, K., Inyangetoh, J. A., Dan, E. E., Okonny, K. E., et al. (2024). Exploring agile methodology in developing a web-based result computation and transcript system: a case study of federal polytechnic Ukana. *Eur. J. Comput. Sci. Inf. Technol.* 12, 1–17. doi: 10.37745/ejsit.2013/vol12n4117
- Făgărăsan, C., Cristea, C., Cristea, M., Popa, O., Mihele, C., and Pîslă, A. (2022). “Key performance indicators used to measure the adherence to the iterative software delivery model and policies,” in *Iop Conference Series Materials Science and Engineering* 1256, 012038. doi: 10.1088/1757-899X/1256/1/012038
- Frisbie, D. A., and Brandenburg, D. C. (1979). Equivalence of questionnaire items with varying response formats. *J. Educ. Measur.* 16, 43–48. doi: 10.1111/j.1745-3984.1979.tb00085.x
- Ganesh, N., and Narayanan, R. C. (2019). Challenges faced in the enterprise resource planning material management section when transitioning towards agile software development. *Int. J. Eng. Adv. Technol.* 8, 3472–3475. doi: 10.35940/ijeat.F9521.088619
- Goyal, A. (2023). Driving continuous improvement in engineering projects with ai-enhanced agile testing and machine learning. *Int. J. Adv. Res. Sci. Commun. Technol.* 3, 1320–1331. doi: 10.48175/IJARSC-14000T
- Gren, L. (2022). “What makes effective leadership in agile software development teams?” in *Proceedings of the 44th International Conference on Software Engineering*, 2402–2414. doi: 10.1145/3510003.3510100
- Gren, L., Torkar, R., and Feldt, R. (2015). “Group maturity and agility, are they connected? A survey study,” in *2015 41st Euromicro Conference on Software Engineering and Advanced Applications*, 1–8. doi: 10.1109/SEAA.2015.31
- Gren, L., Torkar, R., and Feldt, R. (2017). Group development and group maturity when building agile teams: a qualitative and quantitative investigation at eight large companies. *J. Syst. Softw.* 124, 104–119. doi: 10.1016/j.jss.2016.11.024
- Gunsberg, D., Callow, B., Ryan, B., Suthers, J., Baker, P. A., and Richardson, J. (2018). Applying an organisational agility maturity model. *J. Organ. Change Manag.* 31, 1315–1343. doi: 10.1108/JOCM-10-2017-0398
- Hair, J., Ortinau, D., and Harrison, D. (2010). *Essentials of Marketing Research*, volume 2. New York: McGraw-Hill/Irwin.
- Henriques, V., and Tanner, M. (2017). A systematic literature review of agile maturity model research. *Interdisc. J. Inform. Knowl. Manag.* 12, 053–073. doi: 10.28945/3666
- Highsmith, J. A., and Highsmith, J. (2002). *Agile Software Development Ecosystems*. Addison-Wesley Professional. Google-Books-ID: uE4FGFOHs2EC.
- Hu, L., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equat. Model.* 6, 1–55. doi: 10.1080/10705519909540118
- Hukkelberg, I., and Berntzen, M. (2019). “Exploring the challenges of integrating data science roles in agile autonomous teams,” in *International Conference on Agile Software Development* (Cham: Springer International Publishing), 37–45. doi: 10.1007/978-3-030-30126-2_5
- Jin, Z. F. (2024). Withdraw. *Appl. Comput. Eng.* 100, 174–179. doi: 10.54254/2755-2721/65/20240491
- Junker, T. L., Bakker, A. B., Gorgievski, M. J., and Derks, D. (2021). Agile work practices and employee proactivity: a multilevel study. *Hum. Relat.* 75, 2189–2217. doi: 10.1177/00187267211030101
- Kalenda, M., Hyna, P., and Rossi, B. (2018). Scaling agile in large organizations: practices, challenges, and success factors. *J. Softw. Evolut. Pro.* 30:e1954. doi: 10.1002/smr.1954
- Kruyen, P. M., Emons, W. H. M., and Sijtsma, K. (2014). Assessing individual change using short tests and questionnaires. *Appl. Psychol. Measur.* 38, 201–216. doi: 10.1177/0146621613510061
- Latif, F., Bhatti, S. N., Sarwar, S., Mohsen, A. M., and Saboor, A. (2017). Optimized order of software testing techniques in agile process - a systematic approach. *Int. J. Adv. Comput. Sci. Applic.* 8:509. doi: 10.14569/IJACSA.2017.080144
- Leffingwell, D. (2007). *Scaling Software Agility: Best Practices for Large Enterprises*. Pearson Education. Google-Books-ID: N1uf2VRTyGc.
- Leiner, D. J. (2019). Too fast, too straight, too weird: non-reactive indicators for meaningless data in internet surveys. *Surv. Res. Methods*, 13, 229–248.
- Looks, H., Fangmann, J., Thomaschewski, J., Escalona, M.-J., and Schön, E.-M. (2021). “Towards a standardized questionnaire for measuring agility at team level,” in *Agile Processes in Software Engineering and Extreme Programming, Lecture Notes in Business Information Processing*, eds. P. Abrahamsson, M. Marchesi, X. Wang, and P. Kruchten (Cham: Springer International Publishing), 71–85. doi: 10.1007/978-3-030-78098-2_5
- Mahadik, S., Murthy, K. K. K., Cheruku, S. R., Jain, P. A., and Goel, O. (2022). Agile product management in software development. *Int. J. Res. Public. Semin.* 13, 453–467. doi: 10.36676/jrps.v13.i5.1512
- Maharao, C. S. (2024). The influence of agile practices on project outcomes: performance, stakeholder satisfaction, and team dynamics. *Shodhkosh J. Visual Perf. Arts* 5:2284. doi: 10.29121/shodhkosh.v5.i1.2024.2284
- Maneva, M., Koceska, N., and Koceski, S. (2017). Measuring agility in agile methodologies. *J. Appl. Econ. Bus.* 5, 1857–8721.
- Mashmool, A., Khosravi, S., Joloudari, J. H., Inayat, I., Gandomani, T. J., and Mosavi, A. (2021). “A statistical model to assess the team’s productivity in agile software teams,” in *2021 IEEE 4th International Conference and Workshop Óbuda on Electrical and Power Engineering (CANDO-EPE)* (IEEE), 11–18. doi: 10.1109/CANDO-EPE54223.2021.9667902
- Menezes, R., Marinho, M., and Sampaio, S. (2024). “Metrics in large-scale agile software development: a multivocal literature review,” in *Congresso Ibero-Americano em Engenharia de Software (CIBSE)* (SBC), 106–120. doi: 10.5753/cibse.2024.28442
- Moniruzzaman, A. B. M., and Hossain, D. S. A. (2013). Comparative study on agile software development methodologies. *arXiv preprint arXiv:1307.3356*.
- Moraga-Díaz, R., and Piñango, H. (2023). *Continuous improvement in software development and digital product management: addressing the challenges of the digital economy*. Preprints. doi: 10.20944/preprints202307.0285.v1
- Moran, L. A., Guyatt, G. H., and Norman, G. R. (2001). Establishing the minimal number of items for a responsive, valid, health-related quality of life instrument. *J. Clin. Epidemiol.* 54, 571–579. doi: 10.1016/S0895-4356(00)00342-5
- Njanka, S. Q., Sandula, G., and Colomo-Palacios, R. (2021). IT-business alignment: a systematic literature review. *Procedia Comput. Sci.* 181, 333–340. doi: 10.1016/j.procs.2021.01.154
- Nunnally, J., and Bernstein, I. (1994). *Psychometric Theory*. New York: McGraw-Hill 3 edition.
- Ojha, T. R. (2023). Critical success factors of agile software development - a systematic literature review. *Scitech Nepal* 17, 49–57. doi: 10.3126/scitech.v17i1.60467
- Özcan Top, Z., and Demirs, O. (2019). Application of a software agility assessment model - AgilityMod in the field. *Comput. Stand. Interf.* 62, 1–16. doi: 10.1016/j.csi.2018.07.002
- Qumer, A., and Henderson-Sellers, B. (2008). A framework to support the evaluation, adoption and improvement of agile methods in practice. *J. Syst. Softw.* 81, 1899–1919. doi: 10.1016/j.jss.2007.12.806
- Rahman, A. (2024). Agile project management: analyzing the effectiveness of agile methodologies in it projects compared to traditional approaches. *Ajbais* 4, 53–69. doi: 10.69593/ajbais.v4i04.127
- Rahy, S., and Bass, J. M. (2021). Managing non-functional requirements in agile software development. *IET Softw.* 16, 60–72. doi: 10.1049/sfw.12037
- Rathor, S., Xia, W., and Batra, D. (2023). Achieving software development agility: different roles of team, methodological and process factors. *Inf. Technol. People* 37, 835–873. doi: 10.1108/ITP-10-2021-0832
- Royce, W. W. (1970). “Managing the development of large software systems: concepts and techniques,” in *Proceedings of IEEE WESTCON* (Los Angeles, CA), 1–9.
- Sathe, C. A., and Panse, C. (2022). Analyzing the impact of agile mindset adoption on software development teams productivity during COVID-19. *J. Adv. Manag. Res.* 20, 96–115. doi: 10.1108/JAMR-05-2022-0088
- Schriesheim, C. A., Kopelman, R. E., and Solomon, E. (1989). The effect of grouped versus randomized questionnaire format on scale reliability and validity: a three-study investigation. *Educ. Psychol. Measur.* 49, 487–508. doi: 10.1177/001316448904900301
- Shah, D. (2024). Agile methodologies and their impact on software project success with case studies. *Shodhkosh J. Visual Perfor. Arts* 5:1707. doi: 10.29121/shodhkosh.v5.i1.2024.1707
- Sidky, A., Arthur, J., and Bohner, S. (2007). A disciplined approach to adopting agile practices: the agile adoption framework. *Innov. Syst. Softw. Eng.* 3, 203–216. doi: 10.1007/s11334-007-0026-z
- Silva-Martinez, J. (2023). Conceptualization of agile leadership characteristics and outcomes from NASA agile teams as a path to the development of an agile leadership theory. *J. Creat. Value* 10, 173–188. doi: 10.1177/23949643231202894
- So, C., and Scholl, W. (2009). “Perceptive agile measurement: new instruments for quantitative studies in the pursuit of the social-psychological effect of agile practices,” in *Agile Processes in Software Engineering and Extreme Programming, Lecture Notes in Business Information Processing*, eds. P. Abrahamsson, M. Marchesi,

and F. Maurer (Berlin, Heidelberg: Springer), 83–93. doi: 10.1007/978-3-642-01853-4_11

Soundararajan, S. (2013). *Assessing Agile Methods: Investigating Adequacy, Capability, and Effectiveness (An Objectives, Principles, Strategies Approach)*. Doctoral dissertation, Virginia Tech.

Strode, D. E., Dingsoyr, T., and Lindsjörn, Y. (2022). A teamwork effectiveness model for agile software development. *Empir. Softw. Eng.* 27:56. doi: 10.1007/s10664-021-10115-0

Thiyagarajan, S., Saldanha, P., Govindan, R., Leena, K., and Prathyusha, P. V. (2024). Development of agile scrum perception tool to evaluate students' opinions on agile methodology in nursing education. *Int. J. Appl. Basic Med. Res.* 14, 35–41. doi: 10.4103/ijabmr.ijabmr_423_23

Toepoel, V., Das, M., and Van Soest, A. (2009). Design of web questionnaires: the effects of the number of items per screen. *Field Methods* 21, 200–213. doi: 10.1177/1525822X08330261

Tripp, J., Riemenschneider, C. K., and Thatcher, J. (2016). Job satisfaction in agile development teams: agile development as work redesign. *J. Assoc. Inf. Syst.* 17:1. doi: 10.17705/1jais.00426

Uraon, R. S., Chauhan, A., Bharati, R., and Sahu, K. (2023). Do agile work practices impact team performance through project commitment? Evidence from the information technology industry. *Int. J. Product. Perform. Manag.* 73, 1212–1234. doi: 10.1108/IJPPM-03-2023-0114

Yürüm, O. R., Demirörs, O., and Rabhi, F. (2018). “A comprehensive evaluation of agile maturity self-assessment surveys,” in *Software Process Improvement and Capability Determination, Communications in Computer and Information Science*, eds. I. Stamelos, R. V. O'Connor, T. Rout, and A. Dorling (Cham: Springer International Publishing), 300–315. doi: 10.1007/978-3-030-00623-5_21

Zapata, S., Barros-Justo, J. L., Matturro, G., and Sepúlveda, S. (2021). Measurement of interpersonal trust in virtual software teams: a systematic literature review. *Ingeniare Rev. Chilena Ingeniería* 29, 788–803. doi: 10.4067/S0718-33052021000400788