IJTE International Journal of Technology in Education
www.ijte.net

Psychometric Properties of Generative AI Literacy Scale for Preservice Teachers: Evidence from a Scoping Review and Factor Analytic Validation

Joseph C. Pasco

Bukidnon State University, Philippines, 0009-0004-8586-9086
Corresponding author: Joseph C. Pasco (josephpasco@buksu.edu.ph)

Article Info

Article History

Received:
12 November 2025

Revised:
13 March 2026

Accepted:
25 April 2026

Published:
18 June 2026

Keywords

Generative AI
AI literacy
Preservice teachers
Scoping review
Psychometric properties

Abstract

Amid the rapid diffusion of generative artificial intelligence in teacher education, there is a pressing need for a validated measure that captures how future educators understand, apply, and govern these systems in instructional contexts. This study developed and validated the Generative Artificial Intelligence Literacy Scale for Preservice Teachers (GAIL-PT), a psychometrically grounded instrument designed to assess preservice teachers' understanding, application, and ethical engagement with generative artificial intelligence (AI) in educational contexts. Guided by established frameworks for instrument development, the research followed six sequential stages: conceptualization through a systematic scoping review, item generation, expert validation using the Content Validity Index (CVI), pilot testing and refinement, exploratory factor analysis (EFA), and confirmatory factor analysis (CFA). The initial instrument consisted of 42 items distributed across six conceptual domains: conceptual understanding, use and application, evaluation and verification, ethics and responsibility, pedagogical integration, and affective readiness. Data were collected from 402 preservice teachers across various education programs at Bukidnon State University. Exploratory factor analysis using maximum likelihood extraction with Promax rotation confirmed sampling adequacy and factorability, resulting in a refined two-factor model explaining nearly half of the total variance. The emergent factors, Applied Engagement and Instructional Integration and Responsible and Reflective Use, encapsulate both the practical and ethical-pedagogical dimensions of AI literacy. Confirmatory factor analysis supported the adequacy of this two-factor structure, demonstrating acceptable reliability and overall model fit across key indices. The results affirm that generative AI literacy among preservice teachers is a multidimensional construct encompassing cognitive, behavioral, and ethical competencies. The validated GAIL-PT offers a theoretically coherent and empirically robust tool for assessing generative AI literacy among preservice teachers.

Citation: Pasco, J. C. (2026). Psychometric properties of generative AI literacy scale for preservice teachers: Evidence from a scoping review and factor analytic validation. *International Journal of Technology in Education (IJTE), 9*(3), 914-939. https://doi.org/10.46328/ijte.6082

Introduction

Artificial intelligence has become a defining force in 21st century education, profoundly transforming how knowledge is accessed, processed, and applied across all levels of learning. The rapid advancement of generative artificial intelligence technologies such as ChatGPT, Gemini, and other large language models has intensified this transformation, bringing both remarkable opportunities and critical challenges for teachers and students. These technologies enable greater creativity, adaptive instruction, and personalized learning experiences, while also supporting assessment and administrative efficiency (Tapalova et al., 2022; Kamalov et al., 2023; Fan et al., 2022; Msambwa et al., 2025; Bauer et al., 2025; Wang et al., 2024; Walter, 2024; Wu & Zhang, 2025). However, their growing presence in education has also raised complex issues regarding authenticity, ethical use, and the need for a deeper understanding of how artificial intelligence operates within learning contexts (Giannakos et al., 2024; Yan et al., 2023). These developments underscore that technological familiarity alone is insufficient. Educators now require a more comprehensive set of competencies encompassing cognitive, ethical, and pedagogical dimensions that extend beyond traditional forms of digital literacy (Sperling et al., 2024; Tenberga & Daniela, 2024; Long & Magerko, 2020; Ng, 2012).

In response to this educational shift, artificial intelligence literacy has emerged as a critical framework that describes the knowledge, skills, and values necessary to understand, evaluate, and use artificial intelligence responsibly (Zhao et al., 2022; Al Abdullatif, 2025; Coşgun, 2025). Recent research highlights that teachers must be able to integrate artificial intelligence tools meaningfully into their practice while maintaining ethical judgment and pedagogical soundness (Sperling et al., 2024; Abdulayeva et al., 2025). Within teacher education, this need is particularly urgent as preservice teachers form their professional identities in an era of rapidly evolving technological ecosystems. Despite growing exposure to artificial intelligence applications, preservice teachers often rely on self-directed exploration rather than systematic training to develop these competencies (Chiu, 2023; Bower et al., 2024; Zhai, 2024). This pattern suggests a widening gap between access to generative technologies and preparedness to use them critically and reflectively.

The literature shows consistent efforts to conceptualize and measure artificial intelligence literacy through validated scales. Early instruments such as the Artificial Intelligence Literacy Scale and the Meta Artificial Intelligence Literacy Scale established foundational dimensions of awareness, knowledge, and ethics (Wang et al., 2022; Carolus et al., 2023). More recent models like the Artificial Intelligence Literacy Questionnaire incorporate affective and behavioral dimensions, highlighting that literacy involves both motivation and moral reasoning (Ng et al., 2024). Nonetheless, most available instruments were designed for general populations and do not fully address the pedagogical and reflective competencies that teacher education contexts require (Lintner, 2024; Ding et al., 2024; Younis, 2025; Alqarni, 2025; Chan & Tang, 2024). This gap limits the capacity of teacher education institutions to accurately evaluate readiness and to design interventions that strengthen responsible classroom application.

The rise of generative artificial intelligence further complicates the construct of literacy. Engaging with generative models involves skills such as prompt formulation, verification of generated outputs, critical evaluation, and the adaptive integration of these tools into learning environments (Liu et al., 2025; Lee & Park, 2024; Bahroun et al., 2023; Lim et al., 2023; Mishra et al., 2023; Monib et al., 2024). These practices require teachers not only to understand the technology but also to manage issues of bias, transparency, and authorship in ways that uphold ethical teaching and learning. The professional identity of educators is thus evolving from knowledge transmitters to facilitators of human artificial intelligence collaboration, where reflection, creativity, and ethical reasoning are as essential as technical skill (Celik, 2022; Yang et al., 2025).

Despite increasing attention, research on artificial intelligence literacy in teacher education remains in an early phase. While a few validated instruments exist, such as those by Liu et al. (2025), Younis (2025), and Park (2025), none have been systematically developed for preservice teachers or tailored to the educational implications of generative artificial intelligence. The absence of psychometrically sound and context specific tools hinders institutions from assessing preservice teachers' preparedness and designing targeted training programs that align with ethical and instructional expectations. This lack of alignment between technological development and teacher capacity building presents a clear need for an empirically validated measure that captures the multifaceted nature of artificial intelligence literacy among future educators.

To address this gap, the present study developed and validated the Generative Artificial Intelligence Literacy Scale for Preservice Teachers. This instrument is grounded in contemporary theoretical and empirical research and conceptualizes literacy as a multidimensional construct encompassing applied engagement, instructional integration, and reflective responsibility. The scale aims to provide a reliable framework for understanding how future educators conceptualize and employ artificial intelligence in teaching and learning, supporting evidence-based improvements in teacher preparation, curriculum design, and policy development in the age of generative technologies.

Review of Existing Artificial Intelligence Literacy Scales

The growing presence of artificial intelligence in education, work, and everyday life has generated significant interest in how individuals understand, evaluate, and ethically apply intelligent systems. As this field matures, scholars have focused on defining and measuring "AI literacy" through psychometrically validated instruments. A critical examination of existing AI literacy scales reveals a progression from early, general self-report tools toward multidimensional frameworks and, more recently, to context-specific instruments addressing generative AI.

Early empirical efforts to quantify AI literacy were led by Wang, Rau, and Yuan (2022), who developed the Artificial Intelligence Literacy Scale (AILS). This instrument defined AI literacy through four core areas: awareness, use, evaluation, and ethics. It demonstrated strong reliability and factorial validity and was instrumental in shaping how researchers conceptualize human engagement with AI. However, its application was primarily confined to adult populations and lacked emphasis on pedagogical integration, limiting its relevance to educational contexts. Similarly, the Scale for the Assessment of Non-Experts' AI Literacy (SNAIL) by Laupichler, Aster, Haverkamp, and Raupach (2023) advanced the field by introducing a concise three-factor structure encompassing technical understanding, critical appraisal, and practical application. The SNAIL demonstrated robust internal reliability but placed less emphasis on ethical and affective components, thereby maintaining a largely cognitive orientation.

A major development came with the Meta-AI Literacy Scale (MAILS) introduced by Carolus, Koch, Straka, Latoschik, and Wienrich (2023). MAILS extended measurement to include meta-competencies such as self-efficacy, emotional regulation, and reflective judgment in AI contexts. Its strong psychometric evidence, confirmed through both exploratory and confirmatory analyses, made it one of the most comprehensive AI literacy instruments to date. A shorter version validated by Koch, Carolus, Wienrich, and Latoschik (2024) preserved the four main constructs: understanding, detection, use, and ethics, while enhancing efficiency for large-sample applications. MAILS bridged cognitive, ethical, and self-regulatory aspects, though its scope continues to raise theoretical questions regarding the distinction between literacy and meta-competence.

In the educational domain, several instruments have been designed to assess AI literacy among students and teachers. Ng, Wu, Leung, Chu, Qiao, and Li (2024) developed the AI Literacy Questionnaire (AILQ) based on an affective, behavioral, cognitive, and ethical model. The instrument demonstrated excellent reliability and fit indices among university students, reinforcing that AI literacy encompasses not only technical but also emotional and ethical dimensions. Kim and Lee (2022) validated an AI Literacy Scale for middle school students in Korea with four factors: knowledge, skill, attitude, and ethics, each aligned with the developmental needs of adolescents. While primarily focused on secondary education, its strong internal reliability ($\alpha > .85$) offers a foundation for adaptation in higher education research. The emergence of generative AI prompted Lee and Park (2024) to create the ChatGPT Literacy Scale, a 25-item measure that assesses technical proficiency, critical judgment, creativity, and ethical understanding in text-based AI interactions. This scale reflects a shift from static AI use to dynamic, interactive engagement in the classroom.

Alongside self-report measures, scholars have pursued objective or performance-based approaches to mitigate the limitations of perception-based assessments. Hornberger, Bewersdorff, and Nerdel (2023) developed the AI Literacy Test (AILIT), a multiple-choice instrument designed to objectively evaluate factual and procedural knowledge of AI among university students. Using item response theory, the AILIT achieved precise measurement of ability levels. A subsequent short form (Hornberger et al., 2025) retained psychometric rigor while improving testing efficiency. Similarly, Zhang, Lee, Ali, DiPaola, and Breazeal (2024) created the AI Literacy Concept Inventory (AI-CI) for middle school learners, which measured understanding of algorithms and logic systems through objective testing. These efforts represent a methodological shift from self-reported competence toward evidence-based performance measurement.

As generative AI tools become ubiquitous in academic and workplace settings, researchers have developed specialized instruments addressing these systems. Liu, Chen, Wang, and Zhang (2025) designed the Generative AI Literacy Scale to assess competence in prompt design, content evaluation, innovative application, and ethical regulation. The scale demonstrated strong reliability and construct validity and underscored that generative AI literacy involves creativity, judgment, and ethical discernment distinct from traditional AI competencies. Grassini

(2024) contributed the Perceived AI Literacy Questionnaire (PAILQ-6), a brief six-item measure suitable for rapid assessment. Despite its brevity, it correlated well with longer instruments and proved useful for large-scale surveys and longitudinal monitoring of AI literacy development.

A comparative synthesis of these instruments reveals shared dimensions across most models, including conceptual understanding, practical use, critical evaluation, and ethical awareness. More recent tools expand this framework to include affective and creative competencies, reflecting the broader evolution from computational literacy to human-AI collaboration. Reported reliability coefficients across scales consistently range between .80 and .93, underscoring their internal consistency. Nevertheless, methodological gaps remain. The majority of instruments rely on self-report data, which may overestimate perceived competence. Only a few studies incorporate objective assessments or mixed-method validation. Moreover, cross-cultural and longitudinal validations remain scarce, and few instruments test for measurement invariance across gender, academic major, or AI experience.

Within teacher education, AI literacy research remains in its early stages. While instruments such as the AILQ and the Generative AI Literacy Scale offer conceptual foundations for classroom application, none have been systematically validated for preservice teachers. This gap highlights the need for a contextually grounded and psychometrically sound instrument, one capable of assessing the cognitive, ethical, pedagogical, and affective dimensions of generative AI literacy essential for the next generation of educators.

Method

Research Design

This study adopted a quantitative, instrumental design aimed at developing and validating the *Generative AI Literacy Scale for Preservice Teachers (GAIL-PT)*. The methodological sequence adhered to the recommended framework for psychometric instrument construction (Boateng et al., 2018; Worthington & Whittaker, 2006), encompassing conceptualization, item generation, expert review, pilot testing, exploratory factor analysis (EFA), and confirmatory factor analysis (CFA). Each phase followed established best practices for validity evidence as articulated by DeVellis (2016) and Hair et al. (2019). The process emphasized construct clarity, empirical rigor, and replicability, ensuring that the final measure captured both theoretical and applied dimensions of AI literacy in teacher education.

Participants and Sampling

The respondents of this study were preservice teachers enrolled in various undergraduate programs under the College of Education of a certain university in the Philippines. They represented a diverse range of disciplines, including Bachelor of Secondary Education (BSEd) majors in English, Mathematics, Science, Filipino, and Social Studies, as well as Bachelor of Elementary Education (BEEd), Bachelor of Early Childhood Education (BECEd), and Bachelor of Physical Education (BPEd). This diversity provided a broad disciplinary representation and enriched the construct validation process by ensuring that responses reflected multiple pedagogical perspectives and levels of AI familiarity (Boateng et al., 2018; Worthington & Whittaker, 2006).

A total of 412 preservice teachers participated across all year levels, with the majority being in their first and third years of study. This year-level spread enabled the inclusion of both novice preservice teachers, who were still developing technological confidence, and advanced participants with higher classroom exposure. The age of respondents ranged primarily from 18 to 21 years, with a majority being female.

In terms of technology engagement, most respondents reported frequent use of generative AI tools, particularly ChatGPT, Gemini, Perplexity, and Grammarly. These tools were commonly employed for lesson planning, drafting learning materials, writing assistance, and generating creative classroom activities, reflecting a growing pattern of informal AI adoption among preservice teachers (Ng et al., 2024; Carolus et al., 2023). However, only a small proportion had participated in formal AI-related training, implying that their current AI literacy was largely acquired through self-directed exploration rather than structured instruction. Collectively, the sample represents an emerging generation of educators who actively engage with generative AI yet remain in the process of developing a balanced and responsible literacy framework. Their diverse backgrounds and varying levels of exposure provide an appropriate context for validating the *Generative AI Literacy Scale for Preservice Teachers (GAIL-PT)*, ensuring the scale's generalizability across subfields of teacher education and supporting its utility for both instructional design and teacher preparation programs.

Instrument Development

Stage 1. Conceptualization and Scoping Review

The conceptualization stage was grounded in a scoping review conducted to establish a comprehensive theoretical and empirical foundation for developing the *Generative AI Literacy Scale for Preservice Teachers (GAIL-PT)*. The review followed the methodological framework proposed by Arksey and O'Malley (2005), further refined by Levac, Colquhoun, and O'Brien (2010), and guided by the Joanna Briggs Institute (JBI, 2020). This design was chosen to systematically map the breadth and depth of research evidence concerning artificial intelligence (AI) literacy and to identify its conceptual domains, measurement approaches, and operational indicators applicable to preservice teacher education. The framework's five iterative stages, identifying the research question, identifying relevant studies, study selection, charting the data, and collating and summarizing the evidence, were followed rigorously to ensure methodological transparency and reproducibility (Peters et al., 2015).

The guiding research question was formulated to capture both the conceptual diversity and psychometric dimensions of AI literacy: *"What are the existing conceptual dimensions and empirical indicators of AI literacy reported in peer-reviewed literature, and how have these been operationalized or validated in educational or professional contexts?"* This question was designed to focus on the measurement and theoretical articulation of AI literacy rather than general digital competence, recognizing that AI engagement entails distinct cognitive, ethical, and creative proficiencies beyond conventional technology use (Long & Magerko, 2020; Ng, 2012). The scoping review thus served not merely to map existing definitions but to identify the underlying constructs and validation practices that could inform a scale tailored for preservice teachers.

A systematic literature search was conducted across major academic databases, including Scopus, Web of Science, Elsevier's ScienceDirect, SpringerLink, Wiley Online Library, ACM Digital Library, and MDPI Journals. Search terms combined both controlled vocabulary and free-text expressions such as "artificial intelligence literacy," "AI literacy scale," "AI competence," "AI framework," and "AI education." The search covered the years 2020 to 2025, capturing the rapid evolution of AI and generative systems in education while including key pre-2020 foundational works for conceptual grounding (e.g., Ng, 2012; Long & Magerko, 2020). The inclusion criteria required that studies

(a) explicitly define or operationalize AI literacy,

(b) present empirical evidence or validated frameworks, and

(c) be published in reputable, peer-reviewed outlets.

Exclusion criteria omitted grey literature, editorials, and purely theoretical commentaries to maintain methodological rigor and focus on validated or empirically supported work.

The screening process followed the JBI-recommended two-phase protocol: an initial title and abstract screening for relevance, followed by full-text evaluation of eligible papers. Studies were included if they addressed AI literacy within educational, professional, or societal contexts and if they provided conceptual or measurement insights relevant to teaching and learning. Discrepancies during screening were resolved through consensus among reviewers to ensure consistency and reduce bias. Data from the included studies were then charted using an iterative extraction matrix that captured key attributes such as author, publication year, population, context, operational definition, dimensions of AI literacy, type of measurement instrument, and psychometric evidence where available (Arksey & O'Malley, 2005; Levac et al., 2010). This stage laid the methodological foundation for synthesizing conceptual and empirical patterns that would later inform item generation and domain specification for the GAIL-PT scale.

Stage 2. Item Generation

The second stage focused on the systematic development of potential scale items that would represent the construct of generative AI literacy in the context of preservice teacher education. Item generation followed established guidelines for psychometric instrument construction, emphasizing conceptual clarity, simplicity, and contextual appropriateness (DeVellis, 2016; Boateng et al., 2018). The process was both deductive and inductive, drawing from theoretical insights and empirical findings obtained during the scoping phase, as well as from related educational technology and literacy literature.

Stage 3. Expert Validation (Content Validity Index)

The content validity of the draft instrument was established through expert review involving specialists in educational technology, language and communication, and teacher education. These experts were selected for their professional experience in education and their ability to evaluate the linguistic clarity, contextual appropriateness, and pedagogical relevance of the items. Each expert independently assessed every statement using a 4-point rating scale to evaluate clarity, relevance, and representativeness of content, following established procedures for content validation (Lynn, 1986).

Stage 4. Pilot Testing and Refinement

The refined instrument was piloted among a small group of preservice teachers ($n = 10$) to assess clarity, comprehension, and response variability. Participants provided qualitative feedback on item wording and relevance to their academic experience. Quantitative inspection confirmed adequate variance and absence of ceiling or floor effects. Minor revisions improved precision and alignment with classroom realities. This stage confirmed semantic clarity and preliminary reliability, setting the stage for large-scale data collection.

Stage 5. Exploratory Factor Analysis (EFA)

The Exploratory Factor Analysis (EFA) served as the primary statistical technique for identifying the underlying structure of relationships among the observed variables. This stage aimed to determine how the items group together to form latent constructs representing various dimensions of the measured concept. The analysis followed established psychometric procedures for early-stage scale validation (Fabrigar & Wegener, 2012; Hair et al., 2019).

The EFA was performed using Maximum Likelihood Extraction to enable statistical estimation and inferential testing of the factor model. A Promax oblique rotation was applied to allow for correlation among potential factors, as constructs in behavioral and educational research are often theoretically and empirically interrelated. Before conducting the analysis, assumptions for factor analysis were verified, including the normality of item distributions, sampling adequacy through the Kaiser–Meyer–Olkin (KMO) measure (threshold > 0.80), and factorability of the correlation matrix using Bartlett's Test of Sphericity ($p < 0.001$).

Item retention was guided by predefined statistical and conceptual criteria. Items were retained if they had communalities ≥ 0.300, factor loadings ≥ 0.400, and cross-loading differences greater than 0.20, ensuring that each retained item contributed meaningfully to its underlying factor without ambiguity (Worthington & Whittaker, 2006; Boateng et al., 2018). Items failing to meet these thresholds were systematically reviewed and removed. The factor solution was evaluated based on eigenvalues, percentage of explained variance, and theoretical interpretability of the emerging structure.

Stage 6. Confirmatory Factor Analysis (CFA)

The Confirmatory Factor Analysis (CFA) was conducted to statistically validate the factor structure derived from the exploratory phase and to assess the overall model fit, reliability, and construct validity of the instrument. CFA was performed using the Maximum Likelihood (ML) estimation method, a preferred approach in psychometric validation for continuous data under multivariate normality assumptions (Kline, 2021; Brown, 2015). Analyses were executed in AMOS to examine whether the observed data fit the hypothesized measurement model established during EFA.

Model fit was evaluated through multiple indices that jointly assess absolute fit, incremental fit, and parsimonious fit following widely recognized standards (Hu & Bentler, 1999; Schermelleh-Engel et al., 2003). These included the ratio of chi-square to degrees of freedom (χ^2/df), Comparative Fit Index (CFI), Tucker–Lewis Index (TLI), Standardized Root Mean Square Residual (SRMR), and Root Mean Square Error of Approximation (RMSEA). Acceptable thresholds were set as χ^2/df < 3.00, CFI and TLI > 0.95, SRMR < 0.08, and RMSEA < 0.06.

Validity testing included assessments of convergent and discriminant validity to ensure that each construct was both theoretically distinct and empirically consistent. Convergent validity was established through Average Variance Extracted (AVE ≥ 0.50) and Composite Reliability (CR ≥ 0.70), while discriminant validity was confirmed when the square root of AVE exceeded the corresponding inter-factor correlations (Fornell & Larcker, 1981). Reliability was further assessed through Cronbach's alpha and coefficient omega, providing complementary evidence of internal consistency (Hair et al., 2019).

Multiple model comparisons were conducted to determine the most parsimonious and theoretically sound structure. Competing configurations were tested to evaluate whether a correlated-factor model, a second-order model, or a bifactor structure best represented the data. The final model was retained based on both empirical fit indices and conceptual interpretability, providing robust evidence for the construct validity and reliability of the instrument. This stage established the statistical confirmation of the scale's structural integrity, forming the basis for its use in assessing generative AI literacy among preservice teachers.

Data Collection Procedure

Data collection was carried out through a web-based questionnaire administered via Google Forms, ensuring accessibility and efficiency in reaching preservice teachers across multiple education programs. The survey instrument consisted of three sections: (a) an informed consent and study overview that outlined the purpose, confidentiality assurances, and voluntary nature of participation; (b) a demographic profile section capturing age, sex, academic program, and AI usage frequency; and (c) the 42-item Generative AI Literacy Scale for Preservice Teachers (GAIL-PT). The average completion time was approximately 15 minutes, designed to minimize respondent fatigue while ensuring adequate engagement with each item. Prior to data collection, formal permission was secured from the College of Education and other relevant university authorities. The purpose and procedures of the study were explained to the participants through a digital information sheet embedded within the form. Participation was entirely voluntary, and respondents were required to give informed consent before proceeding with the survey. This ethical step ensured that all participants were fully aware of their rights and the intended academic use of the data.

A total of 412 preservice teachers responded to the survey. Upon data screening, 10 cases were excluded due to incomplete responses, straight-lining patterns, or evident inconsistencies, resulting in a final sample of 402 valid responses subjected to statistical analysis. The dataset was exported directly from Google Forms into SPSS version 23 for preliminary screening and descriptive statistics, and subsequently analyzed using AMOS version 23 for structural modeling. Screening procedures followed established recommendations, including the assessment of missing values, univariate and multivariate outliers, and checks for response bias. The cleaned dataset was stored in a password-protected, encrypted file accessible only to the research team, ensuring data security and compliance with ethical standards in educational research.

Results and Discussion

To orient readers to the logic of presentation, the researcher organized this to mirror the sequential stages of the methods. The section opens with the conceptual groundwork derived from the scoping review. Here, the researcher clarifies how the literature mapping yielded six core domains that define generative AI literacy for preservice teachers and explains how these domains informed the initial pool of content. This conceptual frame is followed by a concise description of how items were written to represent each domain through clear, single idea statements appropriate for undergraduate respondents.

The narrative then transitions from content evidence to statistical evidence. The researcher first reports outcomes of expert review and pilot feedback that guided editorial refinement of wording and coverage. This is followed by the exploratory factor analysis, conducted with maximum likelihood extraction and oblique rotation, and evaluated against stated decision rules for communalities, primary loadings, and cross-load differences. Results are interpreted in relation to the six conceptual domains, with an explanation of item retention and trimming decisions. Reliability indices are then reported at the factor level to establish internal consistency.

Finally, the confirmatory factor analysis is presented to test the best fitting structure that emerged from the exploratory phase. Model fit, internal consistency, and evidence for convergent and discriminant validity are summarized using established criteria. Because the confirmatory results show that the six conceptual domains consolidate into two broader latent factors, the researcher concludes by mapping each of the original domains to these factors and clarifying the implications for interpretation and classroom application. Read in this order, the section moves from theory to items, from items to factors, and from factors to practical meaning for teacher education.

Six Core Domains of AI Literacy

The synthesis of twenty-five peer-reviewed studies reveals six recurrent and interrelated domains of artificial intelligence literacy that together define a comprehensive conceptual framework. These domains are *conceptual understanding, use and application, evaluation and verification, ethics and responsibility, pedagogical integration,* and *affective readiness*. Each domain is supported by empirical or theoretical evidence drawn from multiple studies across educational and professional contexts, demonstrating that artificial intelligence literacy is not a single construct but a multidimensional competence integrating cognitive, ethical, and affective dimensions.

Conceptual Understanding

The first domain constitutes the foundation of artificial intelligence literacy. It involves knowledge of what artificial intelligence is, how data-driven models operate, and what their inherent limitations are. Moreover, studies consistently highlight the need for users to grasp how algorithms generate outputs, how machine learning differs from traditional computing, and how biases can emerge from data and model design (Long & Magerko, 2020; Ng et al., 2024; Koch et al., 2024). In addition, conceptual literacy encompasses recognizing the probabilistic nature of generative models and understanding key concepts such as training data, prompts, and model confidence (Lintner et al., 2024; Chiu, 2024). A strong conceptual base thus enables users to approach artificial intelligence critically and prevents misinterpretation or overreliance on algorithmic outputs.

Use and Application

The second domain extend beyond basic operation to the purposeful and contextually appropriate deployment of artificial intelligence tools. This domain captures the ability to plan prompts, interpret generated outputs, and apply them effectively in professional or academic contexts such as lesson planning, content creation, and research support (Carolus et al., 2023; Koch et al., 2024). Similarly, Kong et al. (2023) and Du et al. (2024) demonstrate that training interventions can significantly enhance these application skills, showing that they are teachable and transferable. In the same vein, for preservice teachers, competent use and application involve integrating artificial intelligence into pedagogical design and feedback systems (Ning et al., 2024; Hava et al., 2025). This domain therefore bridges conceptual knowledge with real-world utility, making it central to the practical expression of artificial intelligence literacy.

Evaluation and Verification

The third domain represent the critical-thinking dimension of artificial intelligence literacy. Users must be able to question, verify, and validate the accuracy, appropriateness, and reliability of artificial intelligence outputs. Moreover, the reviewed studies emphasize essential competencies such as bias detection, source triangulation, and logical reasoning when reviewing generated content (Laupichler et al., 2023a; Lintner et al., 2024; Younis, 2025). While objective assessments of these skills remain limited, notably, the literature underscores that the ability to evaluate and verify distinguishes literate users from merely dependent ones (Chiu, 2024). Besides, within teacher education, this evaluative stance aligns closely with academic integrity and the verification of instructional materials. Cultivating this habit of critical scrutiny is thus fundamental to responsible and informed engagement with generative systems.

Ethics and Responsibility

The fourth domain emerge as a moral and professional pillar of artificial intelligence literacy. This domain involves understanding issues related to fairness, transparency, privacy, intellectual property, and accountability (Allen & Kendeou, 2023; Tenberga & Daniela, 2024). Additionally, both teachers and students must learn when and how to disclose their use of artificial intelligence while internalizing principles of responsible and ethical innovation (Ng et al., 2024; Al Abdullatif, 2024). Correspondingly, studies in education reveal that ethical awareness moderates the relationship between knowledge and behavioral intention, suggesting that literacy without moral grounding may lead to misuse, overreliance, or even academic dishonesty (Du et al., 2024; Younis, 2025). A robust ethical framework therefore safeguards the educational benefits of artificial intelligence while preserving professional integrity.

Pedagogical Integration

The fifth domain features most prominently in teacher education research and serves as the conduit through which artificial intelligence literacy becomes transformative for learning. It refers to the capacity to align artificial intelligence use with curriculum goals, instructional strategies, assessment practices, and diverse learner needs (Ning et al., 2024; Hava et al., 2025; Al Abdullatif, 2024). Moreover, the AI TPACK and intelligent TPACK frameworks extend technological pedagogical content knowledge into artificial intelligence contexts, providing a strong theoretical base for this domain. Equally, Sperling et al. (2024) and Allen & Kendeou (2023) emphasize that effective pedagogical integration requires a synthesis of technical expertise and ethical sensitivity to ensure that innovation aligns with policy, curriculum, and student welfare. Hence, pedagogical integration anchors artificial intelligence literacy within educational praxis, transforming it from theoretical awareness into instructional competence.

Affective Readiness

Finally, the sixth domain encompasses the emotional and motivational dimensions of artificial intelligence engagement, including attitudes, curiosity, and self-efficacy. Several studies show that factors such as anxiety, perceived usefulness, and confidence significantly predict willingness to adopt artificial intelligence tools (Du et al., 2024; Younis, 2025; Chiu et al., 2025). In addition, confidence in interacting with artificial intelligence systems correlates positively with both learning outcomes and openness to advanced applications. Conversely, lack of affective readiness may inhibit even well-informed individuals from applying their knowledge in innovative ways. For preservice teachers, therefore, fostering positive affective dispositions becomes a precondition for transforming conceptual understanding and ethical awareness into active, reflective practice.

Collectively, these six domains demonstrate that artificial intelligence literacy is an integrated construct encompassing knowledge, skill, ethics, pedagogy, and emotion. Conceptual understanding and technical proficiency provide structure, yet ethical reasoning, pedagogical intentionality, and affective engagement determine its depth, authenticity, and sustainability. Furthermore, emerging scholarship identifies two auxiliary domains, creative co-production and human–artificial intelligence collaboration, that may enrich future frameworks, particularly as generative systems become increasingly embedded in creative and reflective educational practice (Kong et al., 2024; Yim, 2024; Jia et al., 2025). Overall, the synthesis affirms that genuine artificial intelligence literacy is not merely the ability to use tools but the capacity to think, teach, and act responsibly in an algorithmic world.

Content Development and Pre-Analytic Evidence

Following the scoping review, the researcher translated the conceptual map into a working item pool that could represent the breadth of generative AI literacy in preservice teacher education. Item writing adhered to established guidance for scale construction with emphasis on conceptual clarity, brevity, and contextual appropriateness for undergraduates (DeVellis, 2016; Boateng et al., 2018). A total of forty-two statements were drafted, with seven items aligned to each conceptual domain to preserve balanced representation of cognitive, behavioral, and affective facets. Each statement expressed a single idea and targeted observable beliefs, decisions, or classroom actions related to artificial intelligence. Technical terms were minimized and first-person wording was used to support accurate self-appraisal. Responses were captured on a five-point Likert scale from strongly disagree to strongly agree, consistent with recommendations for clear anchors and unidimensional response processes (Worthington & Whittaker, 2006). Redundant, ambiguous, and double-barreled phrasing was avoided to limit construct irrelevant variance. This content work produced a coherent instrument blueprint suitable for expert review.

The study then shifted from content evidence to structured appraisal by qualified reviewers. Content validity procedures drew on specialists in educational technology, language and communication, and teacher education who evaluated each item independently for relevance, clarity, and representativeness using a four-point scale. Item level and scale level content validity index statistics were computed to quantify agreement, following accepted thresholds for retention and revision (Lynn, 1986). Items with an item level content validity index below point eighty were revised or removed, while items at or above point ninety were retained. Reviewer commentary converged on three areas for refinement. First, language economy and parallel structure were strengthened to improve readability without diluting intent. Second, contextual cues were adjusted so that examples reflected typical coursework, lesson planning, and practicum tasks across elementary and secondary programs. Third, inclusivity and accessibility were emphasized to ensure that items did not privilege a single discipline or prior exposure to advanced tools. These adjustments improved linguistic precision, content alignment with preservice practice, and face validity at the scale level.

A brief pilot administration with ten preservice teachers provided additional evidence on clarity and response behavior. Participants reported that instructions and item wording were clear and that statements reflected realistic study and classroom activities. Item distributions showed adequate variance and no ceiling or floor effects, and time to completion was judged acceptable for course-based data collection. Minor edits addressed local idiom, removed residual redundancies, and sharpened verbs tied to evaluation and responsible use. Collectively, the expert appraisal and pilot feedback confirmed that the forty-two-item instrument was readable, contextually grounded, and ready for quantitative evaluation. These steps complete the pre-analytic phase and justify the transition to statistical testing of dimensionality and reliability through exploratory and confirmatory factor analyses.

Exploratory Factor Analysis

Pre-analysis diagnostics indicated that the data met the core assumptions for common factor modeling with maximum likelihood extraction and an oblique rotation. Sampling adequacy and interitem correlations were sufficient to justify factor analysis, Bartlett's test supported the presence of common variance, and inspection of the anti-image and residual matrices did not reveal offending items or ill conditioned relationships. Given the expectation from prior work that facets of artificial intelligence literacy are correlated rather than orthogonal, an oblique Promax rotation was used to obtain a conceptually faithful solution (Carolus et al., 2023; Ng et al., 2024).

Item retention followed the prespecified rules. Items with communalities below .300 were removed, primary loadings had to be .400 or greater, and items were excluded if a competing loading differed by less than .20. Applying these criteria yielded a two-factor solution with eigenvalues of 7.463 and 2.477, which together accounted for 49.702 percent of the total variance. Factor 1 explained 37.317 percent and Factor 2 explained 12.385 percent of the variance. All retained indicators met both loading and communality thresholds, and no cross-loading violations remained under the stated rule.

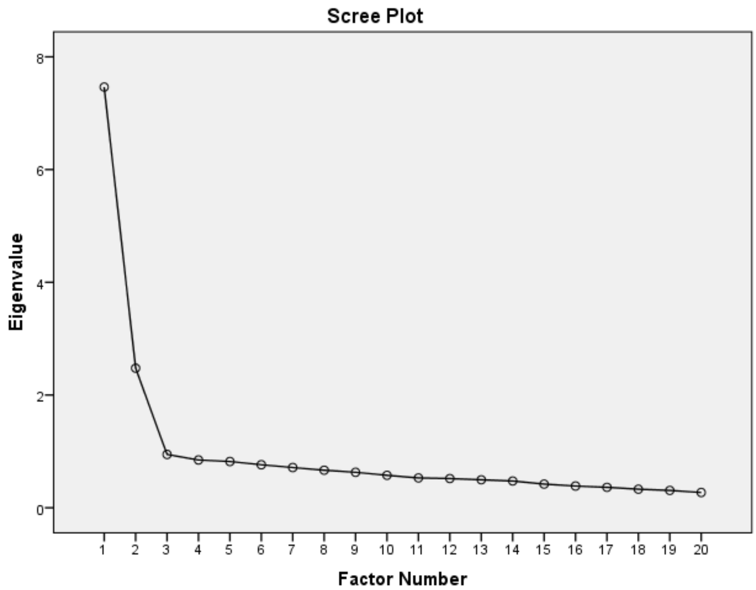

Figure 1. The Scree Plot Diagram of 42-Item Generative AI Literacy for Preservice Teachers

Factor 1 comprised twelve items with loadings from .510 to .770 and communalities ranging from .341 to .593. Item content converged on applied classroom use and exploratory engagement. Respondents reported saving time with artificial intelligence tools, improving lesson plans, designing creative activities, including artificial intelligence topics in class discussions, adapting and editing generated content, writing clearer prompts, and locating examples or materials. These behaviors appeared together with motivational and expectancy statements, including openness to improving skills, interest in learning more, exploring new tools, and believing that artificial intelligence can make teaching more engaging.

Factor 2 consisted of eight items with loadings from .480 to .844 and communalities from .355 to .564. Items expressed a reflective and responsible stance toward model fallibility and the social context of use. Respondents indicated that they check whether artificial intelligence answers are correct, think carefully before accepting outputs, recognize that artificial intelligence can be wrong, double check explanations, handle data responsibly, and consider effects on people and society. One pedagogical item also loaded on this factor, namely showing students how to use artificial intelligence wisely.

The resulting structure therefore reflects two broad and correlated dimensions. The first captures enactment and engagement with artificial intelligence in preservice teaching tasks. The second captures verification and responsibility during evaluation and classroom stewardship.

Table 1. Maximum Likelihood Estimates of the Oblique (Promax) Rotated Factor Loadings for the 42-Item Generative AI Literacy Scale for Preservice Teachers (GAIL-PT)

Item	Factor Loadings		Communalities
	Factor 1	Factor 2	
Factor 1: Instructional Application and Engagement			
Item 5) I can use AI tools to save time on school tasks.	.770		.460
Item 41) I believe AI can make teaching more engaging.	.685		.436
Item 28) I am open to improving my AI skills.	.668		.593
Item 11) I can use AI to make activities more creative.	.658		.378
Item 24) I can use AI to make my lesson plans better.	.654		.471
Item 12) I can include AI topics in my class discussions.	.653		.341
Item 19) I explore new AI tools that can help me as a teacher.	.640		.437
Item 18) I can edit AI-generated content to fit my needs.	.603		.358
Item 31) I can write clear prompts to get better answers from AI.	.568		.417
Item 13) I am interested in learning more about AI.	.566		.461
Item 32) I can use AI to find examples or materials for class.	.565		.499
Item 17) I can use AI tools to help me study or prepare lessons.	.510		.457
Factor 2: Reflective Verification and Responsibility			
Item 33) I always check if AI answers are correct.		.844	.564
Item 34) I think carefully before accepting AI outputs.		.804	.512
Item 37) I think about the effects of AI on people and society.		.713	.543
Item 36) I know that AI data should be used responsibly.		.697	.533
Item 29) I know that AI answers may not always be correct.		.655	.434
Item 21) I double-check AI explanations before using them.		.646	.425
Item 23) I use AI only for learning, not to deceive others.		.535	.355
Item 38) I can show students how to use AI wisely.		.480	.380
Eigenvalues	7.463	2.477	
% of variance	37.317	12.385	
Cumulative %	37.317	49.702	

The emergence of a two-factor solution represents a conceptually coherent consolidation of the initial blueprint. The original instrument separated affective readiness, use and application, evaluation and verification, ethics and responsibility, and pedagogical integration. In the present sample, preservice teachers did not treat these as five distinct strands. Rather, attitudes and use blended into a single applied engagement factor, while evaluation and ethics merged with classroom stewardship into a reflective responsibility factor. This pattern is consistent with findings from recent validation programs where affective and behavioral elements cluster, and where ethical stance is expressed most meaningfully through verification practices during authentic tasks (Carolus et al., 2023; Koch et al., 2024; Ng et al., 2024; Laupichler et al., 2023; Lintner et al., 2024).

The first factor combines motivational energy with practical classroom enactment. Items indicating enjoyment, interest, and openness sit alongside concrete teacher behaviors such as prompt design, lesson planning, creative activity design, content adaptation, and resource finding. The MAILS program reported a similar fusion of meta competencies with applied use, suggesting that for developing users, readiness and action reinforce each other rather than form separate strata (Carolus et al., 2023; Koch et al., 2024). The ABCE based AI Literacy Questionnaire likewise observed that affective and behavioral components tend to cohere in student samples, especially when items are tied to course contexts and routine study tasks (Ng et al., 2024). In light of this content and the literature, Factor 1 is best labeled Instructional Application and Engagement. The label signals that the latent trait captures both willingness to engage and the ability to apply artificial intelligence within teaching and learning work.

The second factor integrates verification behaviors, awareness of model fallibility, responsible data practice, consideration of social impact, and a teacher facing stewardship role. This supports the argument from recent reviews that ethical literacy is most predictive of responsible behavior when it is operationalized through reflective checking, cautious acceptance, and policy aware decision making, rather than measured as detached attitude alone (Lintner et al., 2024). Teacher competence models such as AI TPACK and intelligent TPACK also frame integration as stewardship and alignment that includes modeling responsible practice for students and vetting outputs against curricular aims (Ning et al., 2024; Al Abdullatif, 2024). Given this content, Factor 2 is best labeled Reflective Verification and Responsibility. The label captures the union of critical checking with ethical and pedagogical obligations in classroom practice.

Several items migrated relative to their theorized homes, and these movements are theoretically sensible. The item on showing students how to use artificial intelligence wisely loaded with reflective responsibility rather than with general pedagogical integration. This indicates that preservice teachers conceptualize guiding student use primarily as a responsible practice that is intertwined with accuracy checking and data care, not as a neutral technological integration decision. Similarly, expectancy beliefs about engaging teaching loaded with application and exploration, which aligns with the early adoption profile documented in student and novice teacher samples where enthusiasm and practice are coupled (Carolus et al., 2023; Ng et al., 2024).

From a measurement standpoint, primary loadings exceeded the .400 criterion and communalities for retained items were adequate for maximum likelihood extraction, indicating a well-defined latent structure. The first factor includes a robust set of indicators that range across motivation, skill, and classroom enactment, which supports broad coverage and reliability in subsequent validation. The second factor also demonstrates solid loadings across verification and ethical stewardship behaviors. Together these factors reproduce a pattern frequently noted in the artificial intelligence literacy literature, namely a practical engagement dimension and a reflective responsibility dimension, each of which is essential for responsible use in preservice teacher education (Laupichler et al., 2023; Lintner et al., 2024; Carolus et al., 2023; Ng et al., 2024).

In sum, the exploratory analysis yields a statistically adequate and theoretically persuasive two factor model. Renaming the domains as Instructional Application and Engagement and Reflective Verification and Responsibility provides labels that map precisely onto the observed item content while remaining aligned with current evidence on how learners and novice teachers conceptualize artificial intelligence literacy. This configuration offers a coherent foundation for subsequent validation work and for applied use in diagnosing strengths and needs among preservice teachers.

Confirmatory Factor Analysis

All prerequisite checks indicated that the data met the assumptions for common-factor modeling using maximum likelihood estimation with an oblique rotation. Sampling adequacy and the correlation matrix supported factorability; Bartlett's test indicated sufficient common variance; inspection of anti-image and residual matrices did not reveal ill-conditioned items. Because theory and prior validation work show that facets of artificial intelligence literacy are correlated rather than orthogonal, an oblique (Promax) rotation was retained. Item retention followed the a priori decision rules: communalities below .300 triggered removal; primary loadings had to be at least .400; and cross-loading items were removed when the difference between primary and secondary loadings was not larger than .20. The twenty-item exploratory solution was then subjected to CFA as Model 1, after which items with standardized regression weights below .60 were deleted to yield the final fifteen-item Model 2, in line with common recommendations that prefer loadings near or above .60 to ensure practical significance (e.g., Hair et al., 2019; Kline, 2021).

Table 2. Model Fit Indices from Measurement Model of the Generative AI Literacy Scale for Preservice Teachers (GAIL-PT)

Model	Cmin/df	TLI	CFI	SRMR	RMSEA
1	2.898*	0.891***	.903**	.061*	.069**
2	3.227**	0.907**	.921**	.058*	.060**
Criterion	<3	> .90	> .90	< .08	< .08

Note: Differences between models were considered negligible when ΔCFI ≤ .010 and ΔRMSEA ≤ .015.

*Good fit thresholds: CMIN/df ≤ 3; CFI/TLI ≥ .95; SRMR ≤ .05; RMSEA ≤ .06.

**Acceptable fit thresholds: 3 < CMIN/df ≤ 5; .90–.94 for CFI/TLI; .051–.080 for SRMR; .061–.080 for RMSEA.

***Poor-fit thresholds: CMIN/df > 5; CFI/TLI < .90; SRMR > .080; RMSEA > .080.

Model-fit results show a clear pattern of improvement from Model 1 to Model 2 on the core indices typically reported in scale validation. In Model 1, CMIN/df was 2.898, TLI was 0.891, CFI was .903, SRMR was .061, and RMSEA was .069. In Model 2, TLI increased to 0.907 and CFI to .921; SRMR improved to .058 and RMSEA to .060; CMIN/df rose to 3.227. Interpreted against widely used guidelines, the increases in CFI and TLI move the model beyond the .90 threshold for acceptable incremental fit, while SRMR and RMSEA remain below .08, consistent with acceptable error of approximation; the slight increase in CMIN/df reflects the well-known sensitivity of the chi-square family to model respecification and parsimony trade-offs (Hu & Bentler, 1999; Kline, 2021). Overall, the final model exhibits a more stable and better-fitting structure on the most informative indices.

Construct reliability and convergence are also adequate. For Model 1, composite reliability was .879 and average variance extracted was .447; for Model 2, composite reliability was .857 and average variance extracted improved to .501. The increase of average variance extracted above .50 in the final model indicates that, on average, more variance is captured by the latent factors than is due to measurement error, satisfying the classic convergence criterion even as composite reliability remains comfortably above .80 (Fornell & Larcker, 1981). The report notes that average variance extracted can be conservative and that reliability can be established through composite reliability when average variance extracted is slightly below .50, which is consistent with applied guidance (Malhotra & Dash, 2011). MaxR(H) values were .883 for Model 1 and .861 for Model 2, both consistent with good construct reliability. The latent correlation estimates between the two factors decreased from .708 in Model 1 to .582 in Model 2, suggesting improved discriminant standing between the dimensions in the final specification. All standardized loadings retained in the fifteen-item solution were strong and statistically significant, ranging from approximately .618 to .773 for Instructional Application and Engagement and from approximately .659 to .768 for Reflective Verification and Responsibility, with large critical ratios and p values below .001, which supports the salience of each indicator for its target construct.

Table 3. Unstandardized and Standardized Parameter Estimates for 15-Item Generative AI Literacy Scale for Preservice Teachers (GAIL-PT)

Factor	**Item**	**Unstandardized Regression Weights**	**Standardized Regression Weights**	**Standard Error**	**Critical Ratio (t)**	**p-value**
Instructional Application and Engagement	Item 5	1.174	0.618	0.109	10.748	***
	Item 41	1.041	0.644	0.097	10.777	***
	Item 28	1.244	0.773	0.102	12.209	***
	Item 24	1.059	0.688	0.094	11.310	***
	Item 19	0.977	0.647	0.091	10.719	***
	Item 31	1.012	0.638	0.095	10.677	***
	Item 13	0.998	0.653	0.093	10.767	***
	Item 32	0.949	0.705	0.083	11.459	***
	Item 17	0.852	0.640	0.079	10.748	***
Reflective Verification and Responsibility	Item 21	0.841	0.659	0.065	13.034	***
	Item 29	0.862	0.659	0.068	12.606	***
	Item 36	0.894	0.706	0.068	13.162	***
	Item 37	0.933	0.715	0.070	13.398	***
	Item 34	0.929	0.735	0.062	14.870	***
	Item 33	1.189	0.768	0.091	13.034	***

Note: *All t values are significant at $p<.001$

The confirmatory results show that pruning items with weaker standardized weights produced a more coherent and better-fitting measurement model while preserving the theoretical grain of the two-factor solution. Improvements in comparative fit and non-centrality indices from Model 1 to Model 2, together with SRMR and RMSEA values below conventional .08 thresholds, indicate that the final model reproduces the observed covariance structure with acceptable residual error and acceptable approximation, consistent with current reporting norms in scale validation (Hu & Bentler, 1999; Kline, 2021). The modest increase in CMIN/df in Model 2 is not inconsistent with this conclusion, given the sensitivity of chi-square to degrees of freedom and sample size. Nevertheless, contemporary practice emphasizes the joint interpretation of CFI or TLI with RMSEA and SRMR rather than reliance on a single index (Hu & Bentler, 1999).

Convergent and discriminant evidence are aligned with expectations for a parsimonious two-factor model. The increase of average variance extracted to .501 in the final model indicates adequate convergence at the construct level, and composite reliability remains strong for both factors, satisfying reliability criteria even under conservative standards (Fornell & Larcker, 1981; Malhotra & Dash, 2011). The reduction in the latent correlation from .708 to .582 suggests that the final respecification sharpened the conceptual distinction between the factors and reduced redundancy as reflected in Figure 2. This pattern is theoretically consistent with the emerging literature in artificial intelligence literacy, where preservice teachers tend to organize their competencies into two broad and related strands: a practical strand that combines motivation and classroom enactment, and a reflective strand that couples verification practices with responsible use (Carolus et al., 2023; Ng et al., 2024; Laupichler et al., 2023; Lintner et al., 2024). Figure 2 showing Standardized Regression Weights.

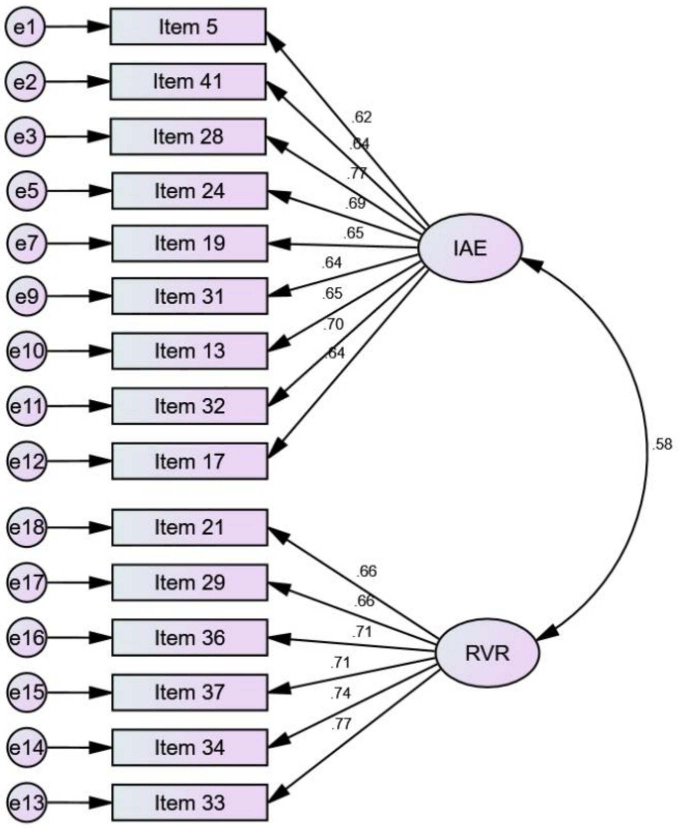

Figure 2. Path Diagram of the Two Factors of Generative AI Literacy for Preservice Teachers

Note: IAE=Instructional Application and Engagement; RVR=Reflective Verification and Responsibility

At the item level, all retained indicators displayed substantial standardized loadings and large, significant critical ratios, lending support to the content validity and statistical salience of the observed indicators. Aside, factor 1 and factor 2 obtained Cronbach alpha of .878 and .856, respectively, indicating strong internal consistency. The *Instructional Application and Engagement* factor is represented by items that capture prompt design, time saving, lesson planning, creative activity design, content adaptation, resource finding, and a willingness to learn and explore. This configuration mirrors prior evidence that affective readiness and classroom application tend to cohere in early adoption contexts, as seen in the MAILS and AILQ programs (Carolus et al., 2023; Ng et al., 2024). The *Reflective Verification and Responsibility* factor is defined by accuracy checking, careful acceptance, recognition of model fallibility, responsible data practices, attention to social impact, and modeling wise use for students, which reflects recommendations to embed ethics within verification and classroom decision making rather than treat ethics as detached belief (Lintner et al., 2024; Laupichler et al., 2023).

In sum, the confirmatory analysis supports a two-factor measurement model with acceptable fit, solid reliability, and evidence of convergent and discriminant validity for the Generative AI Literacy Scale for Preservice Teachers. The final fifteen-item model strengthens model parsimony while capturing the two core dimensions that are most actionable for teacher education: *Instructional Application and Engagement*, and *Reflective Verification and Responsibility*. This structure is well aligned with contemporary theory and with patterns reported in recent validations of artificial intelligence literacy instruments.

Nevertheless, several methodological limitations should be noted. The study involved preservice teachers from a single university, which may restrict the generalizability of findings. Furthermore, although rigorous statistical criteria guided item retention and model fit evaluation, future applications should test the stability of the factor structure across diverse cultural, institutional, and academic contexts. Despite these constraints, the GAIL-PT advances empirical understanding of AI literacy as a multidimensional construct that integrates cognitive, behavioral, and ethical competencies essential for the next generation of educators.

Conclusion

The development and validation of the *Generative Artificial Intelligence Literacy Scale for Preservice Teachers (GAIL-PT)* mark a substantial advancement in the empirical assessment of AI literacy within teacher education. The study established a theoretically anchored and statistically verified framework that captures the complex and evolving nature of generative AI literacy. Through six methodical stages: conceptualization, item generation, expert validation, pilot testing, exploratory factor analysis, and confirmatory factor analysis, the process ensured theoretical coherence, methodological precision, and psychometric integrity.

The initial conceptual structure consisted of six interrelated domains: *conceptual understanding, use and application, evaluation and verification, ethics and responsibility, pedagogical integration,* and *affective readiness*. Collectively, these domains encompass the cognitive, behavioral, and attitudinal competencies required for responsible and pedagogically sound engagement with AI in educational contexts. Factor analyses further refined these domains into two empirically validated factors: *Applied Engagement and Instructional Integration* and *Responsible and Reflective Use*. The two-factor model demonstrated acceptable reliability and model fit, affirming the internal consistency and structural validity of the instrument. Overall, the findings emphasize that generative AI literacy among preservice teachers extends beyond functional expertise to include critical reflection, ethical discernment, and adaptive instructional use. The *GAIL-PT* provides a rigorous diagnostic framework that can inform teacher preparation programs, guide institutional strategies for AI integration, and support the cultivation of educators who are competent, reflective, and ethically grounded in their use of emerging AI technologies.

Recommendations

Further research should continue validating the GAIL-PT across diverse institutional contexts to confirm its consistency and applicability. Employing complementary qualitative methods, such as reflective responses or focus group discussions, may deepen understanding of how preservice teachers interpret and apply AI-related knowledge in practice. Longitudinal investigations are also encouraged to explore changes in AI literacy as preservice teachers gain instructional experience and greater exposure to AI-based tools. At the curricular level, teacher education institutions may use the GAIL-PT to evaluate readiness and design targeted interventions that strengthen ethical, critical, and pedagogical aspects of AI use. Integrating structured AI literacy programs into preservice teacher preparation can foster educators who are not only competent in using AI tools but also discerning, responsible, and innovative in applying them for meaningful teaching and learning.

Statements and Declarations

Acknowledgments/Notes: Not applicable.

During the preparation of this article, the authors did not use ChatGPT.

Supplementary Materials: Not applicable.

Author Contributions: The author has read and agreed to the published version of the manuscript.

Funding: The author received no funding for the research.

Data Availability: Not applicable.

Ethics Approval: The study was performed in accordance with the study protocol and ethical guidelines and regulations.

Informed Consent: Informed consent was obtained from all subjects involved in the study.

Conflicts of Interest: The author declares no conflicts of interest.

References

Abdulayeva, A., Zhanatbekova, N., Andasbayev, Y., & Boribekova, F. (2025). Fostering AI literacy in pre-service physics teachers: Inputs from training and co-variables. *Frontiers in Education*. https://doi.org/10.3389/feduc.2025.1505420

Al-Abdullatif, A. (2025). Auditing AI literacy competency in K–12 education: The role of awareness, ethics, evaluation, and use in human–machine cooperation. *Systems*. https://doi.org/10.3390/systems13060490

Al-Abdullatif, A. M. (2024). Modeling teachers' acceptance of generative artificial intelligence use in higher education: The role of AI literacy, intelligent TPACK, and perceived trust. *Education Sciences, 14*(11), 1209. https://doi.org/10.3390/educsci14111209

Allen, L. K., & Kendeou, P. (2023). ED-AI Lit: An interdisciplinary framework for AI literacy in education. *Policy Insights from the Behavioral and Brain Sciences, 11*(1), 3–10. https://doi.org/10.1177/23727322231220339

Alqarni, A. (2025). Artificial Intelligence-Critical Pedagogic: Design and psychologic validation of a teacher-specific scale for enhancing critical thinking in classrooms. *Journal of Computer Assisted Learning*. https://doi.org/10.1111/jcal.70039

Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology, 8*(1), 19–32. https://doi.org/10.1080/1364557032000119616

Bahroun, Z., Anane, C., Ahmed, V., & Zacca, A. (2023). Transforming education: A comprehensive review of generative artificial intelligence in educational settings through bibliometric and content analysis. *Sustainability*. https://doi.org/10.3390/su151712983

Bauer, E., Greiff, S., Graesser, A., Scheiter, K., & Sailer, M. (2025). Looking beyond the hype: Understanding the effects of AI on learning. *Educational Psychology Review*. https://doi.org/10.1007/s10648-025-10020-8

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health, 6*, 149. https://doi.org/10.3389/fpubh.2018.00149

Bower, M., Torrington, J., Lai, J., Petocz, P., & Alfano, M. (2024). How should we change teaching and assessment in response to increasingly powerful generative artificial intelligence? Outcomes of the ChatGPT teacher survey. *Education and Information Technologies, 29*, 15403–15439. https://doi.org/10.1007/s10639-023-12405-0

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford Press.

Carolus, A., Koch, M. J., Straka, S., Latoschik, M. E., & Wienrich, C. (2023). MAILS—Meta AI Literacy Scale: Development and testing of an AI literacy questionnaire. *Computers in Human Behavior: Artificial Humans, 1*(2), 100014. https://doi.org/10.1016/j.chbah.2023.100014

Celik, I. (2022). Towards Intelligent-TPACK: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education. *Computers in Human Behavior, 138*, 107468. https://doi.org/10.1016/j.chb.2022.107468

Chan, K., & Tang, W. (2024). Evaluating English teachers' artificial intelligence readiness and training needs with a TPACK-based model. *World Journal of English Language*. https://doi.org/10.5430/wjel.v15n1p129

Chiu, T. (2023). The impact of generative AI (GenAI) on practices, policies and research direction in education: A case of ChatGPT and Midjourney. *Interactive Learning Environments, 32*, 6187–6203. https://doi.org/10.1080/10494820.2023.2253861

Chiu, T. K. F. (2024). Define, foster, and assess AI literacy in education: Implications for teaching and learning. *Computers & Education: Artificial Intelligence, 5*, 100223. https://doi.org/10.1016/j.caeai.2024.100223

Chiu, T. K. F., Leung, K. L., & Wong, Y. Y. (2025). Teacher AI competence and self-efficacy (TAICS): Development and validation. *Education and Information Technologies*. https://doi.org/10.1007/s10639-024-13094-z

Coşgun, G. (2025). Artificial intelligence literacy in assessment: Empowering pre-service teachers to design effective exam questions for language learning. *British Educational Research Journal*. https://doi.org/10.1002/berj.4177

DeVellis, R. F. (2016). *Scale development: Theory and applications* (4th ed.). Sage Publications.

Ding, L., Kim, S., & Allday, R. (2024). Development of an AI literacy assessment for non-technical individuals: What do teachers know? *Contemporary Educational Technology*. https://doi.org/10.30935/cedtech/14619

Du, H., Jin, S., & Liu, W. (2024). Exploring the effects of AI literacy on teacher learning and self-efficacy. *Humanities and Social Sciences Communications, 11*(1), 125. https://doi.org/10.1057/s41599-024-03101-6

Fabrigar, L. R., & Wegener, D. T. (2012). *Exploratory factor analysis*. Oxford University Press.

Fan, O., Zheng, L., & Jiao, P. (2022). Artificial intelligence in online higher education: A systematic review of empirical research from 2011 to 2020. *Education and Information Technologies, 27*, 7893–7925. https://doi.org/10.1007/s10639-022-10925-9

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39–50. https://doi.org/10.1177/002224378101800104

Giannakos, M., Azevedo, R., Brusilovsky, P., Cukurova, M., Dimitriadis, Y., Leo, D., Järvelä, S., Mavrikis, M., & Rienties, B. (2024). The promise and challenges of generative AI in education. *Behaviour & Information Technology, 44*, 2518–2544. https://doi.org/10.1080/0144929X.2024.2394886

Grassini, S. (2024). A psychometric validation of the PAILQ-6: Perceived Artificial Intelligence Literacy Questionnaire. In *Proceedings of NordiCHI 2024*. https://doi.org/10.1145/3679318.3685359

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage.

Hava, K., Demir, S., & Aktepe, A. (2025). Digital proficiency and teachers' AI-TPACK competencies: A structural equation approach. *Education and Information Technologies*. https://doi.org/10.1007/s10639-024-12939-x

Hornberger, M., Bewersdorff, A., & Nerdel, C. (2023). What do university students know about artificial intelligence? Development and validation of an AI literacy test. *Computers & Education: Artificial Intelligence, 5*, 100165. https://doi.org/10.1016/j.caeai.2023.100165

Hornberger, M., Bewersdorff, A., & Nerdel, C. (2025). Development and validation of a short AI literacy test (AILIT). *Computers in Human Behavior: Artificial Humans*. https://doi.org/10.1016/j.chbah.2025.100060

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55.

https://doi.org/10.1080/10705519909540118

Jia, K., Sun, X., & Li, T. (2025). Developing a holistic AI literacy framework for children: A design-based research approach. In *Proceedings of the ACM Interaction Design and Children Conference (IDC '25)*. https://doi.org/10.1145/3727986

Joanna Briggs Institute (JBI). (2020). *JBI manual for evidence synthesis*. The University of Adelaide. https://doi.org/10.46658/JBIMES-20-01

Kamalov, F., Calonge, D., & Gurrib, I. (2023). New era of artificial intelligence in education: Towards a sustainable multifaceted revolution. *Sustainability*. https://doi.org/10.3390/su151612451

Kim, S. W., & Lee, Y. (2022). The artificial intelligence literacy scale for middle school students. *Journal of the Korea Society of Computer and Information, 27*(3), 225–238. https://doi.org/10.9708/jksci.2022.27.03.225

Kline, R. B. (2021). *Principles and practice of structural equation modeling* (5th ed.). Guilford Press.

Koch, M. J., Carolus, A., Wienrich, C., & Latoschik, M. E. (2024). Meta AI literacy scale: Further validation and development of a short version. *Heliyon, 10*(21), e39686. https://doi.org/10.1016/j.heliyon.2024.e39686

Kong, S. C., et al. (2023). Evaluating an AI literacy programme for secondary students: Impacts and implications. *Education and Information Technologies, 28*, 5031–5053. https://doi.org/10.1007/s10639-022-11408-7

Laupichler, M. C., Aster, A., Haverkamp, N., & Raupach, T. (2023). Development of the Scale for the Assessment of Non-Experts' AI Literacy: An exploratory factor analysis. *Computers in Human Behavior Reports, 12*, 100338. https://doi.org/10.1016/j.chbr.2023.100338

Laupichler, M. C., Carolus, A., & Koch, M. J. (2023). Establishing a core set of AI literacy items through a Delphi study. *Computers & Education: Artificial Intelligence, 4*, 100126. https://doi.org/10.1016/j.caeai.2023.100126

Lee, S., & Park, G. (2024). Development and validation of ChatGPT literacy scale. *Current Psychology, 43*(21), 18992–19004. https://doi.org/10.1007/s12144-024-05723-0

Levac, D., Colquhoun, H., & O'Brien, K. K. (2010). Scoping studies: Advancing the methodology. *Implementation Science, 5*, 69. https://doi.org/10.1186/1748-5908-5-69

Lim, W., Gunasekara, A., Pallant, J., Pallant, J., & Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *The International Journal of Management Education*. https://doi.org/10.1016/j.ijme.2023.100790

Lintner, T. (2024). A systematic review of AI literacy scales. *NPJ Science of Learning, 9*. https://doi.org/10.1038/s41539-024-00264-4

Lintner, T., Carolus, A., & Koch, M. J. (2024). A systematic review of AI literacy instruments: Dimensions, validity, and challenges. *NPJ Science of Learning, 9*, 1–18. https://doi.org/10.1038/s41539-024-00264-4

Liu, Y., Chen, X., Wang, L., & Zhang, H. (2025). Measuring generative AI literacy: Scale development and validation in workplace contexts. *Computers & Education, 210*, 105001. https://doi.org/10.1016/j.compedu.2025.105001

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–16). https://doi.org/10.1145/3313831.3376727

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research, 35*(6), 382–386. https://doi.org/10.1097/00006199-198611000-00017

Malhotra, N. K., & Dash, S. (2011). *Marketing research: An applied orientation* (6th ed.). Pearson Education.

Mishra, P., Warr, M., & Islam, R. (2023). TPACK in the age of ChatGPT and generative AI. *Journal of Digital Learning in Teacher Education, 39*, 235–251. https://doi.org/10.1080/21532974.2023.2247480

Monib, W., Qazi, A., Apong, R., Azizan, M., De Silva, L., & Yassin, H. (2024). Generative AI and future education: A review, theoretical validation, and authors' perspective on challenges and solutions. *PeerJ Computer Science, 10*. https://doi.org/10.7717/peerj-cs.2105

Msambwa, M., Wen, Z., & Daniel, K. (2025). The impact of AI on the personal and collaborative learning environments in higher education. *European Journal of Education*. https://doi.org/10.1111/ejed.12909

Ng, D. T. K., Wu, J., Leung, J. K. L., Chu, S. K. W., Qiao, M. S., & Li, X. (2024). Design and validation of the AI Literacy Questionnaire: The affective, behavioral, cognitive, and ethical approach. *British Journal of Educational Technology, 55*(3), 915–936. https://doi.org/10.1111/bjet.13411

Ng, W. (2012). Can we teach digital natives digital literacy? *Computers & Education, 59*(3), 1065–1078. https://doi.org/10.1016/j.compedu.2012.04.016

Ning, Y., Lin, Y., & Chen, J. (2024). Teachers' AI-TPACK: Integrating technological, pedagogical, and ethical competencies for AI use. *Sustainability, 16*(3), 978. https://doi.org/10.3390/su16030978

Park, H. (2025). Development of a teacher's AI literacy scale based on meta-analysis. *Korea University Institute of Educational Research*. https://doi.org/10.24299/kier.2025.381.41

Peters, M. D. J., Godfrey, C. M., Khalil, H., McInerney, P., Parker, D., & Soares, C. B. (2015). Guidance for conducting systematic scoping reviews. *International Journal of Evidence-Based Healthcare, 13*(3), 141–146. https://doi.org/10.1097/XEB.0000000000000050

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research, 8*(2), 23–74. *(No DOI available.)*

Sperling, K., Stenberg, C., McGrath, C., Åkerfeldt, A., Heintz, F., & Stenliden, L. (2024). In search of artificial intelligence (AI) literacy in teacher education: A scoping review. *Computers and Education Open*. https://doi.org/10.1016/j.caeo.2024.100169

Tapalova, O., Zhiyenbayeva, N., & Gura, D. (2022). Artificial intelligence in education: AIEd for personalised learning pathways. *Electronic Journal of e-Learning*. https://doi.org/10.34190/ejel.20.5.2597

Tenberga, I., & Daniela, L. (2024). Artificial intelligence literacy competencies for teachers through self-assessment tools. *Sustainability*. https://doi.org/10.3390/su162310386

Walter, Y. (2024). Embracing the future of artificial intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education. *International Journal of Educational Technology in Higher Education, 21*, 1–29. https://doi.org/10.1186/s41239-024-00448-3

Wang, H., Rau, P. L. P., & Yuan, Y. (2022). Measuring artificial intelligence literacy: Scale development and validation. *Computers & Education, 187*, 104515. https://doi.org/10.1016/j.compedu.2022.104515

Wang, X., Xu, X., Zhang, Y., Hao, S., & Jie, W. (2024). Exploring the impact of artificial intelligence application in personalized learning environments: Thematic analysis of undergraduates' perceptions in China. *Humanities and Social Sciences Communications*. https://doi.org/10.1057/s41599-024-04168-x

Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist, 34*(6), 806–838.

https://doi.org/10.1177/0011000006288127

Wu, D., & Zhang, J. (2025). Generative artificial intelligence in secondary education: Applications and effects on students' innovation skills and digital literacy. *PLOS ONE, 20*. https://doi.org/10.1371/journal.pone.0323349

Yan, L., Sha, L., Zhao, L., Li, Y., Maldonado, R., Chen, G., Li, X., Jin, Y., & Gašević, D. (2023). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology, 55*, 90–112. https://doi.org/10.1111/bjet.13370

Yang, H., Lee, W., & Kim, J. (2025). Identification of key factors influencing teachers' self-perceived AI literacy: An XGBoost and SHAP-based approach. *Applied Sciences*. https://doi.org/10.3390/app15084433

Yim, I. H. Y. (2024). Developing an intelligence-based AI literacy framework for primary education. *Computers & Education: Artificial Intelligence, 6*, 100319. https://doi.org/10.1016/j.caeai.2024.100319

Younis, B. (2025). The artificial intelligence literacy (AIL) scale for teachers: A tool for enhancing AI education. *Journal of Digital Learning in Teacher Education, 41*(1), 37–56. https://doi.org/10.1080/21532974.2024.2441682

Zhai, X. (2024). Transforming teachers' roles and agencies in the era of generative AI: Perceptions, acceptance, knowledge, and practices. *arXiv*. https://doi.org/10.48550/arxiv.2410.03018

Zhang, H., Lee, I., Ali, S., DiPaola, D., & Breazeal, C. (2024). Developing and validating the AI Literacy Concept Inventory (AI-CI): An instrument to assess AI literacy among middle-school students. *International Journal of Artificial Intelligence in Education*. https://doi.org/10.1007/s40593-024-00398-x

Zhao, L., Wu, X., & Luo, H. (2022). Developing AI literacy for primary and middle school teachers in China: Based on a structural equation modeling analysis. *Sustainability*. https://doi.org/10.3390/su142114549