Developing and Validating an AI Literacy Scale for English Language Teachers: A Mixed-Methods Study

Kuysin Tukhtaeva[1], Nafisa Ochilova[1], Chen Kaizi[4], Nodira Rakhimova[1] & Yulduz Khasanova[2], Zebiniso Tuychieva[1], Ali Khodi[1,3*]

[1] English Lingusitics Department, Navoi State university, Navoi, Uzbakistan

[2] Department of Practical English, Navoi State university Navoi, Uzbakistan

[3] Institue of Languages, UCSI University, Kuala Lumpur, Malaysia

[4] Faculty of Social Sciences and Liberal Arts, UCSI University, Kuala Lumpur, Malaysia

Correspondence: Ali Khodi, Institue of Languages, UCSI University, Kuala Lumpur, Malaysia, E-mail: alikhodi92@gmail.com

Received: July 21, 2025 Accepted: January 9, 2026 Online Published: February 24, 2026

doi:10.5430/wjel.v16n4p1 URL: https://doi.org/10.5430/wjel.v16n4p1

Abstract

The rapid integration of artificial intelligence (AI) into English language teaching (ELT) necessitates a specialized framework to assess teachers' readiness to navigate these technologies. This study develops and validates an AI Literacy Scale tailored specifically for English language teachers, addressing the gap in existing frameworks that often lack specificity for ELT's linguistic, cultural, and pedagogical demands of ELT. Through a four-phase mixed-methods approach (item generation, content validation, pilot testing, and factor analysis) the scale was refined to a psychometrically acceptable five-factor model: Understanding AI in Education, Proficiency in Using AI, Pedagogical Alignment, Ethical Awareness, and AI for Feedback and Assessment. Participants included 150 in-service English language teachers, all with at least five years of teaching experience, who demonstrated C2 proficiency in English according to the Common European Framework of Reference (CEFR). They were selected using purposive sampling to ensure a diverse representation across urban, semi-urban, and rural settings. Validated with 150 in-service teachers, the scale demonstrated strong fit indices ($\chi^2 = 140.174$, df = 125, p = 0.167; CFI = 0.949; RMSEA = 0.032). This tool not only measures technical and ethical competencies but also guides professional development by identifying gaps in AI literacy. The study used exploratory and confirmatory factor analysis to ensure the scale's construct validity and reliability. By fostering reflective and equitable AI integration, the scale empowers teachers to enhance instruction while preserving learner-centered practices, contributing to the broader discourse on educational innovation in ELT.

Keywords: AI Literacy, English Language Teaching, Scale Development, Teacher Preparedness, Pedagogical Integration, Quality Education

1. Introduction

The field of English language teaching is undergoing a significant transformation, driven by the rapid rise of digital technologies—particularly artificial intelligence (AI). What once relied heavily on printed textbooks and face-to-face instruction is now increasingly shaped by AI-powered tools that support language learning. Applications such as grammar checkers, voice assistants, and adaptive learning platforms are becoming increasingly prevalent and accessible in both formal and informal learning environments. For teachers involved in Teaching English as a Foreign Language (TEFL) and Teaching English to Speakers of Other Languages (TESOL), these developments introduce not only new possibilities for enhancing instruction but also complex challenges regarding integration, ethics, and pedagogy (Widagsa, Senom, & Hutagalung, 2025). This shift necessitates that teachers adapt their methods to incorporate these innovative tools effectively.

As AI becomes increasingly embedded in educational practices, English language teachers are expected to navigate unfamiliar technologies while maintaining the quality and human-centered nature of their instruction (Babanoğlu, Karataş, & Dündar, 2025). However, many professional development programs have not kept pace with these rapid changes. Training often focuses on basic digital skills or general technology usage rather than addressing the unique demands and implications of AI within language teaching contexts. This leaves many teachers unprepared to engage with AI tools critically or to utilize them in ways that align with sound pedagogical principles and ethical responsibilities (Gouseti et al., 2025)

This situation reveals a clear gap: there is currently no widely accepted framework that defines or measures "AI literacy" in a way that is specifically relevant to English language teachers (Zhou et al., 2025). In this study, AI literacy is understood as the ability to make informed, critical, and pedagogically meaningful use of AI technologies in language teaching. It includes technical understanding, awareness of how AI can affect language practices and learner identities, and the capacity to make ethical decisions about AI use in the classroom (Almashour, Aldamen, & Jarrah, 2025). Existing digital literacy frameworks are often too general to capture these specific needs, and they rarely consider the cultural, linguistic, and practical challenges that teachers face when working with AI in real-world ELT education. To address this gap, this research aims to design and validate an AI Literacy Scale tailored to the realities of English language education. This scale is intended to be more than a checklist of digital skills—it is a tool for assessing teachers' readiness to integrate AI in ways that support student learning, preserve teacher agency, and promote reflective, ethical practice. By focusing on how teachers understand and respond to AI's influence on language teaching, the scale also serves as a guide for teacher education programs, curriculum designers, and policymakers (Ahmadi Fatalaki et al., 2025). It can inform the development of targeted training, identify gaps in professional knowledge, and support the broader goal of preparing educators for the future of language instruction. Ultimately, this study contributes to the ongoing conversation about innovation in education by offering a practical and research-based instrument that places teachers at the centre of AI integration. It recognizes the need to prepare teachers not just to use new tools, but to lead pedagogical change with insight, responsibility, and a clear sense of purpose.

2. Literature Review

2.1 Conceptualizing AI Literacy in English Language Teaching

The increasing presence of Artificial Intelligence (AI) in educational contexts is reshaping how we understand literacy in the digital age. While traditional notions of literacy have focused on reading, writing, and text comprehension, contemporary definitions have expanded to include the digital and technological competencies necessary for navigating an AI-driven world (Ng, 2012; Baskara, 2025). Within this broader conceptual shift, AI literacy has emerged as an essential area of focus. It encompasses the ability to comprehend how AI systems function, reflect on their societal and ethical implications, and use them effectively and responsibly (Long & Magerko, 2020).

In ELT, AI literacy is becoming increasingly important. Teachers are encountering a growing range of AI-supported tools, such as intelligent language tutoring platforms, automated feedback systems, speech recognition technologies, and conversational agents. These innovations provide new opportunities to enhance language instruction, personalize learning, and support assessment practices (Chen et al., 2020). However, the potential benefits of these tools can only be realized when teachers possess the knowledge to use them thoughtfully—both in terms of practical application and critical reflection. Importantly, AI literacy in ELT involves more than just learning how to operate educational technologies. Scholars have emphasized its multidimensional nature. For instance, Long and Magerko (2020) describe AI literacy as involving technical knowledge, critical awareness of the broader social and ethical issues surrounding AI, and the ability to interact meaningfully with AI systems. Similarly, Ng et al. (2021) argue for a shift away from basic usage toward a deeper understanding of AI's limitations, biases, and impacts on equity and identity in learning. These perspectives are particularly important in language education, where cultural sensitivity, learner diversity, and communicative interaction are central to effective pedagogy.

For English language teachers, cultivating AI literacy means engaging with technology not only as a tool for instruction but also as a subject of critical exploration (Hao, Fang & Peng, 2024). It requires an awareness of how AI systems shape linguistic norms, influence learner autonomy, and mediate relationships between students and teachers. Given the global and multilingual nature of English teaching, these considerations are especially vital. Teachers must be prepared to guide students in using AI tools responsibly while also questioning how these technologies impact access to language, knowledge, and expression.

Despite its growing significance, AI literacy is still largely lacking in most teacher education and professional development programs. Many English language teachers report limited exposure to AI technologies during their training and few opportunities to develop essential competencies once they enter the profession. Addressing this gap necessitates rethinking how we prepare teachers—not only through technical training but also through critical discussions about the ethical and pedagogical dimensions of AI. Programs should incorporate experiential learning with AI tools, opportunities to analyze algorithmic bias, and strategies for integrating automation without sacrificing the human aspects of language instruction. Moreover, AI literacy must be seen as a shared responsibility between teachers and learners. Recent empirical studies highlight that students who develop AI literacy are not only better at leveraging AI for language learning but also more critical of its limitations (Song & Song, 2023). This suggests that classrooms must function as collaborative spaces where both groups negotiate the affordances and risks of AI. Moreover, developing AI literacy extends beyond teachers; it also involves empowering learners (Zhang & Zhang, 2024). Students need to become discerning users of AI tools, aware of how these technologies affect their learning processes, privacy, and sense of agency. Language classrooms must become environments where both teachers and learners can explore the promises and pitfalls of AI, ensuring that its use supports inclusive, equitable, and reflective practices. Thus, AI literacy in ELT goes beyond digital proficiency to encompass a deeper engagement with how AI technologies intersect with pedagogy, ethics, and language. By fostering this literacy among teachers and learners, we can ensure that the integration of AI into language education aligns with human values and educational goals rather than merely following technological trends (Al-khresheh, 2024). Ultimately, cultivating AI literacy will empower future generations to navigate an increasingly complex digital landscape with confidence and critical insight. Ultimately, a strong conceptualization of AI literacy is not just theoretical—it is increasingly tied to how learners and teachers experience technology in practice. As empirical findings show, AI literacy influences learners' motivation (Sun et al., 2024), writing performance, and even willingness to communicate in English (Zhang et al., 2025). Consequently, the concept must be continuously re-examined in light of evolving classroom realities.

2.2 Emerging Frameworks of AI Literacy

Recent scholarship has increasingly emphasized the need for robust frameworks for AI literacy in education, highlighting the ongoing development required to equip teachers with the knowledge and skills to navigate intelligent technologies responsibly (Zou, Xie, &

Kohnke, (2025). These frameworks, whether established or still under research and development, typically integrate cognitive, ethical, and technical dimensions to support educators across various disciplines. For instance, Allen and Kendeou (2024) introduced the ED-AI Lit model, which highlights the need for foundational knowledge of AI, ethical considerations such as algorithmic bias, and practical skills. While their comprehensive approach advocates for integrating AI literacy across subjects, it lacks specific guidance for English Language Teaching (ELT), where cultural representation and learner identity are vital.

Similarly, Annapureddy et al. (2024) outlined twelve competencies specific to generative AI, encompassing technical skills, critical engagement, and tool-based application. However, only a few of these competencies directly address the pedagogical integration of AI in classroom settings particularly those unique to language instruction. As a result, while the framework is broad in scope, it does not fully respond to the immediate needs of ELT teachers, who must navigate complex instructional, sociolinguistic, and ethical dimensions in multilingual and multicultural environments.

More recently the AI competency framework for teachers developed by Miao and Cukurova, published by UNESCO in 2024, proposed a more adaptable approach that incorporates functional, ethical, rhetorical, and pedagogical literacies (Cukurova & Miao, 2024). These models are based on interdisciplinary consultation and provide a flexible structure for AI literacy. However, they still fall short in addressing specific challenges faced in ELT, such as code-switching, dialect diversity, and the sociocultural nuances of pronunciation and writing feedback. This omission highlights a broader issue: existing frameworks often generalize, failing to fully reflect the realities of teaching English in diverse global contexts.

Empirical research shows that AI is already prevalent in ELT classrooms. A global survey by the British Council (Edmette et al., 2023) involving 1,348 teachers across 118 countries found that English language teachers are among the most frequent users of AI tools. Participants reported using AI mainly for grammar correction (67%), speaking practice (54%), and writing support (42%), with platforms such as Grammarly, Duolingo, and Google Translate. Despite this widespread use, 72% of respondents felt unprepared to use these tools effectively, and 63% expressed concerns regarding ethical issues pertaining to data privacy and algorithmic transparency. These findings reveal a significant gap between practice and preparedness (BERA, 2023). Importantly, many AI tools tend to default to standardized English varieties, which may marginalize non-standard dialects and underrepresented linguistic communities. These concerns reflect broader critiques of bias in AI design and implementation (Lachini, 2024).

Teacher training continues to be a central barrier to effective AI integration. In a study of ESL teachers in Malaysia, Hashemi and Kew (2021) identified limited training time, a lack of contextualized guidance, and low self-confidence as primary obstacles to AI adoption. Mehdaoui (2024) also found that extrinsic barriers significantly hinder the integration of AI technologies. Despite recognizing AI's potential, challenges such as large class sizes, inadequate resources, and insufficient training contribute to resistance. Beyond identifying barriers, empirical studies now point to measurable benefits when AI literacy is systematically cultivated. Song and Song (2023) showed that AI-assisted writing instruction improved coherence, grammar, and vocabulary, while Zhang et al. (2025) demonstrated that AI literacy increased learners' willingness to communicate through higher self-efficacy and reduced anxiety. These findings illustrate how frameworks must link conceptual dimensions with real classroom practices. Despite existing challenges, the potential of AI in ELT is well established. Jiang (2022) found that AI tools boost learner engagement and achievement. AI is also capable to reshape language assessment. To support this, a growing body of research focuses on measuring and fostering AI literacy. For example, Ding et al. (2024) found that case-based professional development significantly improved teachers' understanding and classroom use of AI. Assessment tools like AICOS (Markus et al., 2025) have been developed to evaluate AI competence across domains. However, these instruments lack ELT-specific indicators and often overlook the cultural, linguistic, and pedagogical nuances that shape language education. Taken together, these frameworks and studies underscore the urgent need to create ELT-focused AI literacy models that integrate technical, ethical, and pedagogical perspectives with evidence from classroom research. Without this integration, AI literacy risks remaining a theoretical construct rather than a practical guide for teachers.

In light of these findings, future research and policy should focus on developing AI literacy frameworks tailored to English Language Teaching. These frameworks must promote technical competence while encouraging critical evaluation of AI outputs across various linguistic and cultural contexts, addressing ethical issues like bias and data privacy. They should also provide pedagogical strategies for human-centered AI integration and support professional development grounded in real-world classroom practices, particularly in under-resourced areas. Moreover, creating ELT-specific AI literacy assessment tools is essential for benchmarking teacher preparedness and guiding training programs (Rezvani & Izadi, 2025). As AI becomes integral to ELT, the field faces a critical juncture. Without targeted approaches to AI literacy, language teachers' risk being overwhelmed by technology or falling behind. Investing in AI literacy is vital not only for keeping pace with innovation but also for shaping a future that is equitable and transformative.

While existing interdisciplinary frameworks and assessment tools provide a foundation, they lack responsiveness to the specific demands of ELT. This study addresses this gap by developing and validating an ELT-specific AI Literacy Scale that integrates technical, ethical, and instructional aspects within a language teaching framework. This tool aims to support teacher preparation and ensure that AI integration in language classrooms promotes equitable, reflective, and learner-centered education.

3. Method

This study employed a rigorous four-phase mixed-methods research design to systematically develop and validate an AI Literacy Scale specifically for English language teachers. The process was guided by best practices in scale development and psychometric validation, ensuring both theoretical alignment and empirical robustness in practice. Each phase was executed in sequence, as outlined below. Phase one focused on generating an extensive pool of items by reviewing existing AI literacy frameworks, language teaching literature, and research findings related to ELT. This step aimed to create an initial set of items that reflected the multidimensional nature of AI literacy. It included important aspects like technical skills, ethical considerations, and teaching methods, all tailored to English language education. In the next phase, phase two involved examining the current state of knowledge and practice. Content validation was carried out through expert reviews. Experts were selected based on stringent criteria: they needed to be specialists in applied linguistics, educational technology, and assessment design, have at least 10 years of experience in English language education, and have a strong publication record (with an H-index of 10 or more) in ELT and technology-related areas.

The experts attended a briefing session with the researchers to understand the study's goals and methodological direction. After that, they worked independently to ensure the authenticity and originality of their comments. The initial items developed were shared with them, along with information about the intended users. They had one week to review and provide feedback. After receiving their comments, the researchers held a second round of meetings with the experts clarify any ambiguities in the feedback received. The research team also had the authority to accept, reject, or apply the suggestions to refine the items. This process aimed to improve the clarity, relevance, and representativeness of items, enhancing the content validity of the scale.

Next, phase three involved pilot testing with a randomly selected sample of English language teachers. Participants included both male and female teachers with a Bachelor’s degree or higher and at least 5 years of English teaching experience. They had 60 minutes to complete the tasks and had not received any prior training for the study. The administration of the pilot test was conducted online. This was followed by Exploratory Factor Analysis (EFA) to identify the underlying factor structure and assess the scale’s construct validity. The final phase of the study used Confirmatory Factor Analysis (CFA) on a separate, larger sample to test the stability and goodness-of-fit of the proposed model for the developed instrument. Throughout all stages, the methodological approach followed established protocols for instrument development in educational research (DeVellis, 2016; Boateng et al., 2018) and was based on current understandings of AI literacy within the context of English language teaching. This design ensured that the resulting scale would be both contextually relevant and psychometrically sound, effectively capturing the competencies necessary for the ethical and effective integration of AI in language education.

3.1 Identify Subsections

It is both conventional and expedient to divide the Method section into labeled subsections. These usually include a section with descriptions of the participants or subjects and a section describing the procedures employed in the study. The latter section often includes description of (a) any experimental manipulations or interventions used and how they were delivered-for example, any mechanical apparatus utilized; (b) sampling procedures and sample size and statistical precision; (c) measurement approaches (including the psychometric properties of the instruments applied); and (d) the research design. If the design of the study is complex or the stimuli require a detailed description, additional subsections or subheadings are warranted to help readers locate specific information. Include in these subsections the information essential to comprehending and replicating the study. Insufficient detail leaves the reader with questions; conversely, too much detail burdens the reader with irrelevant information. Consider using appendices and/or a supplemental website for more detailed information.

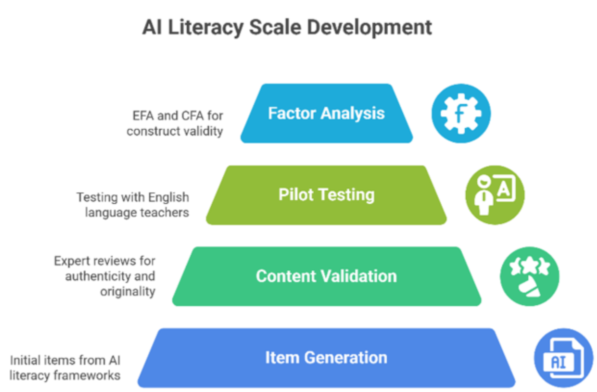

Figure 1. Four-Phased Design of the Study

3.2 Participant

In this study, we recruited participants at four levels: experts for content validation, participants for the pilot study, and those for exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). The panel of experts consisted of six members, purposefully selected for the study, including three females. For the pilot study, 30 English Language Teaching (ELT) teachers randomly selected. These teachers participated online using Google Forms, and the study was open to all ELT teachers who met the minimum requirements. For the EFA and CFA, a total of 150 in-service English language teachers were selected using purposive sampling to ensure relevance and expertise. The inclusion criteria required participants to have at least five years of teaching experience, demonstrate C2 proficiency in English according to the Common European Framework of Reference (CEFR), and currently teach in either public or private institutions. To maintain balanced representation, the sample included approximately 50% male and 50% female teachers, with diverse teaching and cultural backgrounds, including teachers from urban, semi-urban, and rural settings. To ensure methodological rigor in scale development, participants in this study were strategically divided into two groups: approximately 80% (n = 120) were allocated to the exploratory factor analysis (EFA) phase, while the remaining 20% (n = 30) were reserved for confirmatory factor analysis (CFA). This distribution reflects widely accepted practices in instrument validation, where EFA is used first to uncover the latent factor structure, and CFA is then conducted on a separate sample to confirm the fitness and stability of that structure (Worthington & Whittaker, 2006). The rationale for assigning a larger portion of the sample to EFA lies in its data-driven nature—it requires sufficient statistical power to reliably detect and interpret emerging patterns without a priori assumptions (Fabrigar & Wegener, 2012). In contrast, CFA is model-driven and generally less sample-intensive, as it tests a predefined structure against new data. Using independent subsamples for these two analyses helps to avoid overfitting the model to one specific dataset and enhances the generalizability of the findings. Therefore, this 80/20 split not only follows established psychometric guidance but also strengthens the overall validity and reliability of the AI Literacy Scale developed in this study. Although the sample size (N = 150) is modest, it meets the minimum threshold for factor analysis when item communalities are high and factor loadings exceed .40 (Kline, 2023). This size is considered acceptable for preliminary scale validation studies in educational contexts.

3.3 Data Collection

Data collection for the present study was conducted systematically across the four-phase mixed-methods research design, employing both qualitative and quantitative approaches. In Phase 1 (Item Generation), data were gathered through an extensive review of scholarly literature. In Phase 2 (Content Validation), qualitative feedback and numerical ratings were collected from a panel of six experts using a structured review process to evaluate the items overall in terms of clarity, relevance, and representativeness and resolve ambiguities. For Phases 3 (Pilot Testing & EFA) and 4 (CFA), quantitative data were collected to avoid overfitting and enhance generalizability (Alavi, Karami, & Khodi, 2021). These methods were rigorously designed to align with scale development standards, ensuring the data's validity, reliability, and contextual relevance of the data for assessing AI literacy in English language teaching; the details of which are presented in the following sections.

3.3.1 Phase 1: Item Generation

The initial phase of the study focused on the conceptualization and development of survey items aimed at measuring AI literacy in the context of English language education. This phase was structured around five interrelated stages: (1) literature and policy review, (2) expert engagement, (3) initial item drafting, (4) thematic structuring, and (5) iterative refinement. In the first stage, we conducted a comprehensive analysis of scholarly literature, AI literacy frameworks, and institutional policy documents. Key sources included Ng et al. (2021) and Long and Magerko (2020), who emphasized foundational AI literacy competencies such as critical understanding, tool fluency, and ethical judgment. Additionally, we examined 36 institutional AI policies from 12 English-speaking universities, including the University of Melbourne, Stanford University, and the University of Toronto. These policies consistently underscored themes such as responsible AI use, transparency in authorship, and the need for educator oversight. Alongside peer-reviewed literature, we also consulted practical AI education frameworks, such as the OECD (2021) AI and the Future of Skills report and UNESCO's AI Competency Framework. Synthesized insights from these documents formed the theoretical and contextual basis for item generation.

Following this, we engaged in collaborative expert discussions and internal workshops to guide the creation and revision of the item pool. Initial drafting produced 52 items reflecting various facets of AI literacy, such as identifying AI-generated text, interpreting AI feedback, and making pedagogically informed decisions about tool use. These items were informed by both theoretical constructs and practical classroom scenarios, with contributions from six researchers specializing in applied linguistics, educational technology, and AI ethics. The item development process was documented through shared analytic memos, collaborative digital notetaking, and tracked revisions using version-controlled documents. During the third and fourth stages, items were refined through iterative review and inductively grouped using thematic analysis (Braun & Clarke, 2006). Five themes emerged organically from the data: (1) Understanding AI in Education, (2) Proficiency in Using AI, (3) Pedagogical Alignment with Teaching, (4) Ethical Awareness in AI, and (5) AI for Feedback and Assessment. These themes served as both analytical categories and content domains. "Although the five factors in this study emerged inductively through exploratory and confirmatory analyses, their structure resonates with established AI literacy frameworks (Long & Magerko, 2020; Cukurova & Miao, 2024). Specifically, the 'Understanding AI' and 'Technical Proficiency' dimensions correspond to the cognitive and functional components of AI literacy, whereas 'Pedagogical Alignment' and 'Ethical Awareness' extend these concepts into the educational domain. The 'AI for Feedback and Assessment' factor appears unique to English language teaching, highlighting a context-specific dimension that not emphasized in previous general frameworks. Figure 1 illustrates the conceptual correspondence between the empirically derived factors and these broader theoretical models." The final stage involved consolidating and validating the items, resulting in a set of 40 survey items balanced across the five domains. This rigorous, multi-stage development process ensured theoretical coherence, practical relevance, and methodological transparency.

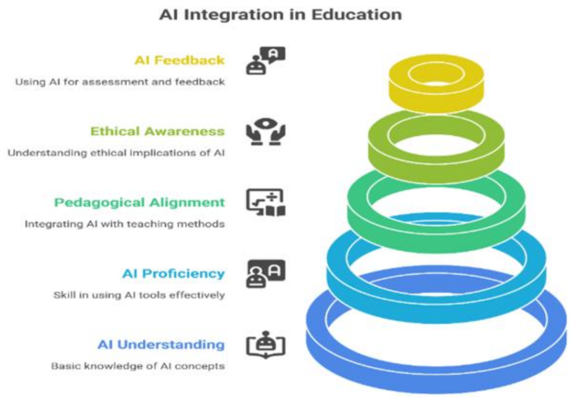

Figure 2. AI Integration Process

3.3.2 Phase 2: Content Validation

To establish content validity, the initial items developed in the item generation phase (Phase I) of the study were evaluated by a panel of six experts in English language teaching (ELT), educational technology, and psychology, with two experts representing each discipline and each gender. In the first step, the experts provided qualitative comments and feedback on the relevance, clarity, and representativeness of the items. The researchers then revised the items based on this feedback. Where discrepancies or ambiguities arose, the researchers consulted the experts for clarification. Next, the revised items were shared again with the experts, who rated the items on a numerical scale. The inter-rater reliability was found to be 0.75, indicating a good level of agreement among the experts. The minimum score for each item was considered for further analysis. Items receiving a Content Validity Index (CVI) below 0.7 were subject to a second round of revision by the researchers. This process was repeated for a third round, where items with a CVI below 0.5 were removed, following Lynn's (1986) guidelines. As a result, from the initial pool of 40 items, the review process led to the modification of 12 items and the elimination of seven items, resulting in a refined pool of 33 items for pilot testing.

3.3.3 Phase 3: Pilot Testing and Exploratory Factor Analysis

In this phase, the 33 items that passed all validity checks were assessed. The revised scale was carefully evaluated for completeness and normality before analysis. The results of the pilot study were obtained by conducting EFA using principal axis factoring with oblique rotation (Promax) to identify the underlying factor structure. This analysis aimed to determine if the theoretically derived dimensions corresponded with the practical items and to establish their construct validity. To this end, the Kaiser–Meyer–Olkin (KMO) measure and Bartlett’s test of sphericity were employed to assess the suitability of the data for factor analysis, as this was necessary for conducting EFA. Following this, items with low item-total correlations (below 0.30) were identified and removed. This 0.30 threshold was used for the further analysis because it guarantees that the items sufficiently contribute to the scale’s internal consistency. Exploratory factor analysis also revealed items with low factor loadings (below 0.40) that needed to be eliminated. This 0.40 cut-off enables the study to retain items with meaningful contributions to the latent constructs aligning with the study’s aim to produce a psychometrically valid and contextually relevant instrument, as supported by prior ELT scale validation studies (Ahmadi Fatalaki et al., 2025). Despite thorough content validity checks and expert validation, some items did not fit the model and contributed minimally to the established theoretical framework in terms of factor loading.

As a result, although efforts were made to revise these items, but ultimately, the cutoff scores determined which items were retained. Additionally, items exhibiting cross-loadings or low communalities were removed iteratively to refine the scale. Based on these assessments, five items were removed from the initial pool, and four items were revised for improved wording. The revised pool of items underwent a second round of EFA, which ultimately supported a five-factor solution aligned with the conceptual framework, leading to a final set of 28 items distributed across the five dimensions. The remaining items were retained for further analysis as part of the CFA, ensuring a solid foundation for subsequent validation.

3.3.4 Phase 4: Confirmatory Factor Analysis

Following the confirmation of the five-factor model for the survey, which corresponded with the theoretical foundation of the instrument developed by the researchers in the Exploratory Factor Analysis (EFA), a Confirmatory Factor Analysis (CFA) was conducted to validate the identified factor structure. The analysis was performed using JASP software (Version 0.19.3), a free and open-source program for statistical analysis supported by the University of Amsterdam. This was done to assess and confirm the fit of the proposed model with the observed data.

The results from the EFA supported the five-factor solution, which was then tested for alignment with the conceptual framework in the CFA. Key fit indices, including the Chi-square statistics, Comparative Fit Index (CFI), and Root Mean Square Error of Approximation (RMSEA), were examined to evaluate the model's goodness-of-fit and to affirm the suitability and validity of the designed instrument. Through this analysis, the final model demonstrated an acceptable fit, therby affirming the validity of the factors and their corresponding items. This robust approach ensured that the measurement instrument effectively captured the intended constructs, laying the groundwork for future research and application. In the following, the results have been presented for the fit of the models and their validity. data (e.g., written questionnaires, interviews, observations) as well as methods used to enhance the quality of the measurements (e.g., the training and reliability of assessors or the use of multiple observations). Provide information on instruments used, including their psychometric and biometric properties and evidence of cultural validity.

4. Results

The findings of this study demonstrate that the dataset was suitable for EFA, as evidenced by the Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy and Bartlett's Test of Sphericity. The KMO value was 0.643, which is considered mediocre (between 0.6 and 0.7). This means that the dataset has some shared variance among the variables, but the connections between them are not very strong. Kaiser (1974) classified KMO values in the 0.60s as "mediocre" (acceptable but not ideal), with values above 0.60 indicating the data is suitable for factor analysis, though higher values (e.g., 0.70s "middling," 0.80s "meritorious") are preferable. This could affect how reliable the factor structure is. In contrast, Bartlett's Test of Sphericity was relatively significant ($\chi^2 = 1017.617$, df = 528, $p < 0.001$), allowing us to reject the idea that the correlation matrix is just random. This indicates that there are significant relationships among the variables, further supporting the use of EFA for our dataset. It also shows that all the items are connected to the factor structure, specifically related to AI literacy.

The results indicate that the data is suitable for factor analysis, which is essential for the next steps. However, the mediocre KMO value requires caution when interpreting the factor structure, as some variables may show weaker correlations. Therefore, a further examination of item-level correlations and factor loadings was conducted to ensure the clarity and reliability of the extracted factors. The results are presented in Table 1.

Table 1. KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		*.643*
Bartlett's Test of Sphericity	*Approx. Chi-Square*	*1017.617*
	df	*528*
	Sig.	*.000*

A Chi-squared goodness-of-fit test was conducted to assess the fit of the proposed five-factor model to the data. The test resulted in a statistic of 389.186 with 373 degrees of freedom. The non-significant p-value ($p > 0.05$) indicates that there is no substantial discrepancy between the observed data patterns and the predictions made by the model. This finding suggests that the five-factor model provides an appropriate fit for the data. Following the assessment of the dataset's suitability for exploratory factor analysis and the evaluation of model fit, a five-factor model was extracted using the promax rotation method, which accommodates correlated factors. The factor loadings, presented in Table 2, elucidate the relationships between the observed variables and the underlying factors, while the uniqueness values indicate the proportion of variance in each variable that is not accounted for by the factors. For clarity, factor loadings below 0.4 have been suppressed.

Table 2. Factor Loadings

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Uniqueness
Item 8	0.894					1.538
Item 6	0.894		0.528			1.783
Item 3	0.817					1.654
Item 4	0.778					1.603
Item 14	0.703					1.021
Item 5	0.703				0.567	1.832
Item 15	0.483					1.534
Item 1	0.481					1.871
Item 2	0.433	0.409				1.370
Item 11	0.424				0.730	1.847
Item 21		1.302				1.597
Item 23		0.755				1.902
Item 22		0.697				0.902
Item 19		0.476			0.554	1.343
Item 16		0.454				1.434
Item 25		0.423				1.024
Item 29			1.016			1.265
Item 33			0.907			1.459
Item 28			0.832			1.385
Item 26			0.649		0.476	1.313

Item 30			0.510	0.675		1.536
Item 27				0.951		0.915
Item 32				0.930		1.171
Item 13				0.629		1.526
Item 18				0.537		0.647
Item 12				0.492		0.843
Item 9					0.891	1.613
Item 17					0.560	1.360
Item 10					0.449	0.901
Item 7						0.582
Item 20						1.648
Item 24						1.598
Item 31						1.198

4.1 Factor Structure Interpretation

A five-factor model was derived, revealing several important insights regarding the relationships among the variables. Factor 1 emerged as a robust construct, showcasing strong loadings from several variables, including Item 8 and Item 6 (both at 0.894). This factor appears to encapsulate a central theme of the theoretical model; however, the presence of cross-loadings from variables such as Item 6 on Factor 3 indicates potential overlap with other factors. Such cross-loadings likely influenced the decision to maintain a five-dimensional structure in the CFA. Factor 2 was primarily defined by a high loading from Item 21(1.302), accompanied by other moderate loadings. A loading exceeding 1.0 raises concerns regarding multicollinearity or redundancy among items, which should be explored further. Additionally, the moderate loading of Item 19 on Factor 5 suggests that this factor may lack distinctiveness. Factor 3 included strong loadings from Item 29 (1.016) and Item 33 (0.907), with the high loading of Item 29 potentially indicating issues with item scaling. Cross-loadings further suggest conceptual overlaps with other factors, warranting careful interpretation. Factor 4 was characterized by significant loadings from Item 27 (0.951) and Item 32 (0.930), indicating a relatively well-defined structure. However, the cross-loading of Item 30 on Factor 3 suggests some ambiguity regarding its distinctiveness. Factor 5 was identified through loadings from Item 9 (0.891) and others, but the presence of multiple cross-loadings indicates that this factor may not be sufficiently distinct, aligning with the initial observations of weaker dimensions. Lastly, several variables (Item 20, Item 24, and Item 31) did not load significantly on any factor, suggesting they are not well aligned with the five-factor structure. These variables may represent dimensions that were excluded during the transition to the final CFA model.

The five-factor EFA model aligns with the acceptable model fit reported previously ($\chi^2 = 389.186$, df = 373, p = 0.271), indicating that the structure is plausible. However, several challenges in the factor structure justified reducing the model to five dimensions in the confirmatory factor analysis (CFA). Issues included cross-loadings, where items such as Item 6 and Item 2 were linked to multiple factors, making it difficult to distinguish some of the original seven dimensions. High uniqueness values for items such as Item 23 indicated that a significant proportion of variance was unexplained, likely due to poor item quality or weak alignment with the intended constructs. Additionally, Items 20, 24, and 31 did not load significantly, suggesting they might belong to dimensions that were dropped from the five-factor model. Finally, some items, such as Item 21 and Item 29, had loadings above 1.0, which could indicate redundancy or issues needing further review.

To better understand the five-factor model derived from the exploratory factor analysis (EFA) using promax rotation, we examined key characteristics including eigenvalues, sums of squared loadings, the proportion of variance explained, and cumulative variance for both unrotated and rotated solutions. These metrics provide insight into the contribution of each factor to the overall variance in the dataset. In the unrotated solution, Factor 1 yielded an eigenvalue of 10.675 and accounted for 13.6% of the total variance, indicating it captures a significant portion of the shared variance. Factor 2 contributed a moderate 5.7%, while Factors 3, 4, and 5 explained 5.1%, 3.9%, and 3.2%, respectively. Overall, the unrotated solution explained only 31.5% of the total variance, suggesting that the factors may not fully capture the underlying complexity of the dataset. This finding aligns with the mediocre KMO value of 0.643, indicating weaker intercorrelations among some variables.

The promax rotation redistributes variance to create a more interpretable structure. In this rotated solution, the sum of squared loadings for Factor 1 decreased to 5.673, explaining 8.4% of the variance, but it remained the strongest contributor. Factor 2 showed a slight improvement, explaining 7.3%, while Factors 3, 4, and 5 accounted for 6.4%, 5.8%, and 5.2%, respectively. The cumulative variance increased to 33.0%, indicating a clearer distribution of variance across factors. The modest 33% cumulative variance explained is considered acceptable for this exploratory scale measuring a complex, multidimensional construct such as ELT-specific AI literacy. Despite this improvement, the cumulative variance remains modest, reinforcing challenges such as cross-loadings and high uniqueness values, which highlight substantial unexplained variance in the model. The low cumulative variance and high uniqueness values indicate potential issues with item quality or measurement error. Furthermore, the presence of cross-loadings implies that some dimensions from the original seven-factor model may have been weak or redundant, thereby justifying their removal in the subsequent CFA.

Table 3. Factor Characteristics

	Unrotated solution				Rotated solution		
	Eigenvalues	SumSq. Loadings	Proportion var.	Cumulative	SumSq. Loadings	Proportion var.	Cumulative
Factor 1	10.675	9.176	0.136	0.136	5.673	0.084	0.084
Factor 2	5.257	3.877	0.057	0.193	4.911	0.073	0.157
Factor 3	4.837	3.434	0.051	0.244	4.295	0.064	0.220
Factor 4	4.051	2.611	0.039	0.283	3.923	0.058	0.279
Factor 5	3.658	2.191	0.032	0.315	3.478	0.052	0.330

Following the EFA, which identified a five-factor structure, a confirmatory factor analysis (CFA) was conducted to validate the revised factor structure. The CFA utilized maximum likelihood (ML) estimation to assess the fit of the five-factor model against the observed data. Table 4 presents the Chi-square test results for the baseline and factor models. The five-factor model produced a χ^2 value of 140.174 with 125 degrees of freedom, resulting in a p-value of 0.167. This non-significant p-value suggests that the model fits the data well, showing no significant discrepancies between the observed and expected patterns. Moreover, the ratio of Chi-square to degrees of freedom (approximately 1.12) is well below the recommended threshold of 2, reinforcing the conclusion of a good model fit according to established guidelines. When comparing the two models, the fit of the confirmatory factor analysis (CFA) model (χ^2 = 140.174, df = 125, p = 0.167) is notably superior to that of the EFA model (χ^2 = 389.186, df = 373, p = 0.271). Despite the CFA employing a more constrained model, it achieves a lower Chi-square value relative to its degrees of freedom. The significant reduction in degrees of freedom—from 373 in EFA to 125 in the CFA—reflects the confirmatory nature of the analysis, which imposes a specific five-factor structure. The non-significant Chi-square in the CFA strengthens the argument that this model adequately captures the data, thereby supporting our decision to reduce the instrument's dimensions from seven to five.

This transition was motivated by both empirical observations and theoretical considerations. The EFA highlighted challenges such as cross-loadings, high uniqueness values, and non-loading items, indicating that some dimensions were weak or redundant. The favorable fit of the CFA suggests that the five-factor model is not only simpler but also more robust, as the two excluded dimensions removed likely contributed minimal unique variance. However, while the Chi-square test points to a good fit, it is important to recognize its sensitivity to sample size. Therefore, additional fit indices such as RMSEA, CFI, TLI, and SRMR—are essential for a more comprehensive evaluation of the model's performance. The moderate KMO value and high uniqueness observed in the EFA uggest that certain items may still possess latent issues that could affect the CFA model's long-term stability.

Table 4. CFA *Chi-square test*

Model	X²	df	p
Baseline model	452.877	153	
Factor model	140.174	125	0.167

In proceeding, Table 5 presents a set of additional fit indices used to further evaluate the goodness-of-fit of the five-factor model, which was derived from the exploratory factor analysis (EFA) and tested through confirmatory factor analysis (CFA) using maximum likelihood (ML) estimation. These indices provide a more comprehensive assessment of how well the model aligns with the data, complementing the previously reported Chi-square test results (χ^2 = 140.174, df = 125, p = 0.167 for the factor model; χ^2 = 452.877, df = 153 for the baseline model). The additional fit indices indicate that the five-factor CFA model demonstrates good to adequate fit. Notably, the Comparative Fit Index (CFI) is 0.949, the Tucker-Lewis Index (TLI) is 0.938, the Bollen's Incremental Fit Index (IFI) is 0.954, and the Relative Noncentrality Index (RNI) is also 0.949. All these indices meet or closely approach the standard threshold for adequate fit, typically set at 0.95 or higher. The non-significant Chi-square test further supports the adequacy of the model. However, certain fit indices signal areas for improvement. The Bentler-Bonett Normed Fit Index (NFI) is lower at 0.690, and both the Parsimony Normed Fit Index (PNFI) at 0.564 and Bollen's Relative Fit Index (RFI) at 0.621 suggest that the model explains only a modest proportion of the covariance among the variables. This aligns with the challenges identified during the EFA, such as a mediocre KMO value of 0.643, high uniqueness values, and the presence of cross-loadings. These findings confirm that the five-factor model is empirically acceptable and more parsimonious while still achieving an acceptable level of accuracy or goodness-of-fit.

Table 5. Fit indices

Index	Value
Comparative Fit Index (CFI)	0.949
Tucker-Lewis Index (TLI)	0.938
Bentler-Bonett Non-normed Fit Index (NNFI)	0.938
Bentler-Bonett Normed Fit Index (NFI)	0.690
Parsimony Normed Fit Index (PNFI)	0.564
Bollen's Relative Fit Index (RFI)	0.621
Bollen's Incremental Fit Index (IFI)	0.954
Relative Noncentrality Index (RNI)	0.949

Table 6 presents the information criteria for the five-factor confirmatory factor analysis (CFA) model, providing evidence regarding its parsimony and overall fit. The model has a log-likelihood of -3652.455 and estimated 46 free parameters, resulting in an Akaike Information Criterion (AIC) of 7396.910, a Bayesian Information Criterion (BIC=7525.135), and a Sample-Size Adjusted Bayesian Information Criterion (SSABIC=7379.705). Lower information-criterion values indicate better fit after accounting for model complexity. Compared to EFA, the five-factor model exhibits improved parsimony with its fewer parameters and favorable fit indices, including a CFI of 0.949, a Chi-square of 140.174 with 125 degrees of freedom, and a non-significant p-value of 0.167. These results indicate that the five-factor model fits the data well while maintaining simplicity.

Table 6. Information criteria

Index	Value
Log-likelihood	-3652.45
Number of free parameters	46.000
Akaike (AIC)	7396.91
Bayesian (BIC)	7525.13
Sample-size adjusted Bayesian (SSABIC)	7379.70

Table 7 demonstrates the additional fit measures for the five-factor confirmatory factor analysis (CFA) model, indicating an acceptable-to-good overall fit. The Root Mean Square Error of Approximation (RMSEA) is 0.032, with a 90% confidence interval of 0.000–0.057 and a p-value of 0.871, which is well below the 0.08 threshold, suggesting acceptable model fit. Additionally, the Standardized Root Mean Square Residual (SRMR) is 0.071, falling within acceptable limits (≤ 0.08). While the Goodness of Fit Index (GFI) is slightly below the preferred level at 0.891, the McDonald Fit Index (MFI) at 0.939 and the Expected Cross Validation Index (ECVI) at 1.935 support a good fit, with lower ECVI values suggesting better expected predictive performance. Furthermore, Hoelter's Critical N values (131.205 at $\alpha = 0.05$ and 141.991 at $\alpha = 0.01$) suggest that the sample size is adequate for reliable results. When combined with prior fit indices, such as a Comparative Fit Index (CFI) of 0.949, a Chi-square of 140.174 with 125 degrees of freedom, and a non-significant p-value of 0.167, these metrics reinforce the five-factor model's superiority over the exploratory factor analysis (EFA) seven-factor structure.

Table 7. Other fit measures

Metric	Value
Root mean square error of approximation (RMSEA)	0.032
RMSEA 90% CI lower bound	0.000
RMSEA 90% CI upper bound	0.057
RMSEA p-value	0.871
Standardized root mean square residual (SRMR)	0.071
Hoelter's critical N (α = .05)	131.205
Hoelter's critical N (α = .01)	141.991
Goodness of fit index (GFI)	0.891
McDonald fit index (MFI)	0.939
Expected cross validation index (ECVI)	1.935

The CFA parameter estimates for the five-factor model provide partial support, with observable loadings in Factors 1-4, particularly strong in Factor 3 (Item 27 = 1.659). However, Factor 5 shows instability due to non-significant loadings (Item 26 = 6.028, p = 0.301). Overall fit indices (CFI = 0.949, RMSEA = 0.032) support the model's adequacy, the data suggests that item refinement is necessary for improved validity. The chi-square value is 140.174 with 125 degrees of freedom (p = 0.167). Table 8 presents the Average Variance Extracted (AVE) values for the five-factor CFA model, which serve as indicators of convergent validity. The AVE values are as follows: Factor 1 = 0.301, Factor 2 = 0.317, Factor 3 = 0.483, Factor 4 = 0.400, and Factor 5 = 0.318. While Factors 3 and 4 approach the acceptable threshold of 0.50, the lower AVEs for Factors 1, 2, and 5 suggest weaker convergent validity, implying that insufficient variance explained by specific factors. The low AVE values likely result from high uniqueness and cross-loadings observed in the exploratory factor analysis (EFA), along with non-significant loadings in Factor 5. Despite the good CFA fit indices (χ^2 = 140.174, df = 125, p = 0.167; CFI = 0.949; RMSEA = 0.032). Although cross-loading items (e.g., Item 6 on Factors 1 and 3) and low Average Variance Extracted (AVE) values (e.g., 0.301 for Factor 1, 0.317 for Factor 2) suggest potential overlap between dimensions like Understanding AI in Education and Pedagogical Alignment, and limited variance explained by some factors, but these issues do not critically undermine the AI Literacy Scale's validity given its strong overall fit indices (CFI = 0.949, RMSEA = 0.032) and alignment with the theoretical framework of ELT-specific AI literacy (Fornell & Larcker, 1981). Such challenges, which are common when developing scales for complex constructs like AI literacy, reflect the multidimensional interplay of technical, pedagogical, and ethical competencies in ELT and can be addressed through targeted item refinement and validation with larger, more diverse samples. Ultimately, this scale offers an actionable tool for assessing and enhancing English language teachers' readiness to integrate AI effectively, notwithstanding these minor limitations.

Table 8. Average variance extracted

Factor	AVE
Factor 1	0.301
Factor 2	0.317
Factor 3	0.483
Factor 4	0.400
Factor 5	0.318

In this study, a comprehensive evaluation of a 33-item scale was conducted which resulted in a refined five-factor model that showcases credible quality. The initial exploratory factor analysis (EFA) confirmed the dataset's suitability for factor analysis, as evidenced by a favorable Kaiser-Meyer-Olkin (KMO) measure, which indicates adequate sampling adequacy, and Bartlett's test of sphericity, which demonstrated significant relationships among the variables. Following the EFA, a confirmatory factor analysis (CFA) was performed to validate the identified factor structure. The CFA results confirmed the five-factor model, achieving a non-significant Chi-square value (χ^2 = 140.174, p = 0.167), which suggests no significant discrepancy between the observed data patterns and the model predictions. Additionally, a Comparative Fit Index (CFI) of 0.949 indicates a strong fit, as values above 0.90 are typically considered acceptable in social science research. These findings underscore the instrument's ability to effectively capture the intended constructs, demonstrating its robustness and reliability. The model's structure aligns closely with theoretical expectations, bolstering confidence in its practical application. Furthermore, the five-factor model reflects a well-structured approach, allowing for a nuanced understanding of the dimensions being measured. This solid foundation lays the groundwork for future research, thereby enhancing our ability to explore related constructs and their implications in various contexts. This study highlights the quality and validity of the revised scale, making it a valuable tool for researchers aiming to assess the relevant dimensions effectively. The rigorous analytical methods employed throughout the study ensure that the findings are both reliable and applicable, supporting the scale's use in further investigations and practical applications.

While the results of the EFA and CFA support the validity of the five-factor AI Literacy Scale, there are potential limitations related to the data and methodological processes. The KMO Measure of Sampling Adequacy yielded a value of 0.643, classified as mediocre (Kaiser, 1974). This indicates moderate intercorrelations among certain variables, suggesting that some items may have weaker shared variance. This could limit the robustness of the factor structure and therefore necessitates cautious interpretation of the extracted factors.

The researchers acknowledge that a higher KMO value (e.g., ≥ 0.80) would provide stronger evidence of data suitability for factor analysis. However, this mediocre KMO value may reflect the complexity of capturing AI literacy's multidimensional nature in English Language Teaching (ELT) contexts. Furthermore, the expert review process in Phase 2 is recognized as a crucial stage in scale development. Although the researchers were rigorous in recruitment, achieving inter-rater reliability of 0.75, potential biases may still exist. The qualifications of reviewers in applied linguistics, educational technology, and psychology could introduce subjective influences that may not fully represent the diverse global ELT landscape. The researchers are aware of this issue and took steps to standardize procedures and share information consistently to minimize the influence of irrelevant variables. Although the five factors identified in this study were derived inductively through exploratory and confirmatory factor analyses, their conceptual configuration aligns closely with established theoretical models of AI literacy. In particular, the dimensions of Understanding AI and Technical Proficiency correspond to the cognitive and functional components proposed in Long and Magerko's (2020) framework, emphasizing the knowledge and practical engagement necessary for interacting meaningfully with AI systems. The factors of Pedagogical Alignment and Ethical Awareness extend these foundational dimensions into the domain of educational practice, reflecting teachers' capacity to integrate AI tools responsibly and pedagogically within English language teaching contexts. The fifth factor, AI for Feedback and Assessment, represents a distinctive, context-specific contribution, highlighting how AI can support formative assessment and learner feedback—an aspect that has received limited attention in broader AI literacy frameworks such as the UNESCO (2024) AI Competency Framework. Collectively, these alignments suggest that while the scale was empirically derived, it is theoretically grounded and contributes to extending existing conceptualizations of AI literacy into language education.

4. Discussion and Conclusion

The development and validation of the AI Literacy Scale for English Language teachers represent a significant advancement in addressing the urgent need for tailored frameworks that support teachers in integrating artificial intelligence into English language teaching. The findings confirm the scale's psychometric robustness, featuring a well-structured five-factor model—Understanding AI in Education, Proficiency in Using AI, Pedagogical Alignment with Teaching, Ethical Awareness in AI, and AI for Feedback and Assessment. This model was validated through confirmatory factor analysis (CFA) with strong fit indices (χ^2 = 140.174, df = 125, p = 0.167; CFI = 0.949; RMSEA = 0.032), demonstrating its effectiveness as a practical tool for assessing teachers' readiness to navigate the complexities of AI integration in ELT contexts.

This scale addresses a critical gap identified in the literature, where existing AI literacy frameworks, such as those by Allen and Kendeou (2024) and Annapureddy et al. (2024), often lack specificity for the linguistic, cultural, and pedagogical demands of ELT. By foregrounding on these dimensions, the scale provides teachers with a framework that is both relevant and actionable, strengthening teachers' capacity to integrate AI in ways that are ethical, pedagogically sound, and centered on learner needs (Mahmoudi-Dehaki, & Nasr-Esfahani, 2025). The study's mixed-methods approach, encompassing item generation, content validation, pilot testing, and factor analyses, ensured methodological rigor and reliability. The iterative refinement process, guided by expert feedback and empirical testing, aligns with best practices in scale development (DeVellis, 2016). While some challenges were noted—such as the KMO value (0.643) and low Average Variance Extracted (AVE) for certain factors—these findings highlight the potential for future refinements. Specifically, focusing on revising or replacing items with low loadings may strengthen the scale's convergent validity (Khodi et al., 2024).

The scale's five-factor structure reflects the multifaceted nature of AI literacy in ELT. Emphasizing Ethical Awareness aligns with concerns raised by the British Council (Edmett et al., 2023), where a substantial proportion of teachers expressed unease about data privacy and algorithmic transparency. Similarly, the emphasis on Pedagogical Alignment addresses the need for AI integration to enhance, rather than replace, human-centered instruction, reinforcing the importance of reflective practice (Ng et al., 2021). Furthermore, the inclusion of AI for Feedback and Assessment acknowledges the growing use use of AI tools like Grammarly and Duolingo, while also highlighting risks associated with over-reliance on standardized outputs that may marginalize non-standard dialects and diverse linguistic identities (Lachini, 2024).

Practically, this scale has significant implications for teacher education and professional development. The lack of AI-specific training reported by Hashemi and Kew (2021) underscores the urgency of equipping teachers with both technical skills and critical perspectives. The scale can inform the design of targeted workshops, enabling teachers to critically evaluate AI tools and align them with pedagogical goals. Its focus on cultural and linguistic nuances makes it relevant for diverse ELT contexts, thereby addressing gaps in frameworks like the ED-AI Lit model (Allen & Kendeou, 2024), which often overlook important aspects like code-switching and dialect diversity.

Although cross-loading items (e.g., Item 6 on Factors 1 and 3) and low Average Variance Extracted (AVE) values (e.g., 0.301 for Factor 1, 0.317 for Factor 2) suggest minor conceptual overlap among dimensions like Understanding AI in Education and Pedagogical Alignment, these do not critically undermine the AI Literacy Scale's validity, given its strong fit indices (CFI = 0.949, RMSEA = 0.032) and alignment with the ELT-specific theoretical framework (Fornell & Larcker, 1981). These psychometric nuances, common in scales for complex constructs, can be addressed through item refinement and broader validation. In low-resource settings, teachers face practical challenges such as unreliable internet and outdated devices, limiting access to AI tools like Grammarly or Duolingo (Hashemi & Kew, 2021). Limited training opportunities and large class sizes further impede effective AI integration, especially in rural ELT contexts (Mehdaoui, 2024). These barriers highlight the need for accessible training and infrastructure support to ensure equitable AI adoption. Notwithstanding these challenges, the scale offers an acceptable framework for assessing and fostering AI literacy, guiding targeted interventions to empower ELT teachers globally.

To enhance the scale's applicability across varied educational contexts, future validation should prioritize testing in diverse settings (Khodi, Khezerlou, & Sahraei, 2021). This should include non-native English-speaking regions and under-resourced educational systems to ensure cultural and contextual relevance (Rezvani & Izadi, 2025). By recruiting larger, more heterogeneous samples, researchers can strengthen statistical power and address potential cultural biases in the current participant pool. Additionally, longitudinal studies are essential to assess how AI literacy evolves with training over time (Zhang & Zhang, 2024). Additionally, cross-cultural psychometric analyses, such as invariance testing, can further confirm the scale's consistency across global ELT contexts, ensuring its adaptability to diverse linguistic and pedagogical environments (Ng et al., 2024). Future research should also explore AI literacy training programs tailored specifically to ELT teachers, examining their impact on technical proficiency and ethical decision-making (Mahmoudi-Dehaki & Nasr-Esfahani, 2025). Investigating how teachers' beliefs, such as skepticism regarding AI's pedagogical value or concerns over data privacy, shape AI adoption can further refine the scale's application and inform targeted interventions (British Council, 2023). These steps will solidify the scale's generalizability and utility in fostering equitable AI integration worldwide.

Limitations of the study include the potential for cultural bias in participant selection, given the diverse global contexts of ELT. Future research should aim to validate the scale across larger and more varied populations, including non-native English-speaking teachers and those in under-resourced settings. Additionally, longitudinal studies could explore how AI literacy evolves with training and classroom experience.

In conclusion, this AI Literacy Scale offers a robust, ELT-specific tool designed to prepare teachers for the evolving landscape of language instruction. By fostering technical proficiency, ethical awareness, and pedagogical alignment, the scale empowers teachers to harness the potential of AI while preserving the human essence of teaching. This study contributes to the broader discourse on educational innovation, providing a foundation for future research and policy to ensure equitable and reflective AI integration in ELT. The AI Literacy Scale, developed through a rigorous four-phase mixed-methods approach, not only addresses the critical gap in teacher preparedness but also sets a high standard for future instrument development in educational research. With strong psychometric properties (e.g. CFI = 0.949, RMSEA = 0.032), this scale provides a reliable and valid tool to assess teachers' readiness to integrate AI, focusing on five key dimensions. Given that a significant percentage of teachers feel unprepared to effectively use AI tools (British Council, 2023), this structured assessment can guide targeted professional development, ensuring that educators are equipped to utilize AI effectively, ethically, and equitably. Moreover, the scale's adaptability to other educational contexts underscores its broader significance, promoting a reflective and equitable approach to AI integration across curricula. This study is not without limitations. The sample size was relatively modest and drawn from a limited geographic and institutional context, which may constrain the generalizability of the findings. Additionally, although the psychometric indicators were within acceptable ranges, some indices such as the KMO and AVE values suggest that further refinement of the scale is warranted. The reliance on self-reported data might also introduce response bias. Future research should validate the scale with larger and more diverse samples, employ longitudinal designs, and explore measurement invariance across different educational settings to strengthen the instrument's general applicability and theoretical robustness.

Acknowledgements

We sincerely thank our colleagues for their invaluable assistance in conducting the study and critiquing the manuscript.

Authors' contributions

All authors contributed significantly to the study. Dr. Ali Khodi supervised the entire research process and ensured methodological rigor. Kuysin Tukhtaeva and Nafisa Ochilova assisted in shaping the research framework. Chen Kaizi focused on data interpretation and analysis, while Nodira Rakhimova contributed to the study's execution. Yulduz Khasanova assisted with data analysis and reporting. Finally, Zebiniso Tuychieva was instrumental in revising the manuscript for clarity and coherence. Each author's contributions were essential to the successful completion of this study.

Funding

The study was funded by the UCSI University through Research Excellence & Innovation Grant (REIG-FOSSLA-2024/050).

Competing Interests

The authors declare no conflicts of interest.

Informed consent

Obtained.

Ethics Approval

This study adheres to the guidelines of the Publication Ethics Committee of the Sciedu Press. The journal's policies comply with the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data Availability Statement

The data supporting the findings of this study are available upon request from the corresponding author, with justification. The data are not publicly available due to privacy and ethical restrictions.

Data sharing statement

No additional data are available.

AI Disclosure

During manuscript preparation, standard digital writing-support tools were used to refine clarity and readability. The scholarly arguments, interpretations, and final text remain the author's responsibility.

References

Ahmadi Fatalaki, J., Saeedi, Z., Sheikhzadeh Marand, E., & Nasiri Firuz, A. R. (2025). Artificial intelligence in education and EFL teachers' attitudes: scale development and validation. Innovation in Language Learning and Teaching, 1-17.

Alavi, S. M., Karami, H., & Khodi, A. (2021). Examination of factorial structure of Iranian English language proficiency test: An IRT analysis of Konkur examination. *Current Psychology*. Advance online publication. https://doi.org/10.1007/s12144-021-01922-1

Al-khresheh, M. H. (2024). Bridging technology and pedagogy from a global lens: Teachers' perspectives on integrating ChatGPT in English language teaching. *Computers and Education: Artificial Intelligence, 6*, Article 100218. https://doi.org/10.1016/j.caeai.2024.100218

Allen, L. K., & Kendeou, P. (2024). 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

Almashour, M., Aldamen, H. A. K., & Jarrah, M. (2025). "They know AI, but they also know us": Student perceptions of EFL teacher identity in AI-enhanced classrooms in Jordan. *Frontiers in Education, 10*, Article 1611147. https://doi.org/10.3389/feduc.2025.1611147

Annapureddy, R., Fornaroli, A., & Gatica-Perez, D. (2024). Generative AI literacy: Twelve defining competencies. *arXiv*. Retrieved from https://arxiv.org/abs/2412.12107

Babanoğlu, M. P., Karataş, T. Ö., & Dündar, E. (2025). Envisioning the future of AI-assisted EFL teaching and learning: Conceptual representations of prospective teachers. *SAGE Open, 15*(2). https://doi.org/10.1177/21582440251341590

Baskara, F. R. (2025). Conceptualizing digital literacy for the AI era: A framework for preparing students in an AI-driven world. *Data and Metadata, 4*(530), 1-13. https://doi.org/10.56294/dm2025530

Butarbutar, R., & Ranta, L. (2024, July 31). Artificial intelligence for language learning and teaching: A narrative literature study. SSRN. https://doi.org/10.2139/ssrn.5276494

Chen, X., Zou, D., Cheng, G., & Xie, H. (2020). Detecting latent topics and trends in educational technologies over four decades using structural topic modeling: A retrospective of all volumes of Computers & Education. *Computers & Education, 151*, Article 103855. https://doi.org/10.1016/j.compedu.2020.103855

Cukurova, M., & Miao, F. (2024). *AI competency framework for teachers*. UNESCO Publishing.

DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications* (5th ed.). SAGE Publications.

Ding, A. C. E., Shi, L., Yang, H., & Choi, I. (2024). Enhancing teacher AI literacy and integration through different types of cases in teacher professional development. *Computers and Education Open, 6*, Article 100178. https://doi.org/10.1016/j.caeo.2024.100178

Edmett, A., Ichaporia, N., Crompton, H., & Crichton, R. (2023). *Artificial intelligence and English language teaching: Preparing for the future*. British Council.

Edmett, A., Ichaporia, N., Crompton, H., & Crichton, R. (2024). The future of AI in English language teaching and learning. British Educational Research Association (BERA).

Gouseti, A., James, F., Fallin, L., & Burden, K. (2025). The ethics of using AI in K-12 education: A systematic literature review. *Technology, Pedagogy and Education, 34*(2), 161-182. https://doi.org/10.1080/1475939X.2024.2428601

Hao, Z., Fang, F., & Peng, J. E. (2024). The integration of AI technology and critical thinking in English major education in China: Opportunities, challenges, and future prospects. *Digital Applied Linguistics, 1*, 2256. https://doi.org/10.29140/dal.v1.2256

Hashemi, A., & Kew, S. N. (2021). The barriers to the use of ICT in English language teaching: A systematic literature review. *Bilgi ve İletişim Teknolojileri Dergisi, 3*(1), 77-88.

Jiang, R. (2022). How does artificial intelligence empower EFL teaching and learning nowadays? A review on artificial intelligence in the EFL context. *Frontiers in Psychology, 13*, Article 1049401. https://doi.org/10.3389/fpsyg.2022.1049401

Khodi, A., Khezerlou, H., & Sahraei, H. (2022). Dependability and utility of using e-portfolios in assessing EFL learners' speaking proficiency. *Computer Assisted Language Learning*. Advance online publication. https://doi.org/10.1080/09588221.2022.2093379

Khodi, A., Ponniah, L. S., Farrokhi, A. H., & Sadeghi, F. (2024). Test review of Iranian English language proficiency test: MSRT test. *Language Testing in Asia, 14*(1), Article 4. https://doi.org/10.1186/s40468-023-00270-0

Kline, R. B. (2023). *Principles and practice of structural equation modeling*. Guilford Publications.

Lachini, K. (2024). Rethinking language assessment in the age of AI: A call for a paradigm shift. https://doi.org/10.2139/ssrn.5043405

Long, D., & Magerko, B. (2020, April). 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

Mahmoudi-Dehaki, M., & Nasr-Esfahani, N. (2025). Utilising GenAI to create a culturally responsive EFL curriculum for pre-teen learners in the MENA region. *Education 3-13*. Advance online publication. https://doi.org/10.1080/03004279.2025.2502016

Markus, A., Carolus, A., & Wienrich, C. (2025). Objective measurement of AI literacy: Development and validation of the AI competency objective scale (AICOS). *arXiv*. https://doi.org/10.1016/j.caeai.2025.100485

Mehdaoui, A. (2024). Unveiling barriers and challenges of AI technology integration in education: Assessing teachers' perceptions, readiness and anticipated resistance. *Futurity Education, 4*(4), 95-108. https://doi.org/10.57125/FED.2024.12.25.06

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence, 2*, Article 100041. https://doi.org/10.1016/j.caeai.2021.100041

Ng, D. T. K., Su, J., Leung, J. K. L., & Chu, S. K. W. (2024). Artificial intelligence (AI) literacy education in secondary schools: A review. *Interactive Learning Environments, 32*(10), 6204-6224. https://doi.org/10.1080/10494820.2023.2255228

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

OECD. (2021). *AI and the future of skills, Volume 1: Capabilities and assessments* (Educational Research and Innovation). OECD Publishing. https://doi.org/10.1787/5ee71f34-en

Rezvani, R., & Izadi, A. (2025). An innovative standards-based post-graduate computer-assisted-language-learning program in a women's university in Iran. *Journal of English Language Teaching and Learning, 17*(35), 301-324. https://doi.org/10.22034/elt.2025.65338.2739

Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology, 14*, Article 1260843. https://doi.org/10.3389/fpsyg.2023.1260843

Sun, J., Motevalli, S., Nee Chan, N., & Khodi, A. (2024). Implementation of self-regulated learning writing module: Amplifying motivation and mitigating anxiety among EFL learners. *Forum for Linguistic Studies, 6*(3), 89-109. https://doi.org/10.30564/fls.v6i3.6600

Widagsa, R., Senom, F., & Hutagalung, F. D. (2025). Teaching out-of-field: Exploring the non-specialist English language teachers' knowledge base and practices. *Cogent Education, 12*(1), Article 2514914. https://doi.org/10.1080/2331186X.2025.2514914

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

Zhang, J., & Zhang, Z. (2024). AI in teacher education: Unlocking new dimensions in teaching support, inclusive learning, and digital literacy. *Journal of Computer Assisted Learning, 40*(4), 1871-1885. https://doi.org/10.1111/jcal.12988

Zhang, Y., Zhang, M., Wu, L., & Li, J. (2025). Digital transition framework for higher education in AI-assisted engineering teaching: Challenge, strategy, and initiatives in China. *Science & Education, 34*(2), 933-954. https://doi.org/10.1007/s11191-024-00575-3

Zhou, X., Li, Y., Chai, C. S., & Chiu, T. K. (2025). Defining, enhancing, and assessing artificial intelligence literacy and competency in K-12 education from a systematic review. *Interactive Learning Environments*. Advance online publication. https://doi.org/10.1080/10494820.2025.2487538

Zou, D., Xie, H., & Kohnke, L. (2025). Navigating the future: Establishing a framework for educators' pedagogic artificial intelligence competence. *European Journal of Education, 60*(2), Article e70117. https://doi.org/10.1111/ejed.70117

Appendix A

Survey

#	Dimension	Statement
1	**Awareness & Understanding of AI in Language Teaching Context**	I can clearly explain to students how AI tools can enhance their English learning.
2		I easily understand how AI supports different language skills.
3		I like to introduce AI tools in class to support student tasks.
4		I farily discuss the limitations of AI with my students.
5		I stay informed and updated about new AI developments in education.
6		I help students understand how AI makes decisions in feedback and scoring.
7		I try to promote digital awareness by teaching how AI tools differ from traditional ones.
8	**Confidence & Proficiency in Using AI in Class**	I can confidently use AI tools during my class activities.
9		I prefer to create engaging classroom activities that involve AI use.
10		I can quicly troubleshoot basic AI-related technical problems in class.
11		I apply knowledge from AI-focused training or professional development to my lessons and instruction.
12		I try my best to explore and test new AI tools to improve learning.
13		I really feel comfortable integrating multiple AI tools my instruction.
14		I can provide guidance to students on how to navigate AI tools effectively.
15	**Pedagogical Integration of AI Tools**	I easily evaluate if an AI tool supports specific learning goals.
16		I can adapt AI activities based on learner needs and levels.
17		I align AI tasks with curriculum outcomes and learning objectives.
18		I can select AI resources appropriate for student proficiency.
19		I fairly reflect on how AI changes my teaching strategies.
20		I blend AI with traditional teaching methods in balanced ways.
21		I modify AI-generated content to better suit my learners' needs and preferences.
22	**Ethical, Inclusive, and Responsible AI Use**	I really consider ethical issues when selecting AI tools.
23		I strongly protect students' data when using AI.
24		I ensure the AI tools I use are inclusive, accessible and sustainable.
25		I guide students in thinking critically about AI-generated content.
26		I usually check AI tools for possible biases or language inaccuracies.
27		I promote fairness by discussing AI's role in assessments and grading.
28	**AI for Feedback and Assessment**	I use AI to provide personalized feedback to students.
29		I use AI to track student progress and adjust instruction.
30		I conduct assessments using AI tools.
31		I interpret AI-generated analytics to inform my teaching.
32		I incorporate AI-based formative assessments into lessons.
33		I encourage students to reflect on AI feedback in their learning process.