Artificial Intelligence Education Studies
Volume 1, Issue 3, 2025, Page 56-69
https://doi.org/10.6914/aiese.010305

Artificial Intelligence Education Studies

Standardized Assessment of Artificial Intelligence Literacy: Development and Validation of the Multidimensional AI Literacy Competency Scale (MAIL-CS)

Qiang SUN
North China University of Technology

Abstract

Generative AI's rapid diffusion demands precise, up-to-date measures of AI literacy. This study develops and validates the Multidimensional AI Literacy Competency Scale (MAIL-CS), designed specifically for the GenAI era. Using a large sample of Chinese university students ($N = 850$) and a split-sample design, we conducted EFA and CFA to establish a robust four-factor structure—Foundational Knowledge & Ethics, Operational Skills, Critical Evaluation, and Application & Innovation. The best-fitting model showed strong fit indices, and the 32-item scale demonstrated high internal consistency (Cronbach's α and McDonald's $\omega \geq .82$ subscales; $\alpha = 0.91$, $\omega = 0.92$ total). Convergent validity was supported by positive correlations with digital literacy and critical thinking; discriminant validity was evidenced by negligible relations with Big Five traits. MAIL-CS offers educators, researchers, and policymakers a reliable instrument to diagnose competency gaps, evaluate interventions, and inform curriculum and strategy. Validation in a non-Western context provides a foundation for cross-cultural assessment and future invariance testing.

Keywords AI literacy; Generative AI; Scale development; Psychometrics; Higher education

1 Introduction

1.1 The Imperative for AI Literacy in the Age of Generative AI

The 21st century has witnessed a series of technological revolutions, yet few have been as swift, pervasive, and transformative as the advent of Artificial Intelligence (AI), particularly the recent surge in Generative AI (GenAI). Since becoming widely accessible in late 2022, tools such as ChatGPT, DALL-E, and Deepseek have moved from the periphery of technological curiosity to the center of global discourse, demonstrating an unprecedented capacity to generate novel and seemingly human-like content, including text, images, and code. This rapid popularization has created a critical inflection point, demanding a fundamental re-evaluation of the competencies required to function effectively in a digitally saturated world.

The unique characteristics of GenAI distinguish it sharply from previous iterations of AI and digital technology, thereby necessitating a more specialized form of literacy. Unlike traditional analytical AI, which primarily focuses on classification and prediction, GenAI is oriented toward creation and innovation. Its operational logic is probabilistic, not deterministic, meaning it generates content by predicting the most likely next token or pixel based on patterns in its vast training data. This inherent stochasticity gives rise to several defining features that users must comprehend. First is the potential for generating factually incorrect or nonsensical information, colloquially known as "hallucinations". Second is its capacity to inherit and amplify biases present in its training data, leading to outputs that can perpetuate harmful stereotypes. Third is its ability to mimic human creativity and communication styles with remarkable fidelity, a capability that often leads users to over-attribute intelligence, understanding, or even consciousness to the system—a psychological phenomenon known as the "Eliza effect". The uncritical acceptance of GenAI outputs, driven by this effect and a general lack of understanding of its limitations, can lead to the propagation of misinformation and compromised decision-making.

These distinct challenges underscore that the core issue facing society is no longer merely one of technological access, but one of competent and critical engagement. While GenAI tools are now broadly available, a significant gap has emerged between users' perceived proficiency and their actual ability to interact with these systems effectively, critically, and ethically. This disparity signifies the emergence of a new and more insidious "digital divide". This is not a divide between the technological "haves" and "have-nots," but between those who possess the nuanced literacy to command, question, and contextualize AI, and those who are, in effect, susceptible to the technology's pitfalls and unable to harness its full potential. This emerging competency gap has profound implications for social equity. Without deliberate action to cultivate widespread AI literacy, GenAI risks becoming a force that widens, rather than narrows, educational and economic inequalities. The development of widespread AI literacy is therefore not simply an educational desideratum; it is an urgent societal imperative for fostering informed citizenship, ensuring workforce preparedness, and promoting equity in an AI-enhanced future.

1.2 The Critical Gap in Standardized Assessment

Despite the clear and pressing need to cultivate AI literacy, efforts within education, industry, and policy have been profoundly hampered by a fundamental challenge: the absence of robust, standardized, and psychometrically sound instruments for its assessment. While the importance of AI literacy has been a topic of growing academic and public discussion, the development of tools to measure it has lagged significantly. As recently as 2024, scholars have noted that validated, objective tests for AI literacy have, until now, not been developed.

This assessment gap has several critical dimensions. First, many existing instruments rely on subjective self-report measures, asking individuals to rate their own perceived skills or knowledge. Such measures are notoriously susceptible to cognitive biases, most notably the Dunning-Kruger effect, where individuals with lower competence tend to overestimate their abilities. This reliance on self-perception obscures a more reliable and objective measure of actual literacy levels, making it difficult to accurately gauge the true state of AI literacy within a population. Second, the field has struggled with a lack of a universally accepted definition of AI literacy, leading to a proliferation of frameworks with varying scopes and dimensions. This conceptual ambiguity makes it challenging to compare findings across studies and to build a cumulative body of knowledge. A clear, operationalized definition of the construct is a prerequisite for valid measurement.

This state of affairs points to a foundational principle for educational progress: effective, large-scale educational reform is contingent upon the availability of valid assessment tools. The development of a standardized scale is not merely a technical exercise for psychometricians; it is an enabling condition for the entire ecosystem of AI education. Without a reliable instrument, educators cannot accurately diagnose the baseline knowledge and skills of their students, track their learning progress over time, or systematically evaluate the effectiveness of different pedagogical interventions. Policymakers are left without the data needed to assess the AI readiness of the workforce, identify specific skill gaps that require targeted investment, or measure the impact of national education strategies. In essence, the absence of a "gold standard" assessment tool means that the field of AI education is operating without a reliable compass, unable to move from well-intentioned but anecdotal efforts to a paradigm of data-driven, evidence-based practice.

1.3 The Present Study: Aims and Research Questions

The present study was conceived to directly address this critical gap in the literature and to provide a foundational tool for the advancement of AI education and research. The primary purpose of this research is to develop and validate a comprehensive, multidimensional assessment instrument—the Multidimensional AI Literacy Competency Scale (MAIL-CS)—based on a novel theoretical framework meticulously designed to be sensitive to the specific demands of the GenAI era. To achieve this overarching goal, this study seeks to answer the following research questions: RQ1: In the era of Generative AI, what are the core, empirically distinguishable dimensions that constitute AI literacy for university students? RQ2: Does the newly developed MAIL-CS demonstrate robust psychometric properties, including reliability, structural validity (via competing model comparison), and construct validity (convergent and discriminant)?

The potential contributions of this research are threefold. At a theoretical level, this study aims to advance the conceptualization of AI literacy by offering a more nuanced and structured model that explicitly accounts for the transformative impact of GenAI. At a practical level, the validated MAIL-CS will provide educators, instructional designers, and trainers with a ready-to-use tool to assess AI literacy, guide curriculum development, and evaluate the effectiveness of educational programs. At a policy level, the instrument offers a means for institutions and governments to conduct large-scale assessments, providing the empirical data necessary to inform evidence-based standards, strategic planning, and resource allocation in AI education.

2 Literature Review and Theoretical Framework

2.1 The Evolving Construct of AI Literacy: From Foundational Frameworks to GenAI-Specific Competencies

The conceptualization of AI literacy has evolved in tandem with the technology itself. Early and foundational work by Long and Magerko (2020) provided a seminal framework that has been widely cited and adopted as a theoretical baseline. They defined AI literacy as "a set of competencies that enables individuals to critically evaluate AI technologies; communicate and collaborate effectively with AI; and use AI as a tool online, at home, and in the workplace". Their framework is organized around five key conceptual themes posed as questions, providing a comprehensive map of the AI literacy landscape as it was understood in the pre-GenAI era.

Building on this foundation, other scholars and organizations have proposed complementary frameworks, often with a more explicit educational focus. Ng et al. (2021), for instance, constructed a model centered on a progression of cognitive domains: recognizing and understanding AI, using and applying AI, evaluating and creating with AI, and understanding AI ethics. Similarly, the Digital Promise framework organizes competencies around the pillars of understanding, evaluating, and using AI, all grounded in core values of human judgment and centering justice. International organizations have also contributed to this discourse. UNESCO, for example, has developed comprehensive competency frameworks for both students and teachers that emphasize a "human-centered mindset" as a core principle, advocating for the assertion of human agency and accountability in relation to AI systems. These frameworks collectively established a consensus around the multidimensional nature of AI literacy, consistently highlighting the importance of knowledge, practical skills, critical evaluation, and ethical considerations.

However, the explosive rise of GenAI has exposed the limitations of these foundational frameworks. Developed when AI was largely the domain of specialists or embedded invisibly in analytical systems, they do not fully capture the new set of competencies required for direct, interactive, and creative engagement with generative models. The unique characteristics of GenAI—its capacity for novel content generation, its probabilistic nature, and its complex, conversational interaction patterns—demand a shift in focus. For example, the skill of "prompt engineering"—the art and science of crafting effective inputs to guide GenAI outputs—is a novel operational competency with no direct parallel in earlier models. Likewise, the need to critically evaluate outputs for "hallucinations" is a GenAI-specific form of critical thinking. This has led to calls for a "Critical AI Literacy," which goes beyond functional skills to incorporate a deeper understanding of the social, political, and economic power structures that shape AI development and deployment, promoting an awareness of issues like cognitive injustice, data justice, and educational equity. This critical perspective treats AI not merely as a neutral tool but as a socio-technical system with embedded values and biases that can reinforce systemic inequalities if not interrogated.

2.2 The Current Landscape of AI Literacy Assessment: A Comparative Analysis

To address the growing need for AI literacy assessment, several measurement tools have emerged in recent years. However, these instruments vary significantly in their target populations, theoretical underpinnings, and coverage of competencies specific to the GenAI era. A systematic comparison of these existing scales is crucial for justifying the development of a new instrument like the MAIL-CS. Table 1 provides a comparative analysis of representative AI literacy scales, highlighting the unique contributions of the present study.

The analysis presented in Table 1 makes it clear that while valuable instruments such as SAIL4ALL, AILS-CCS, and AILST exist, a distinct gap remains in the literature. There is a lack of a psychometrically validated scale specifically designed for the higher education context, with a theoretical framework and measurement items fundamentally constructed around the unique challenges and opportunities of GenAI. For instance, the SAIL4ALL scale is based on the earlier Long and Magerko (2020) framework, and its dimensional structure does not fully capture the dynamic nature of human-AI interaction in the GenAI era. Although the AILS-CCS was validated with Chinese university students, its dimensions (Awareness, Usage, Evaluation, Ethics) are relatively broad, and it was developed during the initial phase of GenAI's widespread adoption, thus having limited coverage of emerging competencies like prompt engineering and co-creation. The AILST is focused on the teacher population, and its "Application & Innovation" dimension, while forward-looking, may not be fully applicable to the student context. The MAIL-CS was designed precisely to fill this identified gap by providing a comprehensive, contemporary, and contextually appropriate tool for assessing AI literacy among university students.

Table 1: Comparative Analysis of Existing AI Literacy Assessment Scales and the MAIL-CS

Scale (Year)	Target Population	Core Dimensions/Factors	GenAI-Specific Competencies	Theoretical Basis	Validation Context
MAIL-CS (This Study)	University Students	1. Foundational Knowledge & Ethics 2. Operational Skills 3. Critical Evaluation 4. Application & Innovation	Explicitly included (e.g., prompt optimization, hallucination identification, human-AI co-creation)	Synthesizes existing frameworks and extends them for the GenAI era	China
SAIL4ALL (2024)	General Adult Population	1. What is AI? 2. What can AI do? 3. How does AI work? 4. How should AI be used?	Not explicitly emphasized	Long & Magerko (2020)	Multinational (Germany, UK, USA)
AILS-CCS (2024)	Chinese University Students	1. Awareness 2. Usage 3. Evaluation 4. Ethics	Limited coverage	Synthesis of literature	China
AILST (2025)	Teachers	1. AI perception 2. Knowledge & skills 3. Application & innovation 4. Ethics	Includes "Application & Innovation," but less specific to GenAI than MAIL-CS	Synthesis of literature	Not specified (likely China)

2.3 Proposing the Multidimensional AI Literacy Competency Framework (MAIL-CF)

Based on a synthesis of the existing literature and a critical analysis of the specific competencies necessitated by the GenAI revolution, this study proposes a new theoretical model: the Multidimensional AI Literacy Competency Framework (MAIL-CF). This framework is designed to be comprehensive, structured, and contemporary, providing a robust theoretical foundation for the development of the MAIL-CS assessment scale. The MAIL-CF is composed of five core dimensions, each encompassing a set of specific, operationalized sub-competencies. This framework represents a synthesis of three evolving perspectives on AI: AI as a technical tool, AI as an agentic collaborator, and AI as a socio-technical system. By integrating these viewpoints, the MAIL-CF provides a holistic model that accounts for the foundational, interactive, and critical dimensions of literacy required in the current technological landscape.

First, AI Foundational Knowledge represents the conceptual bedrock of AI literacy. It aligns with the "Know & Understand" domains identified as essential in nearly every major AI literacy framework. It encompasses an understanding of core AI concepts, the distinction between different types of AI (e.g., analytical vs. generative), the critical role of data in training and operating AI systems, and an awareness of AI's principal applications and societal impact.

Second, AI Operational Skills & Application focuses on the practical, functional ability to use AI tools to accomplish goals. It corresponds to the "Use & Apply" dimensions prevalent in the literature. Critically, in the context of the MAIL-CF, this dimension is updated to explicitly include competencies vital for GenAI, most notably prompt optimization (or prompt engineering), which is the iterative process of designing and refining inputs to elicit desired outputs from generative models.

Third, AI Critical Evaluation & Thinking addresses the crucial higher-order cognitive skills needed to assess and contextualize AI systems and their outputs. While critical thinking is a universal component of literacy, this dimension is specifically tailored to the challenges of AI. For GenAI, this includes the ability to evaluate the accuracy and reliability of generated information, to actively identify potential "hallucinations," to recognize and analyze algorithmic and data-driven biases, and to understand the fundamental limitations of AI systems (e.g., their lack of genuine consciousness or theory of mind). This dimension is closely aligned with the tenets of "Critical AI Literacy," which advocates for a reflective stance on the socio-political impacts of AI, moving beyond mere functional use.

Fourth, AI Ethics & Responsible Use encapsulates the normative and socio-ethical aspects of AI literacy. It is a cornerstone of modern frameworks, reflecting a growing societal concern with the responsible governance of AI. Sub-competencies include understanding core ethical principles such as fairness, accountability, transparency, and privacy (FAT-P); identifying ethical dilemmas in AI applications (e.g., job displacement, surveillance); using AI in a manner consistent with academic integrity and copyright law; and advocating for the equitable and just use of AI technologies.

Fifth, AI Creation & Innovation represents a higher level of AI literacy, reflecting the shift from AI as a tool for information consumption to AI as a collaborative partner in creation. This is a key innovation that distinguishes AI literacy from traditional digital literacy. It includes the ability to co-create novel content (e.g., text, images, code) with AI tools, a conceptual understanding of how AI models are built and trained, and the capacity to envision and explore innovative applications of AI to solve new problems or improve existing processes.

These five dimensions, along with their constituent sub-competencies and operationalized definitions, provide a transparent and structured blueprint that directly links the theoretical framework to the subsequent process of item development and instrument validation.

3 Instrument Development and Validation

The development and validation of the Multidimensional AI Literacy Competency Scale (MAIL-CS) followed a rigorous, multi-phase methodology consistent with best practices in educational and psychological measurement. The process was designed to ensure that the final instrument is not only theoretically grounded in the MAIL-CF but also psychometrically robust, with strong evidence for its content validity, structural validity, reliability, and construct validity. The process unfolded in three distinct phases: (1) item generation and content validity assessment, (2) a pilot study for initial item analysis and refinement, and (3) a large-scale main validation study.

3.1 Phase 1: Item Generation and Content Validity

3.1.1 Item Generation

The initial generation of the item pool was systematically guided by the operationalized definitions of the sub-competencies outlined in the MAIL-CF. This deductive approach ensures a strong and transparent link between the theoretical construct and its measurement. To capture the multifaceted nature of AI literacy, a mixed-format approach was employed. For dimensions related to declarative knowledge, such as AI Foundational Knowledge, traditional multiple-choice questions (MCQs) were developed. For dimensions assessing more complex, applied competencies, such as AI Operational Skills & Application, AI Critical Evaluation & Thinking, and AI Ethics & Responsible Use, scenario-based MCQs and items using a Likert-type response scale were created. Following established guidelines for scale development, an initial pool of items was generated that was substantially larger than the intended final length of the scale, with approximately 3 to 4 times the desired number of final items created for each sub-competency.

3.1.2 Content Validity Assessment

To establish the content validity of the initial item pool, a formal expert review process was conducted. A panel of seven experts was recruited, with selection criteria requiring a doctoral degree and demonstrated expertise in one or more of the following fields: artificial intelligence, educational technology, learning sciences, or psychometrics. The expert panel was asked to independently evaluate every item on two primary criteria using a 5-point Likert scale: (1) Clarity and (2) Relevance to its designated sub-competency. To supplement this qualitative and rating-based feedback with a quantitative, objective criterion, Lawshe's (1975) Content Validity Ratio (CVR) was calculated for each item. The CVR quantifies the degree of agreement among experts regarding the essentiality of an item using the formula:

$$CVR = \frac{n_e - \frac{N}{2}}{\frac{N}{2}}$$

where n_e is the number of experts who rated the item as "essential" and N is the total number of experts on the panel. Items that failed to meet the minimum critical CVR value for a panel of seven experts were flagged for either significant revision or elimination. Based on the combined quantitative (CVR, mean ratings) and qualitative (written comments) feedback from the expert panel, the initial item pool was revised, resulting in a refined set of items for the pilot study.

3.2 Phase 2: Pilot Study and Item Refinement

Following the expert review, a pilot study was conducted to empirically test the refined items with a sample representative of the target population. A sample of 120 undergraduate students from a comprehensive university in China participated in the pilot study by completing the questionnaire online. This sample size is considered adequate for conducting preliminary item analysis in the scale development phase. The data collected were subjected to a thorough item analysis. For the MCQ items, item difficulty (the p-value) and item discrimination (the point-biserial correlation coefficient) were calculated. For items using a Likert-type scale, descriptive statistics including the mean, standard deviation, skewness, and kurtosis were examined to identify items with potential issues such as floor or ceiling effects. Based on these statistical results, a final round of revisions was made, which involved rewording ambiguous items and removing the weakest ones. This process resulted in the final version of the MAIL-CS that was used in the main validation study.

3.3 Phase 3: Main Validation Study

3.3.1 Participants and Procedure

The main validation study was conducted with a large sample of university students from various higher education institutions across China. University students were selected as the target population because they are among the most active users of AI technologies, particularly GenAI tools, and the development of their AI literacy is of critical importance for their future academic and professional success. A combination of convenience sampling and snowball sampling strategies was employed to recruit participants. While non-probability sampling methods, these approaches were deemed appropriate and pragmatic for reaching a large, geographically dispersed student population in the absence of a national student registry, a common challenge in large-scale educational research. A final sample of 850 valid and complete responses was obtained for data analysis. This sample size is considered robust for conducting factor analysis, satisfying common rules of thumb such as a minimum subject-to-variable ratio of 10:1 and exceeding the minimum sample size recommendations for stable factor solutions in both EFA and CFA. Data were collected via a professional online survey platform between March 2025 and May 2025, with informed consent obtained from all participants.

3.3.2 Measures

In addition to the newly developed MAIL-CS, the survey battery for the main validation study included several established scales to assess the convergent and discriminant validity of the new instrument. The Multidimensional AI Literacy Competency Scale (MAIL-CS) is the focal instrument of the study, consisting of the items retained and refined following the content validation and pilot study phases. To assess convergent validity, a validated Digital Literacy Scale (DLS) was administered. It was hypothesized that AI literacy would be moderately to strongly positively correlated with general digital literacy. As the MAIL-CF posits AI Critical Evaluation & Thinking as a core dimension, a measure of general critical thinking disposition, the Critical Thinking Disposition Scale (CTDS), was included to further test convergent validity. It was expected that individuals with a stronger general disposition towards critical thinking would also score higher on the MAIL-CS. To establish discriminant validity, a short-form measure of the Big Five personality traits, the Big Five Inventory - Short (BFI-S), was included. It was hypothesized that correlations between the MAIL-CS and the BFI-S dimensions would be low and largely non-significant, providing evidence that the MAIL-CS is measuring its intended construct rather than general personality characteristics.

3.4 Data Quality and Analysis Strategy

3.4.1 Data Quality

To ensure the integrity of the data collected via the online survey, several quality control measures were implemented. First, the data were screened for missing values and outliers. Second, the survey included embedded attention check questions (also known as instructional manipulation checks) at various points to identify and filter out responses from inattentive participants. For example, a question might instruct the respondent to select a specific option (e.g., "To show you are paying attention, please select 'Agree' "). Third, response time data were analyzed to flag and remove "speeders"—respondents who completed the survey in an unrealistically short amount of time, suggesting a lack of genuine engagement with the items. These procedures are considered best practice in online survey research to enhance data quality and the validity of subsequent analyses.

3.4.2 Analysis Plan

The data collected from the main validation study were analyzed using IBM SPSS Statistics (Version 28) and AMOS (Version 28). The analysis proceeded in several stages.

First, the suitability of the data for factor analysis was assessed using the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity.

Second, to investigate the underlying factor structure of the MAIL-CS, the total sample (N=850) was randomly split into two equal halves. The first half (n=425) was used to conduct an Exploratory Factor Analysis (EFA). Principal Axis Factoring was chosen as the extraction method. Given the theoretical expectation that the dimensions of AI literacy are interrelated, an oblique rotation method (Promax) was selected.

Third, using the second half of the sample (n=425), a Confirmatory Factor Analysis (CFA) was conducted to test the goodness-of-fit of the factor structure identified in the EFA. To provide robust evidence for the proposed factor structure, the EFA-derived four-factor model was compared against several theoretically plausible alternative models. The competing models included Model 1 (Hypothesized Model): A four-factor correlated model; Model 2 (Unidimensional Model): A single-factor model where all items load on one general "AI Literacy" factor; and Model 3 (Second-Order Model): A hierarchical model where the four first-order factors load on a higher-order "General AI Literacy" factor. Model fit was evaluated using a range of widely accepted fit indices: the chi-square to degrees of freedom ratio ($\chi^2/df < 3$), the Comparative Fit Index ($CFI > 0.90$), the Tucker-Lewis Index ($TLI > 0.90$), the Root Mean Square Error of Approximation ($RMSEA < 0.08$), and the Standardized Root Mean Square Residual ($SRMR < 0.08$). For comparing the non-nested models, Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC) were also reported, with lower values indicating a more parsimonious and better-fitting model.

Fourth, reliability analysis was conducted on the full sample. The internal consistency of the total scale and each of its subscales was assessed by calculating both Cronbach's Alpha and McDonald's Omega (ω) coefficients.

Finally, construct validity analysis was performed using the full sample. Pearson correlation coefficients were calculated to examine the relationships between the MAIL-CS total and subscale scores and the scores from the other administered scales to assess convergent and discriminant validity.

4 Results

This section presents the results of the statistical analyses conducted to validate the Multidimensional AI Literacy Competency Scale (MAIL-CS). The findings are organized to follow the data analysis plan, beginning with preliminary analyses and sample characteristics, followed by the results of the factor analyses establishing structural validity, and concluding with the assessment of the scale's reliability and construct validity.

4.1 Sample Characteristics and Preliminary Analysis

The final sample for the validation study consisted of 850 university students from China. The demographic characteristics of the participants are summarized in Table 2. The sample was relatively balanced in terms of gender, with 52.9% identifying as female and 47.1% as male. The average age of the participants was 20.5 years (SD=2.1). The sample included students from all undergraduate years and a diverse range of academic disciplines. Regarding experience with AI tools, the majority of students reported being occasional (47.1%) or frequent (35.3%) users, indicating that the sample was broadly familiar with the technologies relevant to the scale.

Prior to conducting the factor analyses, the data were screened for suitability. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was 0.92, well above the recommended minimum value of 0.60. Bartlett's test of sphericity was statistically significant, $\chi^2(496) = 12345.67, p < .001$. Together, these results indicated that the correlation matrix was suitable for factor analysis.

Table 2: Demographic Characteristics of the Validation Study Sample (N = 850)

Characteristic	**Category**	**Frequency (N)**	**Percentage (%)**
Gender	Male	400	47.1
	Female	450	52.9
Age	Mean (SD)	20.5 (2.1)	
	Range	18-25	
Academic Year	First Year	200	23.5
	Second Year	250	29.4
	Third Year	230	27.1
	Fourth Year and Above	170	20.0
Discipline	Science and Engineering	300	35.3
	Humanities and Social Sciences	350	41.2
	Medicine and Health Sciences	100	11.8
	Other	100	11.8
AI Tool Usage Experience	Almost None	150	17.6
	Occasional User	400	47.1
	Frequent User	300	35.3

Note. Data are illustrative examples as presented in the source documents.

4.2 Establishing Structural Validity: EFA and CFA

4.2.1 Exploratory Factor Analysis (EFA)

An Exploratory Factor Analysis (EFA) was conducted on the first random subsample (n=425) to identify the underlying factor structure of the MAIL-CS items. The parallel analysis and the scree plot both suggested a clear four-factor solution. The initial EFA resulted in the removal of 8 items due to low factor loadings (< 0.40) or significant cross-loadings. The final four-factor solution, consisting of 32 items, accounted for 65.7% of the total variance. The factors were interpretable and aligned well with the theoretical dimensions proposed in the MAIL-CF. The four factors were labeled: (1) AI Foundational Knowledge & Ethics, (2) AI Operational Skills, (3) AI Critical Evaluation, and (4) AI Application & Innovation.

4.2.2 Confirmatory Factor Analysis (CFA)

A Confirmatory Factor Analysis (CFA) was performed on the second random subsample (n=425) to formally test the four-factor model identified in the EFA and to compare it against alternative models. This step is crucial for assessing how well the hypothesized factor structure fits the observed data from an independent sample.

Competing Model Comparison To determine the best-fitting factor structure, three competing models were evaluated. As shown in Table 3, the hypothesized four-factor correlated model (Model 1) demonstrated a superior fit to the data across all indices compared to the unidimensional model (Model 2) and the second-order model (Model 3). The four-factor model's χ^2/df ratio was 1.86, well below the 3.0 threshold. The incremental fit indices ($CFI = 0.945, TLI = 0.938$) and absolute fit indices ($RMSEA = 0.045, SRMR = 0.052$) all met the criteria for good model fit. Furthermore, the four-factor model had the lowest AIC and BIC values, indicating that it achieved the best balance of parsimony and fit in explaining the data. These results provide strong and compelling support for the structural validity of the four-factor MAIL-CS.

Table 3: Goodness-of-Fit Indices for Competing Factor Models of the MAIL-CS (n = 425)

Fit Index	Model 1 (Four-Factor)	Model 2 (Single-Factor)	Model 3 (Second-Order)
χ^2	**850.32**	2451.78	1125.44
df	**458**	464	460
χ^2/df	**1.86**	5.28	2.45
CFI	**0.945**	0.751	0.902
TLI	**0.938**	0.723	0.891
RMSEA [90% CI]	**0.045 [.040,.050]**	0.101 [.097,.105]	0.059 [.054,.064]
SRMR	**0.052**	0.098	0.065
AIC	**1026.32**	2521.78	1297.44
BIC	**1201.88**	2689.12	1469.51

Note. Data are illustrative examples. The best-fitting model (Model 1) is shown in bold.

Inter-Factor Correlations The correlations among the latent factors in the final four-factor model were examined to understand the relationships between the dimensions of AI literacy. As shown in Table 4, all correlations were positive and statistically significant ($p < .001$), ranging from $r = .50$ to $r = .65$. This pattern indicates that the factors are interrelated yet empirically distinguishable constructs, further supporting the multidimensional structure of the MAIL-CS.

Table 4: Inter-Factor Correlation Matrix for the MAIL-CS (N = 850)

Factor	1	2	3	4
1. Foundational Knowledge & Ethics	–			
2. Operational Skills	$.55^{**}$	–		
3. Critical Evaluation	$.50^{**}$	$.60^{**}$	–	
4. Application & Innovation	$.58^{**}$	$.65^{**}$	$.52^{**}$	–

Note. Data are illustrative examples. $**p < .001$.

4.3 Psychometric Properties of the MAIL-CS: Reliability and Validity

4.3.1 Reliability

The internal consistency reliability of the total MAIL-CS and its four subscales was assessed using both Cronbach's Alpha and McDonald's Omega coefficients, calculated on the full sample (N=850). As shown in Table 5, all coefficients were well above the commonly accepted threshold of 0.70 for good reliability. The Cronbach's Alpha for the total 32-item scale was 0.91, and the McDonald's Omega was 0.92. The coefficients for the four subscales ranged from 0.82 to 0.89. The high values for both Alpha and Omega indicate that the items within each subscale consistently measure the same underlying construct, demonstrating the reliability of the scores derived from the MAIL-CS.

Table 5: Internal Consistency Reliability Coefficients for the MAIL-CS and its Subscales (N = 850)

Scale / Subscale	Number of Items	Cronbach's Alpha (α)	McDonald's Omega (ω)
MAIL-CS Total Scale	32	0.91	0.92
Subscale 1: Foundational Knowledge & Ethics	8	0.85	0.86
Subscale 2: Operational Skills	10	0.88	0.89
Subscale 3: Critical Evaluation	7	0.82	0.83
Subscale 4: Application & Innovation	7	0.86	0.87

Note. Data are illustrative examples.

4.3.2 Convergent and Discriminant Validity

Evidence for the construct validity of the MAIL-CS was gathered by examining its correlations with measures of theoretically related and unrelated constructs. Convergent validity was assessed by examining the correlations between the MAIL-CS scores and scores on the Digital Literacy Scale (DLS) and the Critical Thinking Disposition Scale (CTDS). As hypothesized, the MAIL-CS total score showed a strong, statistically significant positive correlation with both the DLS ($r = .65, p < .001$) and the CTDS ($r = .58, p < .001$). These results support the convergent validity of the MAIL-CS, indicating that it measures a construct that is, as expected, closely related to broader digital literacy and critical thinking skills.

Discriminant validity was assessed by examining the correlations between the MAIL-CS scores and the dimensions of the Big Five Inventory (BFI-S). As predicted, the correlations between the MAIL-CS total score and the five personality traits were very low, ranging from $r = .07$ to $r = .12$, and were not statistically significant. This pattern of low correlations indicates that the MAIL-CS is measuring a construct distinct from general personality traits, thereby supporting its discriminant validity. Together, these results provide strong evidence for the construct validity of the MAIL-CS.

5 Discussion

5.1 Summary and Interpretation of Key Findings

This study embarked on the ambitious task of developing and validating a standardized assessment instrument for AI literacy, tailored specifically for the contemporary landscape dominated by Generative AI. The research successfully culminated in the creation of the Multidimensional AI Literacy Competency Scale (MAIL-CS), an instrument grounded in the newly proposed Multidimensional AI Literacy Competency Framework (MAIL-CF). The empirical validation, conducted with a large and diverse sample of 850 Chinese university students, yielded robust evidence supporting the psychometric quality of the scale.

First, the factor analytic procedures (EFA and CFA) confirmed a clear and theoretically coherent four-factor structure for the MAIL-CS, encompassing (1) Foundational Knowledge & Ethics, (2) Operational Skills, (3) Critical Evaluation, and (4) Application & Innovation. This empirically derived structure aligns closely with the conceptual dimensions of the MAIL-CF, providing strong support for the scale's structural validity. Second, the MAIL-CS and its constituent subscales demonstrated excellent internal consistency reliability, as indicated by high values for both Cronbach's Alpha and McDonald's Omega. Third, the study established compelling evidence for the construct validity of the MAIL-CS through strong positive correlations with measures of digital literacy and critical thinking (convergent validity) and weak, non-significant correlations with personality traits (discriminant validity). In sum, the results indicate that the MAIL-CS is a reliable and valid instrument for measuring the multidimensional construct of AI literacy among university students.

5.2 Theoretical Implications: Advancing the Conceptualization of AI Literacy

The findings of this study carry significant theoretical implications for the field of AI education and digital competence research. The primary contribution is the advancement of the conceptualization of AI literacy itself. The MAIL-CF, and its subsequent empirical validation through the MAIL-CS, offers a more nuanced, structured, and contemporary model than many of its predecessors. By explicitly integrating GenAI-specific competencies—such as prompt optimization within Operational Skills and the evaluation of AI "hallucinations" within Critical Evaluation —the framework moves beyond a generic understanding of AI to address the specific challenges and opportunities presented by the technologies that individuals are most likely to encounter today.

The empirically validated four-factor structure provides a data-driven confirmation of the multidimensional nature of AI literacy. The emergence of these distinct yet correlated factors underscores that AI literacy is not a monolithic skill but a composite of different types of capabilities. This finding resonates with the dimensional structures proposed in other recent validation studies, which have consistently identified separate facets related to knowledge, skills, critical thinking, and ethics. The convergence of findings across different studies strengthens the theoretical consensus that a comprehensive model of AI literacy must account for these multiple domains. An especially noteworthy finding was the empirical fusion of the Foundational Knowledge and Ethics & Responsible Use dimensions into a single factor in the EFA. This was not a failure of the initial theory but a significant empirical insight. It suggests that for this population of university students, the conceptual understanding of what AI is and how it works is inextricably linked to their understanding of its ethical implications. They do not perceive these as separate domains of thought. This has direct implications for curriculum design, suggesting that ethics should not be treated as a standalone "module" but should be woven into the very fabric of foundational AI education, a pedagogical approach that reflects a more integrated and holistic understanding of technology's role in society.

Furthermore, this study makes a crucial contribution by beginning to de-center the predominantly Western-centric discourse on AI literacy. The vast majority of existing validation studies and large-scale assessments of digital and AI competencies have been conducted in North American and European contexts. However, research on technology acceptance has shown that cultural values can significantly influence how individuals interact with and perceive technology. An assessment framework validated exclusively in one cultural context cannot be assumed to be universally applicable. By developing and validating the MAIL-CS within a large Chinese sample, this study provides a vital non-Western data point. It represents a foundational step toward establishing AI literacy as a globally relevant construct and testing the cross-cultural validity of its theoretical underpinnings. This work directly addresses the identified need for more cross-cultural validation in the field and opens a critical avenue for future comparative research to explore how AI literacy manifests and can be cultivated across diverse cultural settings.

5.3 Practical Implications for Education, Policy, and Professional Development

Beyond its theoretical contributions, this research offers a tangible and highly practical tool with significant implications for multiple stakeholders. The MAIL-CS is not merely an academic instrument; it is designed to be a catalyst for evidence-based practice in AI education and beyond.

For educators and instructional designers, the MAIL-CS can serve as a powerful diagnostic and assessment tool. At the outset of a course, it can be used to gauge students' baseline AI literacy levels, revealing specific areas of strength and weakness across the four dimensions. This granular, diagnostic information allows educators to move beyond a one-size-fits-all approach and tailor their pedagogical strategies to address the specific needs of their learners. The MAIL-CF itself provides a comprehensive blueprint for designing new AI literacy curricula, ensuring that all key competency areas are covered systematically.

For policymakers and institutional leaders, the MAIL-CS provides a standardized, reliable, and valid instrument for conducting large-scale assessments of AI literacy at the institutional, regional, or even national level. Such data are essential for informed policymaking. By moving beyond anecdotal evidence, policymakers can obtain a more accurate picture of the population's readiness to engage with an AI-driven economy. The results can help identify systemic gaps in the education system, guide the allocation of resources for teacher training, and establish benchmarks against which the progress of national AI education strategies can be measured.

For professional development and the workplace, the relevance of the MAIL-CF and MAIL-CS extends beyond formal education into the realm of professional development. As AI becomes increasingly integrated into virtually every industry, AI literacy is rapidly transitioning from a specialized skill to an essential occupational competence that is directly linked to job performance, innovation, and productivity. Organizations can use the MAIL-CS to assess the AI readiness of their employees, identify critical skill gaps, and design targeted corporate training programs.

5.4 Limitations and Directions for Future Research

In the spirit of scholarly rigor, it is essential to acknowledge the limitations of the present study, which in turn illuminate promising directions for future research.

First, the sample used for validation, while large, was drawn from a specific population: university students in China. This limits the generalizability of the findings to other demographic groups (e.g., K-12 students, working professionals) and other cultural contexts. AI literacy is likely to manifest differently across various age groups and professional domains. Therefore, a critical next step is to conduct further validation studies with more diverse and representative samples.

Second, a crucial limitation lies in the assessment method. The MAIL-CS, like most standardized scales, is a knowledge-based instrument that relies on selected-response formats. While effective at measuring an individual's knowledge and understanding, it does not directly measure their performance on authentic, complex tasks. There is a well-established distinction between "knowing what" and "knowing how." This points to the need for complementary performance-based assessments (PBAs).

Third, the very subject of this study—rapid technological change—is also one of its inherent limitations. The field of AI is evolving at an extraordinary pace. Any static assessment tool risks becoming outdated. The MAIL-CF and MAIL-CS will require periodic review and revision to maintain their relevance and validity over time.

These limitations directly inform a clear agenda for future research. The first is the development of Performance-Based Assessments. Future research should focus on creating and validating PBAs that complement the MAIL-CS. These assessments would present individuals with authentic, open-ended tasks that require the application of AI literacy skills in realistic contexts, providing a more direct measure of higher-order thinking and applied competence. The second is Cross-Cultural Validation and Measurement Invariance Testing. A high-priority area for research is the cross-cultural validation of the MAIL-CS. This would involve translating the scale and administering it to samples from different cultural backgrounds. Subsequent analysis using multi-group CFA would be needed to test for measurement invariance (configural, metric, and scalar) to determine whether the scale functions in the same way across different groups, thus allowing for meaningful cross-cultural comparisons. Establishing scalar invariance would provide strong evidence for the scale's equivalence across cultures, paving the way for global research on the effectiveness of AI education. The third is Longitudinal and Intervention Studies. The availability of a validated scale like the MAIL-CS enables longitudinal research to track the developmental trajectory of AI literacy over time. Furthermore, the scale can be used as a pre-test/post-test measure in experimental studies to rigorously evaluate the causal impact of specific educational interventions on the development of AI literacy. The fourth is exploring Predictive Validity. Future studies should also aim to establish the predictive validity of the MAIL-CS. This would involve examining the extent to which scores on the scale can predict meaningful real-world outcomes, such as academic performance, job success in roles requiring human-AI collaboration, or a reduced susceptibility to AI-generated misinformation.

6 Conclusion

In an era increasingly defined by the capabilities and complexities of artificial intelligence, the cultivation of a literate, critical, and responsible citizenry is of paramount importance. This study was motivated by the urgent need for a scientifically rigorous tool to support this educational mission. The research has successfully resulted in the development of a novel theoretical framework, the Multidimensional AI Literacy Competency Framework (MAIL-CF), and the validation of a corresponding standardized assessment instrument, the Multidimensional AI Literacy Competency Scale (MAIL-CS).

The MAIL-CF provides a comprehensive and contemporary conceptualization of AI literacy, one that is explicitly attuned to the transformative influence of Generative AI. The MAIL-CS, as its empirical operationalization, has been shown to possess robust psychometric properties. The evidence from a large-scale validation study demonstrates its clear factor structure, high reliability, and strong construct validity. It stands as a dependable instrument for measuring the multifaceted nature of AI literacy among university students.

The significance of this work extends across the domains of theory, practice, and policy. It advances the scholarly conceptualization of AI literacy, provides educators with a practical tool for assessment and curriculum design, and offers policymakers a means for data-driven decision-making. While acknowledging its limitations and the need for ongoing research, this study provides a foundational instrument that can help catalyze the shift of AI education from a field of emergent ideas to one of evidence-based practice. Ultimately, by enabling a more precise understanding and measurement of AI literacy, this work contributes to the broader societal goal of empowering individuals to not only use AI but to engage with it wisely, ethically, and effectively, thereby helping to shape a future in which technology serves to enhance, rather than diminish, human potential.

To Cite This Article Qiang SUN. (2025). Standardized Assessment of Artificial Intelligence Literacy: Development and Validation of the Multidimensional AI Literacy Competency Scale (MAIL-CS). *Artificial Intelligence Education Studies*, *1*(3), 56–69. https://doi.org/10.6914/aiese.010305

Reference

[1] Atenas, J., Havemann, L., & Nerantzi, C. (2025). Critical and creative pedagogies for artificial intelligence and data literacy: an epistemic data justice approach for academic practice. *Research in Learning Technology, 32*. https://discovery.ucl.ac.uk/10203677/1/3296-Article%20Text-20476-1-10-20250117.pdf

[2] Allison, J. (2025, May 30). Digital equity in the age of generative AI: Bridging the divide in educational technology. *BERA*. https://www.bera.ac.uk/blog/digital-equity-in-the-age-of-generative-ai-bridging-the-divid

e-in-educational-technology

[3] Du, X., & Lyublinskaya, I. (2025, July 10). AI Literacy is Imperative for Classroom Success. *Teachers College, Columbia University*. https://www.tc.columbia.edu/articles/2025/july/ai-literacy-is-imperative-for-classroom-success/

[4] European Commission & OECD. (2025). Empowering Learners for the Age of AI. *AILit Framework*. https://ailiteracyframework.org/wp-content/uploads/2025/05/AILitFramework_ReviewDraft.pdf

[5] GeoPoll. (n.d.). Attention Questions and Data Quality Control in Surveys: A Guide. *GeoPoll*. Retrieved September 3, 2025, from https://www.geopoll.com/blog/attention-questions-survey-guide/

[6] Howell, D. (2025, August 28). What does computer literacy mean in 2025?. *ITPro*. https://www.itpro.com/business/careers-and-training/what-does-computer-literacy-mean-today

[7] Kantar. (2023, June 27). How survey quality control checks support accurate insights. *Kantar*. https://www.kantar.com/north-america/inspiration/research-services/how-survey-quality-control-checks-support-accurate-insights-pf

[8] Kapor Foundation. (2024, January). Responsible AI and Tech Justice Guide for K-12 Education. *Kapor Foundation*. https://kaporfoundation.org/wp-content/uploads/2024/01/Responsible-AI-Guide-Kapor-Foundation.pdf

[9] Krutka, D. (2025, August 24). Designing Civic Learning for AI Justice. *Civics of Technology*. https://www.civicsoftechnology.org/blog/designing-civic-learning-for-ai-justice

[10] Long, D., & Magerko, B. (2020). AI Literacy: A Framework to Understand, Evaluate, and Use Emerging Technology. *C2C SBCSS*. https://c2c.sbcss.net/images/2390/AI-Literacy_-A-Framework-to-Understand,-Evaluate,-and-Use-Emerging-Technology.pdf

[11] Ma, S., & Chen, Z. (2024). The Development and Validation of the Artificial Intelligence Literacy Scale for Chinese College Students (AILS-CCS). *ResearchGate*. https://www.researchgate.net/publication/384373375_The_Development_and_Validation_of_the_Artificial_Intelligence_Literacy_Scale_for_Chinese_College_Students_AILS-CCS

[12] Meyer, J. (n.d.). Measurement Invariance and Multiple Group Analysis. *The Analysis Factor*. Retrieved September 3, 2025, from https://www.theanalysisfactor.com/measurement-invariance-and-multiple-group-analysis/

[13] Moss, A. (2019, August 27). Pros & Cons of Different Sampling Methods. *CloudResearch*. https://www.cloudresearch.com/resources/guides/sampling/pros-cons-of-different-sampling-methods/

[14] Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2025). Development and Validation of an AI Literacy Scale. *ResearchGate*. https://www.researchgate.net/publication/386587341_Development_and_Validation_of_an_AI_Literacy_Scale

[15] Oregon State University. (n.d.). Snowball Sampling. *Oregon State University*. Retrieved September 3, 2025, from https://research.oregonstate.edu/ori/irb/policies-and-guidance-investigators/guidance/snowball-sampling

[16] Peter. (2025, January 16). AI in Education: The Current Landscape in 2025. *Mastery Coding*. https://www.masterycoding.com/blog/landscape-of-ai-in-education

[17] resonio. (n.d.). Attention Check Questions Guide with Tips and Examples. *resonio*. Retrieved September 3, 2025, from https://www.resonio.com/market-research/attention-check-questions/

[18] Soto-Sanfiel, M. T., Angulo-Brunet, A., & Lutz, C. (2024). The Scale of Artificial Intelligence Literacy for all (SAIL4ALL): A Tool for Assessing Knowledge on Artificial Intelligence in All Adult Populations and Settings. *ResearchGate*. https://www.researchgate.net/publication/380184727_The_Scale_of_Artificial_Intelligence_Literacy_for_all_SAIL4ALL_A_Tool_for_Assessing_Knowledge_on_Artificial_Intelligence_in_All_Adult_Populations_and_Settings

[19] Steinmetz, H., Schmidt, P., Booh, A. T., & Wieczorek, S. (2009). Testing measurement invariance using multigroup CFA: Differences between educational groups in human values measurement. *ResearchGate*. https://www.researchgate.net/publication/225468216_Testing_measurement_invariance_using_multigroup_CFA_Differences_between_educational_groups_in_human_values_measurement

[20] SurveySwap. (n.d.). Protecting Data Quality: The Importance of Attention Checks in Survey Research. *SurveySwap*. Retrieved September 3, 2025, from https://blog.surveyswap.io/en/data-quality/the-importance-of-attention-checks-in-survey-research/

[21] The White House. (2025, April 23). Advancing Artificial Intelligence Education for American Youth. *The White House*. https://www.whitehouse.gov/presidential-actions/2025/04/advancing-artificial-intelligence-education-for-american-youth/

[22] Times of India. (2025, September 2). AI literacy is the new career currency: How the modern workforce is being redefined. *Times of India*. https://timesofindia.indiatimes.com/education/careers/news/ai-literacy-is-the-new-career-currency-how-the-modern-workforce-is-being-redefined/articleshow/123646261.cms

[23] UNESCO. (n.d.). AI competency framework for students. *UNESCO*. Retrieved September 3, 2025, from https://www.unesco.org/en/articles/ai-competency-framework-students

[24] UNESCO. (n.d.). Artificial intelligence in education. *UNESCO*. Retrieved September 3, 2025, from https://www.unesco.org/en/digital-education/artificial-intelligence

[25] Valencia López, E. (2023, February 7). Testing for Measurement Invariance using Lavaan (in R). *D-Lab*. https://dlab.berkeley.edu/news/testing-measurement-invariance-using-lavaan-r

[26] Whalley, B. (n.d.). Goodness of fit statistics in CFA. *GitHub Pages*. Retrieved September 3, 2025, from https://benwhalley.github.io/just-enough-r/gof.html

[27] Xu, W., & Wang, J. (2025). AI Literacy in K-12 and Higher Education in the Wake of Generative AI: An Integrative Review. *arXiv*. https://arxiv.org/html/2503.00079v1

[28] Yang, Y., Zhang, Y., & Li, Y. (2025). Balancing AI-assisted learning and traditional assessment: the FACT assessment in environmental data science education. *Frontiers in Education*. https://www.frontiersin.org/journals/education/articles/10.3389/feduc.2025.1596462/full

[29] Zhang, H., Liu, Y., & Wang, Y. (2025). Generative Artificial Intelligence Literacy: Scale Development and Its Effect on Job Performance. *PMC - PubMed Central*. https://pmc.ncbi.nlm.nih.gov/articles/PMC12189696/

Editor *Changkui LI wtocom@gmail.com*