Contemporary Educational Technology, 2024, **16**(3), ep512
ISSN: 1309-517X (Online)

Research Article

Development of an AI literacy assessment for non-technical individuals: What do teachers know?

Lu Ding [1*]
0000-0002-1454-1532

Sohee Kim [1]
0000-0002-7460-9412

R. Allan Allday [2]
0000-0003-2311-5620

[1] Department of Counselling and Instructional Sciences, University of South Alabama, Mobile, AL, USA
[2] Department of Leadership and Teacher Education, University of South Alabama, Mobile, AL, USA
[*] Corresponding author: luding@southalabama.edu

Citation: Ding, L., Kim, S., & Allday, R. A. (2024). Development of an AI literacy assessment for non-technical individuals: What do teachers know?. *Contemporary Educational Technology, 16*(3), ep512. https://doi.org/10.30935/cedtech/14619

ARTICLE INFO

Received: 10 Jan 2024
Accepted: 17 Apr 2024

ABSTRACT

With the exponential development and vast interest in artificial intelligence (AI), the global economic impact of AI is expected to reach $15.7 trillion by 2030. While AI has infiltrated everyday life, a lack of knowledge of what AI is and how AI works is ubiquitous across all ages and professions. Teaching AI literacy to non-technical individuals has become imperative and requires immediate attention, however, assessing AI literacy has heavily relied on subjective measurements such as qualitative assessment and self-reported surveys, which may lead to biased results. This study contributes to the field by developing and validating an assessment created based on a well-established AI literacy framework. A total of 196 responses were collected from pre-and in-service teachers in the United States, and 186 responses were included in the analysis to validate the assessment. The final assessment comprises 25 objective-based items reduced from an originally 31-item assessment. Both experts' insights were sought, and statistical methodology was employed to ensure the validity of the assessment. The results indicate that pre-and in-service teachers have a moderate level of AI literacy and in-service teachers performed slightly better than pre-service teachers on our assessment. Inconsistent answers across various AI concepts indicate that teachers may possess an even more ambiguous understanding of certain AI concepts.

Keywords: AI literacy, assessment, non-technical, teachers, objective-based

INTRODUCTION

The development of artificial intelligence (AI) has required non-technical individuals to be equipped with the knowledge about AI to employ and maneuver AI-enhanced systems in their everyday lives properly. AI-enhanced technologies exist almost everywhere, from household appliances such as autonomous vacuums to AI-powered drones that can perform surveillance for military use; from personalized recommendation systems in commercial platforms to advanced AI algorithms used in medical diagnoses. The applications of AI-enhanced technologies are vast and continuously expanding, revolutionizing various industries and aspects of our lives. PwC Global (2023) predicts that the global economic impact of AI is expected to reach $15.7 trillion by 2030. While AI has infiltrated everyday life, a lack of knowledge on what AI is and how AI works is a ubiquitous issue across all ages and professions (e.g., Antonenko & Abramowitz, 2023; Maitz et al., 2022; Mertala et al., 2022; Nader et al., 2022).

Due to a lack of knowledge about AI or a superficial understanding of AI, misconceptions toward AI and how AI will impact our lives can cause a false belief that AI will take over the world or an overly enthusiasm toward AI, thus resulting in ignoring ethical issues rooted in AI-enhanced systems such as objectivity and accountability (Cave et al., 2019).

Teaching fundamental topics of AI to non-technical individuals has become an imperative need and requires immediate attention (Ding et al., 2023; Long & Magerko, 2020). Research on how to foster AI literacy among individuals without a technical background has been emerging to respond to this need. Opportunities and challenges of how to teach AI literacy in early childhood have been researched and proposed (Su et al., 2023), how to effectively integrate AI literacy in K-12 education (Yue et al., 2022) and higher and adult education (Laupichler et al., 2022) has also been studies. This effort is not only limited to formal education but has also been extended to informal education (Long et al., 2022) as well as the workforce (Cetindamar et al., 2022).

While considerable effort has been devoted to investigating how to teach AI literacy, assessing non-technical individuals' understanding of AI exhibits substantial variability across formats and evaluating content. Ng et al. (2021a) reported in their literature review study that some of the reviewed articles adopted a qualitative method to evaluate students' understanding of AI, such as interviews, observations, and evaluation of artifacts that students created. Half of the articles used quantitative measurements consisting of both self-reported scales and tests.

Self-reported scales contain items such as those that start with "*how would you rate yourself …*" and require students to self-evaluate their skills in a certain area. Those tests are usually developed by the researchers who conducted the studies and designed for specific programs without a statement of established validity and are usually for individuals with a technical background. It is well-known that self-reported surveys can be subjective and biased (Rosenman et al., 2011). Several other literature review studies have also reported a similar concern about the quality of AI literacy assessments used in the existing research and the scarcity of research in this area (Casal-Otero et al., 2023; Laupichler et al., 2022).

Therefore, in responding to the gap in the literature, this study aims to develop an assessment that measures non-technical individuals' AI literature knowledge. While we were in the process of analyzing our data and writing up our study, Hornberger et al. (2023) published their assessment just in time to fill the gap in the literature. Their assessment comprised 31 multiple-choice questions developed based on AI literacy framework proposed by Long and Magerko (2020), which served as the foundation for the development of our assessment. While the two studies share the same goal of contributing to the field of developing and validating objective AI literacy assessments, our study differs from theirs at both the item development level and the data collection level.

Hornberger et al.'s (2023) assessment consists of 31 multiple-choice questions and one ordering question, whereas the majority of our items are True/False questions, with a few exceptions. Furthermore, while their assessment was originally written in German and conducted with participants from Germany, our assessment items are in English, and our participants are from the United States. Not only are the geographical locations different, but our participants are predominantly pre-service or in-service teachers, whereas their participants were undergraduate students. Lastly, Hornberger et al. (2023) noted that most of their participants studied at technical universities and were enrolled in technical programs, whereas our participants have very limited technical backgrounds in AI. Despite these differences, both studies aim to address gaps in the literature, and we believe that together, they can enrich the field for scholars, practitioners, and individuals interested in AI literacy.

This article is structured, as follows: firstly, it establishes a conceptualization of AI literacy, which serves as the foundation for crafting the assessment. This involves defining AI literature for non-technical individuals and determining the essential components that should be incorporated into AI literature education. Subsequently, the development of assessment items and their evaluation and refinement process will be explained. Lastly, the article concludes with a discussion, followed by the study's limitations and recommendations for future research.

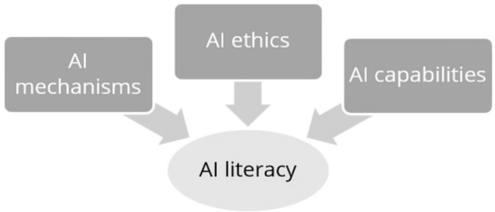

Figure 1. Three dimensions of AI literacy a non-technical individual should possess (Source: Authors)

CONCEPTUALIZE AI LITERACY

Defining Artificial Intelligence Literacy

AI is defined in so many ways. Non-technical individuals view AI as technology, an intelligent machine, computer programs, etc. that mimic human intelligence or an embodied entity that can perform physical actions (Ding et al., In review). However, AI is seen by experts closer to algorithms (McCarthy, 2007) and tools for building computer programs (Chubb et al., 2022). A widely acceptable definition of AI is the science and engineering behind the development of intelligent machines that solve problems through techniques such as machine learning, natural language processing, and reinforcement learning (Mondal, 2020).

AI will eventually impact many aspects of human life through everyday technologies rather than merely a buzzword retained in computer industries and everyone including non-technical individuals should learn AI (Ng et al., 2021b). Recent researchers proposed *AI literacy* as a term to describe individuals' understanding and perception of AI instead of focusing on the technical facets of AI. Literacy is the ability to read and write (McBride, 2015). AI literacy can be broadly defined as the ability to understand and navigate AI-enhanced systems effectively and responsibly. It goes beyond mere familiarity with AI; it empowers individuals to evaluate AI-related information critically, make informed decisions, and contribute positively to AI-driven advancements.

AI literacy encompasses a multifaceted comprehension of the underlying principles, applications, and ethical implications of AI (Ng et al., 2021b). While AI literacy used to be only required by individuals with technical backgrounds, with the increased popularity and the development of AI technology this skill has been extended to non-technical individuals in the recent couple of years (Long & Magerko, 2020). AI-literate non-technical individuals do not need to be programmers or engineers; instead, they should be equipped with the knowledge to make warranted judgments about AI-enhanced products and services, be aware of the ethical considerations of AI technologies, comprehend AI-related news and conversations, and effectively interact with AI-driven systems.

What to Teach in Artificial Intelligence Literacy

Given that AI literacy is multifaceted, several guidelines on what competencies should be taught to non-technical individuals have been proposed. Although they differ slightly, many of the proposed competencies can be categorized into three dimensions as visualized in **Figure 1**:

(1) AI mechanisms,

(2) AI capabilities, and

(3) AI ethics.

AI mechanisms involve the need for non-technical individuals to understand what AI is and how it works. For instance, people need to be equipped with the competence of being able to recognize AI and distinguish between general and narrow AI (Long & Magerko, 2020). In addition, non-technical individuals also need to understand the basic functional concepts behind AI. Kong et al. (2023) argued that laymen need to grasp the fundamental concepts of how AI works, such as machine learning and deep learning. Similarly, Long and Magerko (2020) listed a set of required competencies for non-technical individuals, including understanding the steps and challenges of machine learning, related data literacy concepts, and how AI perceives the world. These concepts were also proposed by Touretzky et al. (2019) in their seminal "big five ideas for K-12." Within the same dimension, another crucial competency, understanding the roles that human beings play in the design, development, and application of AI, has emerged as a vital skill for non-technical individuals to acquire (Long & Magerko, 2020).

The second dimension is AI capabilities, which focus on the competencies of understanding what AI can and cannot do. For instance, being able to recognize the application of AI (Olari, 2023), how to use AI to understand and solve real-world problems (Kong et al., 2023), what are the strengths and weaknesses of AI, and the possibilities of futuristic AI technology as well as their impact (Long & Magerko, 2020). Touretzky et al. (2019) proposed an even more specific competence that K-12 students need to comprehend, which is "making agents interact comfortably with humans is a substantial challenge for AI developers" (p. 9797).

Finally, the sociocultural dimension encompasses the ethical impact of AI on society and has sparked the most concerns among researchers and educators. Issues such as data privacy, algorithmic bias, job replacement, and the impact on social norms have become focal points that non-technical individuals need to be aware of (Long & Magerko, 2020). With the advancement of generative AI (GenAI) and its increased accessibility, there have been suggestions to consider the ethical implications of the artifacts created by GenAI and their potential to spread misinformation (Lee et al., 2021). Additionally, safety concerns arising from AI-enhanced systems have also been brought forefront as core considerations that non-technical individuals need to be acknowledged. For example, how to teach a person to drive a self-driving car without risking one's life (Wong & Huan, 2020).

INITIAL ITEMS DEVELOPMENT

Informed by the three dimensions that constitute AI literacy, we drew upon the work of Long and Magerko (2020) to shape our assessment, which aimed to gauge the essential competencies required by non-technical individuals in realm of AI. Long and Magerko's (2020) extensive review of 150 diverse sources, encompassing conference papers, journal articles, books, and supplementary grey literature, culminated in the identification of 17 distinct competencies. These competencies are organized into five key facets of AI literacy, namely:

(1) understanding AI's nature,

(2) recognizing AI's capabilities,

(3) grasping AI's underlying mechanisms,

(4) discerning appropriate AI utilization, and

(5) comprehending public perceptions of AI.

To address the first facet of AI literacy, four competencies were proposed that focused on the capability of identifying AI-enabled technologies, acknowledging features that entail intelligence, recognizing AI can be developed in many ways, and distinguishing general and narrow AI. The second facet was composed of two competencies and was the ability to identify the strengths and weaknesses of AI as well as the impact of AI applications on the world. The third facet, which was also the most crucial aspect of AI literacy, consisted of nine competencies that mainly addressed individuals' understanding of the following: knowledge representation, how AI makes decisions, the steps involved in machine learning, human roles in AI, how AI processes and interprets data, how AI physically interact with the world, and how AI sense and collect data. A single competency was proposed for each of the fourth and fifth facets to tackle the ethics of AI and the programmability of AI.

The creation of the first version of this assessment entails three steps. First, the primary researcher reviewed all the references cited by Long and Magerko's (2020) work. Subsequently, one to two assessment items were crafted for each competency, often drawing directly from, or revised from, examples found within these references. In instances, where existing references did not yield suitable items for certain competencies, the primary researcher embarked on further academic exploration to construct pertinent assessment items. As a result of this exhaustive process, we successfully generated a total of 32 objective assessment items, thoughtfully designed to encompass true/false statements, multiple-choice questions, and sorting inquiries. Finally, a focus group session was conducted between a professor who researched AI and the primary researcher to review and revise the items. The revised assessment items along with the corresponding competencies proposed in Long and Magerko's (2020) study and their corresponding references are listed in **Table 1**.

Table 1. Assessment items & references

C	Question	Reference
C1	(F) An automatic washing machine is an example of an AI device.	CBSE Question Bank (2023)
C1	(T) Face lock feature in phones is a type of AI.	CBSE Question Bank (2023)
C2	(T) AI cannot solve problems the way humans can.	Antonenko and Abramowitz (2023)
C2	(F) AI systems are always smarter than humans.	Brooks (1991)
C3	(T) ChatGPT is an AI so is Amazon recommendations system.	CBSE Question Bank (2023)
C3	(F) All AIs are created the same way.	Antonenko and Abramowitz (2023)
C4	(F) Voice-activated digital assistants (Alexa & Siri) are examples of general AI.	Touretzky et al. (2019)
C4	(T) Self-driving cars are a type of narrow AI.	Baum et al. (2011)
C5	(F) It is easy for an AI system to recognize objects in unfamiliar or atypical situations, for example, a partially hidden pencil.	Developed by primary researcher & a CS professor researches AI
C5	(T) Walking down a street as well as a 5-year-old can be very difficult for an AI robot.	Goel and Davies (2011)
C6	(F) All human jobs will be replaced by AI in the future.	Antonenko and Abramowitz (2023)
C6	(T) AI can help humans avoid dangerous work (e.g., collecting and packaging of radioactive waste).	CBSE Question Bank (2023)
C7	(F) AI uses same ways to organize & store information for different tasks.	Davis et al. (1993)
C7	(T) How computers store and organize information about the world is not always fully captured.	Long and Magerko (2020)
C8	(T) AIs rely on algorithms to make decisions.	McCarthy (2007)
C8	Matching the techniques that AIs use to the corresponding examples: (a) natural language processing: use a chatbot to respond to customer queries, (b) reinforcement learning: train a robot to navigate a maze, & (c) decision tree: predict whether a customer will buy a product.	Goel and Davies (2011)
C9	(F) AI and machine learning are interchangeable terms.	Antonenko and Abramowitz (2023)
C9	Put the following machine learning steps in order: (a) collect data, (b) train model, & (c) deploy the model.	Ng (2023)
C10	(T) AI cannot learn without human input.	Antonenko and Abramowitz (2023)
C10	(T) AI is not entirely automated & always requiring human decision-making.	Lindner and Berges (2020)
C11	Which one is not type of data: (q) numbers, (b) images, (c) texts, & (d) <u>actions</u>.	Prado and Marzal (2013)
C11	(F) AI algorithms can figure out all your messy data.	Antonenko and Abramowitz (2023)
C12	(T) Machine learning is a kind of statistical inference.	Touretzky et al. (2019)
C12	(F) AI machines cannot keep updating their knowledge by using their own data.	Long and Magerko (2020)
C13	(T) Data is always shaped by decisions and assumptions made during the process of data collection, processing, and analysis.	Hautea et al. (2017)
C13	(T) Data can be error-prone and requires interpretation.	Hautea et al. (2017)
C14	(F) AI robots can walk along a preprogrammed path, but they cannot avoid obstacles on the path when they sense one.	Long and Magerko (2020)
C15	(T) AIs "see" and "hear" the world through the process of extracting information from sensory signals.	Long and Magerko (2020)
C15	(F) Self-driving cars only need object detection sensors for them to drive properly.	Developed by primary researcher & a CS professor researches AI
C16	(F) AI always makes fairness decisions.	Jones-Jang and Park (2023)
C16	Which one of ethical issues is least likely caused by AI if it's used inappropriately: (a) discrimination, (b) lack of accountability, (c) lack of privacy, & (d) <u>lack of compassion</u>	Developed by primary researcher * a CS professor researches AI
C17	(T) AIs are programmable.	Long and Magerko (2020)

Note. C: Competency

ASSESSMENT VALIDATION AND EVALUATION

Recruitment & Data Collection

The developed AI assessment was then administered to in-service and pre-service teachers. The data were collected from June to September 2023 in the Southern region of the United States. Institutional review board approval for this research was obtained before data collection. This process included two phases. During the first phase, four AI experts were invited to review the first version of the assessment. The experts' credentials are shown in **Table 2**. In the second phase, the reviewed and revised assessment was then distributed online via Qualtrics by the provision of a link or a QR code. The participants (i.e., pre- and in-service teachers) were contacted through emails, newsletters, conferences, and instructors who taught courses for pre-service

Table 2. Experts' credentials for the index of IIOC value

Expert	Degree	Areas	Years of experience
Expert 1	Doctoral	Academia/industrial	Over 10 years
Expert 2	Doctoral	Academia/industrial	Over 10 years
Expert 3	Postdoctoral Fellow	Academia	Five years
Expert 4	Doctoral Candidate	Academia	Four years

Table 3. An example of experts' IIOC ratings

Item 1		AI mechanisms	AI ethics	AI capabilities
Evaluator	Experts 1	1	-1.00	-1
	Experts 2	1	0.00	-1
	Experts 3	1	-1.00	-1
	Experts 4	1	-1.00	-1
	Mean	1	-0.75	-1

teachers. All participants were provided with the assessment, a demographic form, and a consent form. AI assessment was expected to be completed within 20 minutes.

Data Analysis

For the current study, diverse approaches were employed to validate and evaluate the developed AI assessment including content validity with experts panel and the index of item-objective congruence (IIOC) value, construct validity with exploratory and CFA, item evaluation with item response theory (IRT). Based on the results of content validity, data screening procedures were executed to check missing data and careless responses. After completing data screening procedures, descriptive statistics and exploratory factor analysis (EFA) were performed using IBM SPSS Statistics 28.0 (IBM Corp, 2021), and CFA and IRT analysis were conducted using Rstudio (RStudio Team, 2020) software with *lavaan* (Rosseel, 2012), and *ltm* (Rizopoulos, 2006) packages, respectively.

Content validity

To ensure content validity, IIOC was employed, derived from feedback provided by four AI experts. This index originated as a procedural tool in the test development phase, specifically during item development, drawing inspiration from the unidimensional assessment framework proposed by Rovinelli and Hambleton (1977). Subsequently, Turner and Carlson (2003) advanced this unidimensional structure into a multidimensional framework, with the corresponding formula, as follows: $IIOC = \frac{(N)\mu_k-(N-p)\mu_l}{2N-p}$, where *N* represents the number of objectives, *p* is the number of valid objectives, μ_k is the experts' mean rating of item *i* on the valid objectives *k*, and μ_l is the experts' mean rating of item *i* on the invalid objective *l*.

Experts were presented with three primary objectives (AI mechanism, AI capacities, and AI ethics) along with a set of items. For each item, experts determined which of the three objectives the specific item was evaluating, and a corresponding score was assigned, utilizing the following three options: +1=congruent (for clearly measuring), 0=questionable (the degree to which it measures the content area is unclear), and -1=incongruent (for not measuring). **Table 3** shows an example of the four experts rating on item 1. For this item, all experts agreed that item 1 was for evaluating AI mechanism. Three of the four experts assented that it was not for measuring AI ethics and all experts consented that it was not intended to evaluate AI capabilities.

Construct validity

To confirm the construct validity, EFA and CFA were carried out. Before conducting EFA, the Kaiser-Meyer-Olkin (KMO>.60; Kaiser, 1974) and Bartlett's test of sphericity (p<.05; Gorsuch, 1973) were evaluated for checking the appropriateness. Additionally, scree plots and explained variance methods were applied to decide the number of factors in AI assessment.

According to EFA outcomes, CFA with a one-factor model was performed to verify whether the developed AI assessment measured AI literacy. As criteria, χ^2 value, root mean square error of approximation (RMSEA; Steiger, 1980), and standardized root mean squared residual (SRMR) were yielded (Hornberger et al., 2023; Kline, 2015).

Item characteristic evaluation using item response theory

In this study, IRT analysis was added for item validation and evaluation. IRT has the advantage of being able to calculate the item difficulty and discrimination regardless of the group that took the test or assessment and to estimate the participants' unique ability scores (De Ayala, 2013). Before performing IRT analyses, two assumptions of IRT were confirmed:

(1) unidimensionality and

(2) local independence.

Then IRT with a parameter logistic model (2PLM) was applied to evaluate individual items in AI assessment. The 2PLM surpasses the 1PLM in flexibility by accommodating variations in both item discrimination and difficulty parameters. Additionally, unlike the 3PLM, the advantage of the 2LM lies in not demanding a large sample size, making it a prevalent choice for IRT analysis.

RESULTS

Data Screening Process

Before conducting the analysis, duplicated cases and missing responses were scrutinized. A total of 196 participants responded to AI assessment and among them, six cases were detected as duplicated cases, and four cases were eliminated as missing responses. Therefore, a total of 186 responses were graded dichotomously (0: incorrect and 1: correct) and employed for the analysis. In the final dataset, 79.0% of participants were female, 18.3% were male, and 1.6% were non-binary/third gender or prefer not to say. 40.3% of participants were pre-candidacy, 38.7% were in candidacy, 19.9% were in-service teachers, and 1.1% were others, respectively.

Descriptive Statistics

Descriptive statistics of all initial items are provided in **Table 4**. The mean of the item, *'AIs are programmable'* was the largest (mean [M]=.989, standard deviation [SD]=.10), and the mean of the item, *'voice-activated digital assistants, such as Alexa and Siri, are examples of general AI.'* was the smallest (M=.065, SD=.25). The overall mean was 0.677 (SD=.10) in the initial AI assessment with 31 items and the final mean value was 0.649 (SD=.11) in the finalized version of AI assessment with 25 items. Based on the final version of AI assessment, for in-service teachers, the average was calculated to be 0.682 (SD=.10), and for pre-service teachers, the average was produced to be 0.646 (SD=.12), hence the average for in-service teachers was slightly higher.

Validity Analyses

Content validity

First, to establish content validity, a panel of four AI experts meticulously assessed all items in the evaluation. The resulting overall IIOC index was calculated, yielding a value of .920. This figure surpassed the predefined criterion of .750, affirming the appropriateness of content validity in the developed assessment. Next, the difficulty and discrimination parameters of each item were yielded based on classical test theory (CTT) and IRT analyses for evaluating each item in AI assessment. The calculated difficulty and discrimination parameters from both CTT and IRT were presented in **Table 4**.

Items were considered as poor quality of items delamination if they were deemed excessively easy, difficult, or demonstrated insufficient discrimination. Through a comprehensive evaluation incorporating the IIOC index, CTT, and IRT analyses, six items (5, 14, 18, 20, 22, and 24) were excluded from AI assessment. These particular items exhibited relatively low IIOC index values, difficulty levels, and inadequate discrimination parameters.

Construct validity

According to content validity outcomes, EFA was carried out with a total of 25 items. KMO and Bartlett's test are commonly used to evaluate the appropriateness of EFA. The value of KMO was .609 and this value

Table 4. Descriptive statistics, item difficulty and discrimination from CTT & IRT

Item	Initial version (31 items)					Final version (25 items)	
	CTT			IRT		IRT	
	Mean/*b*	SD	*a*	*b*	*a*	*b*	*a*
1	0.62	0.49	0.25	-1.55	0.33	-1.48	0.35
2	0.86	0.35	0.06	4.52	0.42	4.15	0.46
3	0.52	0.50	0.14	-0.43	0.20	-0.38	0.23
4	0.82	0.38	0.38	-1.55	1.27	-1.53	1.29
5*	0.85	0.36	0.17	33.65	-0.05	-	-
6	0.92	0.26	0.27	-2.99	0.97	-3.00	0.96
7	0.06	0.25	0.04	4.00	0.72	3.93	0.74
8	0.91	0.29	0.08	4.90	0.49	4.92	0.49
9	0.51	0.50	0.32	-0.06	0.83	-0.05	0.83
10	0.58	0.50	0.22	-1.02	0.30	-1.06	0.29
11	0.80	0.40	0.40	-1.37	1.33	-1.35	1.37
12	0.86	0.35	0.16	4.98	0.18	4.56	0.16
13	0.34	0.47	0.34	0.91	0.86	0.96	0.79
14*	0.87	0.34	0.24	-14.86	0.13	-	-
15	0.90	0.30	0.34	-5.65	0.41	-5.58	0.41
16	0.56	0.50	0.42	-0.35	0.75	-0.36	0.74
17	0.41	0.49	0.18	-5.11	0.07	-6.68	0.05
18*	0.73	0.44	0.22	34.33	-0.03	-	-
19	0.58	0.49	0.07	0.96	0.35	1.03	0.32
20*	0.62	0.49	0.25	-2.49	0.20	-	-
21	0.35	0.48	0.39	0.54	1.65	0.54	1.64
22*	0.82	0.38	0.08	2.83	0.58	-	-
23	0.54	0.50	0.25	-0.40	0.45	-0.42	0.43
24*	0.84	0.36	0.27	18.27	-0.09	-	-
25	0.90	0.30	0.20	-6.63	0.34	-6.98	0.33
26	0.62	0.49	0.29	-0.74	0.72	-0.76	0.70
27	0.82	0.38	0.11	5.71	0.27	5.08	0.31
28	0.70	0.46	0.51	-0.71	1.87	-0.73	1.81
29	0.79	0.41	0.48	-1.22	1.49	-1.19	1.56
30	0.24	0.43	0.25	4.97	0.23	4.62	0.25
31	0.99	0.10	0.15	-4.55	1.14	-4.39	1.19

Note. *Deleted item in final version; *b*: Difficulty parameter; & *a*: Discrimination parameter

indicated that the data were considered appropriate for structure detection (Pett et al., 2003). In the case of Bartlett's test, it was statistically significant at α=.001, representing that EFA might be useful for this data.

To determine factor extraction, the principal axis factoring method, and the 'direct oblimin' rotation method were used. After the determination of the method for factor extraction, Cattell's scree plot (Cattell, 1966), which is a graphical representation of the factors and their corresponding eigenvalues and explained variance (at least 20.0% of explained variance) was considered, as shown in **Figure 2**.

According to the scree plot, we identified the point, where the eigenvalues leveled off, known as the elbow. At this elbow, the total explained variance with three factors amounted to 25.8%. However, the primary objective of this study is to verify whether the developed AI assessment, comprising a total of 25 items, effectively measures one comprehensive latent construct known as AI Literacy. Consequently, CFA was employed to establish construct validity through a one-factor model. This approach aligns with the confirmation of the uni-dimensionality assumption in item characteristic evaluation, as per IRT principles.

Then, CFA with a one-factor model was performed to verify construct validity and to check whether the developed AI assessment measured AI literacy and satisfied the uni-dimensionality assumption of IRT. Following Hornberger et al.'s (2023) and Kline's (2015) recommendations, we used several fit indices and cutoff values: χ^2/df<2-3, RMSEA<.05, and SRMR<.10. CFA outcomes yield appropriate fit indices, representing that this assessment is valid and it measures AI literacy appropriately. Additionally, the assumption of uni-dimensionality was confirmed for IRT analyses (χ^2/df=1.395, RMSEA=.046, and SRMR=.075).

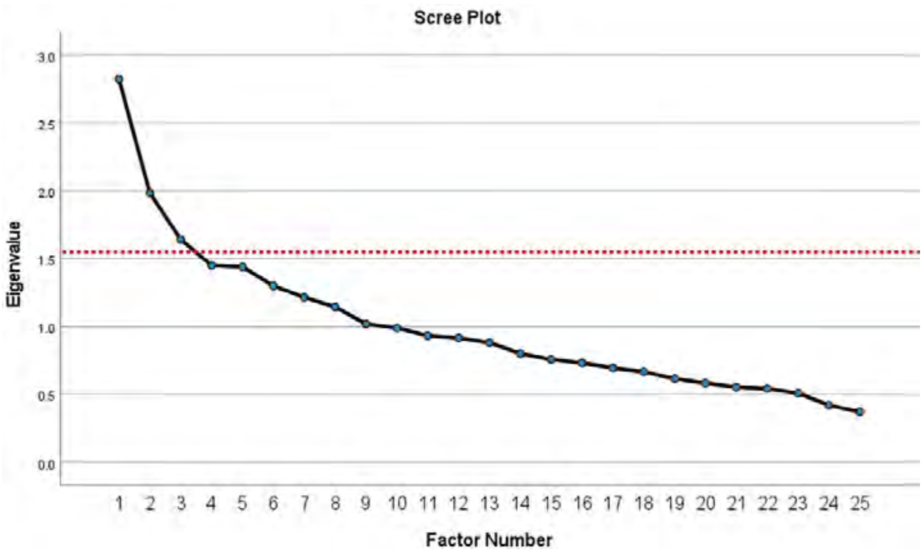

Figure 2. Scree plot (Source: Authors)

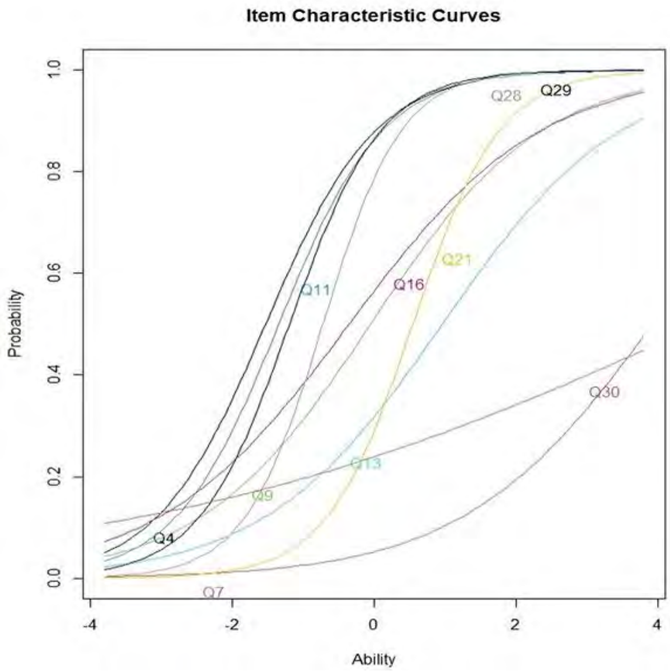

Figure 3. Item characteristic curves (Source: Authors)

Item characteristic evaluation using item response theory

Two crucial assumptions, namely uni-dimensionality and local independence, underwent rigorous testing. As detailed earlier, CFA employing a one-factor model confirmed the successful fulfillment of the uni-dimensionality assumption. To assess local independence, X^2 statistics were computed for the detection of local dependence among items, employing IRT analysis. Based on the X^2 value, the null hypothesis, indicating that items fit the model, was consistently upheld across all individual items, affirming the satisfaction of the local independence assumption.

According to IRT analysis with 2PLM, developed items in AI assessment covered different levels of difficulty and this indicates that a variety of overall difficulty levels are being considered, with no questions being too easy or too difficult. **Figure 3** represented item characteristic curves with 10 items as examples to show the diverse level of difficulty and discrimination in AI assessment and the y-axis of this plot meant the probability of answering correctly and item as a function of ability (Hornberger et al., 2023).

Table 5. Top-5 correctly answered items & top-5 incorrectly answered items

C	Assessment items	Mean	Standard deviation
	Most correctly answered items		
C17	AIs are programmable.	0.99	0.10
C3	All AIs are created the same way.	0.92	0.26
C4	Self-driving cars are a type of narrow AI.	0.91	0.29
C8	AIs rely on algorithms to make decisions.	0.90	0.30
C13	Data can be error-prone and requires interpretation.	0.90	0.30
	Most incorrectly answered items		
C4	Voice-activated digital assistants, such as Alexa and Siri, are examples of general AI.	0.06	0.25
C16	Which one of ethical issues is least likely caused by AI if it is used inappropriately?	0.24	0.43
C7	AI uses the same ways to organize and store information for different tasks.	0.34	0.47
C11	AI algorithms can figure out all your messy data.	0.35	0.48
C9	Put the following Machine Learning steps in order.	0.41	0.49

DISCUSSION & IMPLICATIONS

Contradictories in Understanding of Artificial Intelligence

To better facilitate readability and comparability of the contradictory discussed in this section **Table 5** is created. While the overall mean of the assessment is relatively high, with a correct rate of more than 80.0%, it is interesting to see two contradictions in the top correctly and incorrectly answered items. Two items assessing teachers' knowledge of narrow and general AI (C4) are one of the top correct answered items ("*self-driving cars are a type of narrow AI*") and one of the top incorrect answered items ("*voice-activated digital assistants, such as Alexa and Siri, are examples of general AI*"). Another contradiction shows in their understanding of whether AI or AI-enhanced technology are all the same. One of the highest correctly answered questions is "*all AIs are created the same way* (C3)", which demonstrates that teachers or pre-service teachers can recognize that not all AIs are developed the same way. Nevertheless, one of the most incorrectly answered questions is "*AI uses the same ways to organize and store information for different tasks* (C7)" indicating a lack of understanding of how AI utilizes different knowledge representation strategies for different tasks. Although these two items measure different competencies, at the overarching level they both examine an understanding that AIs are not one uniform entity and that they can be developed and utilized in various ways. These two contradictions in our data revealed that pre-and in-service teachers actually lack a fundamental understanding of how AI works, especially what is general AI and narrow AI, and distinguish between the two, as well as that AIs can be designed in diverse methods.

Additionally, the lack of solid knowledge about how AI works is further demonstrated by the other two pairs of most correctly answered and incorrectly answered items. Most pre-and in-service teachers recognized that "*AIs rely on algorithms to make decisions*;" however, most of them could not identify correct Machine Learning steps ("*put the following machine learning steps in order.*"). This indicates that our participants may have a certain level of knowledge about AI, but only partially. Conceptually they understand that algorithms play a critical role in AI, whereas lack technical knowledge about how algorithms work. This finding corroborates previous research in AI literacy in which individuals are in need of being taught fundamental concepts of how AI works, such as machine learning concepts (Kong et al., 2023). Another pair of items exhibit the same finding centered on data literacy. On one hand, pre-and in-service teachers in our study comprehend that "Data can be error-prone and requires interpretation." On the other hand, they also show a misconception about AI that AI has the superpower to "*[AI algorithms can] figure out all your messy data*" (Ding et al., In review).

Overall, although a relatively high score is achieved in our study, a closer look at the specific items warrants further effort in AI literacy education for pre- and in-service teachers, especially focusing on how AI works. These results are consistent with previous studies and suggest the necessity of AI literacy education, especially considering the rapid advancement of AI technology (Casal-Otero et al., 2023; Laupichler et al., 2022). Given assessment is core in all forms of teaching and learning, this AI literacy assessment has the potential to standardize and enhance AI literacy education, thereby addressing a crucial need in contemporary learning environments.

Development of Assessment for Artificial Intelligence Literacy

In contrast to numerous studies relying solely on content validity verification through expert reviews, this study implemented a more robust methodology by substantiating content validity with objective numerical statistics. By quantifying the evaluation from content experts, this study not only elevates the credibility of the evaluation process but also facilitates a more objective and data-driven integration of their inputs into the development of AI assessment. Furthermore, in the process of validating and evaluating the developed assessment, the application of factor analysis approaches and IRT was undertaken simultaneously. This dual methodology ensures a comprehensive and rigorous examination of the validity and assessment quality of the assessment.

The goal of our study is to build a collective effort with Hornberger et al. (2023) in contributing to the field to provide diverse options for objective-based assessments for educators, scholars, and those who are interested in AI literacy education. While Hornberger et al. (2023) focused on German undergraduate students with a technical background, our assessment targeted pre-service and in-service teachers in the United States, who possess little to no technical background. Furthermore, the items of our assessment and the ones in Hornberger et al.'s (2023) assessments are formatted fundamentally differently. The majority of theirs are multiple-choice questions, whereas ours are true or false questions. In addition, our assessment offers items for measuring the competencies of action/reaction, sensors, and imagine future AI, which are omitted in their assessment. By considering both assessments, one can benefit from a comprehensive resource to cater to specific educational or research needs. Drawing from the items from both assessments enables the creation of a customized assessment. Moreover, Hornberger et al.'s (2023) study and our study offer two baselines for different populations: one for those with a technical background and another for those without. This provides valuable insights into AI literacy across distinct educational contexts and participant backgrounds.

Unlike relying solely on self-reported surveys that may induce measurement bias (Casal-Otero et al., 2023; Laupichler et al., 2022), our assessment allows educators to gauge AI literacy objectively. While acknowledging the value of objective assessments, it is equally crucial to exercise caution as they may inadvertently disregard critical nuances in assessing individuals' comprehension of AI (Saxton et al., 2014). To address this limitation and to ensure a more comprehensive and accurate evaluation of individuals' AI literacy, future research can supplement our assessment with qualitative methods or contextual analyses in events of assessing AI literacy. Additionally, since most of the items in our assessment are true/false questions, and there are typically multiple items measuring the same competency, future educators or scholars can leverage this feature to assess individuals' true understanding of a specific AI concept by comparing responses for the same competency. For instance, if an individual answers one item correctly and another incorrectly within the same competency, caution is warranted regarding the depth of the individual's understanding of that specific competency.

CONCLUSIONS

With the rapid advancement of AI technology and its pervasive integration into everyone's everyday life, fostering AI literacy among individuals without a technical background becomes imperative. Assessments serve as the core of teaching and learning that not only allow teachers to better design instructional materials and evaluate the effectiveness of the instruction but also empower learners to track their progress and deepen their understanding. The lack of a standardized assessment for measuring AI literacy has been reported as a ubiquitous issue in AI literacy education. To fill the gap, we developed an AI literacy assessment grounded in the framework of 17 AI competencies delineated by Long and Magerko (2020). The assessment underwent rigorous evaluation by four AI experts and practical testing by pre-service and in-service teachers with little or no technical proficiency. The final version consisted of 25 items, each crafted to gauge one or two competencies. The final version of this assessment can contribute to the field of AI literacy education, particularly for individuals with non-technical backgrounds.

Limitations

This study had limitations to be strengthened in the future. Firstly, the study is constrained by a relatively small sample size. To enhance the precision and reliability of the findings, future research should consider incorporating sufficiently large and diverse samples. Determining an appropriate sample size is a complex and non-trivial process (Williams, 1978), necessitating additional investigations with expanded participant pools to ensure the robust validity of statistical conclusions. In addition, the current study concentrated on assessing the effectiveness of the initially developed AI assessment, comprised of 25 items, in measuring a singular latent construct known as AI literacy; however, factor analysis identified three sub-factors. This outcome highlights the necessity for future research endeavors to employ multidimensional IRT, thereby deepening our understanding and refining AI assessment's measurement capabilities.

Author contributions: LD: conceptualization, methodology, project administration, writing–original draft, & writing–reviewing & editing; **SK:** methodology, formal analysis, writing–original draft, & writing–reviewing & editing; & **RAA:** investigation. All authors approved the final version of the article.

Funding: The authors received no financial support for the research and/or authorship of this article.

Acknowledgments: The authors would like to thank all the participants who took the time to fill out the survey and the course instructors for pre-service teachers and allowed them to collect data from their courses. The authors would also like to thank SARIC at the University of South Alabama for letting them extend the research to in-service teachers.

Ethics declaration: The authors declared that all research procedures were carried out in compliance with guidelines for ethical research, and measures were taken to ensure anonymity & privacy of the survey participants. The study was approved by the institutional ethics committee of the University of South Alabama on 15 June 2023 (Approval code: 23-204). Consents were obtained from all the participants.

Declaration of interest: The authors declare no competing interest.

Data availability: Data generated or analyzed during this study are available from the authors on request.

REFERENCES

Antonenko, P., & Abramowitz, B. (2023). In-service teachers' (mis)conceptions of artificial intelligence in K-12 science education. *Journal of Research on Technology in Education, 55*(1), 64-78. https://doi.org/10.1080/15391523.2022.2119450

Baum, S. D., Goertzel, B., & Goertzel, T. G. (2011). How long until human-level AI? Results from an expert assessment. *Technological Forecasting and Social Change, 78*(1), 185-195. https://doi.org/10.1016/j.techfore.2010.09.006

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence, 47*. https://doi.org/10.1016/0004-3702(91)90053-M

Casal-Otero, L., Catala, A., Fernández-Morante, C., Taboada, M., Cebreiro, B., & Barro, S. (2023). AI literacy in K-12: A systematic literature review. *International Journal of STEM Education, 10*(1), 29. https://doi.org/10.1186/s40594-023-00418-7

Cattell, R. B. (1966). The scree plot test for the number of factors. *Multivariate Behavioral Research, 1*, 140-161. https://doi.org/10.1207/s15327906mbr0102_10

Cave, S., Coughlan, K., & Dihal, K. (2019). "Scary robots" examining public responses to AI. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 331-337). ACM. https://doi.org/10.1145/3306618.3314232

CBSE Question Bank. (2023). *Introduction to AI: Foundational concepts class 10 artificial intelligence*. Department of Skill Education.

Cetindamar, D., Kitto, K., Wu, M., Zhang, Y., Abedin, B., & Knight, S. (2022). Explicating AI literacy of employees at digital workplaces. *IEEE Transactions on Engineering Management, 71*, 810-823. https://doi.org/10.1109/TEM.2021.3138503

Chubb, J., Reed, D., & Cowling, P. (2022). Expert views about missing AI narratives: Is there an AI story crisis? *AI and Society*. https://doi.org/10.1007/s00146-022-01548-2

Davis, R., Shrobe, H., & Szolovits, P. (1993). What is a knowledge representation? *American Association for Artificial Intelligence*. https://courses.csail.mit.edu/6.803/pdf/davis.pdf

De Ayala, R. J. (2013). *The theory and practice of item response theory*. Guilford Publications. https://doi.org/10.12691/rpbs-7-1-4

Ding, L., Li, T., Jiang, S., & Gapud, A. (2023). Students' perceptions of using ChatGPT in a physics class as a virtual tutor. *International Journal of Educational Technology in Higher Education*, *20*(1), 63. https://doi.org/10.1186/s41239-023-00434-1

Ding, L., Li, T., & Turkson, A. (In review). (Mis)conceptions and perceptions of artificial intelligence: A scoping review. *Manuscript Submitted for Publication*.

Goel, A. K., & Davies, J. (2011). *Artificial intelligence*. Cambridge University Press. https://doi.org/10.1017/CBO9780511977244.024

Gorsuch, R. L. (1973). Using Bartlett's significance test to determine the number of factors to extract. *Educational and Psychological Measurement, 33*(2), 361-364. https://doi.org/10.1177/001316447303300216

Hautea, S., Dasgupta, S., & Hill, B. M. (2017). Youth perspectives on critical data literacies. In proceedings of the *Conference on Human Factors in Computing Systems* (pp. 919-930). https://doi.org/10.1145/3025453.3025823

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

IBM Corp. (2021). *IBM SPSS Statistics for Windows*. IBM Corp.

Jones-Jang, S. M., & Park, Y. J. (2023). How do people react to AI failure? Automation bias, algorithmic aversion, and perceived controllability. *Journal of Computer-Mediated Communication, 28*(1), 1-8. https://doi.org/10.1093/jcmc/zmac029

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*, 31-36. https://doi.org/10.1007/BF02291575

Kline, P. (2015). *A handbook of test construction (psychology revivals): Introduction to psychometric design*. Routledge. https://doi.org/10.4324/9781315695990

Kong, S.-C., Man-Yin Cheung, W., & Zhang, G. (2023). Evaluating an artificial intelligence literacy program for developing university students' conceptual understanding, literacy, empowerment, and ethical awareness. *Educational Technology & Society, 26*(1), 16-30. https://doi.org/10.30191/ETS.202301_26(1).0002

Laupichler, M. C., Aster, A., Schirch, J., & Raupach, T. (2022). Artificial intelligence literacy in higher and adult education: A scoping literature review. In *Computers and education: Artificial intelligence*. Elsevier. https://doi.org/10.1016/j.caeai.2022.100101

Lee, I., Ali, S., Zhang, H., Dipaola, D., & Breazeal, C. (2021). Developing middle school students' AI literacy. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education* (pp. 191-197). ACM. https://doi.org/10.1145/3408877.3432513

Lindner, A., & Berges, M. (2020). Can you explain AI to me? Teachers' pre-concepts about artificial intelligence. In *Proceedings of the 2020 IEEE Frontiers in Education Conference* (pp. 1-9). IEEE. https://doi.org/10.1109/FIE44824.2020.9274136

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

Long, D., Teachey, A., & Magerko, B. (2022). Family learning talk in AI literacy learning activities. In *Proceedings of the Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3491102.3502091

Maitz, K., Fessl, A., Pammer-Schindler, V., Kaiser, R., & Lindstaedt, S. (2022). What do construction workers know about artificial intelligence? An exploratory case study in an Austrian SME. In *Proceedings of the ACM International Conference* (pp. 389-393). https://doi.org/10.1145/3543758.3547545

McBride, C. (2015). *Children's literacy development*. Routledge. https://doi.org/10.4324/9781315849409

McCarthy, J. (2007). From here to human-level AI. *Artificial Intelligence, 171*(18), 1174-1182. https://doi.org/10.1016/j.artint.2007.10.009

Mertala, P., Fagerlund, J., & Calderon, O. (2022). Finnish 5th and 6th grade students' pre-instructional conceptions of artificial intelligence (AI) and their implications for AI literacy education. *Computers and Education: Artificial Intelligence, 3*. https://doi.org/10.1016/j.caeai.2022.100095

Mondal, B. (2020). Artificial intelligence: State of the art. In V. E. Balas, R. Kumar, & R. Srivastava (Eds.), *Recent trends and advances in artificial intelligence and internet of things* (pp. 389-425). https://doi.org/10.1007/978-3-030-32644-9_32

Nader, K., Toprac, P., Scott, S., & Baker, S. (2022). Public understanding of artificial intelligence through entertainment media. *AI and Society, 39*, 713-726. https://doi.org/10.1007/s00146-022-01427-w

Ng, A. (2023). *AI for everyone*. DeepLearning.

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021a). AI literacy: Definition, teaching, evaluation and ethical issues. *Proceedings of the Association for Information Science and Technology, 58*(1), 504-509. https://doi.org/10.1002/pra2.487

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

Olari, V. (2023). Introducing artificial intelligence literacy in schools: A review of competence areas, pedagogical approaches, contexts and formats. In T. Keane, C. Lewin, T. Brinda, & R. Bottino (Eds.), *Towards a collaborative society through creative learning*. Springer. https://doi.org/10.1007/978-3-031-43393-1_21

Pett, M. A., Lackey, N. R., & Sullivan, J. J. (2003). *Making sense of factor analysis: The use of factor analysis for instrument development in health care research*. SAGE. https://doi.org/10.4135/9781412984898

Prado, J. C., & Marzal, M. Á. (2013). Incorporating data literacy into information literacy programs: Core competencies and contents. *Libri, 63*(2), 123-134. https://doi.org/10.1515/libri-2013-0010

PwC Global. (2023). *Global artificial intelligence study: Sizing the prize*. https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html

Rizopoulos, D. (2006). ITM: An R package for latent variable modeling and item response theory analyses. *Journal of Statistical Software, 17*(5), 1-25. https://doi.org/10.18637/jss.v017.i05

Rosenman, R., Tennekoon, V., & Hill, L. G. (2011). Measuring bias in self-reported data. *International Journal of Behavioral and Healthcare Research, 2*(4), 320. https://doi.org/10.1504/IJBHR.2011.043414

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1-36. https://doi.org/10.18637/jss.v048.i02

Rovinelli, R. J., & Hambleton, R. K. (1977). On the use of content specialists in the assessment of criterion-referenced test item validity. *Dutch Journal for Educational Research, 2*(2), 49-60.

RStudio Team. (2020). *RStudio: Integrated development for R*. http://www.rstudio.com/

Saxton, E., Burns, R., Holveck, S., Kelley, S., Prince, D., Rigelman, N., & Skinner, E.A. (2014). A common measurement system for K-12 STEM education: Adopting an educational evaluation methodology that elevates theoretical foundations and systems thinking. *Studies in Educational Evaluation, 40*, 18-35. https://doi.org/10.1016/j.stueduc.2013.11.005

Steiger, J. H. (1980). Statistically based tests for the number of common factors. In *Proceedings of the Annual Meeting of the Psychometric Society*.

Su, J., Ng, D. T. K., & Chu, S. K. W. (2023). Artificial intelligence (AI) literacy in early childhood education: The challenges and opportunities. *Computers and Education: Artificial Intelligence, 4*. https://doi.org/10.1016/j.caeai.2023.100124

Touretzky, D., Gardner-Mccune, C., Martin, F., & Seehorn, D. (2019). Envisioning AI for K-12: What should every child know about AI? In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence* (pp. 9795-9799). https://doi.org/10.1609/aaai.v33i01.33019795

Turner, R. C., & Carlson, L. (2003). Indexes of item-objective congruence for multidimensional items. *International Journal of Testing, 3*(2), 163-171. https://doi.org/10.1207/S15327574IJT0302_5

Williams, B. (1978). *A sampler on sampling*. John Wiley & Sons.

Wong, G. K. W., & Huan, J. (2020). Broadening artificial intelligence education in K-12: Where to start? *ACM Inroads, 11*(1), 20-29. https://doi.org/https://doi.org/10.1145/3381884

Yue, M., Jong, M. S. Y., & Dai, Y. (2022). Pedagogical design of K-12 artificial intelligence education: A systematic review. *Sustainability, 14*(23), 15620. https://doi.org/10.3390/su142315620