Journal of Learning and Teaching in Digital Age, 2025, **10**(1), 58-67
https://dergipark.org.tr/en/pub/joltida
ISSN: 2458-8350 (online)

Research Paper

Artificial Intelligence Literacy Scale: A Study of Reliability and Validity for a Sample of Turkish University Students

Arzu Deveci Topal*[a], Asiye Toker Gökçe[b], Canan Dilek Eren[c], Aynur Kolburan Geçer[d]

[a](ORCID ID: 0000-0001-5090-8592), Faculty of Education, Department of Computer and Instructional Technologies Education, Kocaeli University, Turkey, adeveci@kocaeli.edu.tr
[b](ORCID ID: 0000-0003-1909-1822), Faculty of Education, Department of Educational Science, Kocaeli University, Turkey, asi.gokce@kocaeli.edu.tr
[c](ORCID ID: 0000-0002-7004-5066), Faculty of Education, Department of Science Education, Kocaeli University, Turkey, canandilek@kocaeli.edu.tr
[d](ORCID ID: 0000-0002-2000-9526), Faculty of Education, Department of Computer and Instructional Technologies Education, Kocaeli University, Turkey, akolburan@kocaeli.edu.tr
*Corresponding author

ARTICLE INFO

Received: 23 February 2024
Revised: 06 May 2024
Accepted: 11 June 2024

Keywords:
Artificial intelligence literacy scale
Artificial intelligence competencies
Confirmatory factor analysis
Psychometric measurement
Scale adaptation

doi: 10.53850/joltida.1440845

ABSTRACT

This study aims to adapt to Turkish the "Scale for the assessment of non-experts: AI literacy" developed by Laupichler et al. (2023a). The scale consists of 31 items with three sub-dimensions: technical understanding, critical thinking, and practical applications. The data required for the validity and reliability study of the scale were collected from 642 undergraduate and graduate students studying in different departments of a state university in the fall semester of the 2023-2024 academic year. First, CFA was applied to the data according to the factor structure in the original scale, but as acceptable fit values could not be obtained because of the analysis, exploratory factor analysis was performed. In the reliability analysis of the factor structure determined by EFA, KMO was calculated as =0.948. It was determined that the scale items were collected in three factors and explained 61.1% of the total variance ("critical thinking" is 25.8%, "technical knowledge" is 25.2%, and "practical applications" explains 10.2% of the total variance). As a result of EFA, it was seen that the sub-dimensions of some of the items in the original scale had changed, and since the factor load values of the three items were very close to each other, they were removed from the scale. Because of CFA, which was conducted to evaluate whether the data supported the hypothesized relationships between the measured variables, Cronbach's alpha value was found to be 0.90. As a result of the CFA analysis conducted with the 3 sub-dimensions and 28 items in the scale, the Chi-square value (X^2=2.85; df=345, N=317, p< .001), which is the fit index of the model, has a good fit and is significant, SRMR = 0.0545and RMSEA = 0.077 values and fit indices and the model has an acceptable fit.

INTRODUCTION

The rapid proliferation of artificial intelligence (AI) technologies has brought AI to the forefront of modern human life. AI has become an integral part of various aspects of society, including healthcare, education, finance, and entertainment (Igami, 2020). For example, AI is being used in healthcare to improve diagnostic accuracy and treatment outcomes (Larrazabal et al., 2020). In education, AI is being used to personalize learning experiences and deliver targeted interventions to students (Dai et al., 2020). In finance, AI algorithms are used for fraud detection and risk assessment (Curtis et al., 2022). These examples highlight the need for individuals to become AI literate to be effective members of a society that increasingly relies on AI technologies.

AI literacy refers to the knowledge and skills needed to understand, use, and critically evaluate AI technologies (Casal-Otero et al., 2023). It encompasses several dimensions, including knowing and understanding AI concepts, applying AI in practical settings, evaluating the ethical implications of AI, and being aware of the limitations and biases of AI systems (Zhao et al., 2022). AI literacy is not limited to technical expertise but also includes the ability to think critically, make informed decisions, and engage in ethical discussions related to AI (Faruqe et al., 2021).

The significance of AI literacy is obvious given the necessity for individuals to possess expertise and understanding as both consumers and users of AI technologies. Without AI literacy, individuals may be susceptible to misinformation, biased algorithms, and unethical practices. For example, in the field of medical imaging, a lack of AI literacy can lead to biased classifiers that produce inaccurate diagnoses (Larrazabal et al., 2020). Similarly, in the context of online exam monitoring technologies, a lack of AI literacy can lead to ethical concerns related to privacy and autonomy (Coghlan et al., 2021).

The need for AI literacy scales arises from the recognition that AI literacy is a multidimensional construct that can be measured and assessed. AI literacy scales provide a framework for assessing individuals' knowledge, skills, and attitudes related to AI. These scales can be used to identify gaps in AI literacy, design targeted interventions, and track the effectiveness of AI literacy programs (Casal-Otero et al., 2023). Examples of AI literacy scales include measures of students' perceptions of their knowledge and skills in applying AI technology (Ng et al., 2021) and instruments that assess conceptualizations and competencies about conversational

agents (Wienrich & Carolus, 2021). In addition, Laupichler et al. (2023a) developed a scale to assess the AI literacy of non-experts in a comprehensive study.

The rationale for developing an AI literacy scale is rooted in the need to ensure that individuals are equipped with the knowledge and skills necessary to navigate an AI-driven society. As Faruqe et al. (2021) highlighted, by measuring and assessing AI literacy, educators, policymakers, and researchers can identify areas for improvement and develop targeted interventions to increase AI literacy. In addition, as Ng et al. (2021) pointed out, AI literacy scales can contribute to the development of a standardized framework for AI literacy, which can facilitate cross-cultural and cross-disciplinary comparisons.

The development of an AI literacy scale is essential for promoting equitable access to AI literacy education. By providing a standardized measure of AI literacy, individuals and institutions can identify and address disparities in AI literacy among different populations. This can help ensure that individuals from diverse backgrounds have equal opportunities to develop the knowledge and skills necessary to participate in an AI-driven society. In addition, as highlighted by Cox and Mazumdar (2022), the development of an AI literacy scale can help establish best practices and guidelines for AI literacy education. According to Laupichler vd. (2023b), AI literacy scales developed to assess the status quo of individuals' AI knowledge can be used to evaluate the quality of AI courses.

This study aims to adapt to the scale (Scale for the assessment of non-experts' AI literacy) developed by Laupichler et al. (2023a), which aims to determine the AI literacy of individuals who have "not received formal training in AI and use AI applications rather than developing them" in Turkish. This tool provides a comprehensive assessment in Turkish to measure individuals' skills and understanding of AI technologies that have become irreplaceable in society. The adapted scale is expected to be used in the future by educators, policymakers, and researchers in Turkish-speaking countries to establish a standardized framework for AI literacy and AI courses.

METHOD

In this study, the Scale for the Assessment of Non-Experts' AI Literacy (SNAIL)" developed by Laupichler et al. (2023a) was adapted to Turkish culture by examining the technical features (validity and reliability evidence) of the scale.

Population and Sample

The data required for the validity and reliability study of the scale for assessing the AI literacy of nonexperts were collected from undergraduate and graduate students studying in different departments of a state university in September 2023. The necessary ethical and official permissions (456968 number) were obtained before data collection. A total of 642 people participated in the study. While 325 of the collected data were used in exploratory factor analysis, 317 were used in confirmatory factor analysis. Child (2006) stated that the sample size should be at least five times the number of observed variables, and Büyüköztürk (2002) stated that 200 people are sufficient to obtain a reliable factor analysis result and that a large sample should be used to obtain better results. To ensure the validity of the adapted scale, CFA should be conducted with at least 300 participants (Büyüköztürk, 2012; Seçer, 2015). The demographic information of the students is given in Table 1.

Table 1. *Demographic characteristics of students who participated in the EFA and CFA analyses*

	EFA		CFA			EFA		CFA	
Gender	**N**	**F(%)**	**N**	**F (%)**	**Using an AI application**	**N**	**F(%)**	**N**	**F(%)**
Female	219	67,4	215	67,8	Yes	237	72,9	231	72,9
Male	96	32,6	102	32,2	No	88	27,1	86	27,1
Faculty	**N**	**F(%)**	**N**	**F(%)**	**Taking an AI course**	**N**	**F(%)**	**N**	**F(%)**
Education	160	49,2	164	51,7	Yes	20	6,2	25	7,9
Arts and science	31	9,5	32	10,1	No	305	93,8	292	92,1
Economics and Administrative Sciences	15	4,6	5	1,6	**AI interaction frequency**	**N**	**F(%)**	**N**	**F(%)**
Engineering	33	10,2	41	1 2,9	Never	88	27,1	86	27,1
Technology	30	9,2	30	9,5	Almost never	38	11,7	45	14,2
Medical	34	10,5	30	9,5	Every fortnight	63	19,4	64	20,2
Postgraduate	22	6,8	15	4,7	Once a week	70	21,5	63	19,9
					Every day	66	20,3	59	18,6

As shown in Table 1, most of the participants were female and undergraduate students. In addition, most of the participants were studying at the Faculty of Education.

Adaptation of the Scale to Turkish

For the adaptation of the scale, permission was obtained from the authors via e-mail, and communication and cooperation with the authors was ensured at every step of the adaptation process. To ensure language validity, the original English form was first translated into Turkish by four field experts (working in the fields of educational administration, science education, computer and instructional technology education) who have a good command of the Turkish and English languages. Afterwards, the translators came together to discuss the form, and a common decision was reached on the statements that differed. The agreed Turkish form was examined by two experts from the field of Foreign Language Education, and the Turkish form was finalized by making some corrections to the statements. Based on the expert opinion, the Turkish form was translateback into English by two experts who have a good command of both Turkish and English. The original version of the scale and the translated version of the scale were analyzed by two different field experts, and a common opinion was reached that the two were similar. In addition, the original

German version of the scale was obtained from the researchers who developed it (Laupichler et al. 2023a), and the German items were compared with the Turkish translations.

The pilot application of the scale was conducted with 12 undergraduate and graduate students, and their opinions on the items that were not understood, unclear, or inadequately expressed were considered.

Data Collection Tool

The data required for the study were obtained from the "Scale for the assessment of non-experts' AI literacy (SNAIL)", which was developed by Laupichler et al. (2023a) to determine the AI literacy of individuals with no education in AI or computer science and consists of 31 items. The translated version into Turkish was used. In addition to 31 items, demographic information consisting of gender, age, education level, use of artificial intelligence applications, applications used, receiving training on artificial intelligence, and frequency of interaction with artificial intelligence were included in the scale. The original scale has a 7-point Likert-type rating ranging from "strongly disagree" (one) to "strongly agree" (seven). In the literature, there are different opinions about 5-point and 7-point options in Likert-type scales. In terms of the ease of answering the scale, it is suggested that scales with more options (e.g. 7 options) will take time to fill in (Köklü 1995); in addition, since the psychological distances between the options in the scales are greater than the scales with 4 and 5 options, the use of 7 options is not recommended for socially negative issues (Wakita et al., 2012). In addition, it is emphasized in the literature that having more than 5 options in answering the scale items makes it difficult for the respondents to understand the differences in similar answer options and to choose the appropriate option for themselves (Nadler et al., 2015). Bora Semiz and Altunışık (2016) state that the increase in the number of options decreases the midpoint and endpoint orientations, increases the marking rate of intermediate options, and prevents the respondents from marking "strongly agree" or "strongly disagree" options. In this adapted study, the scale has been used as a 5-point Likert scale with the following order: 1= "strongly disagree", 2= disagree, 3= somewhat agree, 4= agree and 5= "strongly agree". There are no reverse scored items. The scale consists of three subscales: Technical knowledge (14 items), practical applications (7 items), and critical evaluation (10 items). Cronbach's alpha internal consistency coefficients of the subscales are .93, .85 and .91, respectively. The high score obtained from the scale indicates that the level of AI literacy is high.

Analyzing the Data

Exploratory factor analysis is used to reduce variables, identify emerging factors, and reveal whether factors emerging because of factor analysis are similar to latent variables. Confirmatory factor analysis is used to test whether the structure in question can be verified with the data obtained from the measurement tool developed in line with a theoretical structure (Çokluk, Şekercioğlu and Büyüköztürk, 2016). In this context, CFA was first applied to the data according to the factor structure in the original scale. Because of CFA, the factor loading value between critical thinking and practical applications was found to be 1.00 despite the modifications. This situation, which is defined as multiple correlation, shows that there is a high degree of correlation (75% and above) between some of the independent variables (Vupa & Görünlü Alma, 2008). Such high correlation values mean that the sub-factors measure the same skill, i.e., they are combined in the same factor. In addition, the CFI, TLI, NFI, and GFI values, which should be 0.9 and above (Tabachnick and Linda, 2013), were found to be close to 0.8. These values do not correspond to acceptable fit values. Since these reference values could not provide construct validity, in other words, they were not acceptable, the opinions of four academicians working in the field of measurement and evaluation were taken, and it was decided to perform exploratory factor analysis (EFA) by obtaining permission from the owners of the scale in line with the opinions received.

In the reliability analyses of the factor structure determined by EFA, item-total correlations, Cronbach's alpha coefficients, and inter-factor correlations were calculated. To test whether the latent structure was verified with the relevant data set and the suitability of the model, fit indices, kurtosis and skewness coefficients, convergent and divergent validity ratios, Cronbach's alpha internal consistency coefficients for the reliability of the sub-dimensions, item-total correlations, and the significance of the differences between the factor and item mean scores of the upper 27% and lower 27% groups were evaluated using an independent sample t-test.

To obtain accurate and reliable CFA results, the multivariate normal distribution of the data was examined using the maximum likelihood estimation method. With this analysis, we checked whether the observed and latent variables had multiple normal distributions. In cases where multivariate normal distribution is violated, the Chi-square value will be high, the result will be significant, and the model will be rejected even if it is correct (Ayyıldız & Cengiz, 2006). The larger the sample, the higher the probability of significant chi-square analysis results (Büyüköztürk, Akgün, et al. 2004). When a multivariate normal distribution is not provided, the measurement errors in the model will take lower values than normal, so the path coefficients will have more significance values than they should (Ayyıldız & Cengiz, 2006). Data with values in the range of +1.5 or -1.5 for skewness and kurtosis are considered to have a normal distribution (Tabachnick & Linda, 2013). Because of the normality analysis of the AI literacy scale of non-experts, it was determined that the scale had a normal distribution

To ensure the construct validity of the scale, convergent and divergent validity values were calculated in addition to the KMO value. Convergent validity refers to the relationship between the expressions in the variables and the factors they form (Coşkun et al., 2010). To ensure convergent validity, CR > 0.7, AVE > 0.5, and CR>AVE (AVE: average variance extracted, CR: composite reliability). To ensure divergent validity, the items in the factor should be less related to other factors (Yaşlıoğlu, 2017). To ensure divergent validity, maximum shared variance (MSV) and average squared variance (ASV) were calculated and ASV< MSV<AVE<CR. SPSS and AMOS software were used to analyze the data. Details of the analyses used in this study are given in the Findings section.

FINDINGS

The results of exploratory and confirmatory factor analyses conducted to examine the construct validity of the scale are given below.

Exploratory Factor Analysis

In the exploratory factor analysis conducted to examine the construct validity of the scale, Kaiser–Meyer– Olkin (KMO) and Bartlett tests were performed to test the suitability of the data for factorization. The fact that the sample suitability coefficient (KMO) >= .60 and the Bartlett Sphericity test were significant indicates that exploratory factor analysis can be performed (Çokluk, Şekercioğlu, & Büyüköztürk, 2016). When the 31 items in the original scale were subjected to factor analysis, KM =0.951; X2=7317.24, df=465 p=.000 and it was determined that it was suitable for factor analysis.

In the exploratory factor analysis, all vertical and oblique rotation methods were used. The reason for using all rotation methods was to obtain factors with large variance and appropriate to the original scale without losing any items. Because of the trials, it was decided to use the varimax rotation technique, which gave the best result. As a result of this analysis, item 3 (I can explain how artificial intelligence applications make decisions) and item 10 (I can explain how some artificial intelligence systems can move in their environment and react to their environment) in the technical understanding section and item 1 (I can count the weaknesses of artificial intelligence) in the critical thinking sub-dimension were removed from the scale because their factor loadings were very close to each other and the analysis was repeated.

As a result of the second analysis, as a result of KMO and Barlett Sphericity Test, KMO=0,948; X^2= 6466,647; df=378 and p=.000. As a result of this analysis, it was determined that the scale items were gathered in three factors and explained 61.1% of the total variance. When the variances of the factors were analyzed, the first factor explained 25.8% of the total variance, the second factor explained 25.2% of the total variance, and the third factor explained 10.2% of the total variance. The EFA results and factor loadings are presented in Table 2.

Table 2. *EFA results and factor loadings*

Factors	Item numbers as a result of EFA	Items in the original scale	Factor loadings	Factor explainers	Cronbach alpha
Critical thinking	c1	c9	,826	%25,8	0,936
	c2	c8	,791		
	c3	p5	,764		
	c4	c6	,727		
	c5	c7	,718		
	c6	c10	,694		
	c7	p7	,691		
	c8	c5	,654		

	c9	c4	,653		
	c10	c3	,649		
	c11	p3	,555		
	c12	c2	,547		
	c13	p6	,545		
	t1	t8	,834		
	t2	t5	,783		
	t3	t12	,782		
	t4	t7	,781		
	t5	t6	,762		
Technical understanding	t6	t9	,759	%25,2	0,939
	t7	p4	,743		
	t8	t4	,671		
	t9	t13	,670		
	t10	t14	,644		
	t11	t11	,623		
	p1	p1	,736		
	p2	p2	,733		
Practical applications	p3	t1	,647	%10,2	0,788
	p4	t2	,495		
Total factor explainers				**%61,1**	**0,956**

As seen in Table 2, it was observed that some of the items of the original scale were collected in different dimensions; therefore, the dimensions were changed in the Turkish sample. Because of the examination and expert opinions (instructors working in the fields of computer and instructional technologies, science education and measurement and evaluation), it was decided that the skills measured by these items were more appropriate for the new sub-dimensions. The 5th, 6th, and 7th items in the "practical applications" sub-dimension of the original scale were included in the "critical view" sub-dimension, and the 4th item was included in the "technical understanding" dimension. It was determined that the first and second items in the "technical understanding" dimension of the original scale were included in the "practical applications" dimension. It was determined that the "critical view" sub-dimension consisted of 13 items and the load values of the items ranged between.545 and.826, the technical understanding dimension consisted of 11 items and the item load values ranged between.623 and.834, and the practical applications sub-dimension consisted of 4 items and the item load values ranged between.495 and.736. In addition, as shown in Table 3, there was a significant positive relationship between the sub-dimensions of the scale.

Table 3. *Average and standard deviations of the scale and correlation values between the factors*

Factors	**Mean**	**S.D**	**CA**	**PA**	**TU**
Critical appraisal (CA)	38,62	11,47	1	,668**	,627**
Practical application (PA)	11,69	3,51		1	,552**
Technical understanding (TU)	22,43	9,47			1

p<.01

Confirmatory Factor Analysis

On the scale of "Artificial intelligence literacy for non-experts" for which exploratory factor analysis was performed, CFA analysis was performed to evaluate whether the hypothesized relationships between the measured variables were supported by the data. The purpose of CFA is not to define the factor structure but to analyze the extent to which the result obtained by testing all observed and unobserved variables together is consistent with the available data (Özdamar, 2016). The analysis was conducted with another group

consisting of 317 participants different from the EFA group. Because of the reliability analysis conducted with all items before the confirmatory factor analysis, Cronbach's alpha value was found to be 0.90 and interpreted as having a high level of reliability.

Because the model fit criteria obtained as a result of the first CFA analysis with three subdimension and 28 items were not within the desired limits, modification indices were examined. The standardized regression coefficients and fit index values were examined. The values in the first row of Table 4 indicate that the fit is insufficient. Therefore, when the error covariances between items c1–c2 and c9–c10 were analyzed according to the modification values, a significant relationship was determined and a link was established between them (0.53 and 0.42, respectively). This means that the item pairs are under the same latent variable and are close in meaning. It may be considered to remove these items that measure the same feature, but the high error correlations observed between the items whose accuracy was determined by expert opinion were added to the model and the analysis continued. When the fit indices of the model were examined because of the analysis, the Chi-square value (χ^2/df=2,85, sd=345, N=317, p< .001) has a good fit and was significant, and the model had an acceptable fit when SRMR and RMSEA values and fit indices were considered.

Table 4. *Fit values were obtained because of CFA.*

	χ^2/df	RMSEA	SRMR	CFI	TLI	NFI	AIC	ECVI
First Analysis	1135,890/347=3,273	0,085	0.0556	0,893	0,883	0,853	1309,890	4,145
Post-modification analysis	984,616/345=2,854	0.077	0.0545	0.913	0.905	0.873	1162,616	3,679

Because of the CFA, it was determined that the path coefficients of the items in all subdimension were significant and that the factor loadings were at a good level. According to Harrington (2009), factor loadings should be above 0.30. When Table 5 and Figure 1 are analyzed, the factor loading values for all items of the scale vary between 0.57 and 0.89, and these values are significant. According to the standardized path coefficients, c11 (β0 =0.807) has the highest effect in the critical view dimension, t5 (β0 =0.897) in technical understanding and p3 (β0 =0,784) in practical applications dimension.

The critical ratio (C.R.) is determined by dividing the parameter estimate by the standard error and should be greater than +1.96 or -1.96 at the 0.05 significance level (Khine, 2013). The standard error (S.E.) value shows the difference between the actual value of the measured trait and the observed measurement result, and the closer it is to zero, the more accurate the estimate (Office for National Statistics, 2023). The S.E, C.R, and P values in Table 5 and Figure 1 fulfill the desired conditions and show that all parameters are statistically significant (P<0.001).

Table 5. Standard and nonstandard path coefficients obtained from the CFA analysis

Items	Factors	β_0	β_1	S.E.	C.R.	P
c1	CA	0,794	1			
c2	CA	0,784	0,997	0,043	22,919	***
c3	CA	0,767	0,911	0,06	15,195	***
c4	CA	0,804	0,955	0,059	16,184	***
c5	CA	0,731	0,846	0,059	14,284	***
c6	CA	0,782	0,991	0,064	15,595	***
c7	CA	0,801	0,985	0,061	16,095	***
c8	CA	0,719	0,901	0,064	13,985	***
c9	CA	0,778	0,967	0,062	15,481	***
c10	CA	0,767	0,937	0,062	15,187	***
c11	CA	0,807	1,031	0,063	16,259	***
c12	CA	0,704	0,846	0,062	13,624	***
c13	CA	0,771	0,929	0,061	15,292	***
t1	TU	0,885	1			
t2	TU	0,877	0,993	0,043	22,86	***
t3	TU	0,768	0,854	0,048	17,63	***
t4	TU	0,884	0,987	0,042	23,256	***
t5	TU	0,897	1,054	0,044	24,046	***
t6	TU	0,768	0,894	0,051	17,633	***
t7	TU	0,761	0,858	0,049	17,389	***
t8	TU	0,843	0,978	0,047	20,982	***
t9	TU	0,727	0,871	0,054	16,086	***
t10	TU	0,79	0,959	0,052	18,559	***
t11	TU	0,814	1,027	0,052	19,604	***
p1	PA	0,573	1			

p2	PA	0,695	1,348	0,146	9,217	***
p3	PA	0,784	1,494	0,151	9,886	***
p4	PA	0,783	1,545	0,156	9,884	***

β_0 : Standardized path coefficients, β_1: Non-standardized path coefficients, p<.001

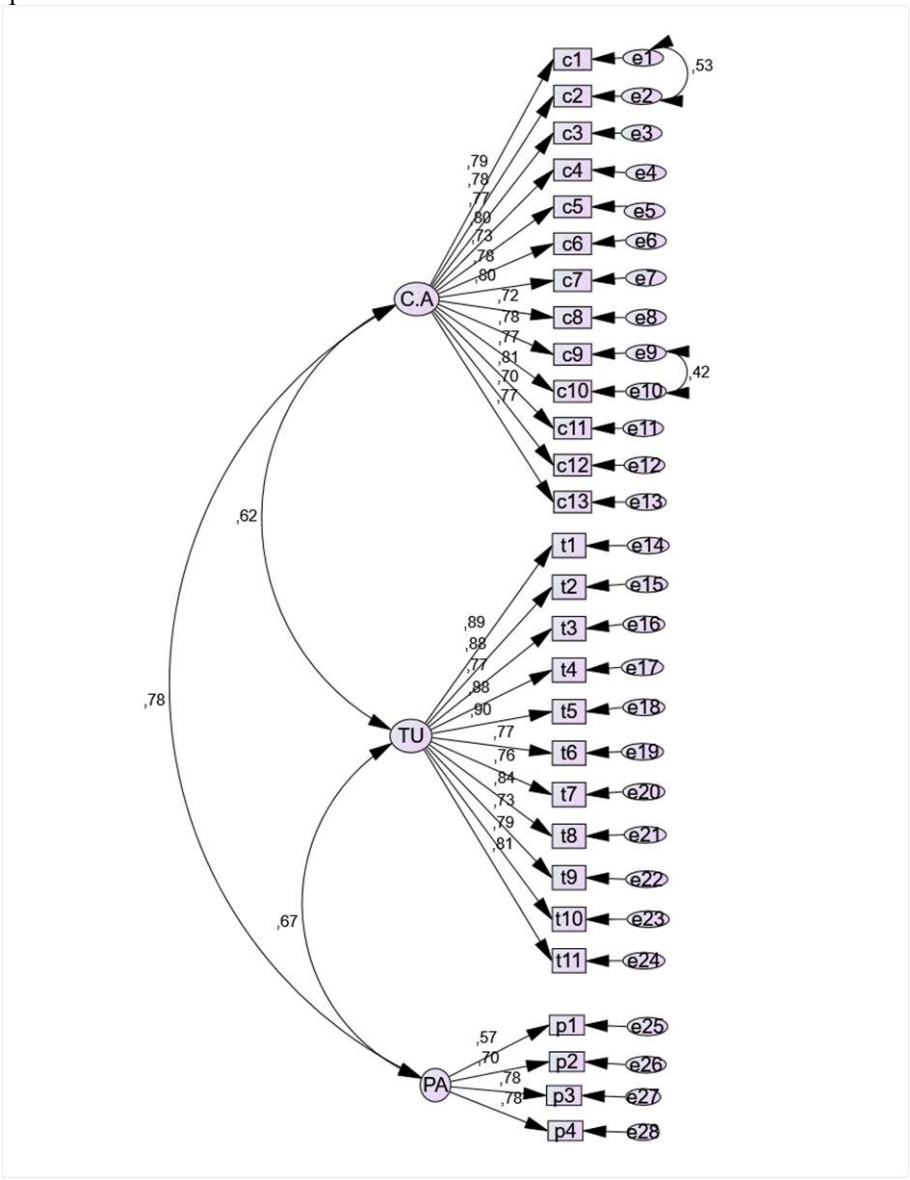

Figure 1. *Standardized path coefficients of the scale*

For the reliability of the sub-dimensions analyzed by CFA, Cronbach's alpha internal consistency coefficients, item-total correlations, and the significance of the differences between the factor and item mean scores of the upper 27% and lower 27% groups were calculated by independent sample t-test (Table 6). According to the results of the analysis, the corrected total correlations of the scale items ranged between 0.44 and 0.78. Because of the t-test between the item mean scores of the lower 27% and upper 27% groups, the differences were found to be significant for all items and total subscale scores. This result shows that all items and subscales in the scale are discriminative.

Table 6. *Cronbach's alpha internal consistency coefficients of the scale, item total correlations, and independent sample t-test findings between the upper 27% and lower 27% scores*

Factors	Item	Corrected item–total correlation	T (Low%27-High%27)	Factors	Item	Corrected item–total correlation	T (Low%27-High%27)
Critical aprasial	c1	0,7	-16,797	Technical understanding	t1	0,726	-14,081
	c2	0,713	-17,829		t2	0,72	-15,165
	c3	0,637	-13,523		t3	0,656	-12,46
	c4	0,691	-14,887		t4	0,715	-14,508
	c5	0,684	-14,563		t5	0,735	-14,773
	c6	0,703	-16,053		t6	0,666	-12,868
	c7	0,712	-15,23		t7	0,628	-11,009
	c8	0,663	-14,174		t8	0,714	-14,847
	c9	0,743	-18,515		t9	0,668	-14,855
	c10	0,748	-16,278		t10	0,667	-12,552
	c11	0,749	-19,803		t11	0,777	-19,77
	c12	0,638	-13,831	Practical applications	p1	0,441	-8,148
	c13	0,707	-16,672		p2	0,585	-12,038
					p3	0,672	-14,126
					p4	0,654	-13,841

The reliability Cronbach's alpha coefficient for the whole scale was 0.963, and the coefficients of the subscales were 0.95, 0.96, and 0.80, respectively (Table 7). The KMO value used to verify the scale was 0.958. These results show that the scale is valid and reliable in this form.

Table 7. *Reliability and validity analysis (convergent and discriminant) of the scale and its sub-factors*

Factors	Cronbach's alpha	AVE	CR	MSV	ASV
Critical appraisal	0,951	0,60	0,95	0,61	0,50
Technical understanding	0,958	0,51	0,80	0,45	0,42
Practical applications	0,806	0,68	0,96	0,61	0,53
Overall scale	0,963				

It was determined that all CR values calculated to ensure convergent validity were greater than 0.7 and AVE values were > .5. While MSV<AVE in Technical understanding and practical applications dimensions and MSV and AVE values in the critical appraisal dimension are very close to each other, all MSV values are greater than ASV values. However, convergent and divergent validity is provided to a great extent by fulfilling other reliability criteria (Fornell & Larcker, 1981). Based on these results, convergent and divergent validity is provided, and the scale is valid and reliable.

DISCUSSION AND CONCLUSION

In today's world where information and communication technologies are rapidly changing and developing, while the concept of AI literacy remains on the agenda, ensuring the integration of AI into learning/teaching processes and making individuals AI literate is seen as an important competence. In this context, the study adapted the 31-item "Artificial Intelligence Literacy Scale" developed by Laupichler et al. (2023) into Turkish.

First, CFA was applied to the data according to the factor structure in the original scale. Because acceptable fit values could not be obtained because of the analysis, we decided to perform exploratory factor analysis (EFA) in line with expert opinions. In the reliability analyses of the factor structure determined by EFA, item-total correlations, Cronbach's alpha coefficients, and inter-factor correlations were calculated. As a result of KMO and Barlett Sphericity Test, KMO=0,948; X2= 6466,647; df=378 and p=.000. As

a result of this analysis, it was determined that the scale items were gathered in three factors and explained 61.1% of the total variance. In the scale developed by Lapuchier et al. (2023), the total variance explained was 57%. It can be stated that the results are similar. When the variances of the factors were analyzed, the first factor, critical thinking, explained 25.8% of the total variance, the second factor, technical knowledge, explained 25.2% of the total variance, and the third factor, practical applications, explained 10.2% of the total variance.

Because of the analysis, it is noticeable that the sub-dimensions of some items in the original scale have changed. Because of the examination and expert opinions, it was decided that the skills measured by these items were more appropriate for the new sub-dimensions. The 5th, 6th, and 7th items in the "practical applications" sub-dimension were included in the "critical view" sub-dimension, and the 4th item was included in the "technical knowledge" dimension. The first and second items in the "technical knowledge" dimension were determined to be in the "practical applications" dimension. This may be due to the differences between German and Turkish cultures.

In the second stage, CFA analysis was conducted with a different group to evaluate whether the hypothesized relationships between the measured variables were supported by the data. Because of the reliability analysis conducted with all items before CFA, Cronbach's alpha value was found to be 0.90 and interpreted as having a high level of reliability. This finding is consistent with Lapuchier et al. Because of the CFA analysis conducted with the 3 sub-dimensions and 28 items in the scale, the model has an acceptable fit when the fit indices of the model, Chi-square value (x^2/df=2.85; sd=345, N=317, p<.001), SRMR=0.0545, and RMSEA=0.077 values and fit indices are considered.

According to the results of the reliability analysis, the corrected total correlations of the scale items ranged between 0.44 and 0.78. Because of the t-test between the item mean scores of the lower 27% and upper 27% groups, the differences were found to be significant for all items and total subscale scores. This result shows that all items and subscales in the scale are discriminative. The reliability Cronbach's alpha coefficient for the whole scale was 0.963, and the coefficients of the subscales were 0.95, 0.96, and 0.80, respectively. These values are at a good level according to the literature (Cortina, 1993). The KMO value used to verify the scale was 0.958. The results of the convergent and divergent validity analyses show that the scale is valid and reliable in this form.

The findings of this study provide evidence for the validity and reliability of the "Scale for Assessing Non-Experts' Artificial Intelligence Literacy" developed by Laupichler et al. (2023a) and adapted into Turkish. Because of the lack of sufficient measurement tools for measuring AI literacy skills and the increasing importance of AI literacy skills today, this scale can be used in future research. The fact that some of the items in the scale developed in Germany are included in different dimensions in Turkey shows that AI literacy changes as we move to the east of Europe. The difference between the findings in the study of Laupichler et al. (2023a) and those in this adaptation study can be attributed to these reasons.

For future studies, it is recommended that the AI literacy scale, which has sufficient validity and reliability evidence, be applied by making measurement invariance in different samples. Thus, the findings to be obtained from different samples regarding AI literacy skills, which have an important place today, will provide support in developing this issue. A validity and reliability study of the scale was conducted on university students. It may be recommended to adapt the artificial intelligence literacy scale for individuals in different age groups.

Ethics Committee: "Ethics Committee Permission" was obtained on 23.08.2023 with the number E-10017888-204.01.07-467360 from Kocaeli University.

REFERENCES

Ayyıldız, H. & Cengiz, E. (2006). A conceptual investigation of structural equation model (SEM) on testing marketing models. *Süleyman Demirel University İ.İ.B.F. 11(*1), 63-84. https://dergipark.org.tr/tr/download/article-file/194860

Bora Semiz, B. & Altunışık R. (2016). An evaluation of various attributes of Likert-likert type scales on response styles in marketing research. *Bartın Üniversitesi İİBF Dergisi, 7* (14), 577-598.

Büyüköztürk, Ş. (2012). Faktör analizi: Temel kavramlar ve ölçek geliştirmede kullanımı. *Kuram ve Uygulamalarda Eğitim Yönetimi, (32),* 470-483.

Büyüköztürk, Ş., Akgün, Ö.E., Özkahveci, Ö. and Demirel, F. (2004). The validity and reliability study of the Turkish version of the Motivated Strategies for Learning Questionnaire. *Educational Scienses: Theory&Practice, 4*(2), 231-239.

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

Child, D. (2006). *The essentials of factor analysis.* Continuum, London.

Coghlan, S., Miller, T., & Paterson, J. (2021). Good proctor or "big brother"? Ethics of online exam supervision technologies. *Philosophy & Technology*, *34*(4), 1581-1606. https://doi.org/10.1007/s13347-021-00476-1

Çokluk Ö, Şekercioğlu G, Büyüköztürk Ş. (2016). *Sosyal Bilimler İçin Çok Değişkenli İstatistik SPSS ve LISREL Uygulamaları* (4. Baskı). Ankara: Pegem Akademi.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied psychology, 78*(1), 98-104.

Coşkun, R., Altunışık, R., Yıldırım, E. & Bayraktaroğlu, S. (2010). *Sosyal Bilimlerde Araştırma Yöntemleri SPSS Uygulamaları*. Sakarya: Sakarya Yayıncılık.

Cox, A. M., & Mazumdar, S. (2022). Defining artificial intelligence for librarians. *Journal of Librarianship and Information Science*, *0*(0). https://doi.org/10.1177/09610006221142029

Curtis, C., Gillespie, N., & Lockey, S. (2023). AI-deploying organizations are key to addressing 'perfect storm'of AI risks. *AI and Ethics*, *3*(1), 145-153. https://doi.org/10.1007/s43681-022-00163-7

Dai, Y., Chai, C. S., Lin, P. Y., Jong, M. S. Y., Guo, Y., & Qin, J. (2020). Promoting students' well-being by developing their readiness for the artificial intelligence age. *Sustainability*, *12*(16), 6597. https://doi.org/10.3390/su12166597

Faruqe, F., Watkins, R., & Medsker, L. (2021). Competency model approach to AI literacy: Research-based path from initial framework to model. *Advances in Artificial Intelligence and Machine Learning; Research, 2*(4), 580-587. DOI: 10.54364/aaiml.2022.1140

Fornell, C.& Larcker, DF. (1981). Evaluating structural equation models with un observable variables and measurement error. *Journal of Marketing Research, 18*(1), 39-50.

Harrington D.(2009). *Confirmatory Factor Analysis.* New York: Oxford University Press; p.21-35.

Igami, M. (2020). Artificial intelligence as structural estimation: Deep Blue, Bonanza, and AlphaGo. *The Econometrics Journal*, *23*(3), S1-S24. https://doi.org/10.1093/ectj/utaa005

Khine, M.S. (2013*). Application of structural equation modeling in educational research and practice.* Sense Publishers, Rotterdam / Boston / Taipei.

Köklü, N. (1995). Tutumların ölçülmesi ve likert tipi ölçeklerde kullanılan seçenekler. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi, 28*(2), 81-93.

Larrazabal, A. J., Nieto, N., Peterson, V., Milone, D. H., & Ferrante, E. (2020). Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis. *Proceedings of the National Academy of Sciences*, *117*(23), 12592-12594.

Laupichler, M.C., Aster, A., Haverkamp, N., & Raupach, T. (2023a). Development of the "Scale for the assessment of non-experts' AI literacy"–An exploratory factor analysis. *Computers in Human Behavior Reports*, *12*, 100338. DOI: 10.1016/j.chbr.2023.100338

Laupichler, M.C., Aster, A., Perschewski, J.O., & Schleiss, J. (2023b). Evaluating AI Courses: A Valid and Reliable Instrument for Assessing Artificial-Intelligence Learning through Comparative Self-Assessment. *Education Sciences*, *13*(10), 978. DOI: 10.3390/educsci13100978

Nadler, J.T., Weston, R. & Voyles, E. C. (2015). Stuck in the middle: The use and interpretation of mid-points in items on questionnaires. *The Journal of General Psychology, 142*(2), 71-89. DOI: 10.1080/00221309.2014.994590

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

Office for National Statistics, (2023). Uncertainty and how we measure it for our surveys. https://www.ons.gov.uk/methodology/methodologytopicsandstatisticalconcepts/uncertaintyandhowwemeasureit

Özdamar K. (2016). *Eğitim, sağlık ve davranış bilimlerinde ölçek ve test geliştirme yapısal eşitlik modellemesi IBM SPSS, IBM SPSS AMOS ve MINITAB uygulamalı.* Eskişehir: Nisan Kitabevi.

Seçer, İ. (2015). Zorbalıkla başa çıkma stratejileri ölçeğinin geliştirilmesi: Geçerlik ve güvenirlik çalışması. *Atatürk Üniversitesi Kazım Karabekir Eğitim Fakültesi Dergisi* (30), 85-105.

Tabachnick, B.G., & Linda S. (2013). *Using multivariate statistics*, Pearson, Boston.

Vupa, Ö., & Gürünlü Alma, Ö. (2008). Doğrusal regresyon çözümlemesinde çoklu bağlantı probleminin sapan değer içeren küçük örneklemlerde bir simülasyon çalışması ile saptanması ve sonuçları. *Selçuk Üniversitesi Fen Fakültesi Fen Dergisi, 2*(32), 41-51.

Wakita, T., Ueshima, N. & Noguchi, H. (2012). Psychological distance between categories in the Likert scale: Comparing different numbers of options. *Educational and Psychological, 72*(4): 533-546. https://doi.org/10.1177/0013164411431162

Wang, B., Rau, P.-L.P., & Yuan, T. (2022). Measuring user competence in using artificial intelligence: validity and reliability of artificial intelligence literacy scale. *Behaviour & Information Technology, 42*(9), 1324–1337. https://doi.org/10.1080/0144929X.2022.2072768

Wienrich, C., & Carolus, A. (2021). Development of an instrument to measure conceptualizations and competencies about conversational agents on the example of smart speakers. *Frontiers in Computer Science*, 70. DOI: 10.3389/fcomp.2021.685277

Yaşlıoğlu, MM. (2017). Sosyal bilimlerde faktör analizi ve geçerlilik: Keşfedici ve doğrulayıcı faktör analizlerinin kullanılması. *İstanbul Üniversitesi İşletme Fakültesi Dergisi, 46*(74), 74-85.

Zhao, L., Wu, X., & Luo, H. (2022). Developing AI literacy for primary and middle school teachers in China: Based on a structural equation modeling analysis. *Sustainability*, *14*(21), 14549.