

Psychometric Properties of the Persian ChatGPT Literacy Scale in Higher Education: A Transcultural Adaptation and Validation Study

Abbas Taghizade^{1*}, Afzal Akbari Balootbangan², Roshan Ahmadi³

¹Department of Educational Sciences, Farhangian University, Tehran, Iran

²Department, of Psychology and Counselling, Farhangian University, Tehran, Iran

³Department of Educational Sciences, Faculty of Humanities, Shahid Rajaei Teacher Training University, Tehran, Iran

ABSTRACT

Background: As ChatGPT becomes embedded in higher education, sound instruments are needed to measure students' literacy in using this technology. Nevertheless, no Persian tool had undergone cultural adaptation and psychometric validation. This study sought to translate and culturally adapt the ChatGPT Literacy Scale (CLS) and examine its psychometric performance among Iranian higher education students.

Methods: A transcultural adaptation and psychometric validation design was implemented from May through December 2025 with 371 students from five Iranian universities. The original 25-item CLS developed by Lee and Park (2024) was translated and culturally adapted according to the World Health Organization procedure for translating and adapting instruments. Two bilingual translators independently produced and reconciled Persian forward translations, after which a third bilingual translator, blinded to the source questionnaire, back-translated the reconciled version into English before expert review and administration. Evaluation included face and content validity, construct validity using exploratory and confirmatory factor analyses, and internal consistency based on Cronbach's alpha and McDonald's omega.

Results: Face validity was supported, as every item had an impact score exceeding 1.5. Content validity was also satisfactory, with mean Content Validity Index (CVI) and Content Validity Ratio (CVR) values of 0.82 and 0.88, respectively. Exploratory Factor Analysis (EFA) identified five factors accounting for 57% of total variance, with item loadings ranging from 0.43 to 0.83. Confirmatory Factor Analysis (CFA) supported the first-order five-factor solution and demonstrated satisfactory model fit ($\chi^2/df = 2.73$, RMSEA = 0.069, GFI = 0.90, CFI = 0.97, NFI = 0.95, NNFI = 0.96, SRMR = 0.05). Internal consistency was satisfactory; Cronbach's alpha and McDonald's omega ranged from 0.72 to 0.89 across subscales, and both were 0.89 for the total scale.

Conclusion: The culturally adapted Persian CLS demonstrated psychometric performance and can be considered a valid and reliable measure of ChatGPT literacy among Iranian higher education students. It can support researchers and educators in assessing AI-related competencies and developing evidence-informed initiatives that promote effective, ethical, and responsible use of ChatGPT in academic environments.

Keywords: Artificial Intelligence; ChatGPT Literacy; Psychometrics; Transcultural Adaptation; Students; Iran

**Corresponding author:*

Abbas Taghizade,
Department of Educational
Sciences, Farhangian
University, P.O. Box: 14665-
889, Tehran, Iran.
Tel: +98 9370497918

Email: A.taghizade@cfu.ac.ir

Please cite this paper as:

Taghizade A, Akbari
Balootbangan A, Ahmadi R.
Psychometric Properties of
the Persian ChatGPT Literacy
Scale in Higher Education: A
Transcultural Adaptation and
Validation Study. *Interdiscip
J Virtual Learn Med Sci.*
2026;17(2):2-18. doi: 10.30476/
ijvlms.2026.111068.1382.

Received: 17-04-2026

Revised: 16-05-2026

Accepted: 26-05-2026

Introduction

Artificial Intelligence (AI) is progressively transforming everyday human activities, with its influence increasingly evident in education, healthcare, and decision-making processes (1–3). Within higher education, generative AI applications, including ChatGPT, are becoming part of teaching, learning, and other academic practices. Although ChatGPT can assist with information retrieval, idea development, and learning, it also creates concerns about the accuracy of generated information, excessive dependence on AI, academic integrity, and ethical practice. These issues make it important for students to be able to use ChatGPT competently, critically, and responsibly (4, 5).

The growing presence of ChatGPT in academic settings has produced both optimism and concern. It can enable new teaching approaches, facilitate self-directed learning, and improve academic productivity. At the same time, dependence on AI-generated material may weaken critical engagement and create challenges for academic integrity (6–8). Consequently, a growing literature has examined both the educational benefits and the potential risks of generative AI (9–13).

One concept that has gained prominence in this discussion is ChatGPT literacy, meaning the capacity to engage with the technology in an effective, critical, and ethical manner (14–16). Such literacy requires users to judge generated outputs carefully, recognize the system's limitations, and employ it responsibly across contexts. ChatGPT literacy is therefore a multidimensional capability that goes beyond basic technical use and includes critical appraisal, effective interaction, creative use, and ethically informed decision-making (16, 17).

The literature generally describes ChatGPT literacy as a multidimensional construct with five main components: Technical Proficiency, Critical Evaluation, Communication Skills, Creative Application, and Ethical Competence. Technical proficiency concerns the practical abilities needed to understand and

operate ChatGPT effectively (18, 19). Critical evaluation refers to judging the accuracy, reliability, completeness, bias, and possible errors in ChatGPT-generated outputs (18–20). Communication proficiency concerns clear and purposeful interaction with ChatGPT (14, 19, 21). Creative application reflects the capacity to use ChatGPT to develop original ideas or innovative solutions (19, 22). Ethical competence, finally, involves recognizing legal or ethical implications and using ChatGPT in a responsible manner (19, 23).

This view is aligned with the broader concept of AI literacy, which holds that effective AI use depends not only on operational knowledge but also on critically appraising AI-generated information and interacting with AI systems responsibly.

ChatGPT literacy is particularly relevant to how students employ generative AI for academic purposes. Limited skill in assessing AI-generated information can increase the likelihood of accepting inaccurate, incomplete, or biased material, whereas stronger critical and ethical abilities may encourage more responsible academic use. Measuring ChatGPT literacy can therefore help identify students' competency gaps and guide educational interventions (24, 25).

Although interest in AI literacy has increased, validated measures focused specifically on ChatGPT literacy are still scarce, especially in non-English-speaking settings. Broad AI or digital literacy measures may fail to capture ChatGPT-specific competencies, including judging generated answers, constructing effective prompts, applying the tool creatively, and recognizing ethical and privacy issues. Lee and Park's (2024) ChatGPT Literacy Scale (CLS) was designed to assess these competencies within a five-dimensional framework. However, validity established in one language and culture cannot automatically be assumed to transfer to another setting (26, 27).

When an established measurement tool is transferred to a new linguistic and cultural setting, transcultural adaptation and psychometric evaluation are essential

to ensure that it retains linguistic, semantic, conceptual, and cultural equivalence for the target population. Such evaluation is particularly relevant to ChatGPT literacy, as the construct may be shaped by language, educational practices, and culturally influenced perceptions of technology.

In Iran, no validated Persian-language instrument has yet been developed specifically to assess ChatGPT literacy among higher education students, despite increasing evidence of ChatGPT use and engagement among Iranian university students across different academic contexts (28-30). This represents an important gap in available measurement resources. Therefore, the present study sought to transculturally adapt the CLS into Persian and assess its psychometric performance among Iranian higher education students. The evaluation included face, content, and construct validity, as well as internal consistency.

Methods

Study Design and Setting

A cross-cultural adaptation and psychometric validation study was conducted from May to December 2025 to assess the Persian version of the CLS among Iranian university students. The research was carried out in two sequential phases: first, translation and cultural adaptation of the instrument, and second, evaluation of its psychometric properties.

Participants and Sampling

Several groups participated in different stages of the study. Initially, bilingual translators and subject-matter experts were involved in the translation and cultural adaptation process. Subsequently, a panel of eight experts, including five specialists in educational sciences and three in artificial intelligence, all with considerable experience in psychometrics, assessed the face and content validity of the scale.

For the psychometric evaluation, 409 students were recruited from five public universities in Iran using convenience

sampling. Eligible participants were required to have prior practical experience with ChatGPT and to provide consent to participate. Questionnaires were excluded if more than 20% of the items were unanswered or if clear straight-lining response patterns were identified.

Following the recommended ratio of 5–10 participants per item for the 25-item instrument (31), a minimum sample of 250 completed questionnaires was considered necessary. To account for potential nonresponse and incomplete data, 409 students were invited to participate. Written informed consent was obtained from all participants, and they were assured that their responses would be kept confidential.

Tools/Instruments

ChatGPT Literacy Scale: The study used the 25-item self-report CLS developed and validated by Lee and Park (19). It measures five domains of ChatGPT literacy: Technical Proficiency, Critical Evaluation, Communication Proficiency, Creative Application, and Ethical Competence. Responses are recorded on a 5-point Likert scale from 1 (not at all) to 5 (very much), yielding total scores from 25 to 125; larger totals indicate greater ChatGPT literacy.

Validity and Reliability - In Lee and Park's (2024) validation study, Cronbach's alpha was 0.85 for technical proficiency, 0.86 for critical evaluation, 0.82 for communication proficiency, 0.79 for creative application, and 0.79 for ethical competence (19). Five demographic questions covering age, gender, field of study, educational level, and frequency of ChatGPT use were also added.

Raters and experts were recruited according to relevant linguistic, methodological, and subject-specific qualifications. During translation and transcultural adaptation, bilingual specialists proficient in Persian and English were used to preserve linguistic and semantic correspondence with the source version. For assessment of content validity, eight doctoral-level experts in educational sciences and/or artificial intelligence

were selected because of their experience with educational technology, AI-related competencies, questionnaire development, and the Iranian educational setting. This expertise supported evaluation of item relevance, clarity, wording, conceptual equivalence, and cultural appropriateness.

Procedures

The study was conducted in two consecutive phases: (1) translation and cross-cultural adaptation of the CLS and (2) psychometric evaluation of the Persian version. The translation and adaptation process preceded the psychometric assessment to ensure that the Persian instrument maintained linguistic, semantic, conceptual, and cultural equivalence with the original questionnaire. The adapted version was administered to students for face-validity assessment and subsequent psychometric analyses only after the translation had undergone expert review and finalization.

Accordingly, the first phase focused on translating and culturally adapting the instrument, while the second phase investigated the psychometric properties of the Persian questionnaire.

A. Translation Technique and Cross-Cultural Adaptation Phase

The original CLS was translated and culturally adapted using forward-backward translation principles consistent with World Health Organization (WHO) translation guidance (32). As the original questionnaire was publicly available, it was used as the basis for the translation process. Established methodological procedures were applied to maintain semantic, conceptual, and cultural equivalence between the original English instrument and its Persian adaptation.

The translation process followed a forward-backward translation approach consisting of three main steps:

1. Forward Translation: Two bilingual translators proficient in Persian and English independently translated the original English questionnaire into Persian, producing two

separate versions.

2. Reconciliation and Back-Translation: Discrepancies between the independently translated Persian versions were reviewed and resolved through discussion, resulting in a unified Persian draft. Subsequently, a third bilingual translator, who had not accessed the original English questionnaire, back-translated the reconciled Persian version into English. The resulting back-translation was compared with the original instrument to identify potential differences in meaning and verify semantic equivalence.

3. Expert Review and Finalization: The translated questionnaire was evaluated by experts to determine its cultural appropriateness and cross-cultural equivalence. Minor modifications were made to the wording of selected items when necessary to improve clarity and suitability for the target population. Following these revisions, the Persian version of the questionnaire was finalized. The final adapted instrument is provided in the Appendix.

B. Psychometric Evaluation Phase

This phase focused on evaluating the instrument's face, content, and construct validity, as well as examining its reliability.

Face Validity – Quantitative face validity was assessed by calculating an impact score for each questionnaire item. A total of 47 students evaluated the 25-item questionnaire, rating the perceived importance of each item on a five-point Likert scale ranging from 1 (Not important at all) to 5 (Very important). The item impact score was determined using the following formula: $\text{Impact Score} = \text{Importance} \times \text{Frequency} (\%)$. Items with impact scores exceeding 1.5 were considered acceptable and retained for further psychometric analyses (33).

Content Validity – Qualitative content validity was assessed by eight doctoral-level experts in educational sciences and artificial intelligence. They reviewed the questionnaire for item placement, grammar, wording, relevance, clarity, and appropriateness to the Iranian cultural context.

Quantitative content validity was then examined using two measures:

Content Validity Ratio (CVR): Experts classified each item as “Essential,” “Useful but not essential,” or “Not essential.” With eight experts on the panel, the CVR acceptance criterion was 0.75; items with values above this cutoff were retained (34).

Content Validity Index (CVI): Each expert rated item relevance, clarity, and simplicity on a 4-point scale. Items were accepted for the final instrument when their CVI exceeded 0.79 (35).

Construct Validity- The scale’s latent structure was explored with principal axis factoring (PAF) and Oblimin rotation. Sampling adequacy was evaluated using the Kaiser–Meyer–Olkin (KMO) statistic, with 0.7 as the target threshold, together with Bartlett’s test, for which statistical significance at $p < 0.05$ was required (36).

Confirmatory Factor Analysis (CFA) was subsequently performed to assess the extent to which the observed items adequately represented the hypothesized latent dimensions. Model fit was evaluated using established goodness-of-fit indices and their recommended cutoff values: the chi-square to degrees of freedom ratio (χ^2/df) < 3 ; the Root Mean Square Error of Approximation (RMSEA) and Standardized Root Mean Square Residual (SRMR) < 0.08 ; and the Non-Normed Fit Index (NNFI), Comparative Fit Index (CFI), Normed Fit Index (NFI), and Goodness-of-Fit Index (GFI) ≥ 0.90 (37).

Reliability- The reliability of the CLS was assessed by examining its internal consistency. Cronbach’s alpha and McDonald’s omega coefficients were computed for the overall scale and its respective subscales. Coefficient values of 0.70 or above were considered indicative of acceptable internal consistency (38).

Data Collection

Data collection was conducted in two consecutive stages, consistent with the overall study design. The first stage involved the cross-cultural adaptation of the instrument through forward translation, reconciliation,

back-translation, and expert evaluation. In the second stage, data were collected from students and experts to assess the validity and reliability of the Persian version of the CLS.

To assess face validity, 47 students evaluated the preliminary 25-item Persian questionnaire by rating the importance of each item using a five-point Likert scale. Independently, eight experts specializing in educational sciences and artificial intelligence conducted qualitative and quantitative assessments of content validity, focusing on item relevance, clarity, simplicity, grammatical accuracy, and cultural suitability.

For the main psychometric evaluation, 409 questionnaires were administered face-to-face during regularly scheduled classes at five public universities in Iran. Participants were recruited through convenience sampling, with eligibility restricted to students who had previous hands-on experience with ChatGPT and voluntarily consented to participate. All returned questionnaires were reviewed and assigned unique sequential identification codes. Of the collected questionnaires, 29 were excluded because respondents indicated that they had never used ChatGPT, while an additional nine were discarded due to incomplete responses or straight-lining patterns. Consequently, 371 questionnaires were retained for the final statistical analyses, representing a usable response rate of 90.7%. Written informed consent was obtained from all participants, and the confidentiality of their information was maintained throughout the data collection process.

Data Analysis

Analysis of the Translation and Transcultural Adaptation Phase

Qualitative material from the translation and transcultural adaptation stage was examined by comparing the translated forms and considering expert feedback. The forward Persian translations were reviewed for differences in wording, meaning, and conceptual interpretation. Once the translations had been reconciled, the back-translated English version was compared with

the source instrument to verify consistency of meaning. Expert comments were then used to identify items needing linguistic, semantic, conceptual, or cultural revision. Items judged semantically redundant were merged or removed. The resulting Persian version was used in the subsequent psychometric assessment.

Analysis of the Psychometric Evaluation Phase

Data from the psychometric phase were processed in SPSS version 27 and LISREL version 8.80. Means, standard deviations, skewness, and kurtosis were computed for questionnaire items. Quantitative face validity was judged from item impact scores, with values above 1.5 treated as acceptable. Content validity was examined with the CVR and CVI using the criteria defined for a panel of eight experts.

Construct validity was investigated through Exploratory Factor Analysis (EFA) and CFA. Principal axis factoring with Oblimin rotation was used to derive the underlying factor pattern. Data suitability for factor analysis was checked with the KMO measure and Bartlett's test of sphericity. The hypothesized first-order five-factor solution was then tested by CFA in LISREL version 8.80. Fit was evaluated using χ^2/df , RMSEA, SRMR, CFI, NFI, NNFI, and GFI.

Internal consistency was estimated for the total scale and its subscales using Cronbach's alpha and McDonald's omega. Coefficients of at least 0.70 were taken to represent acceptable internal consistency. All statistical testing used a significance level of 0.05.

Ethics - Participation was entirely voluntary and responses were collected anonymously. Participants were informed that their information would be kept confidential during data collection, analysis, and reporting. Written informed consent was obtained, and participants could withdraw at any point. The study was carried out in accordance with the research procedures of the Vice Chancellor for Research at Farhangian University, Tehran, Iran.

Results

Demographic Characteristics

A brief researcher-developed demographic questionnaire was included to collect background information about the study participants. The questionnaire comprised five items addressing age, gender, field of study, educational level, and frequency of ChatGPT use. These characteristics were selected based on their relevance to describing the study population and assessing participants' prior exposure to ChatGPT. The demographic questionnaire was used exclusively to describe the sample characteristics and was not included in the psychometric evaluation of the CLS.

During the psychometric evaluation phase, the demographic checklist was administered together with the Persian version of the CLS to obtain descriptive information about the participating students. However, it was not included in the translation and cross-cultural adaptation phase, as that stage focused exclusively on ensuring the linguistic and cultural appropriateness of the instrument through the involvement of translators and expert reviewers rather than the main student sample.

The final study sample consisted of 371 students aged 19–49 years, with a mean age of 28.32 ± 10.22 years. Of these participants, 137 (37%) were male and 234 (63%) were female. Regarding the educational level, 241 students (65%) were enrolled in graduate programs, while 130 (35%) were undergraduate students. In terms of ChatGPT usage frequency, 77 participants (20.7%) reported always using ChatGPT, 117 (31.7%) reported using it often, 121 (32.6%) indicated occasional use, and 56 (15%) reported rarely using it.

Descriptive Statistics

Total questionnaire scores varied from 21 to 94, with an average of 51.29 ± 13.20 , ($N = 371$), indicating a moderate overall level of ChatGPT literacy in the sample. Table 1 summarizes item-level means, standard deviations, impact scores, CVRs, CVIs, skewness, and kurtosis.

Table 1: Descriptive characteristics of quantitative face validity and content validity of the questionnaire

Items	Mean±SD	Impact Score	CVR	CVI	Skewness	Kurtosis
1. I can train and fine-tune ChatGPT for specific purposes or applications.	1.92±0.94	1.85	0.76	0.85	0.667	-0.415
2. I have the ability to identify and solve technical problems of ChatGPT.	1.97±1.14	2.33	0.79	0.86	0.992	0.066
3. I have the ability to use ChatGPT in conjunction with other tools or technologies.	2.42±1.12	2.42	0.81	0.88	0.398	-0.597
4. I can compare and evaluate the functions of ChatGPT and other language models.	2.14±1.09	2.09	0.79	0.86	0.654	-0.367
5. I can understand how ChatGPT generates responses.	2.21±1.03	3.65	0.81	0.85	0.527	-0.354
6. I can understand how ChatGPT works.	Removed judgment	and merged based on expert qualitative				
7. I can evaluate the accuracy of ChatGPT responses.	2.51±1.10	3.71	0.90	0.92	0.189	-0.745
8. I can determine whether ChatGPT's response is true	Removed judgment	and merged based on expert qualitative				
9. I can evaluate the reliability of ChatGPT's responses						
10. I can identify errors in ChatGPT's responses						
11. I can evaluate the completeness of ChatGPT's responses.	2.79±1.03	3.14	0.79	0.88	0.032	-0.446
12. I can recognize and explain bias in ChatGPT's responses.	2.74 ±1.08	3.54	0.77	0.87	0.128	-0.463
13. I can communicate effectively with ChatGPT.	2.69±1.28	3.95	0.82	0.88	0.139	-1.062
14. I can ask appropriate and effective questions to ChatGPT.	2.35±1.12	3.52	0.86	0.89	0.459	-0.553
15. I can use technical terms in conversation situations with ChatGPT.	2.39±1.11	2.95	0.79	0.86	0.344	-0.677
16. I can ask accurate questions to ChatGPT using a rich vocabulary.	2.06±1.19	3.12	0.80	0.88	0.963	0.113
17. I can elicit a response from ChatGPT to suit a particular situation.	2.18±1.23	2.71	0.84	0.90	0.827	-0.184
18. I can use ChatGPT to generate new ideas or solutions	2.78±1.17	3.80	0.90	0.94	0.081	-0.718
19. I can do creative writing or storytelling using ChatGPT.	2.73±1.16	3.76	0.89	0.93	0.203	-0.690
20. I can use ChatGPT to generate insights and trends on large data sets.	2.47±1.14	3.25	0.78	0.90	0.332	-0.649
21. I can improve my creativity or innovation skills using ChatGPT.	2.31±1.19	3.81	0.88	0.86	0.482	-0.705
22. I can identify potential ethical issues associated with the use of ChatGPT.	3.55±1.44	3.48	0.80	0.89	-0.685	-0.874
23. I can explore legal or ethical considerations related to the use of ChatGPT	Removed judgment	and merged based on expert qualitative				
24. I can recognize potential privacy issues related to the use of ChatGPT.	3.53±1.46	3.25	0.89	0.87	-0.701	-0.893
25. I can use ChatGPT ethically.	3.54±1.53	2.91	0.83	0.86	-0.678	-1.059
Total Mean	2.56±1.18	3.16	0.82	0.88	0.27	-0.56

SD: Standard Deviation; CVR: Content Validity Ratio; CVI: Content Validity Index.

Skewness and kurtosis remained within the accepted interval of -2 to $+2.0$, supporting approximate normality of the data.

Face Validity: Quantitative face validity was supported because every item had an impact score above 1.5, with an average impact score of 3.16. These findings justified retaining all questionnaire items at this stage (33).

Content Validity: During qualitative review, the experts identified five items (6, 8, 9, 10, and 23) as semantically redundant because their content substantially overlapped with other statements in the scale. The panel therefore agreed to remove or merge these items to improve parsimony and minimize duplication.

For the 20 retained items, the average CVR was 0.82, with individual item values ranging from 0.76 to 0.90. All values exceeded the minimum acceptable threshold of 0.75 established by Lawshe (34) for a panel of eight experts. Similarly, the mean CVI was

0.88, with item-level scores ranging from 0.85 to 0.90, surpassing the recommended cutoff value of 0.79 proposed by Waltz and Bausell (35). Taken together, these findings indicate that the retained items demonstrated satisfactory content validity, supporting their relevance and appropriateness for measuring the intended construct (Table 1).

Construct Validity: The construct validity of the questionnaire was evaluated using both EFA and CFA to investigate its underlying factor structure and assess the fit of the proposed measurement model.

Exploratory Factor Analysis

Before factor extraction, data suitability was evaluated with the KMO statistic and Bartlett's Test of Sphericity. The KMO was 0.878, indicating adequate sampling for factor analysis. Bartlett's test was also significant ($\chi^2(190) = 3886.647$, $p < 0.001$), showing that the observed correlation matrix differed from an identity matrix.

Table 2: Indices for questionnaire factors after an Oblimin Rotation

Factors	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	7.240	36.198	36.198	6.886	34.429	34.429	2.732	13.658	13.658
2	2.018	10.088	46.286	1.548	7.738	42.167	2.665	13.325	26.983
3	1.754	8.768	55.054	1.323	6.613	48.780	2.334	11.672	38.654
4	1.306	6.530	61.584	1.000	5.002	53.782	2.275	11.373	50.027
5	1.108	5.538	67.122	0.713	3.566	57.349	1.464	7.322	57.349

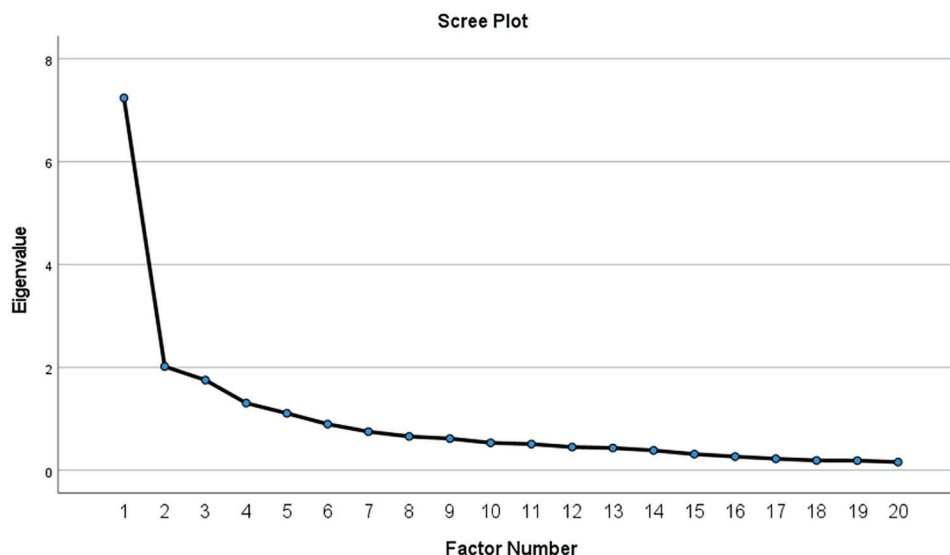


Figure 1: The scree plot

The number of latent dimensions was determined by considering eigenvalues greater than one, the proportion of explained variance, and the scree plot. As illustrated in Table 2 and Figure 1, principal axis factoring and the scree plot both supported a five-factor solution. Together, these factors accounted for 57% of the total variance, supporting the proposed structure of the instrument (Table 2).

Table 3 presents the rotated factor loadings for the 20 retained items. All loadings were acceptable and ranged from

0.427 to 0.834, so no additional items were eliminated. Items 1–5 loaded on Factor 1 (Technical Proficiency); items 6–8 on Factor 2 (Critical Evaluation); items 9–13 on Factor 3 (Communication Proficiency); items 14–17 on Factor 4 (Creative Application); and items 18–20 on Factor 5 (Ethical Competence).

Confirmatory Factor Analysis

A CFA was conducted in LISREL to test the questionnaire's proposed factor structure. The resulting model-fit statistics are reported in Table 4.

Table 3: Factor loadings of items in each subscale

Items	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Q1. I can train and fine-tune ChatGPT for specific purposes or applications.	0.456				
Q2. I have the ability to identify and solve technical problems of ChatGPT.	0.452				
Q3. I have the ability to use ChatGPT in conjunction with other tools or technologies.	0.791				
Q4. I can compare and evaluate the functions of ChatGPT and other language models.	0.833				
Q5. I can understand how ChatGPT generates responses.	0.757				
Q6. I can evaluate the accuracy of ChatGPT responses.		0.680			
Q7. I can evaluate the completeness of ChatGPT's responses.		0.834			
Q8. I can recognize and explain bias in ChatGPT's responses.		0.773			
Q9. I can communicate effectively with ChatGPT.			0.642		
Q10. I can ask appropriate and effective questions to ChatGPT.			0.748		
Q11. I can use technical terms in conversation situations with ChatGPT.			0.697		
Q12. I can ask accurate questions to ChatGPT using a rich vocabulary.			0.500		
Q13. I can elicit a response from ChatGPT to suit a particular situation.			0.427		
Q14. I can use ChatGPT to generate new ideas or solutions.				0.743	
Q15. I can do creative writing or storytelling using ChatGPT.				0.832	
Q16. I can use ChatGPT to generate insights and trends on large data sets.				0.544	
Q17. I can improve my creativity or innovation skills using ChatGPT.				0.516	
Q18. I can identify potential ethical issues associated with the use of ChatGPT.					0.655
Q19. I can recognize potential privacy issues related to the use of ChatGPT.					0.717
Q20. I can use ChatGPT ethically.					0.644

Factor 1: Technical proficiency; Factor 2: Critical evaluation; Factor 3: Communication proficiency; Factor 4: Creative application; Factor 5: Ethical competence.

Table 4: Fit indices for the first-order factor model

	Model	χ^2/df	GFI	NFI	CFI	SRMR	NNFI	RMSEA
Fit indices	First-order factor model	438 /160=2.73	0.90	0.95	0.97	0.05	0.96	0.069
	Optimal fit	$0 \leq \chi^2/df \leq 3$	$0.90 \leq GFI \leq 1$	$0.90 \leq NFI \leq 1$	$0.90 \leq CFI \leq 1$	$0 \leq SRMR \leq 0.08$	$90 \leq NNFI \leq 1$	$0 \leq RMSEA \leq 0.08$

CFI :Comparative Fit Index; GFI :Goodness of Fit Index; NFI: Normed Fit Index; NNFI: Non-Normed Fit Index; SRMR: Standardized Root Mean Square Residual; RMSEA: Root Mean Square Error of Approximation; χ^2/df : Chi-Square to Degrees of Freedom Ratio.

The indices summarized in Table 4 provide support for the first-order factor model. All values satisfied the recommended criteria (37), indicating that the model adequately represents the scale's latent organization. RMSEA was 0.069 (90% CI = 0.061–0.076), a value within the acceptable range. The chi-square/degrees-of-freedom ratio was below 3, and GFI, CFI, NFI, and NNFI were each greater than 0.90. SRMR was also below 0.08. Taken together, these indices indicate satisfactory overall model fit.

Figure 2 displays the standardized measurement parameters for the latent dimensions specified in the model summarized in Table 4. The standardized coefficients are the factor loadings linking each questionnaire item to its corresponding

component and therefore show the item's contribution to variance in that component. Larger loadings indicate greater explained variance. Figure 3 shows that all paths in the five-factor model were statistically significant ($T > 1.96$, $P < 0.01$), further supporting the model. On this basis, the model was considered appropriate for further validation and subsequent analyses.

Reliability

As shown in Table 5, Cronbach's alpha and McDonald's omega ranged from 0.72 to 0.89, with every coefficient exceeding the acceptable cutoff of 0.70 (38). These results indicate satisfactory internal consistency for the subscales as well as for the full instrument and therefore support the reliability of the scale.

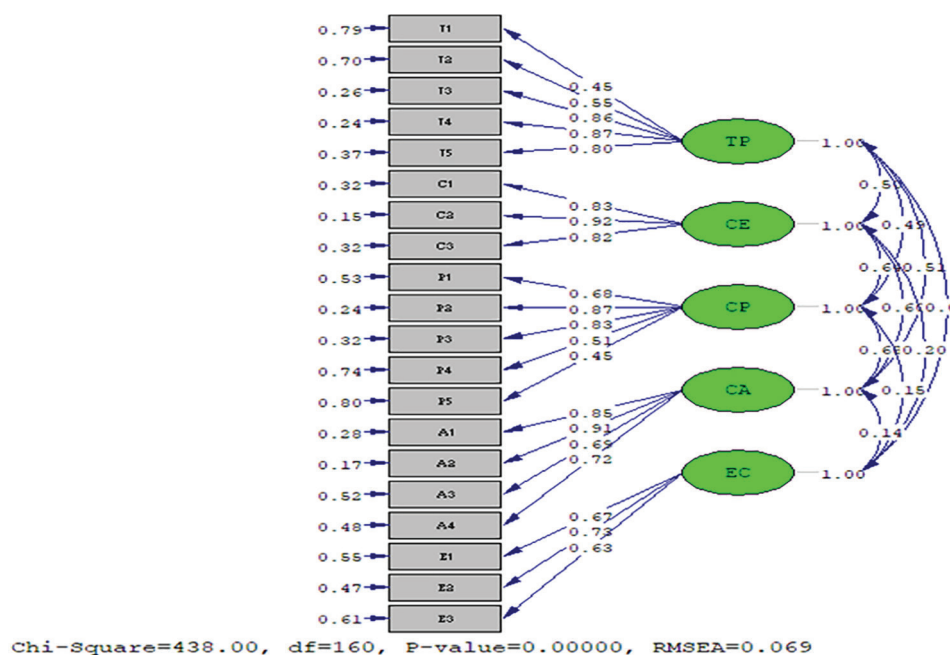


Figure 2: Standardized estimates of the five-factor first-order CFA model of the Persian ChatGPT Literacy Scale. CFA: Confirmatory Factor Analysis; RMSEA: Root Mean Square Error of Approximation; df: Degrees of Freedom Ratio

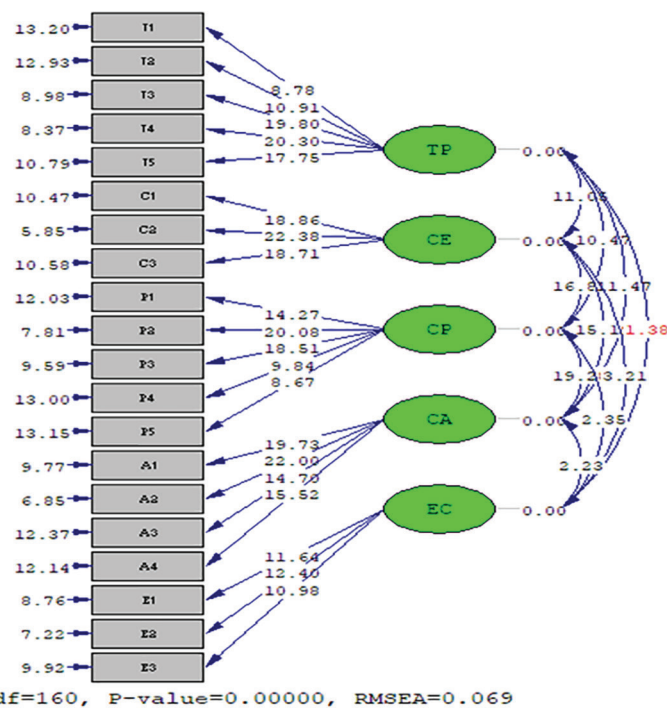


Figure 3: T-values of the five-factor first-order CFA model of the Persian ChatGPT Literacy Scale. CFA: Confirmatory Factor Analysis; RMSEA: Root Mean Square Error of Approximation; df: Degrees of Freedom Ratio

Table 5: Internal consistency of the ChatGPT literacy scale

Subscales	Number of items	Cronbach's alpha coefficient	McDonald's omega coefficient
Technical proficiency	5	0.83	0.84
Critical evaluation	3	0.89	0.89
Communication proficiency	5	0.80	0.80
Creative application	4	0.87	0.87
Ethical competence	3	0.72	0.72
Total scale	20	0.89	0.89

Discussion

In light of the increasing adoption of ChatGPT and its expanding influence on society, the present study aimed to translate, culturally adapt, and assess the psychometric properties of the CLS originally developed by Lee and Park (19) for Persian-speaking university students in Iran. The adaptation process resulted in a 20-item instrument comprising five distinct dimensions: Technical Proficiency, Critical Evaluation, Communication Proficiency, Creative Application, and Ethical Competence. Collectively, these dimensions accounted for 57% of the total variance, while the model demonstrated acceptable goodness-

of-fit indices. Notably, the original five-factor structure was successfully replicated within the Iranian cultural context, providing support for its cross-cultural applicability. Given the previous lack of a psychometrically validated Persian-language instrument specifically designed to assess ChatGPT literacy, this study addresses an important gap in the existing literature.

A structured process of translation, back-translation, expert appraisal, and pilot testing was used to adapt and validate the instrument. These procedures were intended to preserve both linguistic meaning and cultural relevance in the Persian version. This rigorous approach is particularly important in

cross-cultural assessment, as relying solely on literal translation without considering cultural and contextual differences may undermine construct validity, especially when measuring constructs that encompass social and psychological dimensions (39).

Item impact scores ranged from 1.85 to 3.95 and averaged 3.16, exceeding the minimum criterion of 1.5 (33). Thus, participants generally regarded the items as understandable, relevant, and appropriate indicators of ChatGPT literacy. The strong face-validity findings also suggest that the statements reflected students' perceived ability to use ChatGPT effectively and ethically. This interpretation is consistent with earlier work emphasizing the role of respondent comprehension during development of psychometric measures (39).

Expert qualitative review identified five statements (items 6, 8, 9, 10, and 23) from the original 25-item instrument as overlapping in meaning. These items were therefore removed or combined, after which the quantitative content-validity indices supported the adequacy of the revised instrument.

The mean CVR was 0.82, was 0.82, surpassing Lawshe's minimum acceptable threshold of 0.75 for a panel of eight experts (34). Similarly, the mean CVI was 0.88, above the recommended value of 0.79 (35). These findings provide evidence of satisfactory content validity, indicating that the Persian version of the CLS appropriately captures the key domains of ChatGPT literacy, including Technical Proficiency, Critical Evaluation, Communication Proficiency, Creative Application, and Ethical Competence. The resulting five-factor pattern accords with Lee and Park (2024), who developed and validated the original scale, and with the Chilean validation by Del Río and colleagues (2026), which likewise supported a multidimensional ChatGPT literacy construct (19, 40).

The EFA supported five dimensions—Technical Proficiency, Critical Evaluation, Communication Proficiency, Creative Application, and Ethical Competence—which together explained 57% of the variance.

CFA then confirmed the first-order factor structure, with favorable fit statistics and close agreement between the proposed model and the observed data. These findings support treating the five domains as theoretically distinct components that can be examined separately in educational settings. The pattern also parallels Lee and Park's (2024) original five-factor model, which accounted for 61% of the variance, and Del Río and colleagues (2026) Chilean study, where the multidimensional scale explained 73.5% of the variance (19, 40).

The questionnaire's internal consistency was strong, with Cronbach's alpha and McDonald's omega values between 0.72 and 0.89 for the subscales and a coefficient of 0.89 for the total instrument. This degree of reliability indicates that the scale measures its intended domains consistently, an important property for longitudinal and intervention research (38). The findings are also comparable with those of Lee and Park (2024), who reported alpha values of 0.85, 0.86, 0.82, 0.79, and 0.79 across the five subscales, and with Del Río and colleagues (2026), who observed high internal consistency in the Chilean version (19, 40).

Validation of the Persian version of CLS has several important implications for both educational research and practice. First, it provides researchers with a psychometrically supported instrument for assessing ChatGPT-related competencies among university students in Iran. From an educational perspective, its five dimensions—Technical Proficiency, Critical Evaluation, Communication Proficiency, Creative Application, and Ethical Competence—offer a comprehensive framework for identifying areas of insufficient competency and designing targeted educational interventions. Furthermore, the availability of this instrument facilitates future investigations into the relationships between ChatGPT literacy and educational outcomes, including student satisfaction, academic achievement, creativity, and productivity among Persian-speaking learners. In practical settings, the

scale may assist educators and academic institutions in developing strategies to strengthen students' effective and responsible use of artificial intelligence, potentially enhancing learning outcomes and user satisfaction while addressing concerns related to academic integrity, intellectual property, and data privacy. More broadly, the present study contributes to the growing literature on AI literacy by introducing a validated Persian-language assessment tool and providing additional evidence supporting the reliability and construct validity of ChatGPT literacy measures across diverse cultural, educational, and technological environments.

Limitations and Suggestions

Several limitations should be considered when interpreting these findings. First, the use of self-report measures may have introduced biases such as socially desirable responding and inaccuracies in participants' self-evaluations. Future work could address this issue with mixed-methods approaches that combine quantitative questionnaires with qualitative semi-structured interviews and/or objective performance measures, thereby improving triangulation and strengthening validity.

Second, no external or concurrently validated measure of ChatGPT literacy or related AI-literacy constructs was included. As a result, concurrent validity could not be examined fully and direct comparison with established instruments was limited. Future studies should include previously validated AI-literacy or digital-literacy scales to permit stronger cross-instrument validation.

Third, convenience sampling may restrict the extent to which the results can be generalized. Volunteers in this study may differ systematically from the wider Iranian student population with respect to demographic characteristics, digital literacy, access to technology, or previous experience with artificial intelligence tools.

Fourth, expert review resulted in the removal of five redundant items to make the scale more parsimonious. Although

this reduced repetition, it may also have narrowed the content represented by the original instrument or omitted more subtle aspects of ChatGPT literacy. Further studies should investigate whether the excluded content affects the conceptual completeness or predictive validity of the scale.

Fifth, despite the rigorous forward-backward translation procedures and expert-panel evaluations, subtle linguistic, semantic, or cultural differences may persist in the Persian version. These potential discrepancies could influence the interpretation of certain items and limit the comparability of findings across different linguistic, cultural, and educational contexts.

Finally, the cross-sectional nature of the study limits its ability to assess changes in ChatGPT literacy over time or establish causal relationships. Therefore, future longitudinal studies are recommended to examine the development and progression of ChatGPT literacy as students gain greater experience with the technology and as ChatGPT continues to advance and evolve.

Conclusion

The present study provides robust psychometric evidence supporting the Persian adaptation of the CLS among Iranian university students, a population previously underrepresented in ChatGPT literacy research. The findings confirmed a five-factor structure encompassing Technical Proficiency, Critical Evaluation, Communication Proficiency, Creative Application, and Ethical Competence, with satisfactory model-fit indices and internal consistency. These results support the validity, reliability, and practical applicability of the Persian version for assessing ChatGPT literacy within Iranian higher education. The validated instrument may serve as a valuable tool for educational institutions and researchers to evaluate students' competencies in using ChatGPT effectively, critically, and ethically. Furthermore, it provides a foundation for future investigations into the associations between ChatGPT literacy and

important educational outcomes, including academic achievement, student satisfaction, and other aspects of learning and academic performance.

Abbreviations

AI: Artificial Intelligence
CFA: Confirmatory Factor Analysis
CFI: Comparative Fit Index
CI: Confidence Interval
CLS: ChatGPT Literacy Scale
CVI: Content Validity Index
CVR: Content Validity Ratio
df: Degrees of Freedom
EFA: Exploratory Factor Analysis
GFI: Goodness-of-Fit Index
KMO: Kaiser–Meyer–Olkin
NFI: Normed Fit Index
NNFI: Non-Normed Fit Index
PAF: Principal Axis Factoring
RMSEA: Root Mean Square Error of Approximation
SD: Standard Deviation
SRMR: Standardized Root Mean Square Residual
WHO: World Health Organization
 χ^2/df : Chi-square to degrees-of-freedom ratio

Acknowledgements

The authors would like to express their sincere appreciation to all individuals who contributed directly or indirectly to the successful completion of this study and the preparation of the manuscript. Artificial intelligence (AI)-assisted tools were used exclusively to improve the grammatical accuracy, sentence structure, clarity, and overall readability of the manuscript. The authors carefully reviewed and verified all AI-assisted revisions and take full responsibility for the accuracy, originality, and integrity of the final content.

Authors' Contribution

AT conceived and designed the study, supervised the data collection and analysis processes, and coordinated the overall implementation of the research project. AAB contributed to the initial conceptualization

of the study and actively participated in data collection. RA was involved in all stages of manuscript preparation, including drafting, reviewing, and revising the article. All authors critically reviewed the manuscript, approved its final version, and agreed to take responsibility for their respective contributions to the work.

Conflict of Interest

The authors declare that they have no known financial or non-financial conflicts of interest that could have influenced the research, interpretation of findings, or preparation of this manuscript.

Ethical Considerations

Participation in the study was entirely voluntary, and all responses were collected anonymously to protect participants' privacy. Participants were fully informed about the purpose and procedures of the study and were assured that their information would be treated confidentially throughout the data collection, analysis, and reporting processes. Written informed consent was obtained from all participants prior to their participation, and they were informed of their right to withdraw from the study at any stage without any adverse consequences. The study was conducted in accordance with the ethical principles and research guidelines established by the Vice Chancellor for Research at Farhangian University, Tehran, Iran (Reference Code: 52400/308/3060).

Funding/Support

This research received no specific grant or financial support from any public, commercial, or not-for-profit funding agency.

Availability of Data and Materials

The data and materials supporting the findings of this study are available from the corresponding author upon reasonable request, subject to applicable ethical, confidentiality, and institutional requirements. The final Persian version of the questionnaire is available from the supplementary file.

References

- 1 Zhai X, Chu X, Chai CS, Jong MSY, Istenic A, Spector M, Liu J-B, Yuan J, Li Y. A review of artificial intelligence (AI) in education from 2010 to 2020. *Complexity*. 2021;2021:8812542. doi: 10.1155/2021/8812542.
- 2 Reddy S, Fox J, Purohit MP. Artificial intelligence-enabled healthcare delivery. *J R Soc Med*. 2019 Jan;112(1):22-28. doi: 10.1177/0141076818815510. Pubmed PMID: 30507284; PubMed Central PMCID: PMC6348559.
- 3 König PD, Wenzelburger G. Opportunity for renewal or disruptive force? How artificial intelligence alters democratic politics. *Gov Inf Q*. 2020;37(3):101489. doi: 10.1016/j.giq.2020.101489.
- 4 Ouyang F, Jiao P. Artificial intelligence in education: the three paradigms. *Comput Educ Artif Intell*. 2021;2:100020. doi: 10.1016/j.caeai.2021.100020.
- 5 Montenegro-Rueda M, Fernández-Cerero J, Fernández-Batanero JM, López-Meneses E. Impact of the implementation of ChatGPT in education: a systematic review. *Computers*. 2023;12(8):153. doi: 10.3390/computers12080153.
- 6 Hartley K, Hayak M, Ko UH. Artificial intelligence supporting independent student learning: an evaluative case study of ChatGPT and learning to code. *Educ Sci*. 2024;14(2):120. doi: 10.3390/educsci14020120.
- 7 Bin-Nashwan SA, Sadallah M, Bouteraa M. Use of ChatGPT in academia: academic integrity hangs in the balance. *Technol Soc*. 2023;75:102370. doi: 10.1016/j.techsoc.2023.102370.
- 8 İpek ZH, Gözümlü AİC, Papadakis S, Kalogiannakis M. Educational applications of the ChatGPT AI system: a systematic review research. *Educ Process Int J*. 2023;12(3):26-55. doi: 10.22521/edupij.2023.123.2.
- 9 Chan CKY, Hu W. Students' voices on generative AI: perceptions, benefits, and challenges in higher education. *Int J Educ Technol High Educ*. 2023;20:43. doi: 10.1186/s41239-023-00411-8.
- 10 Farrokhnia M, Banihashem SK, Noroozi O, Wals A. A SWOT analysis of ChatGPT: implications for educational practice and research. *Innov Educ Teach Int*. 2024;61(3):460-474. doi: 10.1080/14703297.2023.2195846.
- 11 Kasneci E, Sessler K, Küchemann S, Bannert M, Dementieva D, Fischer F, et al. ChatGPT for good? On opportunities and challenges of large language models for education. *Learn Individ Differ*. 2023;103:102274. doi: 10.1016/j.lindif.2023.102274.
- 12 Mills A, Bali M, Eaton L. How do we respond to generative AI in education? Open educational practices give us a framework for an ongoing process. *J Appl Learn Teach*. 2023;6(1):16-30. doi: 10.37074/jalt.2023.6.1.34.
- 13 Schön EM, Neumann M, Hofmann-Stölting C, Baeza-Yates R, Rauschenberger M. How are AI assistants changing higher education? *Front Comput Sci*. 2023;5:1208550. doi: 10.3389/fcomp.2023.1208550.
- 14 Ng DTK, Leung JKL, Chu SKW, Qiao MS. Conceptualizing AI literacy: an exploratory review. *Comput Educ Artif Intell*. 2021;2:100041. doi: 10.1016/j.caeai.2021.100041.
- 15 Chelawat A, Sant S. Learner motivation through gamification in e-learning: a study on game-based formative assessment in e-learning. In: Lane C, editor. *Handbook of research on acquiring 21st century literacy skills through game-based learning*. Hershey (PA): IGI Global; 2022. p. 810-31. doi: 10.4018/978-1-7998-7271-9.ch041.
- 16 Zan BU, Çolaklar H, Altay A, Taşkın N. A study on digital literacy skills of faculty of letters students: use of university library. *Int J Emerg Technol Learn*. 2021;16(1):152-171. doi: 10.3991/ijet.v16i01.16567.
- 17 Gilster P. *Digital literacy*. New York: Wiley Computer Publishing; 1997.
- 18 Laupichler MC, Aster A, Raupach T. Delphi study for the development and

- preliminary validation of an item set for the assessment of non-experts' AI literacy. *Comput Educ Artif Intell.* 2023;4:100126. doi: 10.1016/j.caeai.2023.100126.
- 19 Lee S, Park G. Development and validation of ChatGPT literacy scale. *Curr Psychol.* 2024;43(21):18992-19004. doi: 10.1007/s12144-024-05723-0.
 - 20 Bawden D. Origins and concepts of digital literacy. In: Lankshear C, Knobel M, editors. *Digital literacies: concepts, policies and practices.* New York: Peter Lang; 2008. p. 17-32.
 - 21 Koc M, Barut E. Development and validation of New Media Literacy Scale (NMLS) for university students. *Comput Human Behav.* 2016;63:834-843. doi: 10.1016/j.chb.2016.06.035.
 - 22 Reddy P, Sharma B, Chaudhary K. Digital literacy: a review of literature. *Int J Technoethics.* 2020;11(2):65-94. doi: 10.4018/IJT.20200701.oa1.
 - 23 Wang B, Rau PL, Yuan T. Measuring user competence in using artificial intelligence: validity and reliability of artificial intelligence literacy scale. *Behav Inf Technol.* 2023;42(9):1324-1337. doi: 10.1080/0144929X.2022.2072768.
 - 24 Fantone A. The new ChatGPT model can be your philosopher, good friend and most dangerous adversary all at once. *J Mater Sci Res Rev.* 2023;6(2):195-208.
 - 25 Bovermann K, Weidlich J, Bastiaens T. Online learning readiness and attitudes towards gaming in gamified online learning: a mixed methods case study. *Int J Educ Technol High Educ.* 2018;15:27. doi: 10.1186/s41239-018-0107-0.
 - 26 Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976).* 2000;25(24):3186-3191. doi: 10.1097/00007632-200012150-00014. PubMed PMID: 11124735.
 - 27 Sousa VD, Rojjanasrirat W. Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. *J Eval Clin Pract.* 2011;17(2):268-274. doi: 10.1111/j.1365-2753.2010.01434.x. PubMed PMID: 20874835.
 - 28 Rahimi AR, Mosalli Z. Language learners' surface, deep, and organizing approaches to ChatGPT-assisted language learning: what contextual, individual, and ChatGPT-related factors contribute? *Smart Learn Environ.* 2025;12:16. doi: 10.1186/s40561-025-00368-3.
 - 29 Mehrvarz M, Salimi G, Abdoli S, McLaren BM. How does students' perception of ChatGPT shape online learning engagement and performance? *Comput Educ Artif Intell.* 2025;9:100459. doi: 10.1016/j.caeai.2025.100459.
 - 30 Shahsavari Z, Kafipour R, Khojasteh L, Pakdel F. Is artificial intelligence for everyone? Analyzing the role of ChatGPT as a writing assistant for medical students. *Front Educ.* 2024;9:1457744. doi: 10.3389/educ.2024.1457744.
 - 31 Hair JF, Black WC, Babin BJ, Anderson RE. *Multivariate data analysis.* Pearson new international ed. Harlow: Pearson Education; 2013.
 - 32 World Health Organization. *Process of translation and adaptation of instruments.* Geneva: World Health Organization; 2023. [Cited 2026 March 08]. Available from: <https://iris.who.int/server/api/core/bitstreams/5650cbda-b968-4f0c-ac0e-102c293229d6/content>
 - 33 Broder HL, McGrath C, Cisneros GJ. Questionnaire development: face validity and item impact testing of the Child Oral Health Impact Profile. *Community Dent Oral Epidemiol.* 2007;35(Suppl 1):8-19. doi: 10.1111/j.1600-0528.2007.00401.x. PubMed PMID: 17615046.
 - 34 Lawshe CH. A quantitative approach to content validity. *Pers Psychol.* 1975;28(4):563-575. doi: 10.1111/j.1744-6570.1975.tb01393.x.
 - 35 Waltz CF, Bausell BR. *Nursing research: design, statistics, and computer analysis.* Philadelphia: FA Davis; 1981.
 - 36 Henson RK, Roberts JK. *Use of*

- exploratory factor analysis in published research: common errors and some comment on improved practice. *Educ Psychol Meas.* 2006;66(3):393-416. doi: 10.1177/0013164405282485.
- 37 Kline RB. Principles and practice of structural equation modeling. 5th ed. New York: Guilford Press; 2023.
- 38 Tavakol M, Dennick R. Making sense of Cronbach's alpha. *Int J Med Educ.* 2011;2:53-55. doi: 10.5116/ijme.4dfb.8dfd. PubMed PMID: 28029643; PubMed Central PMCID: PMC4205511.
- 39 Van de Vijver F, Tanzer NK. Bias and equivalence in cross-cultural assessment: an overview. *Eur Rev Appl Psychol.* 2004;54(2):119-135. doi: 10.1016/j.erap.2003.12.004.
- 40 Del Río CS, Monge-Rogel R, Fuentes-Lama R, Fernández-Ochoa H. Cultural adaptation and validation of a ChatGPT literacy scale for university students in Chile. *Curr Psychol.* 2026;45:778. doi: 10.1007/s12144-026-09314-z.