

A Psychometric Re-evaluation of ChatGPT-Related Constructs Across Demographic Groups in Higher Education

Fabiola G. C. Ribeiro^{1,2}, Gianluca Romano², Sebastian Gombert²,
Leon Camus², Daniel Baumartz³, Maruschka Weber³,
Jan Zottmann⁴, Faiza Mateen², Richard Neumann²,
Hendrik Drachsler²

¹ Instituto Federal Goiano (IF Goiano), Catalão, Goiás, Brazil

²Leibniz Institute for Research and Information in Education, Frankfurt, Germany

³Goethe University Frankfurt, Frankfurt, Germany

⁴Munich Center of the Learning Sciences, Munich, Germany

fabiola.ribeiro@ifgoiano.edu.br

h.drachsler@dipf.de

Abstract. *Generative artificial intelligence systems offer increasing opportunities to support learning and teaching in higher education. However, their effective, ethical, and responsible integration depends strongly on students' AI literacy. This study applies and psychometrically re-evaluates a validated AI Literacy Scale among 270 higher education students in a new empirical context, study design, and broader usage scenario. It further investigates students' factual knowledge, Actual Use, Use Value, and Attitude toward generative AI, as well as differences across gender and age groups. The findings support the use of the instrument in comparative AI literacy studies and provide evidence on demographic variability in students' engagement with generative AI. The findings support the use of the instrument in comparative AI literacy studies and provide evidence on demographic variability in students' engagement with generative AI.*

1. Introduction

Recent advances in AI-driven applications and systems have substantially expanded the adoption of generative AI-based language models across various domains and contexts [Khan et al. 2024, Yusuf et al. 2024]. These technologies are transforming how people work, teach and learn, particularly in educational settings [Wu et al. 2025, Ogunleye et al. 2024]. Adopted models are increasingly recognized for their ability to generate contextualized responses, adapt content to different levels of abstraction, and support learners through interactive and personalized feedback [Shi et al. 2026, Ogunleye et al. 2024]. These capabilities have driven the widespread adoption of Generative Artificial Intelligence (GAI) tools, particularly ChatGPT, in teaching and learning contexts [Ali et al. 2024], while also motivating empirical studies on their educational impact, students' AI literacy and preparedness for effective use, and potential implications for learning outcomes [Bauer et al. 2025, Marzano 2025].

Despite their potential benefits, the effective and responsible use of GAI tools requires sustained cognitive engagement, as well as appropriate guidance, AI literacy training, and institutional support for students and educators [Costa et al. 2024, Jalon and Chua 2024]. Developing AI literacy has therefore become an increasingly relevant competence for educational stakeholders, as it supports both theoretical understanding and practical application of GAI [Jalon and Chua 2024, Jin et al. 2025a]. AI literacy enables users to critically explore the context, limitations, and possibilities of adopting GAI in higher education, particularly in specialized fields such as engineering, medicine, and the natural sciences [Chiu et al. 2024].

Previous studies have explored both students' and educators' perceptions of GAI in educational contexts, as well as its potential impact on teaching and learning activities, including student engagement, language development, and support for academic writing [Costa et al. 2024, Eljack et al. 2025]. Additional research has examined the role of tools such as ChatGPT in cognitive development [Bauer et al. 2025] and academic performance across K-12 educational settings [Chiu 2024, Zhang and Tur 2023, Marzano 2025, Villar Onrubia et al. 2025], as well as in higher education [Köhler and Hartig 2024, McDonald et al. 2025]. Despite these contributions, there is still no consensus on how GAI tools should be pedagogically integrated to effectively address learners' needs and meaningfully enhance learning outcomes [Dempere et al. 2023, Bauer et al. 2025]. Moreover, a large part of the existing literature remains theoretical or primarily perception-based, with limited attention to students' actual and factual knowledge related to these technologies neither their academic or demographic features [Chiu et al. 2024, Chiu and Sanusi 2024]. As a result, it is important to provide a comprehensive empirical evaluation of students' AI literacy levels. Moreover, understanding how students engage with ChatGPT in authentic academic contexts and examining the associations between demographic characteristics and ChatGPT-related constructs remain essential.

The present study builds on the AI literacy framework and measurement instrument developed by [Köhler and Hartig 2024], applying the same core constructs in a distinct empirical context. The initial psychometric re-evaluation becomes necessary due to the application of the instrument to a new population sample with a different demographic and academic composition and, most importantly, within a broader and more established context of GAI use. This psychometric re-examination serves as a foundation for subsequent comparative analyses between the original study and the present dataset, which includes students from medicine and economic academic disciplines. Rather than extending the analytical scope to multiple contextual variables, this study prioritizes an empirical, learner-centered examination of measurement stability and comparability, while focusing on selected demographic characteristics. Specifically, the analyses investigate students' engagement with ChatGPT 3.5 as an AI-enabled educational system, examining differences related to gender and age. A further contribution of this study lies in the re-evaluation of the reliability and dimensionality of the AI literacy instrument proposed by [Köhler and Hartig 2024] under a revised analytical focus. Taken together, these findings may contribute to and inform the development of more ethical, coherent, and pedagogically grounded applications of GAI in higher education [Jin et al. 2025a, Markus et al. 2025].

2. State of the Art in AI Literacy Research

Existing literature defines AI literacy as the ability to effectively interact with AI-powered applications in everyday contexts [Hornberger et al. 2023, Hornberger et al. 2025]. Beyond this basic definition, it is understood as a multifaceted construct consisting of various skills. In [Long and Magerko 2020], 17 core competencies across five domains, including AI usage, capabilities, mechanisms, ethics and perception. Given rapid AI advancements, AI literacy is an important skill to master for the future workforce [Chiu and Sanusi 2024, Sullivan et al. 2023]. Therefore, the integration of AI literacy into education is essential as it allows learners to understand AI systems and critically engage with its ethical impacts [Chiu and Sanusi 2024, Villar Onrubia et al. 2025].

The increasing significance of AI literacy has led to the development of instruments for evaluating this competency in individuals. Recent works include the Generative AI Literacy Assessment Test (GLAT) [Jin et al. 2025a], a performance-based instrument designed to measure students' competencies. Similarly, [Hornberger et al. 2023] introduced a 31-item objective test for university students, showing correlations of AI literacy with AI self-efficacy, interest, and attitudes, though it is limited to investigating a students AI knowledge. [Markus et al. 2025] developed a 51-item AI competency objective scale (AICOS) which tests six sub-competencies (AI application, evaluation, understanding, detection, ethics and literacy) and can differentiate users by prior AI experience and predict performance on a computer science knowledge test.

In [Köhler and Hartig 2024], the authors developed a literacy instrument composed of three scales aimed at assessing knowledge about ChatGPT, its use, and attitudes toward GAI. The Knowledge scale was developed based on scientific articles, common misconceptions, and documented ChatGPT errors, in addition to the authors' teaching experience. The Use and Attitude scales were created respectively from an extension of the TAM and UTAUT models applied to specific tasks, from online media discussions on ChatGPT, and from the authors' experience in formulating items that assess attitudinal perceptions. The results of the scale validation indicated good internal consistency and adequate factorial validity for all scales. Furthermore, knowledge was positively associated with use and slightly negatively associated with attitude, suggesting that better-informed students tend to exhibit a more critical stance toward GAI tools.

Taken together, existing research demonstrates substantial progress in conceptualizing and measuring AI literacy, particularly through the development of both objective and self-report instruments. However, important gaps remain. Empirical studies that simultaneously examine factual knowledge, actual use, Use Value, and attitudes toward generative AI in authentic higher education contexts are still limited. Moreover, comparatively little attention has been given to how these dimensions of AI literacy vary across demographic characteristics, such as gender and age, using validated measurement frameworks. By adopting and re-evaluating the AI literacy instrument¹ proposed by [Köhler and Hartig 2024], the present study seeks to address these gaps. Specifically, it examines multiple dimensions of ChatGPT-related literacy constructs within a diverse student population and provides systematic statistical evidence on how demographic characteristics relate to knowledge, actual usage, use value, and attitudes toward generative AI in higher education.

¹<https://tinyurl.com/AI-Instrument>

3. Research Questions

The main objective of this study is to examine higher education students' knowledge, actual use, use value, and attitudes toward GAI. In addition, the study aims to provide insights into students' engagement with AI-based tools in higher education, considering selected demographic characteristics. This AI literacy study is guided by the following research questions:

RESEARCH QUESTION 1 (RQ1):

To what extent are the AI literacy scale properties and students' scores on knowledge, Actual Use, perceived value, and attitudes toward GAI consistent and comparable to those reported by [Köhler and Hartig 2024]?

RESEARCH QUESTION 2 (RQ2):

To what extent do students' demographic characteristics, specifically gender and age, relate to differences in knowledge, actual usage, use value, and attitudes toward GAI?

4. Methodology

4.1. Data Preparation and Study Design

The original AI Literacy Questionnaire dataset consisted of 908 raw records. An initial data cleaning procedure was applied to remove records that did not contain responses related to the proposed AI literacy questionnaire, thereby restricting the dataset to information relevant to the present study (resulting initially in 280 records and 98 variables). Subsequently, additional cleaning operations were performed, including the removal of empty rows, fully duplicated records, cases with missing identifiers (SERIAL), unfinished responses, and duplicated identifiers. This procedure did not involve any inferential transformations of the responses and was limited to the structural organization of the dataset, resulting in a cleaned version of the database comprising 270 records, which was used as the direct input for subsequent analyses.

Following the preparation of the cleaned dataset, demographic information was integrated from a separate demographic dataset of a longitudinal research program on online reasoning [Molero et al. 2020, Zlatkin-Troitschanskaia et al. 2023]. The correspondence between the two datasets was established using anonymized identifiers (SERIAL). From the initial sample of 270 participants, 14 were excluded due to extremely short completion times (below the 5th percentile of total response time), resulting in 256 valid cases. Among these, 3 participants were excluded from the demographic analyses due to missing information regarding gender, age, academic field, or semester level. This resulted in samples of 253 and 249 participants, respectively, with complete gender and age information. Automated and manual audits confirmed that the integration process did not alter any variables outside the demographic scope and that only the target demographic variables were updated as intended.

In the present study, participants first completed a demographic questionnaire, followed by two academic tasks consisting of both domain-specific and general activities. The AI literacy questionnaire, whose analysis constitutes the primary focus of this study, is described in [Köhler and Hartig 2024] and presented as supplementary material².

²<https://tinyurl.com/SBIE-Supplementary>

4.2. Data Analysis

The data analysis followed a two-step procedure aligned with the research questions. To address the first research question, a systematic re-examination of the measurement scales was performed, following the procedures of [Köhler and Hartig 2024]. Internal consistency was assessed using Cronbach's α and McDonald's ω , and the unidimensionality of the four constructs, knowledge, Actual Use, Use Value, and attitudes, was evaluated using confirmatory factor analysis (CFA). Model fit was assessed using the comparative fit index (CFI), Tucker-Lewis index (TLI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR).

For the second research question, descriptive statistics were computed for all constructs. To address RQ2, group differences across gender and age were examined using one-way analysis of variance (ANOVA). Assumptions of normality, independence, and homogeneity of variance were evaluated through residual diagnostics, Q-Q plots, and Levene's tests. Effect sizes were reported using η^2 to support interpretation beyond statistical significance. All analyses were conducted using the *R* software environment, with the *psych*, *lavaan*, and *car* packages.

5. Results

5.1. Descriptive Analysis by Demographic Data

Tables 1 and 2 present the participants' demographic characteristics, categorized by gender and age group, as well as by academic field and age group. As shown in Table 1, most participants were between 18-20 and 21-23 years of age. In addition, the sample was predominantly female. The most of students enrolled in the academic fields of Economics, Medicine, Physics and Sociology.

Gender	18-20	21-23	24-26	>27	Total
Women	88	27	15	3	133
Men	77	25	10	1	113
Diverse	3	0	0	0	3

Tabela 1. Distribution of participants by gender and age group.

Academic Field	Women	Men	Diverse	Total
Economic	39	50	0	89
Medicine	48	21	0	69
Sociology	25	9	3	37
Physics	5	27	0	32
Not declared	4	5	1	10
Business Education	5	1	0	6
Political Science	3	0	0	3
Social Science	3	0	0	3
Communication	4	0	0	4

Tabela 2. Participants' demographic distribution by academic field and gender.

Table 2 also describes the students' gender distribution by academic field and semester. The sample is predominantly composed of women ($n = 136$), followed by men ($n = 113$), with a very small representation of participants identifying as diverse ($n = 4$). Regarding fields of study, women strongly dominate study fields such as Sociology (25 female vs. 9 male) and Medicine (48 female vs. 21 male), while men outnumber women in Economics and Physics (27 male vs. 5 female). Economic shows a balanced distribution between male ($n = 50$) and female ($n = 39$) students.

5.2. Descriptive Analysis by Constructs

As indicated in the Supplementary Material (Descriptive Data, Table D1³), students reported a generally positive and engaged perception of ChatGPT. On average, they tended to agree that they feel well informed about the tool ($M = 2.93, SD = 0.72$) and reported relatively frequent use in their studies, typically between weekly and almost daily ($M = 3.38, SD = 0.89$). The results indicate strong agreement regarding the need for more institutional information about ChatGPT in higher education ($M = 3.58, SD = 0.64$), suggesting that students value further guidance on GAI. The relatively low standard deviations indicate consistent perceptions about the relevance of ChatGPT in higher education and the importance of learning more about it in academic contexts.

A descriptive analysis was conducted to examine students' perceptions across the constructs knowledge, actual usage, use value, and attitude. This analysis identified and measured the central tendency and dispersion for each construct, providing an overview of students' response patterns and distributions. Figure 1 (a-d) graphically summarizes the results presented in the Supplementary Material (Descriptive Data, Tables D2-D4⁴, providing item-level descriptive statistics for the ChatGPT-related constructs. These results identify which aspects of each construct students understand well and which remain problematic.

For Knowledge, students demonstrated a generally good understanding of ChatGPT's functioning, with high mean scores ($M < 0.7$) on most items. However, accuracy drops sharply for items related to whether "ChatGPT performs web searches⁵" and "retrieves daily updated information", indicating persistent misconceptions about its capabilities. Regarding Actual Use and Use Value, students reported consistent engagement with ChatGPT across academic purposes ($M = 0.4 - 0.8$) and perceived it as useful for learning ($M > 0.7$). Finally, Figure 1 (d) shows that students' attitudes toward ChatGPT were moderate to slightly negative, with mean scores ranging from $M = 1.0$ to $M = 2.0$. This pattern suggests that the students' attitudes toward ChatGPT are moderate and cautious, rather than strongly positive, particularly regarding potential risks related to personal data and privacy.

Figure 2 (a-d) complements the previous analysis by providing descriptive summaries at the construct level. These measures illustrate how students behave within each construct, showing their overall response tendencies. Each histogram represents the distribution of students (y-axis) according to their individual mean scores within each cons-

³<https://tinyurl.com/SBIE-OSF-Artifacts>

⁴<https://tinyurl.com/SBIE-OSF-Artifacts>

⁵The period of this study considered the GPT-3.5 (version without internet access)

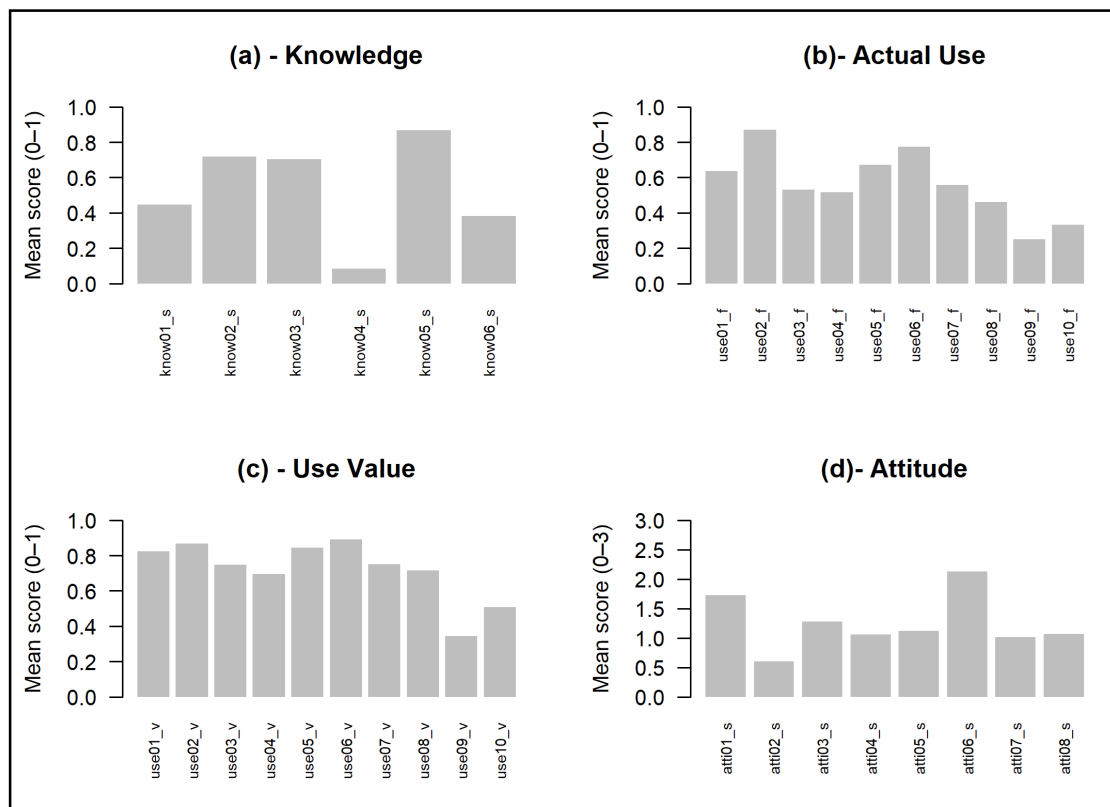


Figura 1. Average response to each construct item across all students.

tract. For Knowledge, the distribution indicates that students possess a moderate level of objective understanding of ChatGPT ($M = 0.4 - 0.5$). Regarding Actual Use and Use Value, students tend to use ChatGPT moderately in their academic activities ($M = 0.4 - 0.8$) and assign high relevance and usefulness to the tool ($M = 0.6 - 1.0$). Finally, the Attitude construct reveals neutral to moderately positive perceptions toward ChatGPT ($M = 1.3 - 1.5$), suggesting that students hold balanced views about its potential benefits and risks.

5.3. Scale Re-Evaluation

We measured the internal consistency and factorial validity of the proposed constructs. Table 3 reports reliability coefficients (α , ω) and goodness-of-fit indices (CFI, TLI, RMSEA, SRMR) for the measurement model, indicating acceptable internal consistency and an adequate model fit for the knowledge, use, and attitude constructs towards ChatGPT.

In the proposed evaluation, the internal consistency for the knowledge scale was lower than in the original study ($\alpha = .57$; $\omega = .84$ [Köhler and Hartig 2024]/ $\alpha = .32$; $\omega = .75$), likely due to the binary scoring and weak inter-item correlations ($r < .20$), indicating that participants mastered different aspects of ChatGPT knowledge rather than sharing a uniform understanding. This pattern is consistent with the broader familiarity with ChatGPT in 2025, where knowledge has become more dispersed but not lower. This suggests that students' attitudes towards ChatGPT have become more coherent and less driven by isolated enthusiasm, reflecting a more consolidated and balanced perception of GAI tools.

Confirmatory factor analyses further supported the unidimensional structure of all

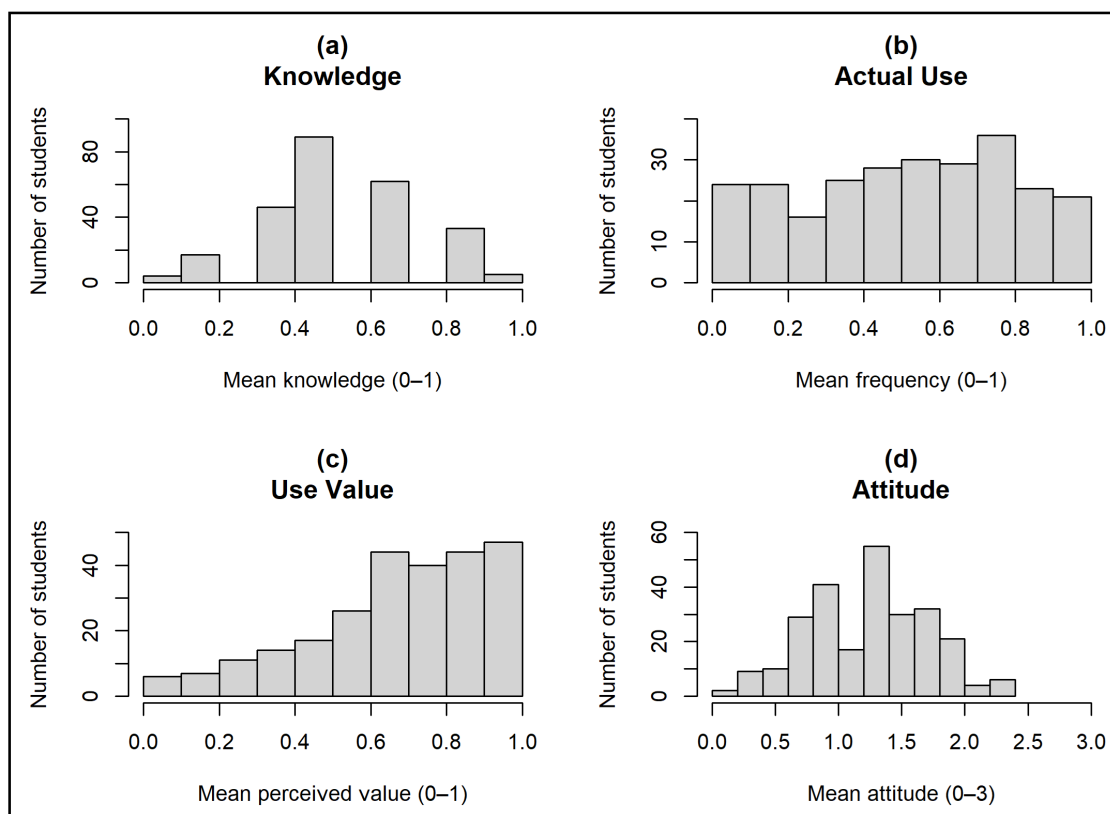


Figura 2. Average Students Answers per Construct.

Item	α	ω	CFI	TLI	RMSEA	SRMR
Knowledge	0.32	0.75	0.93	0.90	0.04	0.10
Actual Usage	0.82	0.94	0.97	0.96	0.09	0.11
Use Value	0.77	0.93	0.96	0.95	0.07	0.12
Attitude	0.72	0.83	0.99	0.98	0.05	0.06

Tabela 3. Internal consistency, reliability, and CFA model fit indicators for the four ChatGPT constructs.

scales, with good to excellent fit indices (CFI = .93-.99; TLI = .90-.98; RMSEA = .04-.09; SRMR = .06-.12). For the knowledge scale, despite the modest internal consistency coefficients, the one-factor model showed an adequate global fit (CFI = .93, TLI = .90, RMSEA = .04, SRMR=.01), indicating that the items still capture a coherent latent dimension of ChatGPT knowledge. This suggests that the variability among items reflects the diversity of knowledge components rather than measurement error. Overall, the CFA results confirm that each scale reliably represents its theoretical construct, supporting the structural validity of the measurement model also in our new sample.

The four-factor model demonstrated an acceptable overall fit. These values suggest that the proposed factor structure reproduces the observed data moderately well, even though they fall slightly below the conventional benchmarks for excellent fit (typically $\geq .95$) and RMSEA (typically $\leq .08$) reflects a good level of approximation error and indicates that the model's discrepancy from the data is small on average. In contrast, SRMR

= .12 is higher than the recommended threshold (.08), suggesting that some residual correlations between items are not fully captured by the model.

5.4. Relations Across Dimensions

Table 4 presents the latent correlations among the four constructs: Knowledge, Actual Usage, Use Value, and Attitude. The data reveal moderate positive linear correlations between Actual Usage and Use Value with Attitude, indicating that participants who use ChatGPT more frequently and perceive it as more valuable tend to hold more positive attitudes toward it.

Item	Knowledge	Actual Usage	Use Value	Attitude
Knowledge	1	[0.06, 0.48]	[−0.26, 0.16]	[−0.60, −0.16]
Actual Usage	0.27 (0.11)	1	[0.75, 0.90]	[0.22, 0.48]
Use Value	−0.05 (0.11)	0.83 (0.04)	1	[0.41, 0.64]
Attitude	−0.39 (0.11)	0.35 (0.07)	0.53 (0.06)	1

Tabela 4. Latent correlations between constructs (r-latent and standard errors below the diagonal, confidence intervals above the diagonal).

Conversely, a strong negative correlation emerged between Knowledge and Attitude constructors, highlighting the findings of [Köhler and Hartig 2024], who observed that individuals who are better informed about ChatGPT tend to adopt a more critical posture toward the tool. Across all latent correlations, the standard errors were low ($SE = .04-.11$), indicating stable parameter estimates and reliable relationships between constructs.

It is therefore possible to infer that, in this sample, frequent and positively valued use of ChatGPT is associated with more favorable attitudes toward the tool, whereas greater factual knowledge about how ChatGPT works tends to foster a more critical or reflective behavior (enthusiasm decreases as understanding increase).

5.5. Differences Considering Demographic Variables

In this section, we examine the extent to which demographic characteristics account for variability in AI literacy constructs. One-way ANOVA tests were conducted to assess differences in Knowledge, Actual Use, Use Value, and Attitude across gender and age groups.

5.5.1. Gender Differences

Gender-related analyses focused on the two most represented groups in the sample (women: $n = 136$; men: $n = 113$). The diverse gender group ($n = 4$) was excluded from the inferential analyses due to its insufficient size for meaningful comparisons.

All assumptions for conducting one-way ANOVA were met, including independence of observations, normality of residuals (see the Q-Q plots in the supplementary material ANOVA Assumptions, Figure A1⁶), and homogeneity of variance (Levene's test:

⁶<https://tinyurl.com/SBIE-OSF-Artifacts>

Construct	Df	F-value	p-value	η^2
Knowledge	1	4.38	0.037	0.02
Actual Use	1	1.74	0.188	0.007
Use Value	1	7.04	0.008	0.03
Attitude	1	7.47	0.006	0.03

Tabela 5. One-way ANOVA results for gender differences across ChatGPT-related constructs

$F(1, 247) = 1.54, p = .216$). As shown in Table 5, the analyses revealed a significant gender effect on factual knowledge, $F(1, 247) = 4.38, p = .037$. Stronger gender differences were observed for Use Value, $F(1, 247) = 7.04, p = .008$, and Attitude, $F(1, 247) = 7.47, p = .006$. In contrast, no significant gender differences were found for actual Use, $F(1, 247) = 1.74, p = .188$, indicating comparable levels of reported use across gender groups.

Gender differences were observed for knowledge (Women: $M = 0.51, SD = 0.20$; Men: $M = 0.56, SD = 0.20$), Use Value (Women: $M = 0.75, SD = 0.21$; Men: $M = 0.67, SD = 0.27$), and attitudes (Women: $M = 1.19, SD = 0.45$; Men: $M = 1.35, SD = 0.48$). In contrast, no significant differences emerged for Actual Use (Women: $M = 0.58, SD = 0.28$; Men: $M = 0.53, SD = 0.29$). This pattern suggests that ChatGPT use is widespread across gender groups, whereas evaluative and reflective dimensions of AI literacy remain more differentiated in both genders. Despite the statistical significance observed in some comparisons, the small magnitude of the effects indicates that further investigation with larger and more diverse samples is necessary.

5.5.2. Age Differences

The age-related analyses were conducted using three age groups to examine differences in constructors' means. The assumptions required for one-way ANOVA were assessed. Visual inspection of Q-Q plots (see supplementary material ANOVA Assumptions, Figure A2⁷) suggested that the residuals followed an approximately normal distribution, and Levene's test indicated no evidence of heterogeneity of variances ($F(2, 243) = 0.47, p = .627$).

Participants were categorized in three age groups, with $n = 165$ participants in the 18-20 group, $n = 52$ in the 21-23 group, and $n = 29$ in the ≥ 24 group. This recoding was based on the demographic distribution, presented in Table 1, and was implemented to create more balanced and statistically reliable categories for the ANOVA.

Construct	Df	F-value	p-value	η^2
Knowledge	2	0.33	0.722	0.003
Actual Use	2	3.25	0.040	0.03
Use Value	2	1.58	0.208	0.01
Attitude	2	0.48	0.616	0.004

Tabela 6. One-way ANOVA results for age group differences across ChatGPT-related constructs.

⁷<https://tinyurl.com/SBIE-OSF-Artifacts>

As shown in Table 6, age did not yield meaningful variability across the constructs. Although a statistically significant effect was observed for Actual Use, $F(2, 243) = 3.25$, $p = .040$, the associated effect size was small ($\eta^2 = 0.026$), indicating that differences between age groups are minimal and unlikely to reflect substantive variation in participants' responses.

6. Discussion

Generative AI (GAI) technologies are increasingly integrated into students' academic contexts and have the potential to shape learners' cognitive processes and academic achievement [Markus et al. 2025, Ali et al. 2024]. Given their widespread use, there is a pressing need to understand, at an institutional level, how these technologies are being integrated, how students engage with them, and how their use can be guided toward safe, ethical, and pedagogically meaningful practices across different educational settings [Annappureddy et al. 2025].

The re-evaluation of the AI literacy scale proposed in [Köhler and Hartig 2024], showed that in average the factual Knowledge and Attitude constructs showed patterns comparable in both studies. However, the Actual Use scale (Likert 1 - 4) exhibited higher mean scores in the present dataset ($M = 1.9 - 3.5$) than in the original study ($M = 1.4 - 2.9$), suggesting weekly or daily based students' academic engagement. It reflects a shift in ChatGPT adoption, as the original validation took place in early 2023, when usage was still emerging, whereas the current data capture a post-adoption phase in which frequent engagement with GAI tools has become commonplace in higher education. The analyses of [Köhler and Hartig 2024] reported that most students felt well informed about ChatGPT ($M = 2.66$, $SD = 0.76$), indicated using the tool only once or twice a month ($M = 2.17$, $SD = 0.91$) and had global information needs ($M = 3.37$, $SD = 0.74$). In the present study, higher mean scores were observed for global self-assessed knowledge ($M = 2.92$, $SD = 0.72$), global Actual Use ($M = 3.37$, $SD = 0.89$), and global information needs ($M = 3.58$, $SD = 0.64$).

Compared to the earlier findings [Hornberger et al. 2023, Zirar 2023, Sallam et al. 2024], our results suggest an increased perceived understanding of ChatGPT alongside a growing demand for information and training related to generative AI. Importantly, the higher mean score on the Actual Use scale corresponds to usage at least on a weekly basis, indicating that engagement with ChatGPT has become more regular and increasingly integrated into students' academic routines. At the same time, while students' self-assessed knowledge appears to have increased over time, the levels of factual knowledge remain remarkably similar across studies and show little variation by gender or age. This discrepancy between perceived and actual knowledge points to a critical challenge for higher education institutions: the need to design strategies that foster not only confidence in using generative AI, but also substantive, practical, and ethically grounded proficiency [Chiu and Sanusi 2024, Jin et al. 2025a, Hornberger et al. 2025]. Such strategies should aim to strengthen students' factual understanding and applied competencies in generative AI, tailored to the specific demands, students profiles, and different disciplinary contexts [Jin et al. 2025b, Hornberger et al. 2025].

Considering the interrelations among ChatGPT-related constructs and students' demographic characteristics, gender emerged as the dimension associated with the lar-

gest number of differences in knowledge, Use Value, and Attitude toward ChatGPT. Men showed marginally higher factual knowledge and more positive attitudes, whereas women reported higher Use Value. However, the magnitude of these effects was modest, pointing to subtle variations rather than pronounced differences across groups. Therefore, these findings should be interpreted with caution and do not support strong generalizations. Instead, they underscore the need for further research examining how gender, academic field, age, semester level, and other contextual factors interact with AI literacy and usage patterns, particularly to inform the design of evidence-based and inclusive educational strategies [Hornberger et al. 2023, Hornberger et al. 2025].

The aforementioned findings also have implications for both research and educational practice. From a research perspective, the demonstrated stability of the AI Literacy Scale supports its use in comparative studies across different educational settings, contributing to the accumulation of evidence on students' AI literacy as generative AI becomes more widespread and mature. From a practical perspective, the instrument has the potential to support higher education institutions in assessing AI literacy and identifying strengths and weaknesses across different dimensions of students' knowledge, use, perceived value, and attitudes toward generative AI. In addition, this study complements the existing literature [Köhler and Hartig 2024, Zirar 2023, Sallam et al. 2024] by examining age and gender related differences and their statistical significance across AI literacy constructs in higher education.

7. Conclusion and Future Work

This study re-examined the AI Literacy Scale proposed by [Köhler and Hartig 2024] within a new empirical context involving 270 higher education students. The findings indicate that the measurement properties of the instrument remained largely consistent with those reported in the original study, supporting its applicability across different samples, academic contexts, and study designs. This result provides additional evidence for the robustness of the scale and its suitability for comparative AI literacy research. Our analysis revealed relatively stable patterns of AI literacy across demographic groups, with gender accounting for the most notable differences across ChatGPT-related constructs. These findings suggest that the widespread adoption of generative AI has contributed to relatively homogeneous patterns of engagement among higher education students, while some evaluative and attitudinal dimensions remain differentiated.

Future research should extend our analyses to broader and more diverse student populations and examine the relationship between AI literacy constructs and academic characteristics such as field of study and semester progression through multivariate approaches. Although the instrument was originally developed during an earlier stage of generative AI adoption, the present findings indicate that its core constructs remain relevant and stable in a broader and more mature usage context. In this sense, further refinements should investigate how AI literacy relates to critical thinking, responsible AI use, and sustained cognitive engagement in higher education.

This study has some limitations, including its cross-sectional design, the use of a single assessment, and its reliance on questionnaire responses, which may not fully capture students' actual AI literacy practices and interactions with generative AI. Consequently, the findings should be interpreted within this context and further validated through longitudinal studies involving more diverse populations and educational settings.

Acknowledgments

This research was funded as part of the Research Unit (FOR) CORE 5404 under grant number 462702138. The authors gratefully acknowledge the anonymous reviewers for their valuable and constructive feedback. The authors also thank the Leibniz Institute for Research and Information in Education (DIPF) for its support during the development of this study.

Generative AI Usage Statement

Claude Sonnet 4.5⁸ developed by Anthropic was used during the preparation of this manuscript to support language revision and to assist with the use of R functions and libraries employed in the statistical analyses and data processing procedures. No artificial intelligence system was involved in the study design, data collection, data analysis, interpretation of results, or scientific decision-making. All analyses, interpretations, and conclusions presented in this work were conducted by the authors, who assume full responsibility for the content of the manuscript.

Ethical Considerations

Ethical review and approval for the study on human participants was not required in accordance with the local legislation and institutional requirements. All participants provided written informed consent, and the data were anonymized to minimize any risk of participant identification or exposure.

Artifact Availability

The study artifacts and supplementary materials are publicly available through the Open Science Framework (OSF): <https://tinyurl.com/SBIE-OSF-Artifacts>.

Referências

- Ali, D., Fatemi, Y., Boskabadi, E., Nikfar, M., Ugwuoke, J., and Ali, H. (2024). Chatgpt in teaching and learning: A systematic review. *Education Sciences*, 14(6):643.
- Annapureddy, R., Fornaroli, A., and Gatica-Perez, D. (2025). Generative ai literacy: Twelve defining competencies. *Digit. Gov.: Res. Pract.*, 6(1).
- Bauer, E., Greiff, S., Graesser, A. C., Scheiter, K., and Sailer, M. (2025). Looking beyond the hype: Understanding the effects of ai on learning. *Educational Psychology Review*, 37:45.
- Chiu, T. K. (2024). The impact of generative ai (genai) on practices, policies and research direction in education: A case of chatgpt and midjourney. *Interactive Learning Environments*.
- Chiu, T. K., Ahmad, Z., Sanusi, T., Sanusi, I., and Ismailov, M. (2024). What are artificial intelligence literacy and competency? a comprehensive framework to support them. *Computers and Education Open*, 6:100171.

⁸<https://claude.ai>

- Chiu, T. K. and Sanusi, I. (2024). Define, foster, and assess student and teacher ai literacy and competency for all: Current status and future research direction. *Computers and Education Open*.
- Costa, A., Lima, N., Viegas, C., and Caldeira, A. (2024). Critical minds: Enhancing education with chatgpt. *Cogent Education*, 11(1).
- Dempere, J., Modugu, K., Hesham, A., and Ramasamy, L. K. (2023). The impact of ChatGPT on higher education. *Frontiers in Education*, 8.
- Eljack, N., Altahir, N., Mohammed, N., and Osman, S. (2025). Chatgpt as an ai assistant tool to enhance university students' teaching, assessment and academic achievements: A case study of an eap lecturer. *Journal of Language Teaching and Research*, 16:1813–1823.
- Hornberger, M., Bewersdorff, A., and Nerdel, C. (2023). What do university students know about artificial intelligence? development and validation of an ai literacy test. *Computers and Education: Artificial Intelligence*, 5:100165.
- Hornberger, M., Bewersdorff, A., Schiff, D., and Nerdel, C. (2025). A multinational assessment of ai literacy among university students in germany, the uk, and the us. *Computers in Human Behavior: Artificial Humans*, 4:100132.
- Jalon, J. and Chua, G. (2024). Chatgpt as a learning assistant: Its impact on students' learning and experiences. *International Journal of Education in Mathematics Science and Technology*, 12:1603–1619.
- Jin, Y., Martinez-Maldonado, R., Gašević, D., and Yan, L. (2025a). Glat: The generative ai literacy assessment test. *Computers and Education: Artificial Intelligence*, 9:100436.
- Jin, Y., Yan, L., Echeverria, V., Gašević, D., and Martinez-Maldonado, R. (2025b). Generative ai in higher education: A global perspective of institutional adoption policies and guidelines. *Computers and Education: Artificial Intelligence*, 8:100348.
- Khan, N., Khan, Z., Koubaa, A., Khan, M. K., and bin Salleh, R. (2024). Global insights and the impact of generative ai-chatgpt on multidisciplinary: a systematic review and bibliometric analysis. *Connection Science*, 36(1):2353630.
- Köhler, C. and Hartig, J. (2024). Chatgpt in higher education: Measurement instruments to assess student knowledge, usage, and attitude. *Contemporary Educational Technology*, 16(1):ep528.
- Long, D. and Magerko, B. (2020). What is ai literacy? competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–16, New York, NY, USA. Association for Computing Machinery.
- Markus, A., Carolus, A., and Wienrich, C. (2025). Objective measurement of ai literacy: Development and validation of the ai competency objective scale (aicos). *Computers and Education: Artificial Intelligence*, 9:100485.
- Marzano, D. (2025). Generative artificial intelligence (gai) in teaching and learning processes at the k-12 level: A systematic review. *Tech Know Learn*.

- McDonald, N., Johri, A., and Collier, A. (2025). Generative artificial intelligence in higher education: Evidence from an analysis of institutional policies and guidelines. *Computers in Human Behavior: Artificial Humans*, 3:100121.
- Molero, D., Zlatkin-Troitschanskaia, O., Nagel, M.-T., Brückner, S., Schmidt, S., and Shavelson, R. (2020). Assessing university students' critical online reasoning ability: A conceptual and assessment framework with preliminary evidence. *Frontiers in Education*, 5.
- Ogunleye, B., Zakariyyah, K. I., Ajao, O., Olayinka, O., and Sharma, H. (2024). A systematic review of generative ai for teaching and learning practice. *Education Sciences*, 14(6).
- Sallam, M., Elsayed, W., Al-Shorbagy, M., Barakat, M., El Khatib, S., Ghach, W., Alwan, N., Hallit, S., and Malaeb, D. (2024). Chatgpt usage and attitudes are driven by perceptions of usefulness, ease of use, risks, and psycho-social impact: A study among university students in the uae.
- Shi, Y., Yu, K., Dong, Y., and Chen, F. (2026). Large language models in education: a systematic review of empirical applications, benefits, and challenges. *Computers and Education: Artificial Intelligence*, 10:100529.
- Sullivan, M., Kelly, A., and McLaughlan, P. (2023). Chatgpt in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning & Teaching*, 6(1):17–34.
- Villar Onrubia, D., Cachia, R., Rietz, C., Feltrero, R., Niemi, H., Hallissy, M., and Reuter, R. (2025). Generative artificial intelligence in secondary education: Uses and perceptions from the perspective of early adopters across five eu member states.
- Wu, F., Dang, Y., and Li, M. (2025). A systematic review of responses, attitudes, and utilization behaviors on generative ai for teaching and learning in higher education. *Behavioral Sciences*, 15(4).
- Yusuf, A., Pervin, N., Román-González, M., and Noor, N. M. (2024). Generative ai in education and research: A systematic mapping review. *Review of Education*, 12(2):e3489. e3489 RoE-2024-01-1127.R3.
- Zhang, P. and Tur, G. (2023). A systematic review of chatgpt use in k-12 education. *European Journal of Education*, pages 1–22.
- Zirar, A. (2023). Exploring the impact of language models, such as ChatGPT, on student learning and assessment. *Review of Education*, 11.
- Zlatkin-Troitschanskaia, O., Fischer, J., and Molero, D. (2023). Critical online reasoning in higher education (core): Overview of the research unit.