EDUR • Educação em Revista.2026;42;e67812
DOI: http://dx.doi.org/10.1590/0102-469867812
Preprint: https://doi.org/10.1590/SciELOPreprints.17329

https://creativecommons.org/licenses/by/4.0/

ARTICLE

GENERATIVE ARTIFICIAL INTELLIGENCE LITERACY AND AUTHENTIC ASSESSMENT IN HIGHER EDUCATION: ACADEMIC INTEGRITY[1]

GERMÁN VINICIO ROBLEZ TORRES[1]
ORCID: https://orcid.org/0009-0003-5061-860X
gvroblez@utpl.edu.ec
ALEXANDRA MARICRUZ ANDRADE APOLO[1]
ORCID: https://orcid.org/0009-0009-4525-1114
amandrade11@utpl.edu.ec
KATHERINE ELIZABETH GONZAGA AGUILAR[1]
ORCID: https://orcid.org/0009-0005-3541-3133
kegonzaga@utpl.edu.ec

[1] Universidad Técnica Particular de Loja. Loja, Loja, Ecuador.

ABSTRACT: Generative artificial intelligence (AI) poses challenges to authorship, learning traceability, and academic honesty in university settings. This study examines the associations among generative artificial intelligence literacy, authentic assessment, and academic integrity in higher education. A quantitative, exploratory, descriptive, and correlational design was adopted, using a nonprobability convenience sample of 603 participants: 482 students and 121 faculty members from an Ecuadorian university. Data were collected through a 30-item questionnaire organized into five domains: generative AI literacy, academic use of generative tools, authentic assessment, academic integrity, and learning traceability. The instrument was reviewed through expert judgment for clarity and content and demonstrated adequate internal consistency within each domain and high internal consistency across the full set of items. Mean scores exceeded the scale midpoint in all domains. Academic integrity had the highest mean score, followed by generative AI literacy, whereas learning traceability had the lowest value. Correlations were positive and statistically significant, particularly among authentic assessment, learning traceability, and academic integrity. These dimensions showed associations within the study sample, suggesting the relevance of considering them when designing more situated, transparent, and verifiable university assessments, without establishing causal relationships.

Keywords: Artificial intelligence; Higher education; Student assessment; Ethics; Educational technology.

[1] Article published with funding from the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* - CAPES/Brazil) for editing, layout, and XML conversion services.

GENERATIVE ARTIFICIAL INTELLIGENCE LITERACY AND AUTHENTIC ASSESSMENT IN HIGHER EDUCATION: ACADEMIC INTEGRITY

ABSTRACT: Generative artificial intelligence (AI) poses challenges to authorship, learning traceability, and academic honesty in university settings. This study examines the associations among generative artificial intelligence literacy, authentic assessment, and academic integrity in higher education. A quantitative, exploratory, descriptive, and correlational design was adopted, using a nonprobability convenience sample of 603 participants: 482 students and 121 faculty members from an Ecuadorian university. Data were collected through a 30-item questionnaire organized into five domains: generative AI literacy, academic use of generative tools, authentic assessment, academic integrity, and learning traceability. The instrument was reviewed through expert judgment for clarity and content and demonstrated adequate internal consistency within each domain and high internal consistency across the full set of items. Mean scores exceeded the scale midpoint in all domains. Academic integrity had the highest mean score, followed by generative AI literacy, whereas learning traceability had the lowest value. Correlations were positive and statistically significant, particularly among authentic assessment, learning traceability, and academic integrity. These dimensions showed associations within the study sample, suggesting the relevance of considering them when designing more situated, transparent, and verifiable university assessments, without establishing causal relationships.

Keywords: Artificial intelligence; Higher education; Student assessment; Ethics; Educational technology.

ALFABETIZACIÓN EN INTELIGENCIA ARTIFICIAL GENERATIVA Y EVALUACIÓN AUTÉNTICA EN EDUCACIÓN SUPERIOR: INTEGRIDAD ACADÉMICA

RESUMEN: La inteligencia artificial (IA) generativa plantea desafíos para la autoría, la trazabilidad del aprendizaje y la honestidad académica en el ámbito universitario. Esta investigación examina las asociaciones entre alfabetización en inteligencia artificial generativa, evaluación auténtica e integridad académica en educación superior. Se adoptó un diseño cuantitativo, exploratorio, descriptivo y correlacional, con una muestra no probabilística por conveniencia de 603 participantes: 482 estudiantes y 121 docentes de una universidad ecuatoriana. Los datos se recogieron mediante un cuestionario de 30 ítems, organizado en cinco dominios: alfabetización en IA generativa, uso académico de herramientas generativas, evaluación auténtica, integridad académica y trazabilidad del aprendizaje. El instrumento fue revisado mediante juicio de expertos respecto de su claridad y contenido, y presentó consistencia interna adecuada por dominio y alta en el conjunto de los ítems. Los resultados mostraron medias superiores al punto medio de la escala en todos los dominios. La integridad académica obtuvo la media más alta, seguida de la alfabetización en IA generativa, mientras que la trazabilidad del aprendizaje presentó el valor

más bajo. Las correlaciones fueron positivas y estadísticamente significativas, especialmente entre evaluación auténtica, trazabilidad del aprendizaje e integridad académica. Se concluye que estas dimensiones presentan asociaciones en la muestra estudiada, lo que sugiere la pertinencia de considerarlas en el diseño de evaluaciones universitarias más situadas, transparentes y verificables, sin establecer relaciones causales.

Palabras clave: Inteligencia artificial; Enseñanza superior; Evaluación del estudiante; Ética; Tecnología educacional.

INTRODUCTION

Generative artificial intelligence has rapidly become integrated into learning practices, academic writing, information retrieval, and the production of evidence in higher education. Its expansion has altered the relationship among students, knowledge, and assessment because it enables the generation of texts, synthesis of sources, completion of tasks, and production of arguments with increasing levels of fluency. This context does not eliminate the need to develop academic competencies; on the contrary, it requires universities to define pedagogical, ethical, and assessment criteria to guide its use (Crompton & Burke, 2023; Chen et al., 2020).

The current debate is no longer limited to whether these tools should be permitted or prohibited. The central issue is how to integrate them without weakening authorship, learning traceability, and assessment validity. UNESCO (2025) states that "artificial intelligence in education should follow a human-centered approach, with attention to equity, data protection, human agency, and the development of competencies for students and faculty". Within this framework, generative artificial intelligence literacy becomes a necessary condition for university stakeholders to understand its possibilities, limitations, and risks.

In higher education, assessment is one of the areas most challenged by generative artificial intelligence (Perkins, 2023). Traditional written assignments, take-home essays, and activities focused solely on final products are more vulnerable when they do not incorporate verifiable processes of development, feedback, and demonstration of learning. In response, authentic assessment becomes particularly relevant because it links academic performance to problems, contexts, and tasks that approximate professional practice. Its contribution lies not only in changing the format of activities, but also in strengthening the relationship among performance, assessment criteria, and evidence of learning (Herrington & Oliver, 2000; Villarroel et al., 2017).

Recent literature identifies authentic assessment as a relevant alternative for addressing the challenges of generative artificial intelligence, particularly when activities require contextual analysis, professional judgment, decision-making, critical reflection, and situated application of knowledge. However, this approach does not by itself resolve academic integrity issues. The authenticity of a task may be compromised if students delegate relevant cognitive processes to automated systems without disclosing their use, understanding the generated outputs, or demonstrating mastery of the process followed (Kasneci et al., 2023).

The intersection of generative artificial intelligence literacy, authentic assessment, and academic integrity remains underexplored in the current literature. Although studies documenting the use of generative tools in university settings and fraud-related risks are abundant, empirical evidence linking students' technological competence to the design of formative and transparent assessments remains limited (Long & Magerko, 2020). Addressing this gap is critical for universities seeking to reformulate their regulatory frameworks beyond the traditional punitive approach based strictly on detection and surveillance software.

Therefore, the objective of this study was to analyze the relationship among generative artificial intelligence literacy, authentic assessment, and academic integrity in higher education in order to provide pedagogical criteria for designing university assessments that are more valid, ethical, and aligned with current challenges in professional education.

METHODOLOGY

Research characterization

Using an empirical quantitative approach with a correlational design, the convergence among generative artificial intelligence literacy, academic integrity, and authentic assessment was evaluated (Bretag et al., 2019). The analysis was based on primary data obtained from the higher education faculty and student community.

The exploratory design made it possible to address an educational phenomenon that is still evolving. The descriptive scope characterized uses, perceptions, and assessment practices associated with generative artificial intelligence tools (Crompton & Burke, 2023). The correlational scope examined associations among the instrument domains.

First, a literature review was conducted on generative artificial intelligence in higher education, artificial intelligence literacy, authentic assessment, and academic integrity (Gulikers et al., 2004). This review made it possible to delimit the research problem, define the instrument domains, and formulate the questionnaire items.

The questionnaire domains were defined on the basis of a literature review on artificial intelligence in higher education, AI literacy, authentic assessment, and academic integrity (Ng et al., 2021). This review made it possible to translate the research problem into observable dimensions: generative AI literacy, academic use of generative tools, authentic assessment, academic integrity, and learning traceability. The initial items were formulated on the basis of these domains and were subsequently reviewed by specialists.

Participants and research setting

The study was conducted at an Ecuadorian higher education institution, identified by the code IES-A to preserve its confidentiality when reporting the results. The institutional name and details regarding location, academic programs, courses, and authorities that could facilitate identification were omitted.

The accessible population consisted of undergraduate students and faculty members involved in assessment activities mediated by digital resources. The sample included 603 participants, comprising 482 students and 121 faculty members. The inclusion of both groups reflected the interest in examining complementary perspectives on the assessment process: students participate in completing tasks and producing evidence of learning, whereas faculty members are involved in designing, administering, and providing feedback on assessments. This complementarity justifies their participation in the study, but does not imply that their responses are interchangeable or should be interpreted as a homogeneous measure of the university community.

The questionnaire included differentiated instructions to contextualize responses according to institutional role. Students responded with reference to assessments they complete in their courses, whereas faculty members considered assessments they design, administer, or provide feedback on. Nevertheless, some items concerning the preparation of assignments, use of feedback, and defense of the process followed may have different meanings across groups. Therefore, the differentiated instructions were not considered sufficient evidence of measurement equivalence.

In view of these differences, descriptive analyses and correlations among domains were conducted separately for students and faculty members, together with exploratory comparisons of the observed scores between the two groups. These comparisons were not interpreted as conclusive differences in the underlying constructs because measurement invariance of the instrument was not assessed. Associations obtained in the combined sample were treated as complementary aggregate results whose interpretation requires comparison with the group-specific analyses.

Participants were selected through nonprobability convenience sampling. Eligible participants were adult undergraduate students and active faculty members with experience in university assessment who had provided informed consent. The protocol established the following exclusion criteria: absence of consent, failure to meet participation requirements, incomplete responses on more than 20% of the items, evident duplicate responses, and inconsistent response patterns. Partial responses within the permitted limit were retained; calculation of each domain score required at least four valid responses among its six items. Because of the sampling procedure and participation from a single institution, the results are limited to the analyzed sample and do not support statistical generalization to Ecuadorian higher education.

Ethical considerations statement

The study was conducted in accordance with the principles of autonomy, voluntariness, privacy, confidentiality, and risk minimization applicable to research involving human participants. An anonymous, voluntary, self-administered survey was used exclusively with adult participants, without experimental interventions, manipulation of variables, invasive procedures, collection of biological samples, or other actions that could affect their physical or psychological integrity. Before accessing the questionnaire, participants received information about the academic purpose of the study, the voluntary nature of participation, data processing, measures intended to preserve confidentiality, and their right to leave the form before submission, after which they electronically provided consent to participate.

These safeguards were established in accordance with Article 66, paragraph 19, of the Constitution of the Republic of Ecuador, which recognizes the right to protection of personal data and provides that its collection and processing require authorization from the data subject or a legal mandate, as well as with the principles set forth in the Organic Law on Personal Data Protection, particularly those concerning free, specific, informed, and unequivocal consent; purpose limitation; relevance and data minimization; confidentiality; and information security (Asamblea Nacional del Ecuador, 2008, 2021). In addition, the database used for analysis consisted exclusively of anonymous responses, and the results

were processed and reported in aggregate form without retaining elements that would enable individual identification or re-identification of participants, in accordance with Ecuadorian provisions on the protection and anonymization of personal data.

Regarding institutional ethical review, the study was not previously submitted to a Human Subjects Research Ethics Committee (CEISH). This circumstance is explicitly disclosed and is not presented as equivalent to approval, waiver, or formal exemption granted by a committee. Nevertheless, based on its methodological characteristics, the study corresponded to a low-risk, observational, non-interventional design based on anonymous collection of information through a questionnaire, without identifying data, sensitive data, biological samples, participation of vulnerable populations, or inquiries into sensitive aspects of behavior. These characteristics are consistent with the criteria established in Ecuadorian regulations for identifying no-risk research potentially eligible for exempt review, pursuant to the Regulation for the Approval, Development, Monitoring, and Control of Observational Research and Intervention Studies in Human Beings and its amendment through Ministerial Agreement No. 00038-2021 (Ministerio de Salud Pública del Ecuador, 2021), as well as with the operational criteria established for this type of research by CEISH committees authorized by the national health authority.

The absence of prior CEISH review therefore constitutes a procedural condition that is reported transparently, without attributing to the study an ethical exemption that was not formally issued. According to Article 47 of the Substitute Regulation for the Approval and Monitoring of Human Subjects Research Ethics Committees, issued through Ministerial Agreement No. 00005-2022, CEISH committees do not review protocols whose implementation is already underway or has been completed; therefore, retrospective ethical approval is not applicable to the research already conducted (Ministerio de Salud Pública del Ecuador, 2022). Consequently, participant protection was based on voluntary participation, electronic informed consent, absence of intervention, anonymous data collection, the impossibility of individually identifying participants, and exclusively aggregate reporting of results.

Research instrument

The research instrument was a structured 30-item questionnaire organized into five domains, with six items in each domain. The domains were generative artificial intelligence literacy, academic use of generative tools, authentic assessment, academic integrity, and learning traceability.

Table 1. *Questionnaire domains*

Domain	Purpose	Guiding items
Generative artificial intelligence literacy	Assess knowledge and criteria for critical use.	- I understand the basic functions of generative AI. - I verify generated information. - I recognize limitations, biases, and risks related to academic work.
Academic use of generative tools	Identify purposes of use in university tasks.	- I use generative AI to organize ideas. - I use it to review texts. - I use it as support for studying or synthesizing information.
Authentic assessment	Assess tasks focused on applied performance.	- Assessments require solving contextualized problems. - Tasks require original argumentation. - Criteria are communicated through rubrics.

Academic integrity	Examine authorship, transparency, and academic responsibility.	- I disclose the use of AI in academic tasks. - I cite the sources used. - I assume responsibility for the submitted content.
Learning traceability	Identify evidence of the process followed by the student.	- I submit progress work or drafts. - I justify decisions made during the task. - I can orally defend the process followed.

Note. Prepared by the authors

The 30 items were answered using a five-category ordered frequency scale: never (1), rarely (2), sometimes (3), frequently (4), and always (5). In addition, the option "I cannot answer" was included; it was recorded in the database as a missing value and excluded from numerical scoring. This option did not constitute an additional scale level and was not interpreted as an intermediate response.

The "I cannot answer" option was coded as NA and treated as missing data; it was not replaced with zero, the scale midpoint, or an imputed value. For each participant, the score for each domain was calculated as the arithmetic mean of valid responses to its six items, provided that at least four items had been answered, in accordance with the coding protocol. The denominator corresponded to the number of items actually answered: four, five, or six, as applicable. The sample mean for each domain was subsequently obtained by averaging the available individual scores.

In the analyzed database, the number of valid responses per item ranged from 582 to 597 because of missing values. However, all 603 participants answered at least four items in each of the five domains and therefore had a valid score for every domain. This explains why item-level analyses had different valid sample sizes, whereas domain-level analyses retained n = 603, without requiring imputation of missing values.

The questionnaire included a final open-ended response section. This section collected observations on the use of generative artificial intelligence in university assessment. The open-ended responses were used as complementary input for interpreting the descriptive results.

Before administration, the instrument underwent face and content validation. Face validation assessed clarity, comprehensibility, and absence of ambiguity. Content validation examined the relevance, representativeness, and sufficiency of the items with respect to each domain. This procedure followed criteria for the development and validation of instruments in educational and social research (Polit & Beck, 2006; Boateng et al., 2018).

The content validity of the questionnaire was established through the judgment of three experts specializing in higher education, assessment, educational technology, and academic integrity, who examined the theoretical correspondence among domains, indicators, and items. Based on their qualitative judgments, the conceptual precision of the items was optimized and syntactic redundancies were removed.

Questionnaire administration

Following institutional authorization, the questionnaire was administered online through Google Forms between April and May 2026. The invitation was sent to faculty members and students through

authorized institutional channels. The access link first presented the informed consent form; only participants who accepted its conditions were able to complete the instrument.

The questionnaire was organized into three sections. The first collected general information: institutional role, age, sex, academic field, teaching or administrative experience, frequency of use of digital tools, and prior experience with generative artificial intelligence. The second included the 30 items distributed across the five study domains. The third included an open-ended question to record observations on the use of generative artificial intelligence in university assessment.

To mitigate ethical risks and ensure confidentiality, data collection omitted explicit personal identifiers and used a self-administered, voluntary, and free-of-charge format. Incomplete responses were handled according to the coding protocol: records within the permitted missing-data limit were retained, and domain scores were calculated when at least four valid responses were available.

Institutional authorization allowed the questionnaire to be distributed through formal channels and defined the academic use of the information. Data were processed in aggregate form to protect the identities of participants and of the institution coded as IES-A.

Statistical analysis

The data were organized in a matrix and cleaned before analysis. Participant characteristics were summarized using frequencies and percentages. Items and domains were described using the mean, standard deviation, and variance.

Internal consistency was examined using Cronbach's alpha for each domain, based on cases with complete responses to the corresponding six items. As a complementary analysis, the coefficient for all 30 items was calculated using cases with complete questionnaire responses. Missing values were not imputed for these calculations. The coefficients were interpreted as evidence of internal consistency in the sample, without attributing to them the ability to demonstrate the instrument's factor structure or construct validity.

Associations among domain scores were examined using Spearman's rank correlation coefficient to describe monotonic relationships among scores derived from ordinal items without requiring bivariate normality. Two-sided tests were performed with a significance level of 0.05. Correlations were calculated separately for students and faculty members; results for the combined sample were considered complementary. The coefficients were interpreted as measures of association without assigning them causal meaning.

Exploratory comparisons between students and faculty members were performed using the two-sided Mann–Whitney U test with asymptotic approximation and correction for ties and continuity. The Holm procedure was applied to the five between-group comparisons and, separately, to the ten correlations within each group. Adjusted p values below 0.05 were considered statistically significant.

RESULTS

Participant characteristics

The sample consisted of 603 participants: 482 students and 121 faculty members. Students represented 79,9% of the sample and faculty members 20,1%. The largest proportion of participants was in the 18–24-year age group. Female participants predominated with respect to sex.

Table 2. *Participant characteristics*

Variable	Category	n	%
Institutional role	Student	482	79,9
	Faculty	121	20,1
Sex	Female	353	58,5
	Male	246	40,8
	Prefer not to answer	4	0,7
Age	18-24 years	390	64,7
	25-34 years	99	16,4
	35-44 years	60	10,0
	45-54 years	42	7,0
	55 years or older	12	2,0
Academic field	Social sciences, education, or humanities	205	34,0
	Administrative or economic sciences	151	25,0
	Health sciences	117	19,4
	Engineering, technology, or exact sciences	95	15,8
	Other	35	5,8
Frequency of digital tool use	Rarely	18	3,0
	Sometimes	105	17,4
	Frequently	281	46,6
	Always	199	33,0
Prior experience with generative AI	I have not used it	39	6,5
	Once or twice	77	12,8
	Occasional	181	30,0
	Frequent	208	34,5
	Routine	98	16,3

Note. Prepared by the authors

Descriptive analysis by domain

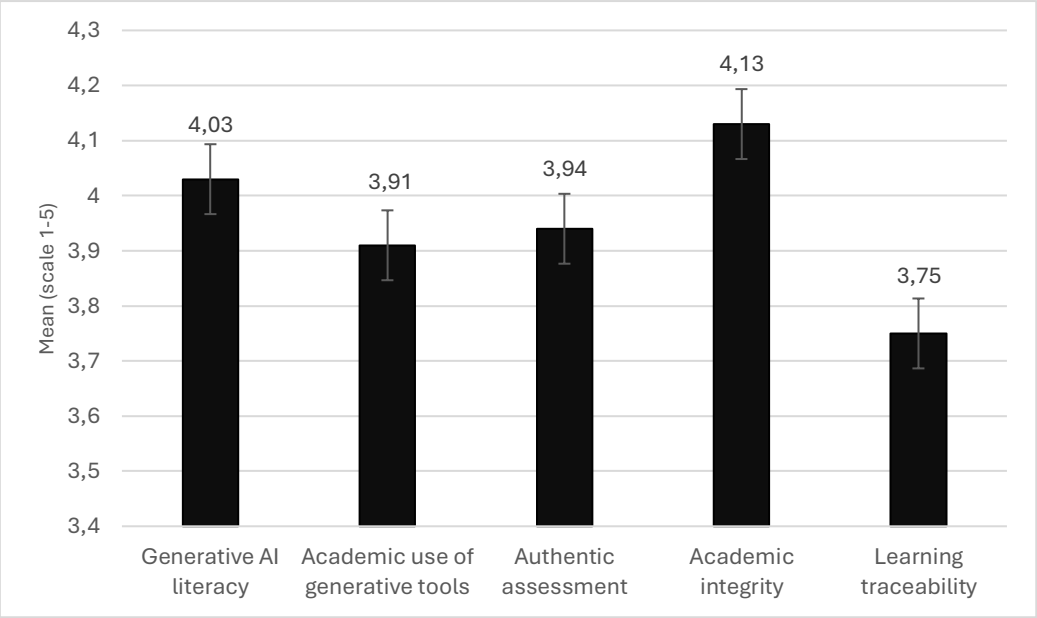
All five domains had mean scores above the midpoint of the scale. Academic integrity had the highest mean (M = 4,13; SD = 0,53), followed by generative AI literacy (M = 4,03; SD = 0,54). Learning traceability had the lowest mean (M = 3,75; SD = 0,59). This pattern suggests a favorable perception of academic integrity and AI literacy, with lower self-reported scores for learning traceability.

Table 3. *Descriptive statistics by domain*

Domain	Valid n	Mean	SD	Variance
Generative AI literacy	603	4,03	0,54	0,29
Academic use of generative tools	603	3,91	0,58	0,34
Authentic assessment	603	3,94	0,55	0,30
Academic integrity	603	4,13	0,53	0,28
Learning traceability	603	3,75	0,59	0,35

Note. Prepared by the authors

Figure 1. *Mean score by questionnaire domain*



Note. Prepared by the authors

Table 4. *Descriptive statistics by domain and exploratory comparison between students and faculty*

Domain	Students: M (SD)	Faculty: M (SD)	Mann–Whitney U	Unadjusted p	Adjusted p
Generative AI literacy	4,03 (0,55)	4,03 (0,49)	29 364,0	0,906	0,906
Academic use of generative tools	3,96 (0,56)	3,75 (0,63)	34 242,5	0,003	0,015
Authentic assessment	3,93 (0,55)	3,99 (0,56)	27 270,5	0,269	0,537
Academic integrity	4,11 (0,54)	4,19 (0,49)	26 856,5	0,177	0,531
Learning traceability	3,73 (0,58)	3,85 (0,61)	25 480,0	0,031	0,125

Note. Prepared by the authors from the study database. Students: n = 482; faculty: n = 121. M = mean; SD = standard deviation. Scores are expressed on a scale from 1 to 5. Comparisons were performed using the two-sided Mann–Whitney U test with asymptotic approximation and correction for ties and continuity. The U statistic reported corresponds to the student group. Adjusted p values correspond to the Holm procedure applied to the five comparisons. The test compares score distributions using ranks; it is not a test of differences between means.

The exploratory between-group comparison showed that, after Holm adjustment, only the academic use of generative tools exhibited a statistically significant difference (adjusted p = 0,015), with higher scores among students. No statistically significant differences were identified in the remaining domains after adjustment, although this result does not demonstrate equivalence between groups. The observed differences are limited to self-report scores within the sample and should be interpreted in light of the different roles of students and faculty members in the assessment process and the absence of evidence regarding measurement equivalence between the two groups.

Item-level descriptive analysis

The item-level analysis described the scores for the items within each domain. In generative AI literacy, means ranged from 3,96 to 4,09. In academic use of generative tools, values ranged from 3,86 to 3,98. In authentic assessment, means ranged from 3,90 to 4,06. In academic integrity, all items remained above 4,08. In learning traceability, means were lower and ranged from 3,70 to 3,82.

Table 5. *Item coding*

Domain	Items
Generative AI literacy	AIG1-AIG6
Academic use of generative tools	UAG1-UAG6
Authentic assessment	EA1-EA6
Academic integrity	IA1-IA6
Learning traceability	TA1-TA6

Note. Prepared by the authors

Table 6. *Descriptive statistics by item*

Code	Domain	Valid n	Mean	SD	Variance
AIG1	AIG	593	4,09	0,72	0,52
AIG2	AIG	595	4,06	0,71	0,50
AIG3	AIG	595	3,96	0,74	0,54
AIG4	AIG	593	4,06	0,69	0,48
AIG5	AIG	594	4,00	0,71	0,50
AIG6	AIG	593	4,02	0,72	0,52
UAG1	UAG	588	3,96	0,75	0,56
UAG2	UAG	588	3,98	0,76	0,57
UAG3	UAG	585	3,86	0,74	0,54
UAG4	UAG	586	3,88	0,76	0,57
UAG5	UAG	592	3,86	0,76	0,58
UAG6	UAG	591	3,95	0,73	0,54
EA1	EA	590	3,90	0,71	0,51
EA2	EA	584	3,93	0,75	0,56
EA3	EA	582	3,93	0,73	0,53
EA4	EA	582	4,06	0,72	0,51
EA5	EA	593	3,92	0,73	0,54
EA6	EA	588	3,90	0,73	0,53
IA1	IA	593	4,19	0,69	0,47
IA2	IA	597	4,11	0,71	0,51
IA3	IA	591	4,09	0,72	0,52
IA4	IA	596	4,19	0,68	0,46
IA5	IA	592	4,09	0,70	0,50
IA6	IA	587	4,09	0,73	0,53
TA1	TA	585	3,72	0,77	0,60
TA2	TA	591	3,76	0,76	0,58
TA3	TA	586	3,77	0,75	0,57
TA4	TA	593	3,72	0,75	0,56
TA5	TA	597	3,70	0,76	0,58
TA6	TA	594	3,82	0,76	0,58

Note. Prepared by the authors

Instrument reliability

Internal consistency was examined using Cronbach's alpha, based, for each domain, on participants with complete responses to all six items and without imputing missing values. Coefficients ranged from 0,846 to 0,867. As a complementary analysis, alpha for the full set of 30 items was 0,918 and was calculated using the 325 participants who answered every item. These results provide evidence of internal consistency for the scores in the analyzed sample, but they do not demonstrate questionnaire unidimensionality, the five-domain structure, or measurement equivalence between students and faculty members.

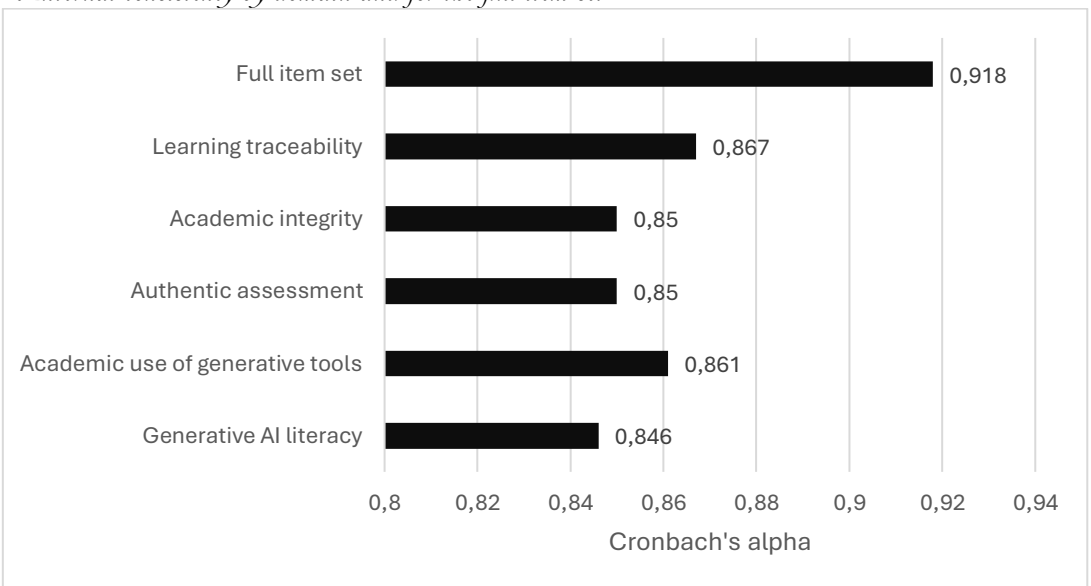
Table 7. *Internal consistency by domain and for the full item set*

Domain or item set	Number of items	Participants with complete responses	Cronbach's alpha
Generative AI literacy	6	550	0,846
Academic use of generative tools	6	522	0,861
Authentic assessment	6	510	0,850
Academic integrity	6	542	0,850
Learning traceability	6	535	0,867
Full item set	30	325	0,918

Note. Prepared by the authors from the study database. Each coefficient was calculated by excluding cases with any missing response within the corresponding item set, without imputation. The analytical sample sizes in this table apply exclusively to the alpha calculation and do not alter the 603 cases available for domain-score analyses. The coefficient for the full item set does not demonstrate the existence of a single dimension and, by itself, does not justify a global score.

Face and content review by specialists and internal consistency analysis address different aspects of instrument evaluation. The former examines clarity and the conceptual correspondence between items and the proposed domains, whereas alpha provides information on their interrelationships. Consequently, the five domains are regarded as a theoretical organization of the questionnaire whose internal structure requires further evaluation through factor analyses; they are not presented as an empirically validated structure on the basis of the coefficients obtained.

Figure 2. *Internal consistency by domain and for the full item set*



Note. Prepared by the authors. Coefficients were calculated using complete cases for each domain and for the full set of 30 items, without imputation. The values do not constitute evidence of construct validity.

Correlations among domains

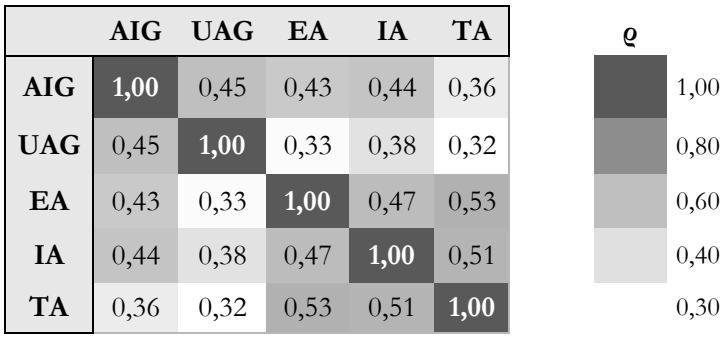
All correlations among domains were positive and statistically significant ($p < 0,001$). The strongest associations were observed between authentic assessment and learning traceability (rho = 0,53), academic integrity and learning traceability (rho = 0,51), and authentic assessment and academic integrity (rho = 0,47). In the combined sample, these results describe positive associations among self-reported scores for authentic assessment, learning traceability, and academic integrity.

Table 8. *Spearman correlations among domains*

Domain	AIG	UAG	EA	IA	TA
AIG	1,00	0,45	0,43	0,44	0,36
UAG	0,45	1,00	0,33	0,38	0,32
EA	0,43	0,33	1,00	0,47	0,53
IA	0,44	0,38	0,47	1,00	0,51
TA	0,36	0,32	0,53	0,51	1,00

Note. Prepared by the authors

Figure 3. *Spearman correlation matrix*



Note. Prepared by the authors

Spearman correlations were calculated separately for students and faculty members, considering the ten possible associations among the five domains. All were positive and statistically significant in both groups, even after Holm adjustment applied to the ten tests within each group (adjusted $p < 0,001$). Coefficients ranged from 0,333 to 0,530 among students and from 0,299 to 0,516 among faculty members.

Table 9. *Spearman correlations among domains, by institutional role*

Association between domains	Students: ϱ	Faculty: ϱ
Generative AI literacy and academic use of generative tools	0,473	0,360
Generative AI literacy and authentic assessment	0,459	0,299
Generative AI literacy and academic integrity	0,468	0,343
Generative AI literacy and learning traceability	0,373	0,320
Academic use of generative tools and authentic assessment	0,334	0,357
Academic use of generative tools and academic integrity	0,366	0,492
Academic use of generative tools and learning traceability	0,333	0,361
Authentic assessment and academic integrity	0,466	0,477
Authentic assessment and learning traceability	0,530	0,516
Academic integrity and learning traceability	0,507	0,500

Note. Prepared by the authors from the study database. ϱ = Spearman correlation coefficient. Students: n = 482; faculty: n = 121 for all associations. Tests were two-sided; all p values were below 0,001 both before and after Holm adjustment applied separately to the ten correlations within each group. The coefficients describe associations within each group and do not constitute a test of differences between student and faculty correlations.

In both groups, the strongest association was between authentic assessment and learning traceability (ϱ = 0,530 among students and ϱ = 0,516 among faculty members), followed by the association between academic integrity and learning traceability (ϱ = 0,507 and ϱ = 0,500, respectively). This pattern indicates that higher scores in authentic assessment and academic integrity tended to be accompanied by higher traceability scores within each group. The descriptive similarity between groups does not demonstrate that their correlations are statistically equivalent and does not permit causal relationships among the domains to be established.

Overall, the survey data showed mean scores above the scale midpoint across all five domains, with the highest value for academic integrity and the lowest for learning traceability. Associations among authentic assessment, academic integrity, and learning traceability were positive for both students and faculty members. These results describe relationships among self-report scores within each group, without establishing causal relationships or demonstrating the construct validity of the instrument.

DISCUSSION

The results show favorable ratings across the five domains analyzed: generative artificial intelligence literacy, academic use of generative tools, authentic assessment, academic integrity, and learning traceability (Ashford et al., 2013). Academic integrity had the highest mean, whereas learning traceability had the lowest.

This pattern indicates that participants recognize the importance of academic responsibility in contexts mediated by generative artificial intelligence, but perceive less consolidated practices aimed at documenting, justifying, and defending the process followed in university assignments (Morris, 2018).

This finding is related to the results reported by Zawacki-Richter et al. (2019) in the article "Systematic review of research on artificial intelligence applications in higher education – where are the educators?". The authors identified that artificial intelligence in higher education has been studied primarily in applications such as assessment, personalization, intelligent tutoring, and academic services; however, they also noted limited critical reflection on its ethical and pedagogical risks. In this regard, the present study provides evidence from the perspectives of faculty members and students in precisely an area that the literature has considered insufficiently developed: the relationship among AI use, assessment, and academic integrity (Chan, 2023).

The high score for generative AI literacy is consistent with the need to develop specific competencies for understanding, using, and critically evaluating these systems. Laupichler et al. (2022), in "Artificial intelligence literacy in higher and adult education: A scoping literature review", argue that AI literacy in higher education still requires greater conceptual precision, systematic training, and validated measurement scales. The results of the present study are consistent with this line of work, showing that AI literacy cannot be reduced to the instrumental use of generative tools; it must also include information verification, recognition of bias, understanding of limitations, and responsibility for disclosing use (Bender et al., 2021; Floridi & Chiriatti, 2020).

The academic use of generative tools domain had a favorable mean, although it was lower than those for AI literacy and academic integrity. This difference is relevant. It suggests that participants recognize ethical and conceptual criteria regarding generative AI, but its practical integration into academic tasks has not yet reached the same level of consolidation (Baidoo & Owusu, 2023; Lo, 2023). This result is consistent with Kasneci et al. (2023) in "ChatGPT for good? On opportunities and challenges of large language models for education", which recognizes that language models can support writing, feedback, information synthesis, and personalized learning, but require competencies to understand their limitations, verify information, and avoid uncritical use.

The highest mean for academic integrity may be interpreted as a positive, although insufficient, signal. Participants appear to value authorship, citation, transparency, and responsibility for the academic product (McCabe et al., 2001; Newton & Lang, 2016). Nevertheless, academic integrity in generative AI environments does not depend solely on awareness of rules. It also requires pedagogical conditions that make it possible to distinguish legitimate technological support from inappropriate delegation of cognitive processes. In this regard, Cotton et al. (2023), in "Chatting and cheating: Ensuring academic integrity in the era of ChatGPT", warn that the availability of ChatGPT requires assessment strategies to be reconsidered because institutional responses focused exclusively on detection or prohibition are limited when faced with tools capable of producing academic texts that appear to be human-authored.

The lowest result for learning traceability represents one of the study's main contributions. Although the mean remained above the scale midpoint, this domain had the lowest mean score. This result reflects lower reported scores on items related to process documentation: drafts, interim submissions, justifications, oral defense, feedback, and reconstruction of decisions. This finding is important because traceability may serve as a bridge between authentic assessment and academic integrity. When the learning process is documented, dependence on external surveillance mechanisms is reduced and pedagogical evidence of how the final product was constructed is strengthened (Tlili et al., 2023).

The strongest correlation was observed between authentic assessment and learning traceability. This result is consistent with the foundations of authentic assessment, which prioritizes contextualized tasks, applied performance, and evidence of learning linked to real-world situations (Wiggins, 1990). The authentic assessment literature maintains that this approach is intended to measure meaningful, situated performance rather than only final responses or reproduction of content. From this perspective, the observed association indicates that higher authentic assessment scores are related to higher learning traceability scores, particularly when assessments incorporate staged submissions, feedback, and reasoned defense.

The positive relationship between authentic assessment and academic integrity is also consistent with studies on assessment design and prevention of academic misconduct. Bretag et al. (2019), in "Contract cheating and assessment design: Exploring the relationship", analyzed the connection between assessment design and conditions that may facilitate or reduce academic dishonesty. The results of the present study align with this perspective: contextualized assessments with explicit criteria and a process-centered approach appear to be associated with more favorable conditions for supporting academic integrity. Nevertheless, this association does not imply that authentic assessment alone eliminates the risk of inappropriate use of generative AI. On the basis of this association, future research may examine the hypothesis of a possible protective role of assessment design when combined with AI literacy and learning traceability; its effectiveness was not evaluated in this study (Newton P., 2016).

The positive correlation between academic integrity and learning traceability reinforces this interpretation. In contexts where generative AI can produce texts, outlines, summaries, or arguments within seconds, evidence of the process acquires methodological and ethical value. Assessing only the final product is insufficient. It is necessary to observe how the student formulated the problem, selected sources, contrasted information, made decisions, incorporated feedback, and justified the outcome. This orientation shifts assessment from a logic centered on suspicion toward one centered on demonstrating learning.

Based on these findings, pedagogical guidelines are proposed whose effectiveness requires subsequent evaluation. First, institutions should move from restrictive policies toward educational policies on the use of generative AI. AI literacy should include technical skills, ethical criteria, information verification, recognition of bias, and responsibility for authorship. Second, assessment redesign should

prioritize authentic, contextualized tasks that are difficult to complete through automated delegation. Third, traceability should be incorporated as a formal assessment criterion through drafts, logs, oral defense, process rubrics, AI-use reports, and evidence of feedback.

The internal consistency analysis provided information on the interrelationships among items in the study sample. Therefore, the coefficients obtained do not demonstrate that the five domains constitute an empirically validated structure or that the full set of 30 items measures a single dimension. The proposed organization requires additional evaluation of its internal structure and measurement equivalence between students and faculty members.

However, the results should be interpreted with caution. The study used nonprobability convenience sampling; therefore, the findings cannot be generalized to all Ecuadorian higher education. In addition, the data are based on self-report and may therefore be subject to social desirability bias, particularly for items related to academic integrity. The difference between what participants report and what they actually do in assessment activities should be considered in future studies. Further research is also needed on differences by institutional role, academic field, level of study, and frequency of generative AI use.

Although the instructions contextualized responses according to institutional role, some items concerning assignment preparation, use of feedback, and defense of academic work may have different meanings for students and faculty members. Measurement invariance between the two groups was not assessed; therefore, observed score differences cannot be interpreted conclusively as differences in construct levels. Separate analysis by role makes it possible to examine associations within each group, but does not eliminate this limitation in comparability.

Future research should incorporate mixed-method designs, comparative analyses across faculties or academic programs, interviews with faculty members and students, and documentary review of assessment tasks. It would also be relevant to analyze actual evidence of traceability, such as drafts, rubrics, logs, AI-use reports, and feedback records. This line of research would make it possible to move from reported perceptions to observation of concrete assessment practices.

CONCLUSION

This study analyzed the relationship among generative artificial intelligence literacy, authentic assessment, and academic integrity in higher education. The findings show positive associations among the scores for these domains in the study sample. The participation of faculty members and students made it possible to examine the phenomenon from the perspectives of two central stakeholders in the educational process, with a sample comprising 482 students and 121 faculty members.

Academic integrity had the highest mean, followed by generative artificial intelligence literacy. This result reflects participants' reported assessments of authorship, transparency, and academic responsibility in relation to the use of generative tools. However, learning traceability had the lowest mean, indicating lower self-reported scores for aspects related to documenting the development process, justifying decisions, presenting progress, and defending the learning achieved.

The positive and statistically significant correlations among the variables show associations among the analyzed domains, but do not by themselves confer construct validity on the proposed model. In particular, authentic assessment was positively associated with learning traceability and academic integrity, although these relationships do not establish that situated, process-oriented tasks produce greater academic transparency. These findings suggest the relevance of exploring institutional responses

that incorporate educational and assessment strategies in addition to surveillance and detection mechanisms.

The instrument showed adequate internal consistency in all domains and high internal consistency across the full set of 30 items. This result provides evidence regarding the internal consistency of scores in the analyzed sample and supports considering the questionnaire as an initial exploratory tool whose five-domain structure requires further evaluation. Future application should be extended to different institutions, fields of knowledge, and educational modalities, together with analyses of its internal structure and measurement equivalence between students and faculty members.

From a pedagogical perspective, the results suggest the relevance of considering generative artificial intelligence literacy in university assessment policies. As guidance for its incorporation, clear criteria are proposed regarding permitted use, disclosure of tools, information verification, citation of sources, and responsibility for the final product. Authentic assessment should also be complemented with process evidence such as drafts, logs, rubrics, oral defenses, and AI-use reports, whose contribution to academic integrity should be evaluated in future studies.

The study provides empirical evidence limited to a nonprobability sample from one institution to support reflection on the redesign of assessment in higher education. Situated, transparent, and traceable assessments constitute a relevant line of work for exploring academic integrity in environments mediated by generative artificial intelligence, although the present study did not demonstrate their effectiveness. Future research should contrast these findings through longitudinal designs, analyses by disciplinary area, comparative studies across institutions, and direct review of assessment evidence produced by students.

REFERENCES

Asamblea Nacional del Ecuador. (2008). Constitución de la República del Ecuador. Registro Oficial No. 449.

Asamblea Nacional del Ecuador. (2021). Ley Orgánica de Protección de Datos Personales. Quinto Suplemento del Registro Oficial No. 459.

Ashford-Rowe, K., Herrington, J., & Brown, C. (2014). Establishing the critical elements that determine authentic assessment. Assessment & Evaluation in Higher Education, 39(2), 205–222. https://doi.org/10.1080/02602938.2013.819566

Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. Journal of AI, 7(1), 52–62. https://doi.org/10.61969/jai.1337500

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 610–623). Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. Frontiers in Public Health, 6, Article 149. https://doi.org/10.3389/fpubh.2018.00149

Bretag, T., Harper, R., Burton, M., Ellis, C., Newton, P., van Haeringen, K., Saddiqui, S., & Rozenberg, P. (2019). Contract cheating and assessment design: Exploring the relationship. Assessment & Evaluation in Higher Education, 44(5), 676–691. https://doi.org/10.1080/02602938.2018.1527892

Bretag, T., Harper, R., Burton, M., Ellis, C., Newton, P., Rozenberg, P., Saddiqui, S., & van Haeringen, K. (2019). Contract cheating: A survey of Australian university students. Studies in Higher Education, 44(11), 1837–1856. https://doi.org/10.1080/03075079.2018.1462788

Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. International Journal of Educational Technology in Higher Education, 20, Article 38. https://doi.org/10.1186/s41239-023-00408-3

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. IEEE Access, 8, 75264–75278. https://doi.org/10.1109/ACCESS.2020.2988510

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. Innovations in Education and Teaching International, 61(2), 228–239. https://doi.org/10.1080/14703297.2023.2190148

Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. International Journal of Educational Technology in Higher Education, 20, Article 22. https://doi.org/10.1186/s41239-023-00392-8

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. Minds and Machines, 30(4), 681–694. https://doi.org/10.1007/s11023-020-09548-1

Gulikers, J. T. M., Bastiaens, T. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. Educational Technology Research and Development, 52(3), 67–86. https://doi.org/10.1007/BF02504676

Herrington, J., & Oliver, R. (2000). An instructional design framework for authentic learning environments. Educational Technology Research and Development, 48(3), 23–48. https://doi.org/10.1007/BF02319856

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Laupichler, M. C., Aster, A., Schirch, J., & Raupach, T. (2022). Artificial intelligence literacy in higher and adult education: A scoping literature review. Computers and Education: Artificial Intelligence, 3, Article 100101. https://doi.org/10.1016/j.caeai.2022.100101

Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. Education Sciences, 13(4), Article 410. https://doi.org/10.3390/educsci13040410

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1–16). Association for Computing Machinery. https://doi.org/10.1145/3313831.3376727

McCabe, D. L., Treviño, L. K., & Butterfield, K. D. (2001). Cheating in academic institutions: A decade of research. Ethics & Behavior, 11(3), 219–232. https://doi.org/10.1207/S15327019EB1103_2

Ministerio de Salud Pública del Ecuador. (2021). Acuerdo Ministerial No. 00038-2021: Reforma al Reglamento para la aprobación, desarrollo, vigilancia y control de investigaciones observacionales y estudios de intervención en seres humanos.

Ministerio de Salud Pública del Ecuador. (2022). Acuerdo Ministerial No. 00005-2022: Reglamento Sustitutivo del Reglamento para la Aprobación y Seguimiento de Comités de Ética de Investigación en Seres Humanos (CEISH) y Comités de Ética Asistenciales para la Salud (CEAS). Quinto Suplemento del Registro Oficial No. 118.

Morris, E. J. (2018). Academic integrity matters: Five considerations for addressing contract cheating. International Journal for Educational Integrity, 14, Article 15. https://doi.org/10.1007/s40979-018-0038-5

Newton, P. (2016). Academic integrity: A quantitative study of confidence and understanding in students at the start of their higher education. Assessment & Evaluation in Higher Education, 41(3), 482–497. https://doi.org/10.1080/02602938.2015.1024199

Newton, P. M., & Lang, C. (2016). Custom essay writers, freelancers, and other paid third parties. In T. Bretag (Ed.), Handbook of academic integrity (pp. 249–271). Springer. https://doi.org/10.1007/978-981-287-098-8_38

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. Computers and Education: Artificial Intelligence, 2, Article 100041. https://doi.org/10.1016/j.caeai.2021.100041

Perkins, M. (2023). Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond. Journal of University Teaching and Learning Practice, 20(2). https://doi.org/10.53761/1.20.02.07

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. Research in Nursing & Health, 29(5), 489–497. https://doi.org/10.1002/nur.20147

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. Smart Learning Environments, 10, Article 15. https://doi.org/10.1186/s40561-023-00237-x

UNESCO. (2025, 29 de septiembre). Qué debe saber sobre la IA y el derecho a la educación. https://www.unesco.org/es/articles/que-debe-saber-sobre-la-ia-y-el-derecho-la-educacion

Villarroel, V., Bloxham, S., Bruna, D., Bruna, C., & Herrera-Seda, C. (2018). Authentic assessment: Creating a blueprint for course design. Assessment & Evaluation in Higher Education, 43(5), 840–854. https://doi.org/10.1080/02602938.2017.1412396

Wiggins, G. (1990). The case for authentic assessment. Practical Assessment, Research, and Evaluation, 2, Article 2. https://doi.org/10.7275/ffb1-mm19

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators? International Journal of Educational Technology in Higher Education, 16, Article 39. https://doi.org/10.1186/s41239-019-0171-0

Submitted: 08/20/2026
Preprint: 08/10/2026
Accepted: 09/19/2026
Section Editor: Suzana dos Santos Gomes

RESEARCH DATA AVAILABILITY STATEMENT

The research materials underlying this study are available as supplementary files in the journal repository. File No. 283536, entitled "base de datos.xlsx", and File No. 278749, entitled "Consentimiento_y_Declaracion_Aspectos_Eticos.pdf", are deposited there.

AUTHOR CONTRIBUTIONS STATEMENT

Germán Vinicio Roblez Torres - Conceptualization, Data Curation, Writing: Review and Editing.
Alexandra Maricruz Andrade Apolo - Investigation, Methodology, Project Administration.
Katherine Elizabeth Gonzaga Aguilar - Software, Validation, Writing: Original Draft.

CONFLICT OF INTEREST STATEMENT

The authors declare that there is no conflict of interest regarding this article.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

ChatGPT (OpenAI) was used exclusively to support improvements in academic writing and the organization of bibliographic references. The tool was not used to define the research problem, design the methodology, develop the instrument, collect or analyze data, interpret the results, or formulate the conclusions. The analyzed data were obtained from the responses to the questionnaire administered to the 603 participants and were not generated or simulated using artificial intelligence. All content supported by this tool was critically reviewed by the authors, who assume full responsibility for the accuracy, originality, and final content of the manuscript.