

INPLASY

INPLASY202650147

doi: 10.37766/inplasy2026.5.0147

Received: 26 May 2026

Published: 27 May 2026

Corresponding author:

Willy Adatao-Medina

wadatao@untels.edu.pe

Author Affiliation:

Universidad Nacional Tecnológica de Lima Sur.

Assessment Instruments for Generative AI Literacy in Higher Education: A Systematic Review of Measurement Approaches, Psychometric Properties, and Implications for SDG 4

Adatao-Medina, W; León-Velarde, C; Fernández-Flores, S; Arévalo-Tuesta, JA; Aybar-Bellido, I; Arones, M; Caycho-Salas, B; Quispe-Andía, A.

ADMINISTRATIVE INFORMATION

Support - This research received no external funding.

Review Stage at time of this submission - The review has not yet started.

Conflicts of interest - None declared.

INPLASY registration number: INPLASY202650147

Amendments - This protocol was registered with the International Platform of Registered Systematic Review and Meta-Analysis Protocols (INPLASY) on 27 May 2026 and was last updated on 27 May 2026.

INTRODUCTION

Review question / Objective General Objective

The aim of this systematic review is to identify, characterize, and critically appraise the instruments and measurement approaches that have been used, developed, or adapted to assess Generative AI literacy in higher education. This review will examine their dimensions, constructs, development processes, application contexts, and psychometric properties, while exploring the potential implications of these instruments for SDG 4 (Quality Education), with emphasis on digital competencies, responsible AI use, and inclusive learning opportunities. This review does not assume that existing instruments directly measure SDG 4.

Review Questions

This systematic review will address the following research questions:

- RQ1: What instruments and measurement approaches have been used to assess Generative

AI literacy in higher education, including their main dimensions and constructs?

- RQ2: What evidence exists regarding the development, application, and psychometric properties of these instruments in higher education contexts?

- RQ3: What implications do the assessed dimensions and reported uses of these instruments have for SDG 4, particularly in terms of quality education, digital competencies, responsible AI use, and inclusive learning?

Methodological Framework

This review will be guided by the PICO framework (Population, Phenomenon of Interest, Context), which is particularly suitable for systematic reviews of educational phenomena and assessment instruments:

- Population (P): Undergraduate and postgraduate students enrolled in higher education institutions. Studies with mixed samples, such as students and faculty members, will be considered only if student data can be clearly identified and analyzed separately.

- Phenomenon of Interest (I): Instruments, scales, questionnaires, survey-based instruments, rubrics, frameworks, and other measurement approaches used, developed, or adapted to assess Generative AI literacy.
- Context (Co): Higher education settings, with emphasis on sustainable education, digital transformation, responsible AI use, and potential implications for SDG 4 (Quality Education).

Rationale The rapid integration of generative artificial intelligence (GenAI) into higher education is transforming teaching, learning, and assessment practices, creating a growing need for educators and students to develop appropriate competencies (Hershkovitz et al., 2025; Ayyoub et al., 2025). GenAI literacy extends beyond basic technical familiarity to encompass critical evaluation of AI-generated content, ethical awareness, responsible use, and the ability to integrate these tools meaningfully into academic work (Almatrafi et al., 2024; Hershkovitz et al., 2025).

Despite growing interest, reliable instruments specifically designed to assess GenAI literacy in higher education remain limited and fragmented. While a few promising tools have recently emerged—including performance-based tests and validated scales focused on learning with GenAI—the evidence regarding their dimensions, development processes, application contexts, and psychometric properties is scattered across individual studies and lacks systematic synthesis (Jin et al., 2025; Gümüş & Kara, 2025). This gap is particularly relevant because broader reviews of AI literacy scales have highlighted inconsistent evidence of validity, reliability, and other measurement qualities, with a heavy reliance on self-report measures (Lintner, 2024).

At the same time, AI literacy is increasingly being discussed in relation to sustainable development in higher education, as students require both conceptual understanding and critical capabilities to harness AI for ethical and sustainable solutions (Alamäki et al., 2024). Emerging research also links AI literacy to key educational factors such as student engagement, self-efficacy, anxiety reduction, and inclusive learning experiences, which hold potential implications for quality education and SDG 4 (Farmanesh et al., 2025). However, no systematic review to date has focused specifically on measurement approaches for Generative AI literacy in higher education contexts.

This systematic review therefore aims to address this specific gap by identifying, characterizing, and

critically appraising existing instruments and measurement approaches, while exploring their potential implications for digital competencies, responsible AI use, and equitable learning opportunities in higher education.

Condition being studied The central phenomenon studied in this systematic review is Generative AI (GenAI) literacy in higher education. It is understood as a multidimensional educational construct comprising the knowledge, skills, attitudes, and critical dispositions that enable higher education students to understand, critically evaluate, responsibly use, and meaningfully integrate generative artificial intelligence tools into academic practices.

This phenomenon is examined primarily through its assessment and measurement. It manifests in the design, validation, adaptation, and application of diverse instruments such as scales, questionnaires, performance-based tests, frameworks, and rubrics. These instruments differ in their theoretical foundations, target populations, scope, and level of psychometric evidence. The phenomenon appears in the context of the rapid adoption of GenAI tools—such as ChatGPT, Gemini, and Claude—across academic writing, assessment, research, personalized learning, and institutional digital transformation.

The assessment of GenAI literacy is particularly relevant because higher education institutions require reliable and context-sensitive ways to identify students' competency gaps, inform curriculum development, support academic integrity policies, and promote equitable access to AI-enhanced learning environments. In this sense, the phenomenon is closely connected to sustainable education, digital competencies, responsible AI use, and inclusive practices, with potential implications for quality education and SDG 4.

METHODS

Search strategy A comprehensive and systematic search strategy will be implemented following PRISMA 2020 and PRISMA-P recommendations. Preliminary exploratory searches using SciSpace informed the identification of key terminology and indexing patterns. The formal search will be conducted in Scopus, ERIC, ScienceDirect, and IEEE Xplore. Search equations will be adapted to the syntax and operational requirements of each database.

Search Parameters

- Publication date: From 2022 to the date of the formal search, planned for July–August 2026. This

timeframe reflects the widespread educational adoption of Generative AI following the public release of ChatGPT in late 2022.

- Language: English.
- Document type: Peer-reviewed scientific journal articles.
- Database filters for publication year, language, and document type will be applied directly within each platform

Search Equations:

Scopus

TITLE-ABS-KEY (("Generative AI literacy" OR "GenAI literacy" OR "AI literacy") AND (assessment OR instrument OR scale OR questionnaire OR measurement OR validation OR psychometric) AND ("higher education" OR universit* OR faculty OR student*))

ERIC

("Generative AI literacy" OR "GenAI literacy" OR "AI literacy") AND (assessment OR instrument OR scale OR questionnaire OR measurement) AND ("higher education" OR university OR faculty OR student*)

ScienceDirect

("Generative AI literacy" OR "GenAI literacy") AND assessment AND university

IEEE Xplore

("Generative AI literacy" OR "GenAI literacy" OR "AI literacy") AND (assessment OR instrument OR scale OR questionnaire OR validation) AND ("higher education" OR university OR faculty)

The search equations will be adapted according to the specific syntax and advanced search options of each database. Truncation (*) and phrase searching (" ") will be used where appropriate.

Complementary Search Strategies:

- Backward and forward citation tracking of included studies.
- Manual screening of reference lists from key articles.

Additional studies identified through complementary strategies will be screened using the same eligibility criteria. Studies focused primarily on prompt engineering, K-12 education, technical AI system development, or general AI adoption without a clear focus on literacy assessment instruments will be excluded during screening.

Formal database searches are planned for July–August 2026, with screening and initial data

extraction expected to be completed by August 20, 2026.

Participant or population The population of interest in this systematic review consists primarily of students enrolled in higher education institutions, including undergraduate and postgraduate students. Eligible studies are those that assess, measure, develop, validate, or apply instruments designed to evaluate Generative AI literacy within university, college, or other tertiary education contexts.

Studies involving students who use or are expected to use Generative AI tools in learning, academic writing, assessment, research, information search, or related academic activities will be included. Studies with mixed higher education populations, such as students together with faculty members or academic staff, may also be considered if student-related data can be clearly identified, separated, or analyzed independently.

Studies focused exclusively on primary, secondary, or K-12 education, corporate training, professional development programs, or non-formal educational settings unrelated to higher education will fall outside the scope of this review. Similarly, studies involving only faculty members, administrative staff, developers, technical professionals, or general adult populations without identifiable higher education student data will not be included.

Intervention The main phenomenon of interest in this systematic review is the assessment of Generative AI (GenAI) literacy in higher education through dedicated measurement instruments and evaluation approaches. The review focuses on tools that have been developed, adapted, or applied to evaluate how higher education students understand, critically evaluate, responsibly use, and meaningfully integrate GenAI technologies in academic contexts.

This phenomenon encompasses a variety of assessment approaches, including psychometric scales, questionnaires, performance-based tests, rubrics, frameworks, and other evaluation measures. These approaches commonly target key dimensions of GenAI literacy, such as conceptual understanding of generative technologies, critical evaluation of AI-generated content, ethical and responsible AI use, academic application of GenAI tools, and awareness of limitations and risks.

Particular attention will be given to the development procedures, application contexts, and reported psychometric properties of these instruments, including evidence of validity and reliability, as well as factor structure. The review will also consider their implementation in learning,

academic writing, assessment, research, and technology-enhanced higher education environments.

Comparator No specific comparator is required for this systematic review. As the review focuses on the identification, characterization, and critical appraisal of measurement instruments and approaches for Generative AI literacy in higher education — rather than on evaluating the effectiveness of interventions — no comparison group or control condition is applicable.

Study designs to be included This systematic review will include empirical studies that report the development, adaptation, application, validation, or psychometric evaluation of instruments designed to assess Generative AI literacy in higher education. Eligible designs include quantitative studies, such as cross-sectional surveys, instrument validation studies, exploratory factor analysis, and confirmatory factor analysis; qualitative studies; mixed-methods research; performance-based assessments; case studies of instrument implementation; and educational evaluation studies.

Eligibility criteria Inclusion criteria

- Publication characteristics: Peer-reviewed scientific journal articles in English, published from 2022 to the date of the formal search.
- Population and context: Empirical studies conducted in higher education or university contexts involving undergraduate or postgraduate students. Mixed samples, such as students and faculty members, will be included only if student data can be clearly identified and analyzed separately.
- Phenomenon of interest: Studies focused on the development, adaptation, application, or validation of instruments, scales, questionnaires, rubrics, frameworks, or other measurement approaches designed to assess Generative AI literacy.
- Study design: Quantitative, qualitative, or mixed-methods empirical studies that report instrument development, psychometric evaluation, or application in higher education contexts.
- Outcomes and measurement evidence: Studies that provide information on instrument dimensions, application contexts, measurement procedures, or psychometric properties, such as validity, reliability, or factor structure.

Exclusion criteria

- Publication characteristics: Non-peer-reviewed documents, conference abstracts or papers, book chapters, theses, dissertations, editorials, opinion

pieces, letters, literature reviews, or purely conceptual papers without empirical data.

- Population and context: Studies conducted exclusively in primary, secondary, K-12, corporate training, professional development, or non-formal education settings, or those without identifiable higher education student data.
- Phenomenon of interest: Studies primarily focused on prompt engineering, technical AI development, software architecture, general AI adoption, attitudes, or perceptions without a clear emphasis on Generative AI literacy assessment instruments.
- Study design: Studies lacking clear empirical methodology or insufficient information on instrument development, application, or validation.
- Outcomes and measurement evidence: Studies that do not report relevant data on instrument dimensions, measurement procedures, or psychometric properties.

Information sources The primary information sources for this systematic review will be four electronic databases:

- Scopus: chosen for its broad multidisciplinary coverage and strong indexing in educational technology, artificial intelligence, psychometrics, and higher education.
- ERIC: included for its specialized focus on education research, higher education, educational assessment, and digital competencies.
- ScienceDirect: selected to capture relevant Elsevier publications in education, artificial intelligence, and learning technologies.
- IEEE Xplore: included due to its strength in AI technologies, digital competencies, educational innovation, and assessment instruments.

These information sources will be complemented by backward and forward citation tracking of included studies and manual screening of reference lists from key articles. All additional records identified through complementary strategies will be screened using the same eligibility criteria.

Main outcome(s) The main outcomes of this systematic review will be extracted and synthesized in direct alignment with the three research questions, focusing on the identification, characterization, and critical appraisal of instruments for assessing Generative AI literacy in higher education.

Outcomes related to RQ1

- Instruments, scales, questionnaires, rubrics, frameworks, performance-based tests, or other measurement approaches identified.
- Main dimensions and constructs assessed by each instrument.

- Theoretical or conceptual foundations reported for the instruments.

Outcomes related to RQ2

- Instrument development, adaptation, validation, and application procedures.
- Reported psychometric properties, including evidence of validity, reliability, factor structure, and other measurement qualities.
- Higher education contexts and student populations in which the instruments were applied.

Outcomes related to RQ3

- Reported implications of the assessed dimensions and instrument uses for digital competencies, responsible AI use, academic integrity, inclusive learning, and quality education.
- Interpretive links between the instruments and SDG 4, only when these links are supported by the extracted evidence.

These outcomes will be extracted using a standardized data extraction form and synthesized through narrative and tabular methods, acknowledging variability in reporting across studies.

Additional outcome(s) In addition to the primary outcomes aligned with the research questions, the following secondary information will be extracted to provide context and support descriptive analysis:

- Study characteristics: author(s), year of publication, country or region, journal, and academic discipline.
- Methodological features: study design, sample size, and participant characteristics.
- Implementation details: higher education setting, administration format of the instrument, and type of GenAI tools referenced.
- Other relevant data: reported barriers, facilitators, challenges, limitations, ethical considerations, or educational implications associated with the instruments.

These additional outcomes will support the identification of patterns and the overall interpretation of findings across studies, without extending the synthesis beyond the evidence reported in the included studies.

Data management Records retrieved from Scopus, ERIC, ScienceDirect, and IEEE Xplore will be exported and initially organized in Zotero to consolidate references and support duplicate identification. Duplicates will be detected automatically in Zotero and then verified manually.

After deduplication, the records will be imported into Rayyan, which will serve as the primary platform for managing the study selection process. Two reviewers will independently screen titles and

abstracts according to the eligibility criteria, followed by full-text assessment of potentially relevant studies. Reasons for exclusion at the full-text stage will be documented. Any disagreements during screening or data extraction will be resolved through discussion or, if needed, consultation with a third reviewer.

Data extraction will be conducted using a standardized matrix developed in Microsoft Excel. This matrix will capture study characteristics, methodological features, instrument details, psychometric properties, application contexts, and outcomes aligned with the research questions.

All records, screening decisions, extraction matrices, full-text articles, and supporting materials will be stored securely in password-protected cloud storage accessible only to the research team, with regular backups maintained to minimize the risk of data loss.

Quality assessment / Risk of bias analysis The methodological quality of included studies will be evaluated using an adapted version of the Critical Appraisal Skills Programme (CASP) framework, tailored to research on the development, adaptation, validation, or application of instruments for assessing Generative AI literacy in higher education. This tool supports a structured yet flexible assessment of the clarity, rigor, and trustworthiness of each

study. Two reviewers will independently appraise each study. Each of the ten domains will be scored from 1 to 5, where 1 indicates very limited or unclear reporting and 5 indicates high clarity and completeness. Disagreements exceeding one point in any domain will be discussed until consensus is reached, with a third reviewer consulted if needed. To assess inter-reviewer agreement, Cohen's Kappa coefficient will be calculated on a random sample of approximately 20% of the included

studies. An overall quality judgment will be assigned based on the total score, with a maximum of 50

points: • Low risk of bias / strong quality: 36–50 points. • Moderate risk of bias / acceptable quality with limitations: 30–35 points. • High risk of bias / limited quality: Below 30

points. Studies with high risk of bias will not be excluded from the review. However, their findings will be interpreted with caution and given less weight during synthesis. The ten domains evaluated are:

- C1. Clarity in the description of the instrument and its intended purpose.
- C2. Transparency in the development, adaptation, or validation process of the instrument.
- C3. Adequacy of the higher education context and participant characteristics.
- C4. Clarity in the theoretical or conceptual foundations underlying the instrument.
- C5. Methodological coherence between study objectives, design, and analytical procedures.
- C6. Quality and completeness of reported psychometric properties.
- C7. Appropriateness of sample size and participant selection procedures.
- C8. Clarity and completeness in the presentation of results and findings.
- C9. Acknowledgment of methodological limitations and potential sources of bias.
- C10. Relevance of the study to the review questions and objectives.

This quality assessment will help interpret the strength of the evidence, give greater weight to higher-quality studies during synthesis, and allow transparent discussion of how methodological limitations may affect confidence in the findings.

Strategy of data synthesis Given the expected heterogeneity in study designs, instrument characteristics, psychometric approaches, and higher education contexts, the findings of this systematic review will be synthesized primarily through a narrative approach. This method will enable an integrated interpretation of the evidence while respecting the methodological and contextual diversity of the included studies.

A descriptive analysis will also be performed to map the characteristics of the evidence base. Frequencies and percentages will be used, where appropriate, to summarize the distribution of instrument types, assessed dimensions, target populations, academic disciplines, methodological features, psychometric evidence, and application contexts.

The synthesis will be organized around the three research questions. For RQ1, the review will summarize the instruments identified, their main dimensions and constructs, and the theoretical foundations reported. For RQ2, attention will focus on instrument development, adaptation, validation, and application procedures, together with the reported psychometric evidence. For RQ3, the synthesis will examine reported implications and evidence-supported links between the instruments and digital competencies, responsible AI use, academic integrity, inclusive learning, quality education, and SDG 4.

Findings will be presented through narrative summaries, descriptive categorizations, and summary tables, where appropriate, to facilitate comparison across studies and support interpretation. A meta-analysis will not be conducted due to the expected methodological, conceptual, and reporting heterogeneity across studies, which would make statistical pooling inappropriate and unreliable.

Subgroup analysis Where feasible and supported by sufficient data, subgroup analyses will be performed on an exploratory basis to examine whether patterns in the characteristics and findings of the included studies vary across relevant factors. Given the anticipated heterogeneity in instruments, populations, and contexts, these analyses will remain descriptive in nature and interpreted with caution.

Potential subgroups may include geographical region, academic discipline, educational level, type of instrument, type of Generative AI tool referenced, study design, target population, and higher education context. When data permit, studies may also be grouped according to whether they address aspects related to digital competencies, responsible AI use, academic integrity, inclusion, quality education, or sustainability.

These analyses will not aim to establish causal relationships. Instead, they will help identify emerging patterns and support a more nuanced interpretation of how different contexts and approaches may relate to the assessment of Generative AI literacy. If the number of studies within any subgroup is limited, the results will be presented as exploratory and interpreted conservatively.

Sensitivity analysis Given the narrative nature of the synthesis, sensitivity analysis will be conducted in a descriptive and interpretive manner. The aim is to explore whether the main patterns and interpretations remain consistent when studies with higher risk of bias or lower methodological quality are considered separately from those with stronger quality.

Where the data allow, the synthesis will examine whether key findings appear to depend heavily on particular types of instruments, study designs, or educational contexts. No formal statistical sensitivity analysis will be performed, as no meta-analysis is planned. Instead, these considerations will support a cautious and transparent interpretation of the evidence, taking into account the expected diversity across studies.

Language restriction Only studies published in English will be included. English continues to be the predominant language in the scientific literature on Generative AI, educational technology, psychometric validation, and higher education research.

Country(ies) involved This review is being conducted by researchers based in Peru, within the context of higher education and educational technology. The research team is affiliated with Peruvian higher education institutions.

Other relevant information This review focuses specifically on instruments and measurement approaches used to assess Generative AI literacy in higher education. It does not address general AI attitudes, prompt engineering, technical AI system development, or educational contexts outside higher education unless they provide clearly relevant evidence on GenAI literacy assessment instruments. Due to the emerging and rapidly evolving nature of this field, the review is expected to be largely exploratory. Variability in instrument design, psychometric reporting, terminology, and educational contexts is anticipated. No meta-analysis will be conducted because of this heterogeneity. Any amendments made to the protocol during the review process will be documented and reported transparently. The review is planned for completion during 2026.

Keywords Generative AI literacy; higher education; assessment instruments; psychometric properties; digital competencies; responsible AI use; systematic review; SDG 4.

Dissemination plans The results of this systematic review are primarily intended for publication in the journal *Sustainability* (MDPI), specifically within the Special Issue “Generative AI Literacy in Education and Innovation: Fostering Sustainable Futures”, given its strong alignment with the review’s focus on assessment instruments, psychometric properties, and implications for sustainable higher education. The review will be disseminated through open-access publication to maximize its reach among researchers, educators, policymakers, and higher education institutions. In addition, the findings may be presented at relevant academic conferences and seminars in the fields of educational technology, higher education, artificial intelligence, and digital competencies. Selected results may also be shared through institutional reports or other academic dissemination activities within participating universities.

Contributions of each author

Author 1 - Willy Adauto-Medina - Will contribute to Conceptualization, Methodology, Investigation, Writing – Original Draft, and Project Administration. Email: wadauto@untels.edu.pe

Author 2 - Cesar León-Velarde - Will contribute to Methodology, Validation, Investigation, and Writing – Review & Editing. Email: cleon@unfv.edu.pe

Author 3 - Silvia Fernández-Flores - Will contribute to Investigation, Data Curation, Validation, and Writing – Review & Editing. Email: c29752@utp.edu.pe

Author 4 - José Antonio Arévalo-Tuesta - Will contribute to Investigation, Formal Analysis, Methodology, and Writing – Review & Editing. Email: jarevalotu@unfv.edu.pe

Author 5 - Irma Aybar-Bellido - Will contribute to Investigation, Data Curation, Validation, and Visualization. Email: irma.aybar@unica.edu.pe

Author 6 - Maritza Arones - Will contribute to Investigation, Formal Analysis, Writing – Review & Editing, and Visualization. Email: marones@unica.edu.pe

Author 7 - Beatriz Caycho-Salas - Will contribute to Investigation, Supervision, Validation, and Writing – Review & Editing. Email: bcaycho@une.edu.pe

Author 8 - Adrián Quispe-Andía - Will contribute to Investigation, Supervision, Methodology, Writing – Review & Editing, and Formal Analysis. Email: aquispe@une.edu.pe

References

- Alamáki, A., Nyberg, C., Kimberley, A., & Salonen, A. O. (2024). Artificial intelligence literacy in sustainable development: A learning experiment in higher education. *Frontiers in Education*, 9, 1343406. <https://doi.org/10.3389/educ.2024.1343406>
- Almatrafi, O., Johri, A., & Lee, H. (2024). A systematic review of AI literacy conceptualization, constructs, and implementation and assessment efforts (2019–2023). *Computers and Education Open*, 6, 100173. <https://doi.org/10.1016/j.caeo.2024.100173>
- Ayyoub, A. M., Khlaif, Z. N., Shamali, M., Abu Eideh, B., Assali, A., Hattab, M. K., Barham, K. A., & Bsharat, T. R. K. (2025). Advancing higher education with GenAI: Factors influencing educator AI literacy. *Frontiers in Education*, 10, 1530721. <https://doi.org/10.3389/educ.2025.1530721>
- Farmanesh, P., Vehbi, A., & Solati D.N. (2025). AI literacy in achieving sustainable development goals: The interplay of student engagement and anxiety reduction in Northern Cyprus universities.

Sustainability, 17(11), 4763. <https://doi.org/10.3390/su17114763>

Gümüş, M. M., & Kara, M. (2025). Development and validation of the Generative AI Literacy for Learning Scale (GenAI-LLs). *Australasian Journal of Educational Technology*, 41(4). <https://doi.org/10.14742/ajet.10236>

Hershkovitz, A., Tabach, M., Reich, Y., Lurie, L., & Cholcman, T. (2025). Framing and evaluating task-centered generative artificial intelligence literacy for higher education students. *Systems*, 13(7), 518. <https://doi.org/10.3390/systems13070518>

Jin, Y., Martinez-Maldonado, R., Gašević, D., & Yan, L. (2025). GLAT: The generative AI literacy assessment test. *Computers and Education: Artificial Intelligence*, 9, 100436. <https://doi.org/10.1016/j.caeai.2025.100436>

Lintner, T. (2024). A systematic review of AI literacy scales. *npj Science of Learning*, 9, 50. <https://doi.org/10.1038/s41539-024-00264-4>

O'Dea, X., Ng, D. T. K., O'Dea, M., & Shkuratsky, V. (2024). Factors affecting university students' generative AI literacy: Evidence and evaluation in the UK and Hong Kong contexts. *Policy Futures in Education*. <https://doi.org/10.1177/14782103241287401>