Article

Journal of Interdisciplinary Studies in Education

Volume 15, Issue 4 (2026), pp. 45-82
Journal of Interdisciplinary Studies in Education
ISSN: 2166-2681Print 2690-0408 Online | https://ojed.org/jise
https://doi.org/10.32674/nrmwyw65

AI-Assisted Learning and the Illusion of Competence: Measuring the AI-Learning Gap Through Random Knowledge Verification

Tran Minh Duc
Faculty of Cultural Industries, Thu Dau Mot University, Vietnam
ORCID: 0009-0003-9359-3000

ABSTRACT

This study examines the AI–Learning Gap (ALG), defined as the difference between the quality of AI-assisted academic work and students' independently demonstrated knowledge mastery. Using data from 1,498 undergraduate students across 38 classes at a Vietnamese public university, the study developed a random knowledge verification (RKV) approach combining oral questioning and written recall tasks under device-restricted conditions. The results showed a substantial gap between assignment quality (M = 7.62) and knowledge mastery (M = 5.55), producing an average ALG of 2.07. The findings suggest that AI-assisted tasks may improve the quality of academic outputs without necessarily strengthening conceptual understanding. The study proposes RKV as a practical assessment approach for evaluating learning in AI-integrated higher education.

Keywords: AI–Learning Gap, assessment, generative artificial intelligence, higher education, illusion of competence, knowledge mastery

© Author(s), 2026. Published by Star Scholars Press.
This article is distributed under the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. https://creativecommons.org/licenses/by/4.0/

INTRODUCTION

The integration of generative artificial intelligence (GenAI), particularly large language model (LLM) applications such as ChatGPT, impacts the way academic practices are conducted in higher education. Instead of augmenting information retrieval processes such as earlier iterations of digital tools, GenAI systems can retrieve and summarize, outline and explain, and write entire academic papers, all based on prompt input from the user. Consequently, AI-facilitated activities are now commonplace in academic exercises such as brainstorming, drafting, and revising (Cotton et al., 2023; Farrokhnia et al., 2024; Michel-Villarreal et al., 2023). Recent bibliometric evidence further demonstrates the rapid global expansion of research on generative AI in higher education, particularly in areas related to assessment, academic integrity, student learning, and educational governance (Dai et al., 2026). Alongside the promise of augmenting students' writing abilities and making learning processes more efficient, legitimate concerns exist regarding the degree to which AI-assisted assignments demonstrate students' true mastery of course material (Kasneci et al., 2023; Selwyn, 2024). Generative AI can help with structuring ideas, writing more coherently, and completing scholarly writing assignments (Adeshola & Adepoju, 2024; Baek et al., 2024; Mollick, 2024). However, the degree of educational value is determined by the manner in which students engage with AI. AI can function as a cognitive tool for scaffolding learning when students engage with it by evaluating and revising its outputs and incorporating its ideas into their own work.

On the other hand, if AI outputs are incorporated into assignments without any direct engagement with the material, the opposite may happen, and learning may be more impaired than enhanced (Holmes & Miao, 2023; Levine et al., 2024). In these instances, students may produce generic academic writing that meets superficial structural and cohesion requirements but fails to demonstrate an adequate grasp of the underlying concepts. This presents a serious challenge for assessment processes in higher education. For example, most higher education assignments are geared toward the submission of essays, briefs, or presentations and are treated as evidence of learning and mastery of concepts. However, with the presence of generative AI, it may be the case that educational outputs are of a higher level than the actual educational attainment of the student. This has been noted in educational psychology, where the student may experience an ‘illusion of competence’ when they avoid the learning effort required to actually understand the material and instead demonstrate only academic presentation (Bjork et al., 2013). In the case of learning with AI, this illusion may be even greater because, as Albadarin et al. (2024) and Darvishi et al. (2024) demonstrate, AI-generated papers are often the most fluent, cohesive, and academically presented. This is the challenge that universities now face: the need to differentiate between fabricated academic work and academic understanding demonstrated by students and between work completed with and without students' actual comprehension.

Most studies focus on student perceptions and sentiments rather than on academic measures of actual learning (Kuhail et al., 2023; Labadze et al., 2023; Wang et al., 2024). Research pertaining to educational tools that utilize AI regularly employs surveys based on models of technology acceptance. Although such surveys have merit, they cannot capture whether students can articulately discuss the overall purpose and content of AI-enhanced assignments. Research that employs log and prompt analysis also tends to avoid the question of whether students are able to independently articulate the AI-enhanced work they submit. This illustrates the disparity between AI-assisted academic work and the work students can produce on their own.

This study addresses interdisciplinary scholarship in higher education by integrating educational technology, learning sciences, writing studies, and assessment validity while providing evidence from a Vietnamese public university as a cross-cultural case study of a global problem in AI-integrated learning environments. Recent scholarship has stressed that higher education institutions need to redesign assessment and shift from forms of technological surveillance to practices that document authentic learning and conceptual understanding (Chan, 2023). Recent studies have also highlighted the importance of generative AI competence, learner autonomy, and patterns of AI utilization in shaping student engagement and learning experiences in higher education (Sampah et al., 2026).

This study aims to define that disparity by using the term AI–Learning Gap to denote the gap between the level of work students submit and the level of understanding they can demonstrate. Empirically, the study is based on classroom data collected over two academic years at a large public university in Vietnam involving 1,498 undergraduate students across 38 classes in the social sciences, humanities, and communication-related fields.

To assess knowledge in an environment where AI tools are unavailable, the study uses a classroom-based assessment technique called random knowledge verification (RKV). This technique integrates two verification methods under device-restricted conditions: random oral questioning before the presentation and a brief handwritten recall task in class. These methods facilitate analysis of the perceived quality of coursework and the student's ability to reconstruct and articulate the conceptual framework of the coursework without external assistance. This study seeks to understand whether AI-supported tasks create a definable AI–Learning Gap between the quality of the task and the extent of knowledge demonstrated by students in higher education. The remainder of the article is structured as follows. The next section summarizes the literature regarding the use of generative AI in higher education, AI-assisted academic work, and learners' illusion of competence. The third section presents the conceptual framework. The fourth section presents the study's research design, data, measurement methods, and analytical approach. The fifth section presents the empirical findings concerning the quality of assignments, knowledge mastery, and the size of the AI–Learning Gap. The final sections discuss the findings, the study's limitations, directions for future research, and the conclusion of the article.

LITERATURE REVIEW

Cognitive offloading is an example of a related theory in educational psychology. Cognitive offloading occurs when a person uses various tools or means to help complete a task, or a number of tasks, that involve the mental processes of remembering, organizing, or evaluating a large amount of information. When students are able to use cognitive offloading to free mental space for more complex reasoning, cognitive offloading is a positive thing in terms of supporting learning. On the other hand, a person who relies excessively on means of cognitive offloading gives up opportunities to engage in effortful cognitive processes, which are essential to complex conceptual understanding and enduring learning. This dynamic could be heightened with the use of generative AI since it can help with not only information retrieval but also generating ideas, organizing arguments, and writing complete, coherent texts (Molenaar, 2022; Michel-Villarreal et al., 2023). In these cases, learners may skip important cognitive processes that are traditionally involved in deep learning, such as synthesis, explanation, and revision. The same problems are discussed in relation to scaffolding and learning. Scaffolding in education can support learning activities as long as students are engaging in tasks that are just slightly more complex than those they are able to do on their own. In contrast, scaffolding can be counterproductive when it replaces, rather than supports, the amount of cognitive labor that the learner must exert (Belland, 2013; Van de Pol et al., 2010; Wood et al., 1976). In the same way, generative AI can be understood as a positive factor that supports learning when it promotes cognitive activities such as elaboration, reflection, and evaluation. When it substitutes for these cognitive processes, AI may diminish active cognitive engagement and further impede the growth of critical thinking and abstract comprehension.

In the context of AI-enabled education, the differences between surface learning and deep learning become highly relevant. Surface learning is typically considered a form of cognition focused on completing tasks, reproducing knowledge, and formulating responses deemed academically appropriate. On the other hand, deep learning involves understanding a concept, explaining it, and being able to synthesize, reflect on, and defend the construction of knowledge. In writing-intensive courses, students may be able to produce assignments that are fluent and structurally sophisticated but lack an understanding of the associated concepts. Deep learning involves cognitive effort, whereas generative AI may diminish this cognitive effort. More recently, scholars have suggested that AI responses that are fluent and well structured may lead to cognitive offloading and cognitive overreliance to the extent that students are unable to appraise or even verify AI responses, therefore accepting them as truth (Kasneci et al., 2023; Selwyn, 2024; Wang et al., 2024). In some cases, students may be able to perform simple editing or simple reproduction of AI outputs that are fluent and plausible and, as a result, presume that they have understood the underlying concepts. This phenomenon is referred to as surface-level performance and is reflective of cognitive offloading and a lack of cognitive engagement. In this article, 'illusion of competence' is used to mean that students believe they have a deep understanding of concepts when they are actually unable to explain, reconstruct, or justify those concepts independently. Because the outputs are generated by AI, students take a passive role, cognitive effort is diminished, and deep learning may be weakened.

These concerns have generated renewed interest in an area of research concerning the role of oral assessment, viva voce examinations, oral defenses, and explanation-based assessment in higher education (Kehm, 2001; Joughin, 2010). In oral assessment, unlike traditional product-based written assignments, educators are able to evaluate reasoning processes, reconstructive and creative abilities, authorship, and students' abilities to explain and defend their understanding of the question or problem at hand. The traditions of oral assessment at the tertiary level of education have emphasized students' abilities to explain their reasoning, defend their interpretation, and illustrate their understanding of concepts beyond written presentation (Joughin, 2010; Kehm, 2001). These methods complement assessment scholarship focusing on reasoning processes and evaluative judgment, in which students' reasoning processes must be made visible and polished academic work alone must not be relied upon in assessment (Tai et al., 2018). The concerns discussed are also connected to the broader educational psychology scholarship on the illusion of competence. Research in educational psychology indicates that learners' understanding of material may be overestimated when learning activities emphasize fluency and ease of access rather than retrieval and explanation activities (Bjork et al., 2013). With generative AI, this phenomenon may be more pronounced because AI responses are generally fluent, organized, and academically persuasive. Reading, editing, and submitting a text that is academically sophisticated may give students the illusion of mastery when the fluency of expression is a poor indicator of conceptual understanding. For example, learners may be able to identify the language and pattern of an explanation but be unable to reproduce the argument independently of the text or articulate the reasoning behind the ideas presented. Recent research on generative AI in education emphasizes that AI-assisted tasks may uphold traditional academic standards but conceal a fundamental lack of understanding (Albadarin et al., 2024; Cotton et al., 2023; Farrokhnia et al., 2024). This is a particular concern in communication, cultural studies, and the social sciences, where the desired learning outcomes focus on explanation, interpretation, and argumentation. When academic assessment places greater weight on well-developed written work, educational institutions are likely to value linguistic proficiency over cognitive understanding.

The increasing integration of generative AI into society brings fundamental challenges to assessment methods in higher education. Recent evidence from higher education instructors indicates growing concerns regarding the educational meaning of ChatGPT-generated assignments and whether such outputs can continue to serve as valid indicators of student learning and academic achievement (Dhamija & Dhamija, 2025). In the past, written assignments such as essays, reports, and take-home exams have been the main way educators have gauged a student's comprehension and level of critical thinking. This becomes very problematic as AI technologies are able to quickly produce high-quality written assignments. Recent investigations continue to point toward generative AI technologies driving a more profound shift in the use of written discourse in learning and teaching while also raising concerns related to academic honesty, learning authenticity, and the relevance of construct-based assessment strategies (Baig & Yadegaridehkordi, 2024; Bobula, 2024; Dempere et al., 2023; Mai et al., 2024; Montenegro-Rueda et al., 2023; Munaye et al., 2025; Tillmanns et al., 2025). Many researchers, therefore, posit that learning assessment methods in higher education need to be reconceptualized in the presence of AI.

More recent studies indicate that assessment frameworks can benefit from incorporating students' abilities to describe and reconstruct knowledge, as opposed to assessing ability based only on the quality of submitted artifacts (Chan, 2023; Holmes & Miao, 2023; Association for Writing Across the Curriculum, 2025; MLA-CCCC Joint Task Force on Writing and AI, 2024). Oral questioning, in-class writing, reflective explanation, and process-oriented assessment give faculty the opportunity to evaluate students' abilities to demonstrate conceptual understanding on their own. These methods do not exclude AI-supported teaching but place AI outputs in the context of a richer learning process, where understanding and reasoning beyond AI must be demonstrated. Recent systematic review evidence further emphasizes the growing importance of transparency, disclosure, and verification mechanisms in AI-supported assessment environments, as higher education institutions seek ways to distinguish authentic learning from AI-assisted academic production (Pérez-Pérez et al., 2026). Although the scholarship on generative AI in education is growing, the specific intersection between AI-assisted tool use in assignments and students' knowledge when demonstrated independently is an underresearched area. Most studies focus on the adoption, attitudes, and/or perceptions of AI and the perceived technological functionality of AI (Adeshola & Adepoju, 2024; Labadze et al., 2023). Most studies of interaction behavior also focus on the degree to which students use AI tools rather than the degree to which students synthesize and explain AI-generated ideas.

As such, the literature still does not adequately address the gap concerning the relationship between the perceived quality of academic tasks and the understanding students independently demonstrate. In the literature, studies focus on both the increasing importance of generative AI in learning, the risks of cognitive nonengagement, the concern over the illusion of competence, and the validity of assessing students based on produced work. However, studies on the relationship between AI-empowered assignments and students' independently demonstrated understanding are still very limited. Consequently, the purpose of this study, based on the AI learning literature, is to focus on the relationship between assignment performance and demonstrated knowledge mastery.

Given the current literature and the conceptual framework delineated above, the study explores the following empirical expectations and theoretical propositions:

H1: The quality of assignments submitted exceeds students' independently demonstrated knowledge mastery in an AI-assisted learning environment.

H2: There is a gap between assignment performance and the results of knowledge verification under device-restricted conditions.

P1: In AI-assisted learning environments, a larger AI–Learning Gap may indicate a stronger risk of illusion of competence, reflected in the discrepancy between apparent academic performance and independently demonstrated conceptual understanding.

Rather than identifying the illusion of competence through direct psychological constructs, the AI–Learning Gap is treated here as an observable element that signals the discrepancy between apparent academic performance and a learner's demonstrated understanding in the context of autonomous learning. Therefore, the study does not attempt to identify students' subjective estimations of competence but rather interprets the gaps between assignment output and learners' demonstrated understanding in AI-assisted environments as a possible indicator of the illusion of competence.

CONCEPTUAL FRAMEWORK

This study intends to offer a definable conceptual model in higher education to identify instances in which students present advanced academic work with the aid of AI but show substantially weaker conceptual understanding when their work is verified independently. Instead of framing the situation as one of authorship or academic dishonesty, the model aims to address the challenges of assessment caused by generative AI within higher education. This approach coincides with emerging scholarship arguing that the primary educational dilemma is not whether students utilize AI; rather, the dilemma is whether existing assessment practices remain valid and allow students to demonstrate learning in the presence of highly advanced generative AI capable of performing functions associated with academic writing and reasoning (Holmes & Miao, 2023; MLA-CCCC Joint Task Force on Writing and AI, 2024; Association for Writing Across the Curriculum, 2025).

The framework is based on three constructs that are related to each other: assignment quality (AQ), knowledge mastery (KM), and AI–learning gap (ALG). These constructs help the research explore the extent to which the quality of students' academic work matches their conceptual understanding. Assignment Quality refers to the evaluative quality of coursework products submitted by students, including written reports and presentation slides, and is used as an indicator of the evaluative quality of students' academic work.

Knowledge Mastery refers to students' ability to demonstrate comprehension of the same work without AI assistance. The AI–Learning Gap is the difference between these two dimensions, that is, the gap between product-level performance and demonstrated conceptual mastery. In this framework, assignment quality (AQ) is defined as observable product-level academic performance, while knowledge mastery (KM) is defined as independently demonstrated conceptual understanding when AI is restricted. Thus, the AI–Learning Gap (ALG) captures the gap between these two dimensions rather than students' psychological perceptions of learning or feelings of competence. The main point of the framework is that generative AI could enhance the general superficial quality of academic outputs, specifically organization, clarity, and rhetorical style, without improving students' understanding, reasoning, and conceptual grasp of the work submitted. This aligns with the observation that outputs from large language models may appear rational and authoritative to users, particularly those who do not apply critical evaluation. This remains the case even when such outputs contain weak reasoning, inaccurate information, or fabricated references (Albadarin et al., 2024; Farrokhnia et al., 2024; Michel-Villarreal et al., 2023). This also aligns with writing studies suggesting that when students do not engage in the necessary cognitive processes involved in articulating written arguments, polished writing may not indicate genuine learning (Graham, 2019; Langer & Applebee, 1987).

For this study, assignment quality (AQ) involves the academic evaluation of coursework artifacts that have been submitted, including written works and slide presentation decks. Artifacts created for coursework in the social sciences, humanities, and communication disciplines typically require constructing a conceptual framework, developing arguments, and incorporating relevant ideas. Submissions may demonstrate familiarity with the conventions of academic writing and even the operational logic of generative AI systems, without significant conceptual input from students. As such, this study defines AQ in terms of the extent to which a piece of coursework conveys rational thought, relevant evidence, clarity, and competence in academic articulation.

The main concern when assessing a student's writing is their reasoned argumentation, evidence use, and the strength of their claims, and not solely surface linguistic and writing fluency (Poe & Elliot, 2019; Hodges et al., 2019). In the case of AI-enabled learning, it becomes crucial to assess the accuracy and synthesis of provided sources, as AI-generated documents may contain incorrect or fabricated references (Adeshola & Adepoju, 2024; Williamson et al., 2012). Hence, for the purposes of this study, the assignment quality (AQ) of submissions measures the academic merit of the work and not the level of conceptual understanding itself. Improved assignment appearance due to the use of generative AI tools may result from the ability of these tools to assist students in presenting better-organized documents, clearer logical flows, and enhanced academic expression (Wang & Tian, 2025; Levine et al., 2024). Unfortunately, these assignment enhancements may contribute to a phenomenon referred to as product inflation, which occurs when academic coursework is submitted with higher visible quality than the level of understanding behind it.

Knowledge Mastery (KM) is the ability of students to articulate, restate, and explain the content of their submissions without the use of digital or AI tools. KM should not be confused with self-measured learning or self-assessed confidence, as it requires demonstrable understanding under specific assessment conditions.

An emphasis on learning science and writing-to-learn research highlights the consolidation of learning through retrieval, explanation, and restructuring, rather than through mere recognition of completed texts (Bjork et al., 2013; Nückles et al., 2020). KM therefore describes the extent to which students grasp the conceptual framework of the work they have submitted.

In this research, KM is operationalized through random knowledge verification (RKV), which includes two complementary device-free assessment formats. The first assessment type is oral verification and involves students explaining certain details of their work, recalling the conceptual framework, and justifying the main points in response to unexpected questions related to what they submitted. The second assessment type is a brief in-class device-free handwritten recall task, where students reconstruct the assignment structure and paraphrase the main points. Both procedures create a situation in which external devices, including cell phones, are not available and assess students' ability to explain, justify, and demonstrate understanding of the material. These procedures emphasize the need for assessment to focus on learning agency and conceptual understanding rather than authorial intent (Holmes & Miao, 2023; Association for Writing Across the Curriculum, 2025).

The central concept of this research is the AI–Learning Gap (ALG). This is understood as the difference between the quality of submitted coursework and the amount of knowledge students demonstrate when the task is performed under AI-restricted conditions. In other words, ALG represents the gap between the apparent level of academic performance and the level of conceptual understanding demonstrated independently.

The gap can formally be expressed as:

$$ALG = AQ - KM$$

Where: AQ represents assignment quality, and KM represents knowledge mastery.

In empirical analysis, KM can also be calculated using scores received in oral verifications, scores from written recall tasks, or the mean score from both measures. This operational flexibility allows researchers to examine the extent to which the observed gap persists across multiple measures of knowledge mastery. This study defines the AI–Learning Gap (ALG) as a measurable difference between the standard of academic work students submit and their mastery of a subject under verification conditions. The framework does not define ALG as a direct psychological assessment of the illusion of competence. A large AI–Learning Gap is perceived as a potential indicator of the illusion of competence when students produce advanced academic outputs but demonstrate weaker skills in explaining, reconstructing, or retrieving the work once AI is removed. The framework distinguishes between measurable performance and the psychological factors that are likely to accompany such performance. ALG therefore has a relatively straightforward interpretation. A small gap suggests that the quality of the assignment and the level of understanding demonstrated are in close proximity. This would indicate that AI may be acting as a learning scaffold or productivity tool that streamlines the work without compromising learning. The opposite is true for a large positive gap. In that case, there is a substantial difference between the quality of the submitted work and the student's explanatory or reconstructive performance without assistive AI. AI-enhanced coursework could showcase impressive academic performance even when the student has an insufficient grasp of the underlying concepts.

The gap can be explained through two mechanisms: product inflation and reduced effortful processing. Product inflation refers to situations in which generative AI systems produce essays that fulfill a large number of academic conventions (e.g., balanced and coherent essay structure, fluent and grammatically correct language, and an authoritative academic tone), which may result in assignments obtaining high evaluative scores even when students exhibit comparatively weaker conceptual reconstruction or explanatory skills under verification conditions (Levine et al., 2024; Wang et al., 2024). The mechanism of reduced effortful processing may also occur.

Bypassing the elaboration and self-explanation processes that foster durable learning occurs when students engage minimally with AI-generated text and skip processes of revision, evaluation, or integration of the text into their own reasoning. Instructional scaffolding theory is particularly relevant here, as it describes the type of support that is beneficial for learning: support that pushes learners beyond their current capabilities while still requiring cognitive effort (Belland, 2013; Van de Pol et al., 2010). Writing-to-learn research makes a similar argument, emphasizing the importance of actively organizing, testing, and refining one's own ideas rather than relying on finished texts that remove the learner from the meaning-making process (Graham & Harris, 2006; Graham et al., 2018).

The framework also does not assume that generative AI always has a negative impact on learning. Studies on the beneficial use of AI show that when students critically engage with and evaluate AI-generated responses, deeper learning engagement can occur (Long & Magerko, 2020; Ng et al., 2021). The AI–Learning Gap framework should not be read as a critique of generative AI itself. Rather, it identifies the extent to which AI-supported academic production may diverge from independently demonstrated understanding under assessment conditions requiring retrieval, explanation, synthesis, and conceptual reconstruction. A larger gap may indicate a greater risk of illusion of competence, but the framework does not claim to measure students' psychological experiences directly.

Table 1 presents the conceptual framework's key constructs and their operational indicators. The table details how each theoretical construct corresponds to observable and operationalizable indicators within the proposed framework. This assists future researchers in replicating or expanding the framework in comparable educational settings.

Table 1: Constructs and operational indicators

Construct	Definition	Primary indicators (0–10 scale)	Example evidence
Assignment Quality (AQ)	Quality of submitted coursework artifacts	Structure and organization; argumentation and conceptual reasoning; references; presentation quality	Submitted assignments and slide decks
Knowledge Mastery (KM)	Demonstrated understanding under AI-restricted conditions	RKV–Oral; RKV–Written; or composite of both	Oral responses; short handwritten recall tasks
AI–Learning Gap (ALG)	Discrepancy between assignment quality and knowledge mastery	$ALG = AQ - KM$	Gap index calculated per student

Note. The table illustrates the operational definitions and conceptual metrics used to assess the fundamental constructs of this research.

According to this conceptual framework, the next section delineates the study's empirical design and the sequential steps for collecting data on AQ, KM, and the resulting AI–Learning Gap in classroom contexts.

MATERIALS AND METHODS

Research context and design

This study utilized a classroom-based empirical methodology to investigate the possibility of an AI–Learning Gap (ALG). Specifically, the focus was on students using AI to complete assignments and later not demonstrating an adequate level of understanding of the material when AI was not available. The empirical methodology was designed to be as close as possible to natural classroom conditions. The study purposefully avoided artificially created lab-style empirical settings and instead assessed the natural course of an authentic higher education setting over multiple semesters. This design choice also reflects the primary focus of engaging with the emerging generative AI literature. The primary concern has not only been whether AI can generate text that meets academic credibility standards but also whether prevailing classroom assessment practices can still generate authentic evidence of learning when students have the ability to use AI to complete assignments (Holmes & Miao, 2023; Michel-Villarreal et al., 2023; Association for Writing Across the Curriculum, 2025).

This study was conducted over a two-year period at a public university in Vietnam. The population for this study consisted of 1,498 students distributed across 38 classes, with approximately 40 students per class. Students came from different faculties, including Tourism, Graphic Design, History Education, Multimedia Communication, Music, Public Administration, and International Studies. The range of these faculties is significant because students are now beginning to use generative AI in areas beyond traditional writing-intensive courses. It is evident that students are increasingly using large language models to explain and interpret concepts across different fields and to engage in synthesis and conceptual reasoning (Albadarin et al., 2024; Farrokhnia et al., 2024; Wang & Tian, 2025).

Assignment structure and AI-integrated learning context

The participants in the study were students already participating in customary interactions associated with an educator’s daily responsibilities. The study did not design artificial research assignments but instead built upon assessment tasks that were already being practiced in the participants’ classes. The assignments were derived from typical forms of assessment used across Vietnamese social sciences, humanities, communication, and applied social science disciplines. The students participated in individual and group assignments that involved topic selection, literature review, literature analysis, the development of structured academic writing, and the preparation of oral presentations. Each assignment required the completion of a written report and a set of presentation slides, which were presented in class followed by a question-and-answer session. The assignments were prescriptive, output-driven, and intended to be completed outside class. The term "prescriptive" was used in this study to refer to assignments that were structured and involved the provision of a specific topic, a description of the required sections, and an expected work submission, which could include a written report and presentation slides. Such assignments are pedagogical in nature because they guide students toward structured academic work. However, prescriptive assignments are particularly susceptible to "product inflation" in the context of academic integrity and AI-integrated learning environments. In such contexts, generative AI can easily assist students in developing outlines and section headings, summarizing literature, drafting thesis statements and structured paragraphs, and preparing presentation slides. This may result in assignments that appear academically comprehensive in both content and presentation. However, such assignments may be produced without students independently learning and internalizing the concepts and ideas underlying the work.

At the same time, the study does not presume that the use of AI is damaging to learning in all cases. The learning impacts of generative AI depend on how students use the technology. To the extent that students use AI to scaffold learning activities, such as brainstorming, clarifying ideas, comparing ideas, or improving drafts, it can support learning. On the other hand, if students use AI to outsource critical tasks such as planning, synthesis, drafting, or explanation, the outputs generated by AI may not represent the student's understanding (Mollick, 2024; Mollick & Mollick, 2023; Ng et al., 2021). Thus, the study examines the extent to which assignment quality is aligned with demonstrated understanding after digital assistance is removed.

Random Knowledge Verification procedure

The study's primary methodological approach was Random Knowledge Verification (RKV). Knowledge mastery in RKV was operationalized through two device-free verification formats: random oral questioning and a short, closed-device written recall task. This approach was developed independently of AI detection mechanisms, which are often criticized for their lack of reliability, fairness, and interpretive validity. Rather than focusing on the detection of AI use, the method evaluates whether students are able to describe and defend the reasoning processes underlying their work and whether they are able to do so without the assistance of technology (Liang et al., 2023; Walters, 2023; MLA-CCCC Joint Task Force on Writing and AI, 2024). Therefore, the research shifts the focus from AI detection to the demonstration of reasoning.

To enhance procedural uniformity across the 38 classes, RKV was standardized. In all participating classes, verification was conducted under device-free conditions. This meant that students had no access to mobile phones, laptops, tablets, computers, or AI tools during verification. The RKV procedure concentrated on the same key dimensions across classes: (a) students' recall of assignment structure, (b) their understanding of key concepts, theories, or arguments, and (c) their ability to explain and justify the relationships among the main sections, examples, evidence, and conclusions of the submitted work.

The first component, RKV–Oral, took place during class presentations. Randomly selected students or groups of students responded to assignment-related questions either before or during the presentation session. In a typical class, approximately ten students were selected for the oral task. The questions were not intended to elicit memorized factual answers but rather to determine whether students could justify how their assignment was designed and explain the reasoning behind the selection of particular concepts, theories, examples, or arguments. The questions were largely developed from a standardized question framework rather than from a fixed question list. This allowed the instructor to maintain a degree of standardization in assessment while still adapting the task to the specific assignment context. Examples of such questions included: "What is the main structure of your assignment?"; "Which theory or concept did your group use?"; "Why did you arrange the argument in this order?"; "Can you explain the relationship between this section and the next?"; and "What is the main evidence supporting your conclusion?"

The second component, RKV–Written, consisted of a brief handwritten recall task conducted in class under closed-device conditions. Within approximately 10 minutes, each student was required to reconstruct the structure of the assignment and identify the main claims, concepts, theories, and examples used in the major sections. In individual assignments, all students completed the written recall task. In group assignments, depending on class size and assignment type, either a representative group member or a specifically selected student completed the task.

The prompt followed a standardized format across classes. Students were asked to list the parts of the assignment they had submitted, summarize the main argument or idea presented in each section, and briefly explain the relationship between the different sections.

Research in learning science supports written recall activities because it provides evidence that retrieval activities are more effective than recognition activities in demonstrating knowledge internalization (Bjork et al., 2013; Nückles et al., 2020). In the context of AI-supported learning environments, device-free recall tasks are especially insightful because they can reveal differences between refined assignment outputs and students' independent understanding.

Measurement of Assignment Quality

All variables were recorded on a 0–10 scale, following the university's grading protocol. Maintaining consistency in the grading scale applied to assignments and assessments enables comparisons between the quality of assignment outputs and the level of knowledge demonstrated. Analytic rubrics were utilized in the evaluation of assignment quality (AQ) based on four criteria: structural clarity, reasoning, quality of evidentiary support, and presentation. These criteria, while addressing the challenges posed by generative AI, are focused on expectations within the social sciences and humanities. AI has the potential to produce fluent academic prose, but such text can lack conceptual clarity and may even include fabricated references unless it has been substantially revised (Adeshola & Adepoju, 2024; Albadarin et al., 2024). As a result, the rubric focuses on reasoning and evidential support rather than merely superficial fluency.

Measurement of Knowledge Mastery

Knowledge mastery (KM) was derived from the two RKV procedures. While oral verification measured a student's ability to articulate a conceptual explanation and defend an argument, written verification measured a student's ability to reconstruct structure and recall key ideas. For the purposes of analysis, KM was defined operationally as the oral score, the written score, or the mean of the two.

To enhance scoring transparency, common scoring descriptors were used to assess RKV responses. The scoring process considered four dimensions: (a) recall of assignment structure, (b) accuracy of key concepts or theories, (c) explanation of the main argument, and (d) ability to justify examples, evidence, or conclusions. A participant who received scores in the range of 0–2 could neither explain nor recall the submitted work. A participant who received scores in the range of 3–4 provided fragmented recall and a weak conceptual description. Scores in the range of 5–6 indicated that a participant possessed basic and partial understanding. Participants who scored in the range of 7–8 provided coherent explanations along with reasonable conceptual reconstruction. Scores in the range of 9–10 reflected a participant's ability to provide a strong and independent explanation of the submitted work.

The researcher could not conduct interrater reliability analysis for the full dataset because of the single-evaluator scoring system. To reduce evaluator bias, the researcher applied the same scoring system, question framework, and scoring descriptors across all classes. The researcher acknowledges the dual role of instructor and evaluator as a methodological limitation and suggests that future studies strengthen reliability by involving independent evaluators and reporting interrater reliability measures (e.g., intraclass correlation coefficients).

According to the conceptual framework, the ALG index was calculated as follows:

$$ALG = AQ - KM$$

The higher the ALG is, the greater the gap between assignment quality and demonstrated knowledge. Rather than searching for AI-generated text, the index examines evidence of learning through retrieval practices. In learning environments saturated with AI technologies, this method offers a stronger assessment of assessment validity (American Educational Research Association et al., 2014; Association for Writing Across the Curriculum, 2025).

Measurement of AI use

At the course level, students' patterns of AI use were tracked through self-reporting. Self-reported AI access and usage data were used to categorize the students. These categories included students reporting the use of paid AI subscriptions, students reporting the use of free AI resources, and students who reported AI use as rare and/or irregular. All reports were anonymized and compiled in aggregate form prior to analysis. The self-reporting process was not intended for disciplinary purposes and did not affect students' academic assessment.

As in many higher education contexts, the use of AI by students was prevalent in the study. Approximately 46% of the students reported that they had paid subscriptions to AI tools, generally reported as subscriptions costing approximately USD 20 per month; 37% reported using free AI tools; and 17% reported occasional or non-regular use. As the use of AI was almost ubiquitous in the context of the study, the focus was placed on access mode and intensity as a way of distinguishing between levels of AI engagement, rather than making a simple distinction between AI users and nonusers. The empirical reality was that a separate nonuser group was either absent or too small to serve as a viable comparison group. While self-reported AI access should not be equated with the absence of AI engagement, paid access does not necessarily guarantee deeper or more frequent use, and free access does not mean that AI was not used extensively. Occasional users may still use AI for substantial portions of an assignment. For these reasons, AI access was treated as a category of reported AI access mode rather than a direct measure of AI engagement behavior. In future research, self-reported data should be complemented with process-oriented empirical evidence, such as prompt histories, AI-assisted drafting records, reflective statements, revision histories, and learning logs.

Summary of measures and scoring

The measurement techniques employed in this study are summarized in Table 2.

Table 2: Summary of measures and scoring

Measure	Scale	Description	Source
Assignment Quality (AQ)	0–10	Rubric-based evaluation of submitted assignments and presentation slides	Submitted coursework files
RKV–Oral	0–10	Random oral verification based on standardized question categories	In-class questioning
RKV–Written	0–10	Ten-minute handwritten recall task without digital devices	In-class written recall
Knowledge Mastery (KM)	0–10	Oral score, written score, or composite mean	RKV scores
AI–Learning Gap (ALG)	−10 to +10	Difference between assignment quality and knowledge mastery (AQ − KM)	Computed variable
AI access mode	Categorical	Self-reported paid, free, or occasional/nonregular AI use	Anonymized student self-report

Note. AQ = assignment quality; KM = knowledge mastery; ALG = AI–learning gap; RKV = random knowledge verification.

Data analysis

The analysis of the data took place in four stages. The first stage involved the use of descriptive statistics to study the distributions of AQ, RKV–Oral, RKV–Written, KM, and the ALG index. Means, standard deviations, selected percentiles, and distributional patterns were calculated to examine whether assignment performance and knowledge mastery showed systematic gaps (Bjork et al., 2013; Nückles et al., 2020).

Comparisons among groups considered variables related to ALG within the parameters of disciplinary clusters, assignment formats, types of courses, and reported AI access mode. To this end, independent-samples t tests and ANOVA, combined with effect size measures, were used to balance statistical significance and practical significance (Cohen, 1988; Field, 2018). These comparisons were made in response to concerns pertaining to AI-assisted coursework and product inflation (Cotton et al., 2023; Michel-Villarreal et al., 2023).

In the third stage, regression models were constructed to estimate the impact of AI access parameters on the magnitude of ALG, controlling for assignment format and course characteristics. If AI acts primarily as a learning scaffold, increased access should support assignment production without widening the gap between product quality and demonstrated understanding. Conversely, if AI replaces significant cognitive functions such as synthesis or reasoning, increased access may worsen the gap between outputs and understanding (Mollick, 2024; Holmes & Miao, 2023).

Given the nesting of students within classes and the hierarchical nature of the data, multilevel modeling was used. Classroom datasets frequently violate independence assumptions because students share the same instructional context.

Multilevel models have the capacity to analyze variance at different levels, which helps determine statistical precision as well as the role that different contexts play in influencing outcomes (Raudenbush & Bryk, 2002; Snijders & Bosker, 2012). If the variation between classes is sufficiently large, it suggests that the way assessments are constructed may be a significant factor in the observed gaps.

Ethical considerations

The study was conducted following the principles of educational research ethics. The study was observational in nature and nonexperimental and did not include any activities outside the normal scope of the course. Random knowledge verification was implemented as a learning accountability tool rather than as an act of surveillance (Holmes & Miao, 2023; MLA-CCCC Joint Task Force on Writing and AI, 2024).

As the researcher also acted as the instructor, several steps were taken to mitigate the potential risks associated with self-evaluation. First, data were collected from routine, nonintrusive learning activities. Second, class and individual identities were anonymized prior to analysis and presentation. Third, the analysis was limited to aggregated and anonymized datasets. Fourth, students' permission was obtained post hoc, after grades had been finalized, for the use of anonymized learning data to ensure that participation in the research would not affect their academic assessment. Finally, self-reported AI use data were used solely for research classification purposes and were not considered for any disciplinary or punitive actions.

AI-generated text that simulates the form of academic writing expected from students may appear coherent, but this does not necessarily mean that it demonstrates critical thinking and conceptual understanding (Cotton et al., 2023; Albadarin et al., 2024). Considering the logic of the study, it is appropriate to note the importance of one of the most fundamental principles of educational measurement. This principle states that the most important function of tests and other forms of assessment is to determine whether students have mastered the required concepts and are able to apply higher-order thinking skills rather than merely producing text that meets superficial requirements (American Educational Research Association et al., 2014).

This leads to the next section, where the empirical results are presented and analyzed across different teaching contexts, together with the AI–Learning Gap and the levels of knowledge verification.

RESULTS

According to recent studies on generative AI and assessment redesign (Chan, 2023; Cotton et al., 2023; Farrokhnia et al., 2024; Kasneci et al., 2023; Selwyn, 2024), it is vital to differentiate between the perceived quality of academic outputs and independently demonstrated knowledge mastery. Thus, this study examines assignment quality (AQ), knowledge mastery (KM) through random knowledge verification (RKV), the AI–learning gap (ALG), and AI access modes across different instructional contexts.

In the full dataset (N = 1,498 students nested across 38 classes), assignment quality was assessed using the rubric described in the Methods section on a 0–10 scale. Overall, assignment quality was high (M = 7.62, SD = 1.05), with a range of 3.10–9.90. Most AQ scores fell between 7.0 and 8.9. The findings also show that even under the influence of generative AI tools, many students submitted assignments with good structural quality, argumentation, referencing, and presentation quality.

This trend aligns with emerging scholarship suggesting that generative AI may improve the perceived quality of academic work by aiding structural organization, editing, and rhetorical fluency (Cotton et al., 2023; Selwyn, 2024; Farrokhnia et al., 2024). The variation across assignment types was relatively small; however, group assignments tended to have higher AQ scores than individual assignments. This may be due to the phenomenon of collaborative production, where the division of labor and the interplay of different team members' skills can elevate the quality of the collective outcome. Research on collaborative academic work has found that the division of labor and the cooperation of different contributors can improve the quality of the presentation of results (Levine et al., 2024). However, the absence of an explicit collaboration design and the presentation of results at an appropriate AQ level do not capture the extent to which students can reformulate and articulate the conceptual framework contained in their submissions. Descriptive statistics for assignment quality across the total sample and across different assignment types are presented in Table 3.

Table 3: Descriptive statistics for assignment quality (AQ)

Outcome	N	Mean	SD	Min	Max
AQ (overall)	1,498	7.62	1.05	3.10	9.90
AQ (individual assignments)	720	7.48	1.10	3.10	9.80
AQ (group assignments)	748	7.76	0.98	3.40	9.90

Note. AQ = assignment quality. All scores are evaluated on a 0–10 scale. The aggregate counts of the subgroups may not match the total sample size because 30 assignments were ambiguous in the classification of individual versus group submissions.

Knowledge mastery was evaluated through a combination of two random knowledge verification procedures. The average score of the oral task (RKV–Oral) was M = 5.88 (SD = 1.42), while the score obtained through the written recall task (RKV–Written), in which participants were not allowed to use any devices, was lower at M = 5.21 (SD = 1.50). The two measures had an average difference of Δ

= 0.67, with RKV–Oral results consistently higher than those of RKV–Written. The difference between the RKV formats reflects, among other things, the different cognitive demands of the two procedures. During the oral component, students may have been supported by interactional cues, as well as by the opportunity to reconstruct reasoning conversationally. Rapid recall retrieval and the reconstruction of structure, independent of conversational scaffolds, were the primary cognitive demands of the written component. These types of tasks are common in educational research because they help distinguish long-term durable learning from performance based largely on surface familiarity (Bjork et al., 2013; Nückles et al., 2020). The correlation between RKV–Oral and RKV–Written scores was moderate (r = .56), indicating that the two measures, while both capturing knowledge mastery, reflect partly distinct dimensions. Oral reasoning emphasizes explanation, while the written component emphasizes recall, retrieval, and structure.

The two verification formats allow for a composite KM indicator that combines explanation-based and retrieval-based evidence of learning. This composite measure works to counteract single-method bias and, situationally, may more accurately capture students' independently demonstrated understanding when AI-supported production potentially clouds the connection between product quality and underlying conceptual understanding (Kasneci et al., 2023; Chauncey & McKenna, 2023; Farrokhnia et al., 2024). Descriptive statistics for the knowledge verification measures, including oral score, written recall score, and the composite Knowledge Mastery indicator, are presented in Table 4.

Table 4: Descriptive statistics for knowledge verification measures (KM components)

Outcome	N	Mean	SD	Min	Max
RKV–Oral	1,498	5.88	1.42	1.40	9.40
RKV–Written	1,498	5.21	1.50	1.10	9.10
KM composite (mean of oral and written scores)	1,498	5.55	1.33	1.35	9.05

Note. All measures were evaluated on a 0–10 scale. KM = knowledge mastery; RKV = random knowledge verification.

The AI–Learning Gap (ALG) is defined as:

$$ALG = AQ - KM$$

Knowledge mastery in this study's main specification was measured as the average of the two verification measures. Across the entire sample, the AI–Learning Gap had an average of M = 2.07 (SD = 1.33), with values ranging from −1.90 to 6.50. The distribution was positively skewed rather than bimodal. The alignment between assignment quality and knowledge mastery was high for some students, while for a smaller number, negative values indicated that their demonstrated mastery exceeded assignment quality. A considerable part of the sample demonstrated positive gaps, where an assignment was academically strong, but the student failed to describe, explain, or reconstruct the ideas once the device was no longer available during assessment. In this analysis, KM was the average of the scores on the oral and written verification tasks. Separating the measures showed the same pattern, with positive AI–Learning Gaps recorded in both.

To offer a clearer classification on the 0–10 scale, threshold values were applied to gaps of different magnitudes. Alignment was categorized as follows: moderate gaps when $1.0 < ALG \leq 2.0$ and large gaps when $ALG > 2.0$. Given this classification, 34.1% of students were classified as having large gaps. This pattern of distribution constitutes an important measurement contribution of the study. It shows the significant divergence between product quality and demonstrated mastery across numerous classes and subjects and strengthens the claims made about the need for assessment systems to show evidence of learning outcomes instead of inferring learning from the quality of academically produced artifacts (Chan, 2023; Selwyn, 2024; Kasneci et al., 2023; Farrokhnia et al., 2024). The distributional characteristics of the AI–Learning Gap across the sample are presented in Table 5.

Table 5: Distribution summary of the AI–Learning Gap (ALG)

Metric	Value
ALG mean	2.07
ALG standard deviation (*SD*)	1.33
ALG median	1.92
ALG 25th percentile	1.12
ALG 75th percentile	2.86
Students with ALG > 2.0 (large gap)	34.1%

Note. ALG = assignment quality − knowledge mastery. Thresholds reflect interpretable differences on a 10-point scale: $|ALG| \leq 1.0$ = alignment; $1.0 < ALG \leq 2.0$ = moderate gap; $ALG > 2.0$ = large gap.

To clarify the distributional pattern of the AI–Learning Gap, students were placed into interpretable categories reflecting alignment, moderate discrepancy, and large discrepancy between assignment quality and independently demonstrated mastery. The distribution is presented in Table 6.

Table 6: Frequency classification of the AI–Learning Gap (ALG)

ALG category	Interpretation	Percentage
ALG ≤ 1.0	Alignment between assignment quality and demonstrated mastery	28.4%
1.0 < ALG ≤ 2.0	Moderate gap	37.5%
ALG > 2.0	Large gap	34.1%

The distribution of ALG scores indicated a positive skew rather than a bimodal distribution. While a small group of students showed close alignment between the quality of their assignments and their level of mastery, more than two-thirds of students demonstrated moderate to large discrepancies between academic performance and independently demonstrated understanding.

The next part of the analysis focused on the relation between self-reported AI access type and the extent of the AI–Learning Gap. Students who paid for AI subscriptions reported an average gap of $M = 2.20$ ($SD = 1.32$). For the free-tools group, the average was $M = 1.78$ ($SD = 1.26$). The occasional-use group recorded an average of $M = 1.70$ ($SD = 1.22$). There was a mean difference of $\Delta = 0.42$ between the paid-access and free-access groups, which corresponds to a small-to-moderate effect size ($d \approx 0.33$). The average patterns of AI-assisted assignment completion suggest that paid subscriptions reduce usage friction, such as response limits and processing delays. However, this does not seem to improve knowledge mastery. Many studies on generative AI in education fail to measure this directly (Cotton et al., 2023; Chauncey & McKenna, 2023; Baek et al., 2024; Walter, 2024).

Since AI use was nearly universal in the study context, a separate nonuser group was not meaningful for comparative purposes. Therefore, the analysis focused on reported AI access modes and levels of engagement rather than on a simple user/nonuser distinction. To clarify the relationship between AI access modes and the AI–Learning Gap, descriptive comparisons of AQ, KM, and ALG across the different AI access groups are presented in Table 7.

Table 7: Assignment Quality (AQ), Knowledge Mastery (KM), and AI–Learning Gap (ALG) by AI access mode

AI access mode	*N*	AQ Mean	KM Mean	ALG Mean	*SD*
Paid AI access	689	7.82	5.62	2.20	1.32
Free AI access	554	7.49	5.71	1.78	1.26
Occasional/nonregular AI use	255	7.31	5.61	1.70	1.22

The characteristics of the data indicate that students with paid AI access generally submitted assignments that were evaluated as higher in quality, despite demonstrating knowledge mastery levels similar to, or only slightly higher than, those observed among students using free or occasional AI access. Consequently, students in the paid-access group exhibited the highest average AI–Learning Gap. To account for the hierarchical structure of the data and contextual variations across assignment formats, ALG was set as the outcome variable in the estimation of multilevel random-intercept models.

When control variables for assignment type and disciplinary cluster were considered, paid AI access was positively associated with larger AI–Learning Gaps

($\beta = 0.38$, $SE = 0.08$, $p < .001$) relative to free AI access or occasional/nonregular AI use. Group assignments were also found to contribute to greater gaps ($\beta = 0.24$, $SE = 0.07$, $p = .001$). This finding reinforces the idea that in group work, the splitting of tasks may enhance the submission quality of the group while leaving individual understanding of the concepts at varying levels. The estimated class-level random-intercept variance was $\tau_{00} = 0.18$, which resulted in an intraclass correlation coefficient ($ICC \approx 0.09$). This suggests that a nontrivial amount of the variance in ALG was between classes rather than within classes. Such variance indicates that the design of instruction and assessment may affect the degree to which assignment outputs reflect students' independently demonstrated mastery of content. These findings support recent discussions about assessment redesign in the context of generative AI (Raudenbush & Bryk, 2002; Selwyn, 2024; Chan, 2023). The primary model can be succinctly stated as:

$$ALG_{ij} = \gamma_{00} + \gamma_{10} PaidAI_{ij} + \gamma_{20} GroupAssignment_{ij} + u_{0j} + e_{ij}$$

Where: i indexes students and j indexes classes.

Overall, the findings demonstrate specific patterns across the entire dataset. The submitted assignments had relatively high scores, while the Knowledge Mastery checks, in which students completed tasks without access to devices, revealed a substantially larger share of students with limited demonstrated mastery. This disparity is reflected in the right-skewed distribution of the AI–Learning Gap, suggesting that polished academic work may not necessarily correspond to independently demonstrated understanding of the underlying concepts. Furthermore, the size of the gap varied according to differences in AI access conditions and assignment type, indicating that Random Knowledge Verification offers strong pedagogical value because it assesses students' ability to explain, articulate, and justify the ideas in the work submitted (Chan, 2023; Kasneci et al., 2023; Selwyn, 2024; Raudenbush & Bryk, 2002).

DISCUSSION

The findings supported nearly all of the study's empirical expectations and theoretical proposition. They supported the expectation that the quality of AI-assisted assignments would be appreciably higher than students' ability to demonstrate knowledge mastery independently. The measured AI–Learning Gap (ALG = 2.07) also supported the expectation that there would be a significant disparity between assignment performance and knowledge verification under device-restricted circumstances. Instead of capturing learners' individual perceptions of competence, the findings suggest that noticeable gaps between assignment quality and knowledge mastery demonstrated through independent tasks may indicate an illusion of competence in AI-assisted learning environments.

The study's findings reveal that polished coursework and substantial mastery of the underlying concepts may diverge when generative AI tools are within easy reach. The primary implication of the evidence is not that students employed AI but that many of them produced coursework that, on the surface, appeared to be of good quality, while independently demonstrated understanding was significantly weaker when they were required to reproduce, clarify, or justify the assignment without assistance from a digital device. This shifts the debate from authorship detection toward the consideration of assessment. The most significant question is not whether AI has increased academic output but whether traditional assignment formats still demonstrate learning in environments where AI is omnipresent.

This explanation fits with the literature showing that generative AI may improve the superficial quality of academic work by enhancing coherence, structure, and rhetorical flow (Cotton et al., 2023; Farrokhnia et al., 2024; Selwyn, 2024). These findings contribute to the literature by showing that improvements in product-level quality do not always result in improvements in demonstrated understanding. In this way, the study adds classroom-level empirical evidence on the relationship between assignment quality and individually demonstrated understanding. This is especially important in light of the fact that other literature has examined the potential, risks, and student perceptions, but not the degree to which students can articulate the conceptual framework of AI-generated tasks when external support is absent.

The results are also in line with research on cognitive offloading. Learning science research shows that when cognitive tasks are shifted to electronic devices, learners may lose the need to engage in the kind of effortful processing that typically leads to deep understanding, unless they remain actively engaged in evaluating and integrating the information (Risko & Gilbert, 2016; Molenaar, 2022). The results of the present study fit this pattern. Assignment products were often stronger than performance on the oral and written verification tasks. This suggests that some of the cognitive work required to organize, express, or explain ideas has been offloaded to AI use. This explanation fits the data, especially given that knowledge mastery scores were lower across both oral and written device-free verification tasks. This was reflected in a discrepancy that was not specific to one measurement mode.

It is also important to clarify that the data do not suggest that engagement with AI always has a detrimental impact on learning. Considerable evidence indicates that the impact of generative AI on learning is determined by the way learners engage with the technology. For instance, AI engagement involving brainstorming, comparative idea generation, outlining, critique, language refinement, or revision may provide instructional scaffolding that supports learning and interaction. In these situations, learners are still required to integrate, evaluate, and revise ideas independently. Conversely, the use of AI for full-text generation, idea substitution, automated argument construction, and the submission of AI-generated texts with minimal or no revision may increase the risk of cognitive offloading. In the distinction between constructive scaffolding and cognitive offloading, the determining factor is not whether AI is involved in a task but whether learners remain actively engaged in the cognitive processes of retrieval, explanation, evaluation, and conceptual construction throughout the learning activity.

Previous research has underscored that when students use generative AI to scaffold their learning through brainstorming, critiquing, and revising, it can facilitate deeper learning (Chan, 2023; Chauncey & McKenna, 2023; Farrokhnia et al., 2024). The current study does not contradict this perspective. Instead, the current findings indicate that AI-assisted learning becomes educationally concerning when students can no longer independently develop outlines, explain relationships between concepts, verify references, paraphrase ideas through their own reasoning, or demonstrate the ability to reconstruct the logic of the submitted work once AI is removed. In such cases, generative AI may shift from functioning as a learning support system to functioning as a mechanism of cognitive replacement.

This implies that assessments focusing on refined final products may largely overlook the learning processes in which AI may or may not engage the student. The observed gap illustrates cases where superficial academic achievement may overshadow deeper conceptual deficits. Current findings also speak to the body of work associated with the illusion of competence. There is a long tradition in educational psychology that documents the phenomenon in which learners confuse contextual competence with independently demonstrated mastery of a skill (Bjork et al., 2013; Fiorella & Mayer, 2015; Rittle-Johnson et al., 2017), especially when the context does not require the learner to retrieve or explain knowledge. In the context of learning supported by artificial intelligence, the illusion of competence may be greater because AI outputs are seamless. Therefore, students may feel a high degree of understanding of the content because they can identify AI outputs and make superficial edits even when they cannot explain the underlying rationale. The combination of high scores on assignments and low scores on verification assessments indicates that, for many learners, the superficial completion of the task exceeds their actual understanding of it.

The relationship between access to paid AI services and widening AI–Learning Gaps is also significant. Students who hold paid subscriptions exhibit greater gaps on average than those who use free services or paid services on an ad hoc basis, and this pattern remains even stronger in the multilevel models. One possible explanation is that unrestricted access to a sophisticated AI system eliminates friction in AI-supported production and fosters students' overreliance on AI for preparing assignments. This resonates with research showing that the conditions surrounding access to AI tools influence the frequency and depth of their use (Adeshola & Adepoju, 2024; Baek et al., 2024; Wang et al., 2024).

The findings also raise broader concerns regarding equity issues in AI-enhanced higher education. Subscription-based AI services that students pay for may improve the quality of completed assignments by providing faster, more advanced, and more sophisticated responses with fewer usage restrictions. However, enhanced AI access does not necessarily correspond to more advanced learning. This creates a paradox of assessment in which students with access to more advanced AI tools may produce higher-quality academic work while demonstrating comparatively weaker independently demonstrated understanding. In such cases, students may increasingly rely on advanced AI systems to complete substantial portions of their assignments. The paradox lies in the possibility that, under conventional assessment systems, students may be rewarded for polished academic outputs that are increasingly supported by AI rather than by independently developed understanding and reasoning.

In addition, the results provide evidence that adds to this niche of research by showing that wider access to advanced AI tools may be associated with a wider gap between product quality and the level of understanding that is demonstrated. Group assignments also exhibit greater gaps. This is understandable, given the collaborative nature of the coursework. Group assignments may enhance the overall quality of the final output owing to organizational structure and the division of tasks. However, that division of tasks may also mask the understanding of each individual, especially when AI tools are used to enhance the quality of the final output.

The findings are consistent with the literature arguing, for instance, that while collaborative writing may improve the quality of the end product, the same is not necessarily true for the quality of engagement of all authors (Levine et al., 2024). This research has several important implications for the evaluation of learning in higher education. For example, this research supports the idea that universities may need to shift beyond a focus on detection and implement evaluation practices that make learning more visible (Chan, 2023; Kasneci et al., 2023; Farrokhnia et al., 2024; Ahangama, 2026). If AI-assisted coursework is allowed in the future, polished written assignments may not be sufficient evidence of content understanding. Instead, a balanced approach to assessment may retain take-home assignments while complementing them with some form of verification, whether that be a requirement to explain the work, short in-class writing, prescribed reflective exercises, or some form of closed-device demonstration of understanding.

The results indicate that the successful incorporation of RKV and similar assessment methods requires both institutional and faculty support. While RKV is less resource-intensive than formal oral examinations and AI-detection systems, additional forms of institutional support are still necessary. This may include developing AI-informed assessment training programs for faculty, creating standardized rubrics for explanation-based assessment, and establishing shared assessment question banks to reduce faculty workload while maintaining assessment quality and integrity. Sampling-based verification approaches, in which only a subset of students in large classes are required to complete oral or written verification tasks, may also reduce the instructional burden associated with large enrollments. In addition, retrieval-based writing tasks requiring minimal class time and disruption can be integrated into routine instructional practices. RKV is intended to enhance the visibility of reasoning and conceptual understanding within AI-supported educational environments rather than to foster a culture of surveillance and distrust.

The outcome of this research is applicable to fields that rely on learning through explanation, interpretation, and argument construction, primarily the social sciences and the humanities. These disciplines are particularly at risk in the context of generative AI since they can convincingly simulate discourse within a specific field. Hence, the concern extends beyond plagiarism to the exacerbation of product-level metric inflation, which is classically employed as a proxy for learning. Currently, assessment systems must go beyond the submission of well-worded papers to determine whether students are able to articulate the necessary conceptual distinctions, support their assertions, and defend their interpretations. In this regard, the study's RKV is relevant. It is simple and easy to use. It can be employed in the classroom with minimal disruption, unlike formal oral exams, and unlike AI detection, which requires intricate technology. The approach, which combines brief handwritten recall tasks with oral questions, shows whether students can describe and elaborate on their submission.

The findings assist in understanding the possible effectiveness of RKV. Different aspects of mastery were captured by oral verification and written recall, and both were lower than assignment quality. This indicates that explanation- and retrieval-based tasks assess understanding that may be hidden by refined coursework. In the context of assessment, tasks that necessitate reconstruction, justification, and explanation may be able to restore the value of coursework assessment from a pedagogical perspective without requiring institutions to ban AI. This means that RKV is not merely a means of surveillance but rather a means of illuminating understanding. Last, the class-level variance present in the multilevel models indicates that the design of instruction is also important. A nontrivial amount of the variance in the AI–Learning Gap occurred between classes, meaning that the gap is not solely a characteristic of the student and/or the technology individually but is also related to assignment design, the degree to which students are required to elaborate in their work, and the extent to which course designs prioritize reasoning visibility over output. This suggests that there is potential for institutional- and course-level changes to address the problem of superficial learning in AI-supported learning environments.

In addition, the transferability of the findings must be interpreted in relation to the specific educational context of Vietnamese higher education. Many

Vietnamese university environments emphasize structured written assignments, presentation-based assessment, and clearly defined output requirements. Under such conditions, conventional coursework formats may be particularly compatible with generative AI-assisted production. Furthermore, levels of digital literacy and AI literacy vary considerably across institutions and student populations. Therefore, although the findings may have broader relevance for AI-integrated higher education contexts, future research should examine whether comparable patterns emerge in private universities, research-intensive institutions, STEM and professional disciplines, and educational systems beyond Vietnam.
Overall, the results align with prior studies on cognitive offloading, illusion of competence, and writing-to-learn, while augmenting classroom-based evidence on the operation of these dynamics in AI-integrated higher education systems. The findings do not justify a simplistic anti-AI perspective. Rather, they suggest that the impact of AI on education is contingent on the interplay among technological support, cognitive involvement, and assessment structure. Given these conditions, the most important question is not whether AI can assist students in creating more aesthetically pleasing assignments but whether universities retain the ability to differentiate between superficial academic performance and independently demonstrated conceptual understanding.

LIMITATIONS AND FUTURE RESEARCH

The scope within which the results and their implications can be interpreted is subject to several limitations. First, this study was conducted within a single institutional context, namely, a large public university in Vietnam, and focused primarily on courses within the social sciences, humanities, communication, and related applied social science disciplines. Although a dataset of this scale remains meaningful for examining student engagement with generative AI (GenAI) and associated assessment outcomes, the institutional and disciplinary context may shape both patterns of AI use and the resulting AI–Learning Gap (ALG). The results, of course, should be examined in terms of the specific customs of teaching and assessment in the particular context of the study. Assignment types in given programs that are more “report”, “presentation”, or “performance” based may be more or less interactive with generative AI tools than those in the more “laboratory”, “quantitative”, or “highly procedural” disciplines. For this reason, the patterns observed may not be applicable to the STEM fields, clinical and laboratory instructional disciplines, or professional programs where teaching and assessment customs are more different from the disciplines of the study. Future research should therefore examine whether similar patterns emerge across private universities, research-intensive institutions, STEM disciplines, professional education programs, and educational systems outside Vietnam (Kasneci et al., 2023; Selwyn, 2024; Farrokhnia et al., 2024).

The second limitation concerns the measurement of AI use. Compared to the actual complexity of student interaction with AI systems, the measure of AI use in this study relied primarily on self-reported modes of AI access and use, including paid subscriptions, free access, and occasional or nonregular use. Although these measures provide some indication of AI accessibility, they do not fully capture the depth, frequency, or nature of student engagement with generative AI systems. There is a wide range of ways in which students may utilize AI technology. Brainstorming, outlining, editing, summarizing, reference checking, prompting, idea generation, and producing entire essays are some of the ways students may engage with AI. Each of these represents a different mode of engagement, which may influence learning outcomes in different ways. Self-reported measures are also vulnerable to recall bias, interpretive ambiguity, and underreporting or overreporting of actual AI use (Baek et al., 2024; Levine et al., 2024). The study also lacked process-based evidence, such as prompt logs, revision histories, interaction traces, or developmental drafts, which could provide deeper insight into students' cognitive engagement with AI-supported writing processes. Future studies may therefore benefit from combining self-reported AI-use measures with process-oriented evidence such as reflective AI-use statements, prompt histories, tracked revisions, portfolio drafts, and learning logs.

The third limitation concerns the possibility of measurement error within the Random Knowledge Verification (RKV) procedures. Oral verification may be affected by classroom anxiety, communication confidence, social pressure, or situational stress associated with unexpected questioning. Similarly, the written recall task may partially reflect time pressure, memory retrieval speed, or writing fluency rather than purely conceptual understanding. The verification tasks were intentionally designed to be brief and classroom-based to encourage practical implementation in authentic teaching contexts rather than laboratory-style control over all testing variables. Although the study attempted to improve procedural consistency through common scoring descriptors, shared question categories, and standardized verification principles, variability in student performance across different classroom conditions remains possible.

Another limitation concerns the instructor-researcher role in the study. The instructor taught the students, implemented the RKV procedures, assessed students' work, and verified the completion of guideline tasks. Although efforts were made to apply common scoring descriptors and standardized verification procedures to reduce inconsistencies, evaluator bias cannot be fully excluded. Furthermore, interrater reliability was not assessed across the full dataset because scoring was conducted by a single instructor-researcher. To strengthen measurement reliability, future studies should involve multiple independent raters and report formal interrater reliability measures, such as intraclass correlation coefficients (ICC) and other appropriate reliability indicators.

The correlational relationships identified in this study are also limited in the scope of conclusions that can be drawn. The study employed a classroom-based observational design rather than a controlled experimental design; therefore, conclusions regarding the effects of AI on learning outcomes should be interpreted with appropriate caution. Although larger AI–Learning Gaps were associated with certain forms of AI access and assignment structures, the evidence does not support the claim that the use of generative AI directly causes a decline in conceptual understanding. Unmeasured factors, such as students' prior academic preparation, motivation, instructional quality, assessment culture, and levels of digital literacy, may also help explain the observed patterns.

These limitations suggest that future AI–Learning Gap research should broaden both its methodological approaches and research settings. For example, cross-national comparative research involving multiple countries should examine both public and private higher education institutions with varying levels of research intensity to assess the generalizability and reliability of the findings. Comparative studies across STEM, laboratory-based, professional, and humanities disciplines may also help clarify how the relationship between assignment quality and independently demonstrated understanding differs across disciplinary traditions.

Finally, longitudinal and quasiexperimental research designs should be prioritized to evaluate the effectiveness of Random Knowledge Verification, retrieval-based assessment, and AI literacy interventions in narrowing the AI–Learning Gap over time. Future studies should also investigate how different instructional and assessment interventions influence students' ability to integrate AI responsibly while minimizing excessive cognitive offloading. In addition, both traditional assessment approaches and digitally integrated assessment systems incorporating AI verification practices should be examined further.

Comparisons among oral verification, handwritten recall tasks, reflective explanations, portfolio-based assessments, AI interaction logs, and revision-tracking systems may help identify the assessment practices that most effectively reveal independently demonstrated understanding in AI-supported educational environments. The present study should therefore be understood as an initial empirical attempt to investigate the relationship between AI-supported academic production and independently demonstrated knowledge mastery. The findings contribute to emerging debates on assessment validity, AI-integrated higher education, and authentic learning while also highlighting the need for broader, longitudinal, multimethod, and interdisciplinary investigations into the impact of generative AI on academic work and independently demonstrated understanding.

CONCLUSION

This study aimed to establish the extent to which the growing adoption of generative artificial intelligence (AI) in higher education may widen the gap between the apparent quality of student assignments and students' independently demonstrated learning. Through the lens of the AI–Learning Gap (ALG), defined as the disparity between Assignment Quality (AQ) and Knowledge Mastery (KM) under device- and internet-restricted conditions, the study analyzed classroom-based data from 1,498 undergraduate students across 38 classes at a public university in Vietnam. The data revealed a clear pattern. A significant number of students received relatively high rubric-based scores on the structure, argumentation, and presentation of assignments, yet when asked to rationalize or explain the conceptual framework of their assignments, many students underperformed significantly. This gap produced an average AI–Learning Gap of approximately two points out of ten, and over one-third of students exhibited large gaps. Therefore, the study found that in AI-assisted learning environments, high-quality academic work does not necessarily represent a student's own cognitive mastery and may create an illusion of competence when polished academic products are mistaken for independently demonstrated understanding. The study seeks to add to the still limited body of knowledge on generative AI and higher education through two distinct contributions.

First, it presents the AI–Learning Gap framework, which can be used to differentiate product-level academic achievement from independently demonstrated conceptual understanding. Most of the literature analyzes technological features, ethical issues, or policy discussions, but few studies explore the extent to which students can autonomously articulate AI-supported tasks when the technology disappears. Second, the study introduces random knowledge verification (RKV) as an example of an assessment strategy. RKV, which integrates concise oral questions and short written recall exercises, provides a way for educators to assess whether students can articulate and defend the conceptual basis of their submissions independently of technological support.

The results of this study also have theoretical and practical implications. In line with cognitive offloading and the illusion of competence, it appears that students may not develop strong conceptual understanding when they delegate argument structuring or explanation generation to AI. Nonetheless, the findings should not be interpreted as evidence that AI necessarily has a negative impact on learning. Rather, they suggest that the educational consequences of AI largely depend on whether learners engage with it as a tool that supports thinking or as a substitute for thinking. There will be less reliance on traditional product-focused models of assessment when there is ready access to AI; thus, assessment practices will have to shift toward approaches that determine whether understanding and mastery are actually visible. This could take the form of oral verification, in-class analytical writing, and structured reflective exercises in which students must demonstrate their reasoning without AI. Limitations include the reliance on self-reported students. The detected use of AI was contextualized primarily within a single institution and focused mainly on a sample drawn from the social sciences, humanities, and communication disciplines. The study was observational and conducted within classroom settings, meaning that the findings should be interpreted as evidence of association rather than as evidence of direct causation. Future studies should build on this work through comparative studies across multiple institutions and disciplines, longitudinal studies of AI-assisted learning behavior, and studies examining whether AI literacy and assessment redesign can narrow the gap between the quality of work submitted and the quality of understanding.

Overall, the findings suggest that a productive institutional response to AI in higher education may not be possible through technological surveillance or total prohibition alone. Rather, it requires a rethinking of assessment. By redesigning assessments to emphasize reasoning and understanding, higher education institutions can safeguard the academic integrity of assessment while simultaneously enabling students to engage with the positive potential of new AI technologies. There are three practical implications for educators and course developers. First, the widespread availability of generative AI means that instructors can no longer treat polished take-home assignments as sufficient evidence of students' learning. Second, take-home assignments may still serve valuable educational purposes when combined with verification tasks that require students to explain, revise, and justify their submissions independently. Finally, random knowledge verification represents a practical, low-cost, and nonpunitive approach for making understanding more visible without requiring educators to prohibit AI use or rely primarily on technological detection systems. The issue is not that AI itself is inherently harmful; rather, assessment practices must be redesigned so that learning with AI remains connected to reasoning, explanation, and conceptual understanding.

Funding: The author received no external funding for this research.

Institutional Review Board Statement: Ethical review and approval were not required for this study because the research involved minimal risk and relied on the analysis of routine course-based learning activities conducted during normal university instruction.

Informed Consent Statement: Informed consent was obtained from all participants involved in the study. To avoid any potential influence on academic assessment, consent for the use of anonymized learning data was obtained only after the course grades had been finalized.

Data Availability Statement: The data supporting the findings of this study consist of anonymized records of classroom assessments and results from verification tasks collected during regular course activities. Due to privacy and institutional considerations, the full dataset is not publicly available but may be obtained from the author upon reasonable request.

Conflicts of Interest: The author declares no conflict of interest.

REFERENCES

Adeshola, I., & Adepoju, A. P. (2024). The opportunities and challenges of ChatGPT in education. *Interactive Learning Environments, 32*(10), 6159–6172. https://doi.org/10.1080/10494820.2023.2253858

Ahangama, N. (2026). Designing assessments in the generative AI era: A tailored assessment framework for ICT tertiary education. *International Journal of Educational Technology in Higher Education, 23*, Article 9. https://doi.org/10.1186/s41239-026-00582-0

Albadarin, Y., Saqr, M., Pope, N., & Tukiainen, M. (2024). A systematic literature review of empirical research on ChatGPT in education. *Discover Education, 3*(1), 60. https://doi.org/10.1007/s44217-024-00138-2

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. AERA.

Association for Writing Across the Curriculum. (2025). *AWAC statement on AI and writing across the curriculum*. Retrieved May 21, 2026, from https://wacassociation.org/statement-on-ai-writing-tools-in-wac/

Baek, C., Tate, T., & Warschauer, M. (2024). ChatGPT seems too good to be true: College students' use and perceptions of generative AI. *Computers and Education: Artificial Intelligence, 6*, Article 100294. https://doi.org/10.1016/j.caeai.2024.100294

Baig, M. I., & Yadegaridehkordi, E. (2024). ChatGPT in higher education: A systematic literature review and research challenges. *International Journal of Educational Research, 127*, 102411. https://doi.org/10.1016/j.ijer.2024.102411

Belland, B. R. (2013). Scaffolding: Definition, current debates, and future directions. In M. J. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology* (pp. 505–518). Springer. https://doi.org/10.1007/978-1-4614-3185-5_39

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology, 64*, 417–444. https://doi.org/10.1146/annurev-psych-113011-143823

Bobula, M. (2024). Generative artificial intelligence (AI) in higher education: A comprehensive review of challenges, opportunities, and implications. *Journal of Learning Development in Higher Education*, (30)**,** Article 1137. https://doi.org/10.47408/jldhe.vi30.1137

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

Chauncey, S. A., & McKenna, H. P. (2023). A framework and exemplars for ethical and responsible use of AI chatbot technology to support teaching and learning. *Computers and Education: Artificial Intelligence, 5*, Article 100182. https://doi.org/10.1016/j.caeai.2023.100182

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2023). 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

Dai, K., Liu, Y., & Zhang, X. (2026). Generative AI in higher education: A bibliometric review of emerging trends, power dynamics, and global research landscapes. *Computers and Education: Artificial Intelligence, 10*, 100544. https://doi.org/10.1016/j.caeai.2026.100544

Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D., & Siemens, G. (2024). Impact of AI assistance on student agency. *Computers and Education, 210*, Article 104967. https://doi.org/10.1016/j.compedu.2023.104967

Dempere, J., Modugu, K., Hesham, A., & Ramasamy, L. K. (2023). The impact of ChatGPT on higher education. *Frontiers in Education, 8*, 1206936. https://doi.org/10.3389/feduc.2023.1206936

Dhamija, A., & Dhamija, D. (2025). Understanding teachers' perspectives on ChatGPT-generated assignments in higher education. *Journal of Interdisciplinary Studies in Education, 14*(1). https://doi.org/10.32674/ptf9yd75

Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE Publications.

Fiorella, L., & Mayer, R. E. (2015). *Learning as a generative activity: Eight learning strategies that promote understanding*. Cambridge University Press. https://doi.org/10.1017/CBO9781107707085

Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2024). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International, 61*(3), 460–474. https://doi.org/10.1080/14703297.2023.2195846

Graham, S. (2019). Changing how writing is taught. *Review of Research in Education, 43*(1), 277–303. https://doi.org/10.3102/0091732X18821125

Graham, S., & Harris, K. R. (2006). Strategy instruction and the teaching of writing. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 187–207). Guilford Press.

Graham, S., MacArthur, C. A., & Hebert, M. (Eds.). (2018). *Best practices in writing instruction* (3rd ed.). Guilford Press.

Hodges, T. S., Wright, K. L., Wind, S. A., Matthews, S. D., Zimmer, W. K., & McTigue, E. (2019). Developing and examining validity evidence for the writing rubric to inform teacher educators (WRITE). *Assessing Writing, 40*, 1–13. https://doi.org/10.1016/j.asw.2019.03.001

Holmes, W., & Miao, F. (2023). *Guidance for generative AI in education and research*. UNESCO Publishing. https://doi.org/10.54675/EWZM9535

Joughin, G. (2010). *A short guide to oral assessment*. Leeds Met Press in association with University of Wollongong.

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

Kehm, B. M. (2001). *Oral examinations at German universities*. *Assessment in Education: Principles, Policy & Practice, 8*(1), 25–31. https://doi.org/10.1080/09695940120033234

Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. (2023). Interacting with educational chatbots: A systematic review. *Education and Information Technologies, 28*(1), 973–1018. https://doi.org/10.1007/s10639-022-11177-3

Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: Systematic literature review. *International Journal of Educational Technology in Higher Education, 20*(1), 56. https://doi.org/10.1186/s41239-023-00426-1

Langer, J. A., & Applebee, A. N. (1987). *How writing shapes thinking: A study of teaching and learning* (NCTE Research Report No. 22). ERIC.

Levine, S., Beck, S. W., Mah, C., Phalen, L., & Pittman, J. (2024). How do students use ChatGPT as a writing support? *Journal of Adolescent & Adult Literacy*. Advance online publication. https://doi.org/10.1002/jaal.1373

Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against nonnative English writers. *Patterns, 4*(7), 100779. https://doi.org/10.1016/j.patter.2023.100779

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

Mai, D. T. T., Da, C. V., & Hanh, N. V. (2024). The use of ChatGPT in teaching and learning: A systematic review through SWOT analysis approach. *Frontiers in Education, 9*, 1328769. https://doi.org/10.3389/feduc.2024.1328769

Michel-Villarreal, R., Vilalta-Perdomo, E., Salinas-Navarro, D., Thierry-Aguilera, R., & Gerardou, F. (2023). Challenges and opportunities of generative AI for higher education as explained by ChatGPT. *Education Sciences, 13*(856), 1–18. https://doi.org/10.3390/educsci13090856

MLA-CCCC Joint Task Force on Writing and AI. (2024). *Generative AI and policy development: Guidance from the MLA-CCCC task force*. Retrieved May 24, 2026, from https://aiandwriting.hcommons.org/working-paper-1/

Mollick, E. (2024). *Cointelligence: Living and working with AI*. Random House UK.

Mollick, E., & Mollick, L. (2023). *Assigning AI: Seven approaches for students, with prompts*. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4475995

Molenaar, I. (2022). Toward hybrid human–AI learning technologies. *European Journal of Education, 57*(4), 632–645. https://doi.org/10.1111/ejed.12527

Montenegro-Rueda, M., Fernández-Cerero, J., Fernández-Batanero, J. M., & López-Meneses, E. (2023). Impact of the implementation of ChatGPT in education: A systematic review. *Computers, 12*(8), 153. https://doi.org/10.3390/computers12080153

Munaye, Y. Y., Admass, W., Belayneh, Y., Molla, A., & Asmare, M. (2025). ChatGPT in education: A systematic review on opportunities, challenges, and future directions. *Algorithms, 18*(6), 352. https://doi.org/10.3390/a18060352

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*, 100041. https://doi.org/10.1016/j.caeai.2021.100041

Nückles, M., Hübner, S., & Renkl, A. (2020). The self-regulation view in writing-to-learn: Using journal writing to optimize cognitive load in self-regulated learning. *Educational Psychology Review, 32*(3), 753–778. https://doi.org/10.1007/s10648-020-09528-1

Pérez-Pérez, I., González-Afonso, M. C., Plasencia-Carballo, Z., & Pérez-Jorge, D. (2026). Transparency mechanisms for generative AI use in higher education assessment: A systematic scoping review (2022–2026). *Computers, 15*(2), Article 111. https://doi.org/10.3390/computers15020111

Poe, M., & Elliot, N. (2019). Evidence of fairness: Twenty-five years of research in *Assessing Writing*. *Assessing Writing, 42*, 100418. https://doi.org/10.1016/j.asw.2019.100418
Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Sage Publications.
Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences, 20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002
Sampah, S. N. A., Nabang, M., Krampa, E. K., Issah, M., Koduah, F., Essel, H. B., & Adu Agyem, J. (2026). Generative AI competence and student engagement in higher education: Mediating roles of utilization, autonomy, and formal learning. *Journal of Interdisciplinary Studies in Education, 15*(2), 109–146. https://doi.org/10.32674/3s3cx241
Selwyn, N. (2024). On the limits of artificial intelligence (AI) in education. *Nordisk Tidsskrift for Pedagogikk og Kritikk, 10*(1), 3–14. https://doi.org/10.23865/ntpk.v10.6062
Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). SAGE Publications.
Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). *Developing evaluative judgment: Enabling students to make decisions about the quality of work*. *Higher Education, 76*(3), 467–481. https://doi.org/10.1007/s10734-017-0220-3
Tillmanns, T., Salomão Filho, A., Rudra, S., Weber, P., Dawitz, J., Wiersma, E., Dudenaite, D., & Reynolds, S. (2025). Mapping tomorrow's teaching and learning spaces: A systematic review on GenAI in higher education. *Trends in Higher Education, 4*(1), 2. https://doi.org/10.3390/higheredu4010002
Van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in teacher–student interaction: A decade of research. *Educational Psychology Review, 22*(3), 271–296. https://doi.org/10.1007/s10648-010-9127-6
Walter, Y. (2024). Embracing the future of artificial intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education. *International Journal of Educational Technology in Higher Education, 21*(1), 15. https://doi.org/10.1186/s41239-024-00448-3
Walters, W. H. (2023). The effectiveness of software designed to detect AI-generated writing: A comparison of 16 AI text detectors. *Open Information Science, 7*(1), 20220158. https://doi.org/10.1515/opis-2022-0158
Wang, C., & Tian, Z. (2025). *Rethinking writing education in the age of generative AI*. Taylor & Francis. https://doi.org/10.4324/9781003426936
Wang, C. Z., Aguilar, S. J., Bankard, J. S., Bui, E., & Nye, B. (2024). Writing with AI: What college students learned from utilizing ChatGPT for a

writing assignment. *Education Sciences, 14*(9), Article 976. https://doi.org/10.3390/educsci14090976

Williamson, D. M., Xi, X., & Breyer, F. J. (2012). A framework for evaluation and use of automated scoring. *Educational Measurement: Issues and Practice, 31*(1), 2–13.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*(2), 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

---

TRAN MINH DUC, PhD, is a Lecturer in the Faculty of Cultural Industries, Thu Dau Mot University, Vietnam. His research interests include higher education, assessment and learning, generative artificial intelligence in education, cultural studies, and Vietnamese studies. His recent research focuses on AI-assisted learning, assessment validity, cognitive offloading, authentic learning, and the development of innovative assessment approaches in higher education. Email: ductm@tdmu.edu.vn

---