Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Random Knowledge Verification as a Scalable Assessment Strategy in AI Rich Higher Education

Tran Minh Duc

ductm@tdmu.edu.vn

Thu Dau Mot University

Case Report

Keywords: cognitive offloading, generative artificial intelligence, higher education assessment, learning verification, retrieval practice

Posted Date: March 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9103013/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

The rapid advancement of generative artificial intelligence (GenAI) is altering student workflows within higher education. Concerns have arisen regarding large language models' ability to produce articulate essays, reports, and presentation materials. Many institutions have attempted to respond by implementing strategies to identify AI-generated work and increasing the enforcement of academic integrity. In the end, detection methods have not resolved the primary concern of whether students truly comprehend course content. This article posits that the greatest concern with the integration of AI into education is not detecting the use of AI, but rather confirming the presence of learning. The article presents Random Knowledge Verification (RKV) as a practical assessment method that focuses on confirming students' understanding. The study is a classroom-based analysis conducted over a period of two academic years at a public university in Vietnam with 1,498 undergraduate students across 38 classes in the social sciences, humanities, and communication fields. Students participated in individual and group work, and their performance was evaluated through standard grading and knowledge verification exercises that incorporated technology restrictions, random verbal questioning, and brief written recall activities. The data show a persistent gap between Assignment Quality and Knowledge Mastery, with an average AI-Learning Gap of 2.07 out of 10. Even though students submitted assignments that appeared sophisticated, many struggled to describe or reconstruct the core ideas when other resources were unavailable. Thus, these findings indicate that sophisticated academic outputs, in this case, do not necessarily demonstrate genuine understanding in an AI-augmented learning environment. The study suggests that assessment reform should focus on learning verification rather than AI detection. Additionally, Random Knowledge Verification is a practical, low-cost method for increasing the visibility of learning in AI-rich higher education classrooms.

1. Introduction

1.1. The emerging assessment crisis in AI-rich higher education

The recent proliferation of Generative Artificial Intelligence (GenAI) tools, specifically large language models like ChatGPT, has fundamentally altered the way students engage with academic work in higher education (Montenegro-Rueda et al., 2023). Activities that previously entailed considerable investments of time and cognitive energy can now be completed in minutes with the help of AI. These activities may include generating ideas, organizing arguments, drafting text, language editing, and even creating presentation slides (Munaye et al., 2025). It is no surprise, then, that GenAI is now integrated into the academic work of many students in writing-intensive classes and in other fields of study that require explanation, interpretation, and structured reasoning (Kasneci et al., 2023; Michel-Villarreal et al., 2023; Rahman & Watanobe, 2023). GenAI's rapid adoption in higher education has both positive and negative aspects: it supports brainstorming, feedback, and revision, while also raising concerns about authorship, originality, and academic performance (Baig & Yadegaridehkordi, 2024; Bobula, 2024; Dwivedi et al.,

2023). The changes that GenAI tools are bringing about, and their impact on teaching and learning, have created an assessment crisis in AI-rich higher education. Traditional assignment-based assessment has assumed that the work students submit demonstrates the quality of their learning.

The increasing fragility of this assumption lies in students' ability to complete polished essays, reports, and presentations with substantial assistance from generative systems. In this case, the visible quality of academic work may become detached from students' own understanding. This issue has become pivotal in discussions regarding academic integrity, the originality of students' work, and the evolving nature of assessment in higher education (Chan, 2023; Cotton et al., 2023; Dempere et al., 2023). Institutional responses may differ, but most recent approaches still predominantly focus on identifying whether AI was used rather than determining whether any meaningful learning took place.

1.2. AI detection approaches' limitations

The most common institutional response to GenAI has been the expansion of detection-focused strategies, which include AI detection software, plagiarism detection systems, and more stringent monitoring of students' writing. There are, however, important conceptual and practical limitations to these approaches.

To begin with, research has shown that AI detection tools are quite unreliable, often producing inconsistent outcomes across different platforms and different styles of writing (Walters, 2023; Walter, 2024). Second, AI detection tools have, in some cases, wrongly classified the writing of non-native speakers of English as AI-generated, raising serious concerns about fairness and validity (Liang et al., 2023). Third, AI-generated writing that has been partially or substantially edited can often evade detection, which makes such systems easier to circumvent than institutions may assume (Perkins et al., 2024a,b). In general, detection-based methods are addressing the wrong problem. In educational contexts, the central issue is not simply whether AI was used, but whether students have actually learned or understood the content they submitted (Lodge et al., 2023; Perkins & Roe, 2024). From this perspective, the more appropriate shift is away from AI surveillance and toward learning verification, with a stronger focus on students' actual understanding and knowledge retention.

1.3. The need for mastery-visible assessment

Learning science strongly supports this shift. Numerous studies show that understanding is better demonstrated through explanation, recall, and reasoning than through the production of final outputs. Active recall is far more likely to result in durable learning than passive recognition (Karpicke & Blunt, 2011; Roediger & Karpicke, 2006). In addition, studies have shown that learners often incorrectly self-assess, and that cognitive illusions may lead them to mistake fluency for mastery (Bjork et al., 2013; Kornell & Bjork, 2008). In circumstances where AI is involved, text may appear to reflect a high level of comprehension, coherence, and persuasive argumentation, while the learner's actual understanding remains weak. The concept of cognitive offloading further supports this concern. When students use external tools to organize, explain, or generate content, they may not engage in the cognitive processes associated with deeper learning. In AI-assisted learning situations, this offloading may create a

misleading sense of competence (Risko & Gilbert, 2016). In digitally supported environments, particularly those shaped by GenAI, students may believe they "know" the material simply because they can reproduce generated text, even though they may not be able to explain the ideas in their own words.

In this context, high-quality coursework may mask a disparity between observable performance and true mastery.

1.4. Research gap

The growing body of literature on GenAI and education demonstrates the absence of practical strategies for verifying learning in the classroom that are feasible, inexpensive, and scalable. Current discussions center around policy, ethics, or the adoption of AI, while giving little attention to assessment processes that can be applied in typical university classrooms, especially those that lack advanced classroom technology and cannot rely on extensive administrative support (Farrokhnia et al., 2024; Holmes & Miao, 2023). This gap is especially relevant in large classroom settings where teachers are in urgent need of strategies that are both conceptually sound and operationally practical.

1.5. Purpose of the study

In response to this challenge, the current study introduces the concept of Random Knowledge Verification (RKV) as a scalable assessment mechanism suitable for higher education. Instead of focusing on ascertaining the potential use of AI, RKV aims to make learning visible by creating situations in which students, under conditions of restricted device use, are required to explain, recall, and reconstruct the reasoning behind the knowledge reflected in their submissions. Thus, RKV is conceptualized as a mastery-visible evaluative approach and as a constructive alternative to AI detection mechanisms.

1.6. Research questions

Thus, this study seeks to answer the following research questions:

RQ1: In what ways can Random Knowledge Verification serve as a meaningful evaluative approach within the context of AI-rich higher education?

RQ2: To what extent can RKV identify gaps between independently demonstrated knowledge and the knowledge exhibited in submitted assignments?

RQ3: What benefits does RKV provide in terms of manageable assessment in large classrooms?

2. Literature Review

The swift proliferation of generative artificial intelligence (GenAI) within higher education has heightened scrutiny regarding how learning should be assessed when students are able to produce sophisticated academic work with external assistance. While the debate often centers on academic misconduct, the issue is more fundamentally epistemological and pedagogical: in AI-enhanced environments, does

visible and evaluable academic performance still constitute proof of cognitive mastery? This review approaches the issue through four interrelated theoretical pillars: the illusion of knowledge and perceived understanding; cognitive offloading in digital environments; retrieval practice and learning verification; and assessment in the context of AI-embedded education.

2.1. Illusion of knowledge and perceived understanding

A primary focus of learning research is how students come to believe that they understand material. Relevant concepts include illusions of competence, fluency-based misjudgment, and the illusion of explanatory depth. Each of these suggests that learners may mistake familiarity, readability, or coherence for true understanding, even when they struggle to explain the material or the logic of what they have studied (Bjork et al., 2013; Fisher et al., 2015). Research in educational psychology has shown that material that appears smooth, accessible, or easy to process can mislead students into overestimating their understanding. In learning situations involving reading, writing, and structured explanation, students may recognize key concepts, retain terminology, or even follow a strong and well-structured argument. However, those same students may still lack the ability to define core ideas, articulate the relationships among claims, or elaborate the reasoning in their own words. Research on learning and metacognitive judgment similarly shows that learners' judgments of mastery are often inaccurate when they rely too heavily on surface fluency rather than retrieval and transfer (Kornell & Bjork, 2008). In simple terms, the more familiar students become with a format, the more likely they are to believe they have understood the material. Yet the more they rely on that format to demonstrate understanding, the more likely it is that they are reproducing, paraphrasing, or restructuring material without truly mastering it.

The problem becomes more pronounced in academic settings that incorporate AI tools. GenAI systems produce fully formed sentences that appear coherent and fluent. Because of this, both students and teachers may interpret polished text as evidence of the author's understanding of the underlying ideas. Recent studies have shown that AI-generated academic work may satisfy traditional structural and stylistic criteria, while the author demonstrates little or no genuine understanding of the content (Albadarin et al., 2024; Farrokhnia et al., 2024). This suggests not only that students are able to submit AI-assisted work, but also that polished documents can create a false impression of competence. As a result, higher education faces a significant dilemma. The traditional product-based evaluative approach to learning may unintentionally reinforce forms of academic dishonesty by assuming that fully formed and well-structured work necessarily demonstrates understanding of academic concepts.

2.2. Cognitive offloading in digital environments

The second theoretical pillar is cognitive offloading, which refers to the use of external aids to reduce cognitive demands on the user. Offloading has long been part of human problem-solving. People have used tools such as notepads, calculators, flowcharts, search engines, and digital devices to support memory, organization, and reasoning. In some situations, these tools enhance efficiency and free up cognitive resources for more complex tasks (Risko & Gilbert, 2016). However, their educational impact is

highly contextual and depends on what is being offloaded and whether the learner remains actively engaged in the reasoning process. Earlier digital tools typically supported the storage or retrieval of information. By contrast, generative AI can perform various higher-order cognitive functions, including drafting explanations, structuring logical sequences, summarizing texts, generating and interpreting examples, and proposing possible explanations (Baek et al., 2024). This shifts offloading from memory support to the offloading of synthesis, expression, and conceptual organization.

Research on hybrid human-AI learning environments suggests that benefits are more likely when learners actively critique and revise AI-generated outputs. However, when the technology substitutes for cognitive work rather than supporting it, the result may be detrimental to learning (Molenaar, 2022). This distinction echoes earlier scholarship on scaffolding, which emphasizes that supports contribute to learning only when they extend rather than replace the learner's own intellectual activity (Belland, 2013; Van de Pol et al., 2010). In educational contexts, the learning potential associated with elaboration, self-explanation, and knowledge restructuring is weakened when cognitive offloading becomes excessive or substitutes for active thought. Recent findings on AI writing tools resonate with concerns about shallow or non-durable learning under conditions of heavy cognitive offloading, suggesting that the educational value of GenAI depends strongly on how it is used (Levine et al., 2024; Wang et al., 2024; Wang & Tian, 2025). GenAI is more educationally effective when it is used in guided ways for drafting, critique, and revision, rather than as a tool that replaces students' own cognitive work by allowing them to transfer AI-generated responses directly into assignments.

In these conditions, the cognitive task of structuring and articulating knowledge shifts from the student to the system, potentially widening the discrepancy between demonstrated performance and true mastery.

2.3. Retrieval practice and learning verification

The third pillar comes from learning science, which has repeatedly shown that retrieval is one of the strongest indicators of learning. Unlike passive review or recognition-based familiarity, retrieval requires learners to reconstruct knowledge from memory, reorganize it, and prepare it for production under conditions of limited external cues. For this reason, recall, explanation, and spontaneous reconstruction are especially strong indicators of whether understanding has been internalized (Roediger & Karpicke, 2006). Retrieval is not only an indicator of learning, but also a mechanism for strengthening learning because it requires memory to access relationships among ideas that constitute meaningful knowledge. Subsequent studies further support the value of retrieval: across a range of contexts, retrieval practice, as opposed to elaborative restudy, leads to stronger learning, especially when learners are required to produce explanations or reconstruct conceptual structures (Karpicke & Blunt, 2011). In writing-to-learn studies, knowledge has been shown to be more stable when learners actively reframe, summarize, and reorganize ideas, rather than simply encounter complete formulations (Nückles et al., 2020). These studies provide important insight for assessing learning in AI-rich classrooms. If students are able to produce polished content with the assistance of AI, a central challenge for assessment is determining whether they can independently retrieve and explain the ideas contained in that work.

Assessment design must account for the possibility that product quality may reflect access to tools, language fluency, division of labor, or surface-level language polishing. By contrast, retrieval-based performance is more closely tied to cognitive availability. For this reason, the design of learning verification must move beyond the apparent complexity of submitted assignments. It should include opportunities for students to demonstrate memory retrieval, explanation of underlying concepts, and reasoning that cannot easily be outsourced to external systems. In instructional settings augmented by AI, such activities become crucial for keeping assessment focused on understanding and reasoning rather than on merely mechanistic production.

2.4. Assessment challenges in AI-rich education

Most current institutional responses to generative artificial intelligence (GenAI) tend to center on detection, prohibition, and containment. These include the use of AI-detection software, revisions to plagiarism policies, controlled in-class examinations, and requirements for disclosing AI use. While these actions reflect legitimate concerns about academic integrity, they also have important practical limitations. For example, detection tools have been described as inconsistent, fragile, and at times biased against non-native speakers of English (Liang et al., 2023; Perkins et al., 2024a,b), and they can often be evaded through basic paraphrasing. More critically, these measures tend to address concerns about authorship rather than the more important educational question of whether students actually comprehend the material they submit. A growing number of scholars and policy-oriented reports have argued that these responses are less constructive than assessment reform and have suggested placing greater emphasis on redesigning assessment rather than expanding surveillance.

If institutions wish to remain focused on student learning in contexts shaped by AI-supported writing and thinking tools, evaluation practices should be redesigned (Holmes & Miao, 2023; Lodge et al., 2023). This aligns with guidance on the responsible use of AI, which emphasizes the importance of accountability and the alignment between learning outcomes and assessment methods (Chan, 2023; Perkins et al., 2024a,b; Perkins & Roe, 2024). The key issue is not merely whether AI use can be identified, but whether assessment design remains appropriate, equitable, and sustainable in classrooms where AI is increasingly integrated. Sustainability is particularly important. Many proposed solutions to the challenges posed by GenAI are difficult to sustain in large classrooms because they require substantial administrative labor, specific technological infrastructure, and continuous oversight. For assessment in higher education to be sustainable, it should be feasible in ordinary classroom settings while still producing evidence of meaningful student learning. Assessment redesign is therefore more important than ever in AI-rich higher education, especially if educators want to adopt simple and scalable approaches that verify learning rather than merely verify the use of AI tools.

This approach requires students to demonstrate recall, explanation, and reasoning without the assistance of external supports. It is this gap that the current study seeks to address by proposing the Random Knowledge Verification (RKV) model as a workable assessment tool for higher education.

3. Conceptual Framework: Random Knowledge Verification

This study's core value proposition is that in AI-enhanced higher education, the key assessment challenge extends beyond the question of whether students are using external resources, to the question of whether learning is obscured, and to what degree, as a result of tool availability. As elaborated on in the previous chapter, polished academic outputs do not demonstrate evidence of ownership of a conceptual understanding because learners can rely on generative AI to create outputs that are coherent, well-developed, and rhetorically effective. To meet this challenge, this study proposes Random Knowledge Verification (RKV) as a mastery-visible assessment strategy that seeks to determine whether students can explain, remember, and reconstruct the knowledge that undergirds the work that was submitted.

3.1. Conceptual definition of RKV

Random Knowledge Verification can be explained as an assessment method that evaluates student knowledge by requiring the spontaneous explanation and retrieval of information while restricting the use of devices. The primary goal of this method is not to determine whether students are using AI technologies to generate responses, but to assess whether they can demonstrate ownership of the ideas, reasoning, and conceptual structure associated with the assignment without external assistance. Therefore, Random Knowledge Verification shifts the evaluative focus from authorship to knowledge. This distinction is very important from an educational perspective. The literature on assessment validity states that evaluative practices must correspond to the construct being measured (American Educational Research Association et al., 2014). When the construct being measured is cognitive understanding, the evaluative method must provide evidence of students' ability to explain, structure, and retrieve information autonomously. In AI-rich classrooms, this is an essential requirement because assignment quality may reflect either subject mastery or the technological support being used (Holmes & Miao, 2023; Lodge et al., 2023). Therefore, Random Knowledge Verification is viewed as a teaching-oriented assessment method that brings the focus back to learning rather than functioning as a tactic for controlling student behavior.

3.2. Core principles of RKV

The principles behind RKV are interlinked. First is the principle of spontaneous knowledge recall. For the RKV method to be applied, learners must demonstrate an immediate understanding of the learning task. They should be able to articulate the essential ideas, reformulate a line of reasoning, or clarify the rationale behind the work they are submitting. This is in line with learning science research showing that retrieval-based learning is more effective than passive exposure or repeated review of completed answers (Roediger & Karpicke, 2006; Karpicke & Blunt, 2011). The second principle is knowledge verification under conditions of restricted device use. The emergence of tools capable of generating explanations, summaries, and arguments in real time makes knowledge verification more difficult when students have unrestricted access to digital devices. Therefore, the temporary absence of digital tools does not indicate hostility toward technology; rather, it structures an environment in which the learner's understanding becomes more observable. This principle is also supported by research on cognitive

offloading, which shows that reliance on external support tools can reduce internal cognitive engagement (Risko & Gilbert, 2016). The third principle is randomized questioning. The RKV method assumes that verification should not be predictable in relation to either timing or content.

This approach reduces students' ability to rely on narrow, memorized responses that may be disconnected from actual understanding. It increases the likelihood that explanations reflect conceptual flexibility rather than repetition. RKV aligns with the literature on the illusion of knowledge, in which learners may appear knowledgeable without actually demonstrating deep understanding (Bjork et al., 2013; Fisher et al., 2015). The fourth principle is the demonstration of mastery. RKV operates on the assumption that comprehension must be made visible through explanation, restructuring, and justification. This principle responds to calls for assessment reform in AI-rich educational contexts, where the challenge is not merely to manage technology, but to capture evidence of students' thinking and learning processes (Chan, 2023; Perkins et al., 2024a,b). For a clearer understanding of how these principles interact, Table 1 presents the conceptual rationale of the RKV model.

Table 1
Core principles of Random Knowledge Verification

Principle	Conceptual meaning	Assessment implication
Spontaneous knowledge recall	Understanding is demonstrated through on-the-spot retrieval and explanation	Students are asked to restate or explain ideas without extended preparation
Device-restricted verification	External digital assistance is temporarily removed	Evaluation occurs under conditions where AI or online support is unavailable
Randomized questioning	Verification is not fully predictable in timing or exact prompt structure	Students must show flexible understanding rather than rehearsed responses
Mastery-visible demonstration	Learning should be observable through explanation, reasoning, and reconstruction	Assessment emphasizes demonstrated understanding rather than product polish alone

Note. The table summarizes the conceptual principles underlying Random Knowledge Verification as a mastery-visible assessment strategy in AI-enriched higher education.

3.3. Components of the RKV model

Operationally, RKV has two main components: RKV-Oral and RKV-Written. RKV-Oral uses random oral questioning in which students explain concepts, clarify conceptual structures, justify examples, or restate arguments. This format requires students to reason and articulate ideas in real time. It tests students' ability to explain the logic of the work presented in their assignments without relying on prompts or notes.

RKV-Written consists of a brief handwritten recall task completed under device-restricted conditions. In this format, students are asked to reconstruct the structure of an assignment and summarize its most

important components or restate the main ideas in their own words. In contrast to oral questioning, RKV-Written is designed to strengthen retrieval and the internal organization of knowledge. It examines whether students can independently reproduce the intellectual architecture of their submitted work in a coherent manner.

The two components of RKV serve complementary functions. RKV-Oral captures spontaneous reasoning, verbal explanation, and responsiveness to questioning. RKV-Written captures recall, structural reconstruction, and the ability to organize knowledge independently without prompts.

When combined, these two formats provide a broader range of evidence for assessing mastery than either format alone.

3.4. Why combining oral and written verification matters

Oral and written verification each capture different dimensions of understanding. Oral verification reflects students' ability to think flexibly, respond to unexpected questions, and explain or defend conceptual choices in real time. These elements reflect cognitive spontaneity. At the same time, momentary communicative pressure may sometimes influence oral performance.

Written verification, by contrast, allows students to reconstruct ideas more deliberately, demonstrating their ability to recreate mental structures and organize their reasoning in a coherent form. However, written verification does not capture the same degree of cognitive spontaneity required for real-time explanatory reasoning.

Combining the two formats therefore provides a more complete representation of students' understanding and reasoning, offering stronger justification for their use in assessment. Empirical research supports the use of multiple assessment formats when evaluating students' understanding, reasoning, and knowledge transfer (Poe & Elliot, 2019; Williamson et al., 2012).

In AI-supported educational contexts, the combined use of oral and written verification is particularly valuable because it helps distinguish genuine understanding from superficial performance. A student may appear capable of presenting synthesized ideas orally yet struggle to reconstruct those ideas independently in writing, or vice versa.

By requiring both spontaneous articulation and structured recall, RKV reduces the likelihood that polished assignments will be mistaken for genuine learning.

3.5. Scalability of RKV in higher education

A key advantage of RKV is its scalability. RKV does not rely on technology-based monitoring systems and therefore requires minimal technological infrastructure. It does not depend on specialized software and does not require substantial financial resources.

The oral component can be integrated into existing presentations, seminars, or discussion sessions, while the written component can be implemented as a short in-class handwritten exercise under controlled conditions. This is particularly useful in higher education environments where instructors face large class sizes and diverse assessment requirements.

RKV also aligns well with existing teaching practices. It can be incorporated into seminars, project presentations, group reports, or assignment-based activities without requiring major changes to existing course structures. For this reason, RKV is not intended to replace traditional assignments but rather to function as an additional assessment layer that strengthens verification of learning.

This flexibility is important because many current strategies proposed to address AI use in education are difficult to enforce at scale, particularly those that rely heavily on unreliable detection technologies or high-surveillance monitoring systems (Liang et al., 2023; Walters, 2023). In this context, RKV offers a meaningful, low-cost, and realistic approach to addressing the challenges posed by AI in higher education.

4. Methods

4.1. Research design

The purpose of this research was to understand how Random Knowledge Verification (RKV) can be integrated as a form of assessment within AI-rich higher education classrooms. The focus of this research was on naturalistic educational settings rather than laboratory settings. This research purpose determined the choice of educational environment. From an educational perspective, assessment practices are best understood in naturalistic settings. Laboratory-based educational studies, for example, are often criticized for lacking the ecological and instructional authenticity of real classroom assessment activities (Cohen et al., 2018; Creswell & Creswell, 2018; Long & Magerko, 2020; Mai et al., 2024). Classroom research therefore provided the most suitable context for data collection in this study. Data collection took place in classrooms over the course of two academic years at a public university in Vietnam with a diverse population of undergraduate students across the social sciences, humanities, and communication studies. The study adopted a mixed-methods classroom-based observational design in which regular teaching activities served as the primary context for research. The Random Knowledge Verification method was integrated into ordinary teaching practice in order to assess whether students could demonstrate their understanding independently.

This method enabled the researcher to compare the apparent quality of submitted coursework with students' knowledge performance under device-restricted verification conditions.

4.2. Participants

Data were collected from 1,498 undergraduate students across 38 university classes over several semesters. All participants were students enrolled in classes taught by the researcher during the data

collection period. The courses covered various areas in the social sciences, humanities, and communication studies, all of which involved interpretation-, theory-, and argument-oriented coursework. Students participated in the research by virtue of taking the course. No additional coursework or alternative assessment tasks beyond the usual course requirements were introduced. Participation therefore reflected typical classroom engagement rather than a recruited experimental sample. This helped preserve the naturalness of student behavior during both the preparation and verification phases of assigned work. The distribution of participants across assignment types and class configurations is presented in Table 2.

Table 2
Overview of participants and class distribution

Category	Number
Total students	1,498
Total classes	38
Individual assignments	720
Group assignments	748
Unclassified assignments	30

Note. Each assignment format was determined according to the specifications of the course. A small number of assignments could not be clearly classified as either individual or group submissions because of overlap across project types.

4.3. Assignment tasks and learning activities

As part of normal course assessment, students completed both group and individual assignments. In most cases, students were assigned individual reports that involved researching a course-related topic, reviewing relevant literature, and writing a report in an academic format. An academic report typically included an introduction, an analytical discussion, and a concluding interpretation. In addition to the written report, students were also required to prepare a presentation to be delivered in class. These assignments were designed to assess skills such as information synthesis, concept explanation, and disciplinary communication (Biggs & Tang, 2011). At the same time, a growing body of research suggests that the increasing availability of AI writing tools has made the interpretation of assignment performance more complex than in the past. In many cases, it is unclear how much conceptual understanding students actually possess, since large language models (LLMs) can generate text with structured arguments and explanations at a level comparable to student work (Kasneci et al., 2023; Dwivedi et al., 2023). In this way, the assignments in the current study served two roles. First, they functioned as conventional measures of academic achievement. Second, they served as the basis for knowledge verification through RKV, which aimed to determine students' ability to articulate and reconstruct the concepts reflected in their submitted work.

4.4. Random Knowledge Verification procedures

Random Knowledge Verification was carried out using two complementary methods: RKV-Oral and RKV-Written. To reduce the influence of real-time digital assistance, both methods were conducted in device-free environments. The RKV-Oral method involved brief, random, and unscripted questioning of students in order to elicit spontaneous responses rather than memorized ones. Students were asked about key arguments in their papers, definitions of central concepts, and the analytical rationale underlying their work. The RKV-Written method involved a short written recall exercise in which students, without the aid of electronic devices, summarized arguments, identified the main concepts in their work, and described the structure of their reasoning. Through the combination of oral and written procedures, students were required to engage both in real-time reasoning and in the more organized reconstruction of knowledge. The combined oral and written format allowed the researcher to document spontaneous thinking as well as the level of reasoning demonstrated. In this way, the dual method captured two dimensions of unscripted reasoning and aligned with research supporting the use of multiple forms of assessment to evaluate learning outcomes (Pellegrino et al., 2001).

4.5. Measures

The study utilized three primary measures: Assignment Quality (AQ), Knowledge Mastery (KM), and the AI-Learning Gap. Assignment Quality was the instructor's evaluation of submitted coursework using the 0–10 grading system commonly applied in Vietnamese higher education. Scores reflected criteria such as clarity of argument, relevance, organization, and use of evidence. Knowledge Mastery assessed students' independent understanding as demonstrated during the RKV procedures. Oral responses and written recall tasks were evaluated on the basis of accuracy, coherence, and the ability to reconstruct the main ideas of the assignment. The AI-Learning Gap was defined as the difference between Assignment Quality and Knowledge Mastery. This measure was used to assess the gap between the perceived quality of submitted work and students' independent understanding under verification conditions.

4.6. Data analysis

Data analysis was conducted in several steps. First, descriptive statistics were used to summarize Assignment Quality and Knowledge Mastery scores across the full sample. This included mean scores, standard deviations, and score ranges. Second, comparative analysis was conducted to examine differences between assignment performance and knowledge verification results. This step helped estimate the average extent of the gap between completed coursework and the independent understanding demonstrated by students. Finally, the study explored variation across classes to determine whether the observed gap patterns were consistent across different learning environments. Class-level analysis helped clarify whether these patterns reflected only isolated cases or formed part of a broader assessment phenomenon in AI-rich teaching contexts.

Ethical considerations. This classroom-based study used assessment data generated through normal course activities. Students participated in the assignments and verification tasks as part of standard teaching and assessment procedures. All data used in the study were anonymized prior to analysis, and

results were reported only in aggregated form to protect student confidentiality. The study did not introduce additional interventions beyond regular classroom practice.

5. Results

This section presents the empirical findings of the study in four parts. First, it describes the profile of Assignment Quality (AQ) across the sample. Second, it reports the findings from the two Random Knowledge Verification (RKV) procedures—oral verification and written recall. Third, it analyzes the AI-Learning Gap in cases where product-level performance and independently demonstrated understanding are misaligned. Finally, it evaluates the practical usefulness of RKV as a classroom-based assessment tool. This structure aligns with recent trends in higher education assessment research, which emphasize the integration of descriptive performance data and assessment practice within authentic teaching contexts (Lodge et al., 2023).

5.1. Assignment Quality across the sample

Across the total sample of 1,498 students, Assignment Quality was moderate to relatively strong, with a mean AQ score of 7.62 (SD = 1.05) on a 0–10 scale and a range from 3.10 to 9.90. This suggests that, at the level of submitted coursework, most students produced assignments that met at least a reasonable standard in terms of structure, organization, and presentation. As shown in Table 3, group assignments yielded slightly higher mean AQ scores than individual assignments. The mean score for individual assignments was 7.48 (SD = 1.10), whereas the mean score for group assignments was 7.76 (SD = 0.98). Although the difference is relatively small, it may indicate that group assignments tend to result in higher-quality final products, possibly because the workload is distributed and allows students to contribute in different ways, including content development, design, and language refinement. This pattern is also consistent with recent discussions of AI and collaborative academic work, in which the final product may appear more polished despite varying levels of individual conceptual involvement (Levine et al., 2024; Wang et al., 2024). Table 3 further presents the distribution of Assignment Quality across the sample by assignment format.

Table 3
Descriptive statistics for Assignment Quality (AQ)

Outcome	N	Mean	SD	Min	Max
AQ (overall)	1498	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 scores were measured on a 0–10 scale. **N** represents the number of observations in each category. Because a small number of assignments (**n** = 30) could not be reliably classified as either individual or group submissions, the sum of subgroup counts may not equal the total sample size.

Taken together, these findings suggest that product-level coursework performance was relatively satisfactory across the sample. Yet, as noted in recent contributions on AI-rich assessment, the quality of a submitted assignment does not necessarily imply that students are able to autonomously articulate or reconstruct the understanding embedded in their work (Kasneci et al., 2023; Perkins & Roe, 2024). For this reason, AQ was treated primarily as a measure of product-level performance rather than as a sufficient indicator of underlying understanding.

5.2. Knowledge verification results

Knowledge Mastery was captured through the two RKV methods: RKV-Oral and RKV-Written. The mean score for the oral verification task was 5.88 (SD = 1.42), while the mean score for the handwritten recall task was lower, at 5.21 (SD = 1.50). The average difference between the two measures was 0.67 points, with oral verification consistently producing higher scores. This pattern is substantively important. Oral prompts may allow students to respond through explanation, while real-time reasoning can benefit from interactive cues, prompting, or the step-by-step reconstruction of reasoning during questioning. Written recall tasks, by contrast, require students to generate and organize reasoning more independently, which is often more demanding under conditions that restrict access to mobile devices and other learning supports, as well as under time constraints (Karpicke & Blunt, 2011; Nückles et al., 2020). The lower scores for written recall reflect the greater difficulty many students faced in restating the conceptual structure of their work in a fully self-generated format, rather than responding orally in what may still be perceived as a supported interaction. Descriptive statistics for the two RKV components and the composite Knowledge Mastery indicator are summarized in Table 4.

Table 4
Descriptive statistics for knowledge verification measures

Outcome	N	Mean	SD	Min	Max
RKV–Oral	1498	5.88	1.42	1.40	9.40
RKV–Written	1498	5.21	1.50	1.10	9.10
KM composite (mean of oral & written)	1498	5.55	1.33	1.35	9.05

Note. All measures were scored on a 0–10 scale. The KM composite represents the mean of the oral and written verification scores.

The discrepancy between AQ and KM is evident. While assignment scores clustered in the upper-middle range, knowledge mastery scores were noticeably lower, indicating that for a large proportion of students the quality of the submitted product and their independently demonstrated understanding were not aligned. This pattern is consistent with broader findings in educational psychology suggesting that polished or fluent outputs may overstate actual mastery when processes of explanation and retrieval are not directly engaged (Bjork et al., 2013; Risko & Gilbert, 2016).

5.3. Evidence of discrepancies between output quality and knowledge mastery

o quantify the discrepancy between assignment performance and verified knowledge, the study used the AI-Learning Gap (ALG), defined as:

ALG = AQ − KM

Using the composite KM indicator, the mean AI-Learning Gap was 2.07 (SD = 1.33). The median AI-Learning Gap was 1.92, with the 25th and 75th percentiles at 1.12 and 2.86, respectively. These results indicate that, for a substantial proportion of students, assignment quality tended to exceed the level of knowledge they were able to demonstrate under device-restricted verification conditions. Table 5 presents the main distributional characteristics of the AI-Learning Gap.

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	34.1%

Note. ALG = Assignment Quality − Knowledge Mastery. The more positive this value is, the greater the gap between the presumed quality of submitted coursework and the student's independently demonstrated knowledge.

The study's focus on this gap illustrates the diagnostic value of RKV. Importantly, the presence of an AI-Learning Gap does not demonstrate, nor is it intended to imply, that the use of AI directly causes poor learning or academic misconduct. Rather, it identifies situations in which the apparent quality of an assignment cannot be treated as a reliable proxy for a student's conceptual mastery. These findings are consistent with theoretical perspectives suggesting that polished academic outputs may mask gaps in understanding when assessment relies primarily on produced work rather than on explanation or retrieval-based tasks (Fisher et al., 2015; Holmes & Miao, 2023).

5.4. Classroom feasibility of RKV

Beyond its diagnostic function, RKV also proved workable as a classroom-based assessment strategy. The RKV-Oral component was integrated into presentation sessions and typically required only a small extension of regular class time. The RKV-Written component was implemented as a short, 10-minute in-class handwritten recall task. Because both procedures used formats already familiar in instructional activities and did not require specialized educational technologies, they could be easily incorporated into the normal course flow.

In terms of student responses, RKV appeared to increase accountability regarding how coursework was prepared. Students were expected not only to produce academic work but also to demonstrate ownership of the ideas underlying that work. In this sense, the approach shifted evaluative attention away from surface polish and toward conceptual preparedness. From the instructor's perspective, the workload associated with RKV remained manageable. Oral verification activities could be conducted during scheduled presentations, and written recall tasks were short enough to administer and review without major disruption.

These findings are particularly relevant in discussions about responses to AI use in higher education, many of which are difficult to scale—especially when they rely on unreliable detection systems, intrusive monitoring, or extensive institutional restructuring (Liang et al., 2023; Walters, 2023). In contrast, the present findings suggest that RKV offers a practical, low-tech classroom approach for generating meaningful evidence of student learning. The feasibility of RKV is both pedagogically and logistically reasonable, as it encourages forms of assessment in which students' understanding becomes visible through explanation and recall rather than through the evaluation of polished coursework alone.

6. Discussion

The purpose of this study is to advance the view that the central assessment challenge in AI-rich higher education extends beyond students' use of generative tools to a broader problem: the increasing invisibility of student learning in environments where such tools are widely available. While the sample of submitted assignments demonstrated a moderate level of quality, the results of knowledge verification were considerably lower, revealing a measurable gap between product-level performance and independently demonstrated knowledge. Using the conceptual framework outlined earlier in the study, this gap illustrates the need to move away from detection-focused strategies and toward assessment approaches that capture explanation, memory retrieval, and reasoning. In this regard, Random Knowledge Verification (RKV) represents more than a procedural classroom technique; it functions as an assessment approach aimed at making mastery visible.

6.1. RKV as an alternative to AI detection

A central contribution of this study is the shift in emphasis from AI detection to evidence of learning. Current responses to generative AI in higher education frequently focus on identifying AI-generated text, strengthening anti-plagiarism mechanisms, and imposing restrictions on the ways students produce written work. Although these responses are understandable, prioritizing the detection of AI use rather than the evaluation of students' understanding represents a significant pedagogical limitation. Research on AI detection tools suggests that such systems are often unreliable, may misclassify student writing, and can be circumvented using relatively simple techniques (Liang et al., 2023; Walters, 2023). Even a highly accurate detection tool would still fail to address the more fundamental question of whether a student truly understands the work they submit.

RKV offers an alternative perspective. Rather than attempting to determine whether students used AI, it evaluates whether students can independently articulate and reconstruct the concepts underlying their assignments. Knowledge-centered assessment approaches are therefore better positioned to evaluate what students know, how they reason, and how effectively they command disciplinary concepts, rather than focusing on the mere presence of technological assistance.

Recent discussions among scholars and practitioners have emphasized that assessment reform in the age of AI must address more than questions of authorship. A central challenge is maintaining credible evidence of learning in increasingly technology-supported educational environments. In this regard, RKV aligns with the broader call to shift evaluation away from authorship verification and toward demonstrated understanding (Lodge et al., 2023). RKV also helps address a practical dilemma faced by many educators. Strict prohibitions on AI use are difficult to enforce, while completely unrestricted AI use can undermine the evidential value of assignments. Verification-based assessment directly confronts this tension. It acknowledges that students may use AI during preparation while still requiring them to demonstrate their understanding under conditions that make mastery observable. In this way, RKV may be more compatible with contemporary higher education than strategies that rely primarily on detection or prohibition.

6.2. RKV and mastery-visible assessment

The results of the present study support the broader concept of mastery-visible assessment. In many disciplines, particularly in the social sciences, humanities, and communication studies, assessment relies primarily on written reports, take-home assignments, and presentation slides. These assessment formats, however, increasingly risk conflating refined performance with genuine comprehension. Durable mastery is more reliably evidenced through retrieval, explanation, and conceptual reconstruction rather than through exposure to well-articulated formulations produced by others (Karpicke & Blunt, 2011; Roediger & Karpicke, 2006). The discrepancy observed in this study between Assignment Quality and Knowledge Mastery is consistent with existing literature and supports the argument that product quality should not be treated as a sufficient proxy for learning.

RKV addresses this issue by making learning visible through two complementary modes: oral explanation and written recall. The oral component assesses whether students can think through and articulate the logic of their work in real time, whereas the written component assesses whether they can reconstruct that logic independently in a device-restricted environment. When combined, these forms of evidence reduce the likelihood that assessment rewards rhetorical fluency rather than conceptual understanding.

In this respect, RKV implements a mastery-visible approach where understanding must be demonstrated, not just inferred from the quality of a final product, a concern long emphasized in educational measurement standards regarding the validity of assessment evidence (American Educational Research Association et al., 2014). This is particularly important in AI-assisted contexts, where generative systems can intensify what educational psychology describes as the illusion of

knowledge. Students may experience a heightened sense of mastery when interacting with coherent and authoritative text, even when the underlying concepts are not fully understood (Bjork et al., 2013; Fisher et al., 2015). By requiring explanation and retrieval, RKV disrupts this illusion and enables a more accurate evaluation of what students actually know.

6.3. Implications for higher education assessment

The implications of RKV extend beyond the specific courses examined in this study. First, the method appears to integrate naturally with presentation-based assessments, where oral questioning can be incorporated into normal classroom routines. Verification can be conducted immediately before or after student presentations, allowing instructors to evaluate whether students' explanations genuinely reflect conceptual understanding with minimal disruption to the class flow.

Second, RKV may be particularly useful in research and report-writing assignments, where students are typically expected to synthesize, articulate, and defend their ideas. In such contexts, structured recall tasks or targeted oral questioning can help instructors determine whether students are able to independently articulate the intellectual structure of the work they submit. This becomes especially important in environments where AI tools can assist with drafting, summarizing, and stylistic refinement.

Third, RKV aligns well with seminar-based and discussion-oriented teaching formats. In these contexts, knowledge verification can be integrated into routine academic dialogue and function simultaneously as both formative and summative assessment.

The findings of this study also support the integration of open-tool learning environments with moments of independent verification within hybrid assessment models (Holmes & Miao, 2023). Such models may help institutions balance the pedagogical benefits of AI-supported learning with the need to ensure that evaluation remains aligned with students' conceptual understanding.

6.4. Limitations

Several limitations should be acknowledged. First, the study was conducted within a disciplinary context focused on the social sciences, humanities, and communication studies. These disciplines emphasize explanation, interpretation, and writing, which may make the discrepancy between polished outputs and genuine understanding particularly visible. It is possible that RKV may operate differently in disciplines where assessment relies more heavily on numerical problem-solving, technical construction, or laboratory-based work.

Second, the research design is cross-sectional. The study captures patterns of discrepancy within existing classroom assessment practices rather than changes in learning over time. Consequently, the findings cannot determine whether repeated use of RKV would influence students' learning strategies, strengthen conceptual understanding, or reduce the observed gap between assignment quality and demonstrated knowledge. These questions require longitudinal investigation. Furthermore, the observed AI-Learning Gap should not be interpreted as direct evidence that the use of AI causes reduced learning;

rather, it indicates that the quality of submitted coursework alone may not reliably reflect students' independently demonstrated understanding.

6.5. Future research

Future research should examine how RKV can be adapted to STEM disciplines, where verification logic may differ due to the prominence of problem-solving, symbolic reasoning, and design-oriented tasks. If mastery-visible assessment proves effective in these contexts, its value as a broader strategy for higher education assessment would be further strengthened.

Additional studies should also test the model across multiple institutions and educational systems in order to evaluate both its generalizability and its contextual sensitivity. Another promising direction concerns the integration of RKV with AI-informed learning design. In the present study, RKV was primarily examined as an assessment strategy. However, its presence may also influence how students prepare for assignments, revise their work, and engage in study practices.

Future research could therefore investigate how RKV interacts with AI literacy instruction, reflective AI-use statements, or scaffolded drafting processes. Such studies may clarify the extent to which RKV, combined with responsible AI integration in learning environments, contributes to improved learning outcomes (Ng et al., 2021a,b). In this sense, RKV may play a role not only in assessment reform but also in rethinking how learning itself is structured in an educational landscape increasingly shaped by generative AI.

7. Conclusion

The rapid introduction of generative artificial intelligence into higher education has opened new opportunities for learning, but it has also exposed weaknesses in traditional assessment methods. As students are increasingly able to use large language models to produce essays, reports, and presentations, the quality of submitted work may no longer accurately reflect their actual level of understanding. This places higher education in a position where the central concern is not simply the detection of AI-generated submissions, but whether student learning remains visible and assessable. This study addresses that concern by presenting Random Knowledge Verification (RKV), a scalable classroom-based assessment approach designed to evaluate students' understanding independently. Based on evidence from 38 classes involving 1,498 undergraduate students, the study found that while many assignments appeared to be of relatively high quality, knowledge verification scores were substantially lower. The AI-Learning Gap captures this pattern, suggesting that polished academic submissions may conceal limited understanding and that assessment focused primarily on finished products may fail to capture the quality of students' actual learning.

The study offers three contributions to current discussions on assessment reform in the era of generative AI. First, it shifts the assessment focus from AI detection to learning verification, emphasizing that the central educational issue is whether students can demonstrate understanding of the knowledge

reflected in their work. Second, it shows how mastery-visible assessment can be implemented through oral explanation and written recall as complementary forms of evidence. Third, it demonstrates that RKV can function within existing teaching practices without requiring substantial technological infrastructure, making it usable even in large classes with limited resources. Taken together, these findings support the view that verification-based approaches can help preserve assessment validity in AI-rich educational environments. Rather than attempting to ban emerging technologies, a more constructive response in higher education may lie in designing assessment frameworks that require students to demonstrate understanding even when AI tools are available. In this context, Random Knowledge Verification offers a practical form of assessment flexibility for teaching and learning in the age of generative AI.

Declarations

Funding: This research received no external funding.

Institutional Review Board Statement: The Ethics Committee of Thu Dau Mot University waived review and approval because the study posed minimal risk and used anonymized classroom assessment data collected during regular teaching activities. All procedures involving human participants were conducted in compliance with institutional policy and applicable law and in accordance with the Helsinki Declaration.

Informed Consent Statement: All participants provided informed consent to participate in this study. Participants were informed of the study's purpose, their right to decline participation, and the use of anonymized data for research purposes. To prevent any potential impact on students' academic evaluation, informed consent for the use of anonymized classroom data was obtained after course grades had been released.

Consent to publish: Informed consent for publication was obtained from all participants. The manuscript does not contain any identifiable personal data, and all data used in this study were fully anonymized.

Data Availability Statement: Anonymized records of classroom assessments and knowledge verification tasks conducted as part of regular course activities form the basis of this study's findings. Selected anonymized records are included in the Supplementary Materials. Due to privacy and ethical considerations, the complete dataset is not publicly available; however, the author may consider sharing it upon reasonable request.

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

References

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

2. American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. Standards for educational and psychological testing. AERA; 2014.
3. Baek C, Tate T, Warschauer M. ChatGPT seems too good to be true: College students' use and perceptions of generative AI. Computers Education: Artif Intell. 2024;6:100294. https://doi.org/10.1016/j.caeai.2024.100294.
4. Baig MI, Yadegaridehkordi E. ChatGPT in higher education: A systematic literature review and research challenges. Int J Educational Res. 2024;127:102411. https://doi.org/10.1016/j.ijer.2024.102411.
5. Belland BR. (2013). Scaffolding: Definition, current debates, and future directions. In M. J. Spector, M. D. Merrill, J. Elen, & M. J. Bishop, editors, *Handbook of research on educational communications and technology* (pp. 505–518). Springer. https://doi.org/10.1007/978-1-4614-3185-5_39
6. Biggs J, Tang C. Teaching for quality learning at university: What the student does. 4th ed. Open University Press/McGraw-Hill Education; 2011.
7. Bjork RA, Dunlosky J, Kornell N. Self-regulated learning: Beliefs, techniques, and illusions. Ann Rev Psychol. 2013;64:417–44. https://doi.org/10.1146/annurev-psych-113011-143823.
8. Bobula M. Generative artificial intelligence (AI) in higher education: A comprehensive review of challenges, opportunities, and implications. J Learn Dev High Educ. 2024;30. https://doi.org/10.47408/jldhe.vi30.1137. Article 1137.
9. Chan CKY. A comprehensive AI policy education framework for university teaching and learning. Int J Educational Technol High Educ. 2023;20. https://doi.org/10.1186/s41239-023-00408-3. Article 38.
10. Cohen L, Manion L, Morrison K. (2018). *Research methods in education* (8th ed.). Routledge. https://doi.org/10.4324/9781315456539
11. Cotton DRE, Cotton PA, Shipway JR. Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. Innovations Educ Teach Int. 2023;61(2):228–39. https://doi.org/10.1080/14703297.2023.2190148.
12. Creswell JW, Creswell JD. Research design: Qualitative, quantitative, and mixed methods approaches. 5th ed. SAGE; 2018.
13. Dempere J, Modugu K, Hesham A, Ramasamy LK. The impact of ChatGPT on higher education. Front Educ. 2023;8:1206936. https://doi.org/10.3389/feduc.2023.1206936.
14. Dwivedi YK, Kshetri N, Hughes L, Slade EL, Jeyaraj A, Kar AK, Baabdullah AM, Koohang A, Raghavan V, Ahuja M, Albanna H, Albashrawi MA, Al-Busaidi AS, Balakrishnan J, Barlette Y, Basu S, Bose I, Brooks L, Buhalis D, Carter L, Wright R. So what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. Int J Inf Manag. 2023;71:102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642.
15. Farrokhnia M, Banihashem SK, Noroozi O, Wals A. A SWOT analysis of ChatGPT: Implications for educational practice and research. Innovations Educ Teach Int. 2024;61(3):460–74. https://doi.org/10.1080/14703297.2023.2195846.

16. Fisher M, Goddu MK, Keil FC. Searching for explanations: How the Internet inflates estimates of internal knowledge. J Exp Psychol Gen. 2015;144(3):674–87. https://doi.org/10.1037/xge0000070.
17. Holmes W, Miao F. Guidance for generative AI in education and research. UNESCO Publishing. 2023. https://doi.org/10.54675/EWZM9535.
18. Karpicke JD, Blunt JR. Retrieval practice produces more learning than elaborate studying with concept mapping. Science. 2011;331(6018):772–5. https://doi.org/10.1126/science.1199327.
19. 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, Stadler M, Kasneci G. ChatGPT for good? On opportunities and challenges of large language models for education. Learn Individual Differences. 2023;103:102274. https://doi.org/10.1016/j.lindif.2023.102274.
20. Kornell N, Bjork RA. Learning concepts and categories: Is spacing the enemy of induction? Psychol Sci. 2008;19(6):585–92. https://doi.org/10.1111/j.1467-9280.2008.02127.x.
21. Levine S, Beck SW, Mah C, Phalen L, Pittman J. How do students use ChatGPT as a writing support? J Adolesc Adult Lit. 2024. https://doi.org/10.1002/jaal.1373. Advance online publication.
22. Liang W, Yuksekgonul M, Mao Y, Wu E, Zou J. GPT detectors are biased against non-native English writers. Patterns. 2023;4(7):100779. https://doi.org/10.1016/j.patter.2023.100779.
23. Lodge JM, Howard S, Bearman M, Dawson P, Associates. (2023). *Assessment reform for the age of artificial intelligence*. Tertiary Education Quality and Standards Agency. https://www.teqsa.gov.au/sites/default/files/2023-09/assessment-reform-age-artificial-intelligence-discussion-paper.pdf
24. 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
25. Mai DTT, Da CV, Hanh NV. The use of ChatGPT in teaching and learning: A systematic review through SWOT analysis approach. Front Educ. 2024;9:1328769. https://doi.org/10.3389/feduc.2024.1328769.
26. Michel-Villarreal R, Vilalta-Perdomo E, Salinas-Navarro D, Thierry-Aguilera R, Gerardou F. Challenges and opportunities of generative AI for higher education as explained by ChatGPT. Educ Sci. 2023;13(9):856. https://doi.org/10.3390/educsci13090856.
27. Molenaar I. Towards hybrid human–AI learning technologies. Eur J Educ. 2022;57(4):632–45. https://doi.org/10.1111/ejed.12527.
28. Montenegro-Rueda M, Fernández-Cerero J, Fernández-Batanero JM, López-Meneses E. Impact of the implementation of ChatGPT in education: A systematic review. Computers. 2023;12(8):153. https://doi.org/10.3390/computers12080153.
29. Munaye YY, Admass W, Belayneh Y, Molla A, Asmare M. ChatGPT in education: A systematic review on opportunities, challenges, and future directions. Algorithms. 2025;18(6):352. https://doi.org/10.3390/a18060352.

30. Ng DTK, Leung JKL, Chu KWS, Qiao MS. (2021a). AI literacy: Definition, teaching, evaluation and ethical issues. *Proceedings of the Association for Information Science and Technology, 58*(1), 504–509. https://doi.org/10.1002/pra2.487
31. Ng DTK, Leung JKL, Chu SKW, Qiao MS. Conceptualizing AI literacy: An exploratory review. Computers Education: Artif Intell. 2021b;2:100041. https://doi.org/10.1016/j.caeai.2021.100041.
32. Nückles M, Hübner S, Renkl A. The self-regulation view in writing-to-learn: Using journal writing to optimize cognitive load in self-regulated learning. Educational Psychol Rev. 2020;32(3):753–78. https://doi.org/10.1007/s10648-020-09528-1.
33. Pellegrino JW, Chudowsky N, Glaser R, editors. Knowing what students know: The science and design of educational assessment. National Academy; 2001. https://doi.org/10.17226/10019.
34. Perkins M, Furze L, Roe J, MacVaugh J. The Artificial Intelligence Assessment Scale (AIAS): A framework for ethical integration of generative AI in educational assessment. J Univ Teach Learn Pract. 2024b;21(6). Article 06. https://doi.org/10.53761/jutlp.2024.21.6.06.
35. Perkins M, Roe J. The use of generative AI in qualitative analysis: Inductive thematic analysis with ChatGPT. J Appl Learn Teach. 2024;7(1). Article 1. https://doi.org/10.37074/jalt.2024.7.1.1.
36. Perkins M, Roe J, Vu BH, Postma D, Hickerson D, McGaughran J, Khuat HQ. Simple techniques to bypass GenAI text detectors: Implications for inclusive education. Int J Educational Technol High Educ. 2024a;21(1):53. https://doi.org/10.1186/s41239-024-00584-w.
37. Poe M, Elliot N. Evidence of fairness: Twenty-five years of research in *Assessing Writing*. Assess Writ. 2019;42:100418. https://doi.org/10.1016/j.asw.2019.100418.
38. Rahman MM, Watanobe Y. ChatGPT for education and research: Opportunities, threats, and strategies. Appl Sci. 2023;13(9). https://doi.org/10.3390/app13095783. Article 5783.
39. Risko EF, Gilbert SJ. Cognitive offloading. Trends Cogn Sci. 2016;20(9):676–88. https://doi.org/10.1016/j.tics.2016.07.002.
40. Roediger HL, III, Karpicke JD. Test-enhanced learning: Taking memory tests improves long-term retention. Psychol Sci. 2006;17(3):249–55. https://doi.org/10.1111/j.1467-9280.2006.01693.x.
41. Van de Pol J, Volman M, Beishuizen J. Scaffolding in teacher–student interaction: A decade of research. Educational Psychol Rev. 2010;22(3):271–96. https://doi.org/10.1007/s10648-010-9127-6.
42. Walter Y. Embracing the future of artificial intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education. Int J Educational Technol High Educ. 2024;21(1):15. https://doi.org/10.1186/s41239-024-00448-3.
43. Walters WH. The effectiveness of software designed to detect AI-generated writing: A comparison of 16 AI text detectors. Open Inform Sci. 2023;7(1):20220158. https://doi.org/10.1515/opis-2022-0158.
44. Wang C, Tian Z. (2025). *Rethinking writing education in the age of generative AI*. Taylor & Francis. https://doi.org/10.4324/9781003426936

45. Wang CZ, Aguilar SJ, Bankard JS, Bui E, Nye B. Writing with AI: What college students learned from utilizing ChatGPT for a writing assignment. Educ Sci. 2024;14(9). https://doi.org/10.3390/educsci14090976. Article 976.
46. Williamson DM, Xi X, Breyer FJ. A framework for evaluation and use of automated scoring. Educational Measurement: Issues Pract. 2012;31(1):2–13.