

Research Article

Lecturers' Challenges in Evaluating Student Essays in the Era of Generative AI: A Case Study of the Writing 3 Course at the Faculty of Foreign Languages, Dong Nai University

Thu Ha Vu^{1,*} ¹ Dong Nai University, Vietnam

* Correspondence: vuha.dnu@gmail.com

<https://doi.org/10.59652/4xfr5h65>

Abstract: The emergence of Generative Artificial Intelligence (AI) tools such as ChatGPT and Gemini has created new challenges for writing assessment in higher education. This study investigates the challenges faced by English lecturers in evaluating student essays in the Writing 3 course at the Faculty of Foreign Languages, Dong Nai University. The study focuses on three areas: assessing students' authentic writing ability, differences between take-home and in-class writing performance, and difficulties related to AI policies and detection tools. A mixed-methods descriptive case study was conducted with ten English lecturers through a Likert-scale questionnaire and six semi-structured interviews. Quantitative data were analyzed using descriptive statistics, while qualitative data were examined through thematic analysis. The findings indicate that AI-generated writing often masks students' actual language weaknesses, making it difficult for lecturers to assess genuine writing competence. A significant gap was observed between AI-assisted take-home assignments and supervised handwritten in-class tests. In addition, the absence of an institutional AI policy and the limited reliability of AI detection tools increase grading difficulties and teacher stress. The study recommends increasing the weight of supervised in-class writing, incorporating oral defenses, and establishing clear AI guidelines to enhance assessment validity and academic integrity.

Keywords: Generative AI; writing assessment; academic integrity; EFL lecturers; higher education

1. Introduction

In English Language Teaching (ELT), academic writing has always been a challenging skill for students to master and for educators to evaluate. At the university level, proficiency in writing serves as a key indicator of a student's overall language competence, significantly influencing their academic performance and future career opportunities.

At Dong Nai University, English majors are required to meet high standards in academic writing, particularly in the Writing 3 course. This course is specifically designed to prepare students for the international IELTS Academic Writing examination. Traditionally, lecturers assigned unsupervised take-home essays and drafts to monitor student progress and provide formative feedback for improvement. However, this conventional approach to writing assessment faces a major challenge due to the rapid rise of Generative Artificial Intelligence (AI) tools, such as ChatGPT and Gemini. These advanced AI tools can generate well-structured academic essays and reports within seconds based on simple user prompts. While these tools can support the writing process, their integration requires strict ethical considerations and close teacher oversight to maintain academic integrity (Madhavi, 2025).

While Generative AI can support autonomous learning, its unmonitored use at home diminishes the pedagogical value of take-home assignments, as educators can no longer accurately gauge a student's independent writing ability. This unmonitored use creates serious challenges in foreign language (L2) assessment, making it difficult for teachers to verify work originality and distinguish AI-generated content from authentic student writing (Dau & Lam, 2026). Consequently, writing instructors now face significant difficulties in maintaining academic integrity, distinguishing authentic student work from AI-generated content, and evaluating true linguistic improvement.

Received: April 14, 2026

Accepted: June 21, 2026

Published: June 24, 2026



Copyright: © 2022 by the authors.

Submitted for open access publication

under the terms and conditions of the

Creative Commons Attribution (CC BY)

license

(<https://creativecommons.org/licenses/by/4.0/>).

Within the Faculty of Foreign Languages at Dong Nai University, Writing 3 is recognized as a highly demanding course for second-year English Teacher Education students. Utilizing the textbook *Mindset for IELTS 1* (Crosthwaite et al., 2017), the curriculum requires learners to master both IELTS Writing Task 1 (Data Description) and Task 2 (Social Essay Writing). These tasks demand advanced vocabulary, complex grammatical structures, and coherent argumentation, which places considerable pressure on students. Due to the rigorous nature of the course, an increasing number of students have turned to Generative AI tools to complete their unsupervised take-home assignments.

This shifting dynamic creates serious complications for lecturers during evaluation. Instructors have observed a stark discrepancy between assignments completed at home and those written under supervised classroom conditions. Homework submissions often appear flawless and error-free; however, when the same students sit for monitored, handwritten in-class tests without access to digital devices, their actual language proficiency is noticeably weaker. This mismatch highlights a major risk where a heavy reliance on AI tools makes students dependent, leading to weaker independent thinking and a lack of academic self-discipline (He, 2024).

Furthermore, addressing this issue is highly problematic because Dong Nai University currently lacks an official institutional policy regarding AI plagiarism. At the same time, commercial online AI detection tools remain unstable and inaccurate. This issue becomes more complicated because Generative AI outputs themselves are often unreliable, meaning that tools like ChatGPT and Gemini are not yet dependable enough for academic writing tasks (Garg et al., 2024).

While a growing body of international literature addresses the impact of Generative AI in education, research focusing on the specific assessment challenges and grading anxieties faced by EFL lecturers at provincial universities in Vietnam remains limited. Although recent studies in Vietnam show that structured AI training can improve students' academic writing performance at major universities (Ta et al., 2025), empirical research on the unsupervised challenges faced by instructors in provincial contexts is still limited. Furthermore, while global and regional studies have mapped the wider challenges of the AI era (Dau & Lam, 2026; He, 2024; Madhavi, 2025), empirical data about the practical grading anxieties of writing instructors in provincial Vietnamese higher education remains a distinct gap. This study aims to fill this empirical gap.

This study aims to answer the following three research questions: (1) What are the primary challenges in assessing real student ability faced by English lecturers at Dong Nai University in the Writing 3 course due to Generative AI? (2) To what extent do lecturers observe a difference in student performance across tasks, specifically between unsupervised take-home assignments and monitored, handwritten in-class tests? (3) What difficulties with AI policy and detection tools do lecturers experience when trying to manage AI-generated writing and enforce academic honesty?

2. Literature Review and Conceptual Framework

2.1. *The Writing 3 Course Context*

The Writing 3 course is a required subject for second-year English Teacher Education students at the Faculty of Foreign Languages, Dong Nai University. The main goal of this course is to help students reach the standards of the IELTS Academic Writing exam. This exam has two parts: Task 1 (Report Writing) and Task 2 (Essay Writing). Mastering these two tasks is very important for future English teachers because it shows their English skills and future teaching capacity.

The course follows the *Mindset for IELTS 1* (Crosthwaite et al., 2017) textbook from Cambridge University Press. The syllabus has 8 units divided into two parts. The first part, Units 1-4, focuses on IELTS Task 1 (Report Writing). In the first half of the semester, students learn how to describe visual data like diagrams, maps, bar charts, line graphs, and tables. They practice through 4 topics, particularly, Relationships (Unit 1), Places and Buildings (Unit 2), Education and Employment (Unit 3), and Food and Drink (Unit 4).

The second one include Units 5 to 8. They focus on IELTS Task 2 (Social Essays). In the second half, students learn to write long essays. They practice three types of essays, such as Opinion, Discussion, and Two-part questions (including Problem-Solution). They learn how to open an essay, connect paragraphs, and give examples. The topics include Consumerism (Unit 5), Leisure Time (Unit 6), Fame and The Media (Unit 7), and the Natural World (Unit 8).

2.2. *Challenges in Evaluating Academic Writing in the Era of Generative AI*

Normally, writing courses use homework and essay drafts to give students time to search for information, fix grammar, and arrange ideas. But the fast growth of AI tools like ChatGPT and Gemini makes grading very difficult for teachers.

For IELTS Task 1, students can easily use generative AI tools to transform charts, tables, or other visual data into complete written reports. Because AI-generated texts are often grammatically accurate and well organized, lecturers may find it increasingly difficult to determine whether students can independently analyze information and apply appropriate report-writing conventions. This raises concerns about the authenticity of student work and the validity of writing assessment (Deep & Chen, 2025).

Chen and Gong (2025) reported that AI-assisted writing can substantially improve the quality of students' written products, raising concerns about whether assessment results accurately reflect learners' independent writing competence. This discrepancy challenges the validity of assessments. According to Messick (1989), test scores must accurately reflect the specific construct being measured. When Generative AI produces unearned text, this relationship is disrupted, meaning the results fail to represent the student's independent writing competence. This problem is also supported by Dau and Lam (2026), who showed that separating AI-generated text from real student work is now a major practical challenge for foreign language teachers.

For IELTS Task 2, AI tools can easily write structured essays with many difficult words about topics like Consumerism or Media. It is very hard for teachers to know if the essay is written by the student or AI. Chen and Gong (2025) found that AI-generated support often produces linguistically polished texts, making it more difficult for teachers to distinguish students' independent performance from AI-assisted writing.

In addition, Ray (2023) pointed out some hidden errors of AI text: AI writing often misses traditional styles, changes tone suddenly, and matches Western viewpoints more than Eastern ones. Most dangerously, AI often creates fake academic citations and references (Yan, 2023).

When students submit these essays, teachers have to waste a lot of time. The nice words of AI hide the real weak English of students. Teachers also have to spend time checking if the data and sources are real or fake (Yan, 2023). This widespread use of Generative AI severely threatens academic integrity. As noted by Vangibhurathachhi (2025), the evolution of language models forces educators to face a dual burden: modifying teaching strategies while policing AI-generated plagiarism. This shift necessitates a critical re-examination of institutional policies and detection tools.

Therefore, unsupervised take-home assignments may no longer provide a fully accurate indication of students' actual writing ability, often leading to inflated grades (Lo, 2023) and reducing students' independent thinking (Warner, 2018). This issue becomes worse because Generative AI outputs themselves are often unreliable, meaning that tools like ChatGPT and Gemini are still not dependable enough for serious academic writing tasks (Garg et al., 2024). Furthermore, relying too much on these tools makes students dependent, leading to weaker independent thinking and a lack of self-discipline (He, 2024).

2.3. *Artificial Intelligence, Academic Integrity, and Writing Assessment*

Many researchers in the world have studied how AI affects language learning and testing. Bui and Tong (2025) showed that while AI tools bring numerous benefits to English language practices, they present significant concerns regarding academic integrity. Similarly, Cotton et al. (2024) argued that traditional take-home assessments are heavily challenged because universities face immense difficulties in detecting and preventing AI-driven academic dishonesty. They emphasized that the ease of using tools like ChatGPT outside the classroom creates an urgent crisis for educators, forcing them to completely rethink how to design and grade assessments to minimize cheating.

Nguyen and Barrot (2024) also discussed this challenge in Assessing Writing. They found that teachers often read student essays without immediately suspecting AI involvement. L2 students (people learning English as a second language) usually make basic grammar mistakes or use vocabulary influenced by their mother tongue (Vietnamese). However, AI-generated texts often do not contain these natural learner errors. As a result, teachers may find it increasingly difficult to rely on their intuition when evaluating student writing.

Recently, using too much AI in writing has caused two big problems: homogenization (all essays look the same) and the loss of student identity. According to Sun (2026), AI tools change students' diverse writing into a single, standardized style, making different essays look

very uniform. This flawless AI voice also makes EFL students feel anxious, because they worry that AI will completely erase their own personal style and authentic voice (Yan, 2023).

To address this cheating issue, AI detection tools (like Turnitin) are often used, but they are not reliable. These tools frequently make mistakes by flagging real student work as AI text (false positives). They also fail if students change a few words. Furthermore, institutional preparedness remains a critical issue; as Chan (2023) emphasizes, universities face immense challenges in developing official policies and guidelines for generative AI.

In the absence of clear regulations and reliable detection software, individual teachers experience significant difficulties and stress when grading and trying to protect academic honesty. To handle this problem, recent studies emphasize that higher education needs clear ethical frameworks and active human oversight from instructors to keep evaluation quality (Madhavi, 2025). Therefore, current research suggests focusing on the real daily difficulties of teachers rather than technology. Although some recent studies in Vietnam show that structured AI training can improve students' academic writing performance at major universities in big cities (Ta et al., 2025), there is still a lack of empirical research about this topic in IELTS writing courses at provincial universities. This study will focus on this gap.

2.4. Conceptualizing the Impact of Generative AI on Writing Assessment

The conceptual framework of this study integrates principles of assessment validity with contemporary literature on Generative AI in education. It is designed to examine how AI influences the assessment of writing performance among students at Dong Nai University. At the core of the framework is the challenge of assessing students' actual writing ability in the era of generative AI. According to Messick's (1989) theory of validity, assessment tasks should accurately measure the knowledge and skills that students genuinely possess.

However, the increasing use of AI-assisted writing tools complicates this objective. Recent studies have highlighted that generative AI poses significant threats to academic integrity by facilitating academic dishonesty and obscuring authentic student performance (Vangibhurathachhi, 2025). When students rely heavily on AI-generated content for take-home assignments, their written work may appear grammatically accurate and stylistically sophisticated despite not reflecting their true language proficiency. Consequently, such assignments may lose their validity as indicators of actual writing development.

The framework also considers differences in student performance across assessment contexts. Previous research suggests that discrepancies often emerge between unsupervised take-home assignments and supervised handwritten classroom tests (Nguyen & Barrot, 2024; Lo, 2023). These differences may indicate the extent to which external technological assistance influences students' written production and assessment outcomes.

A further component of the framework concerns institutional and pedagogical challenges associated with AI use. Universities continue to struggle with the development of comprehensive policies governing generative AI in teaching and assessment (Chan, 2023). At the same time, existing AI-detection tools remain unreliable and frequently produce inaccurate results. The absence of clear institutional guidelines, combined with limitations in detection technologies, places substantial pressure on instructors who are responsible for evaluating student work while maintaining academic integrity. These circumstances create additional challenges in ensuring fair, valid, and trustworthy assessment practices.

The framework explains how generative AI affects assessment validity, influences performance differences across assessment formats, and creates institutional challenges related to policy development and academic integrity. These interconnected dimensions provide the theoretical foundation for investigating the impact of AI on IELTS writing assessment at Dong Nai University.

3. Materials and Methods

This study employed a mixed-methods descriptive case study design to investigate lecturers' perceptions and experiences regarding the impact of Generative AI on writing assessment. The use of both quantitative and qualitative approaches enabled a more comprehensive understanding of the phenomenon. Quantitative data were collected through a Likert-scale questionnaire to identify general patterns and trends in lecturers' perceptions, while qualitative data were obtained through semi-structured interviews that provided detailed insights into their experiences, concerns, and assessment practices.

The study involved ten English lecturers from the Faculty of Foreign Languages at Dong Nai University. All participants held at least a Master's degree in TESOL or Applied

Linguistics and had between five and more than fifteen years of experience teaching academic writing. In addition, all lecturers had taught the Writing 3 course using the Mindset for IELTS 1 textbook during the previous two academic years. Their professional experience and direct involvement in IELTS-oriented writing instruction made them well positioned to provide informed perspectives on the influence of Generative AI on writing assessment.

Data were collected through a questionnaire and semi-structured interviews. The questionnaire consisted of fourteen Likert-scale statements developed from the study's conceptual framework. The items examined three major areas: challenges in assessing students' actual writing ability, differences between take-home and in-class writing performance, and difficulties related to AI policies and detection tools. Participants indicated their level of agreement using a five-point scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree). The questionnaire provided an overall picture of lecturers' perceptions regarding the impact of AI on writing assessment.

To complement the survey findings, six lecturers participated in individual semi-structured interviews conducted via Zoom. Each interview lasted approximately 20–30 minutes and was conducted in Vietnamese. With participants' consent, all interviews were audio-recorded for subsequent analysis. The interview protocol included six questions.

The first five questions explored issues directly related to the study's research questions, including the influence of AI on writing assessment, discrepancies between take-home assignments and supervised classroom tests, and challenges associated with AI policies and detection tools. The final question invited participants to provide recommendations for improving writing instruction and assessment practices in the context of increasing AI use.

Data analysis was conducted using both quantitative and qualitative techniques. Given the small sample size ($N = 10$), the quantitative data were analyzed descriptively through the calculation of frequencies, percentages, and mean scores to identify overall patterns of agreement across the fourteen survey items.

The qualitative data were analyzed using the six-step thematic analysis framework proposed by Braun and Clarke (2006). Interview recordings were transcribed and carefully reviewed to identify recurring patterns and themes. The analysis focused on lecturers' perceptions of challenges in assessing students' genuine writing ability, differences between take-home and in-class writing performance, and difficulties associated with AI detection tools and institutional policies. In addition, participants' recommendations for improving the Writing 3 course were examined. These qualitative findings were used to enrich, explain, and contextualize the survey results, thereby providing a more comprehensive understanding of the research problem.

4. Results

4.1. Respondent Information

The first part of the survey collected personal information from the 10 teachers to check their background. Table 1 shows their teaching experience, teaching history of the Writing 3 course, and gender.

Table 1. Background profile of the survey respondents.

Demographic category	Options	Frequency (n)	Percentage (%)
Teaching experience	Under 5 years	0	0.0
	5-10 years	2	20.0
	11-15 years	3	30.0
	Over 15 years	5	50.0
Taught Writing 3 at DNU	Yes	10	100.0
	No	0	0.0
Gender	Male	5	50.0
	Female	5	50.0
	Prefer not to say	0	0.0

Regarding the data presented in Table 1, the demographic profile of the participants shows that the majority of the teachers are highly experienced. Specifically, half of the respondents (50.0%) have been teaching for over 15 years, followed by 30.0% with 11 to 15 years of experience, and the remaining 20.0% having 5 to 10 years in the profession. Notably, there are no teachers with under 5 years of experience. Additionally, there is absolute

uniformity concerning course alignment, confirming that 100.0% of the participating teachers possess direct, practical experience in teaching the Writing 3 course at Dong Nai University.

Finally, the gender distribution of the sample is perfectly balanced, consisting of exactly 50.0% male and 50.0% female teachers. In short, these demographic results demonstrate that the respondents in this study constitute a highly qualified, extensively experienced, and directly relevant group of educators for evaluating the Writing 3 course.

4.2. *The Impact of Generative AI on Writing Assessment*

4.2.1. Challenges in Measuring Authentic Student Performance

Table 2 presents the key findings regarding the challenges of assessing students' real writing ability. First, Item 3 receives the highest average score ($M = 4.30$, $SD = 0.82$). Specifically, 80% of the teachers agree or strongly agree that they are evaluating the capability of AI tools instead of the students' actual skills. Second, 90% of the participants agree that AI creates flawless homework, which hides the students' real grammatical errors (Item 2, $M = 4.20$, $SD = 0.63$). Third, a majority of the teachers (70%) state that it is difficult to verify if students write the assignments by themselves (Item 1, $M = 3.90$, $SD = 0.99$).

Consequently, while a substantial proportion (60%) of respondents expressed agreement or strong agreement that the unmonitored use of Generative AI leads to artificially high grades, this statement also yielded a notable variance in responses (Item 4, $M = 3.70$, $SD = 1.16$). Finally, a cumulative 60% of the surveyed lecturers acknowledged that unsupervised take-home assignments have lost their effectiveness in identifying student errors due to AI interference (Item 5, $M = 3.60$, $SD = 1.07$). This quantitative finding directly accounts for the grading inflation observed during the instructional process, where unsupervised assignments fail to reflect the students' actual independent competence.

Table 2. Challenges in assessing real student ability.

Item	Survey statement	SA (%)	A (%)	N (%)	D (%)	SD (%)	Mean	St. deviation
1.	I find it difficult to check whether take-home assignments are written entirely by the students themselves.	3 (30%)	4 (40%)	2 (20%)	1 (10%)	0 (0%)	3.90	0.99
2.	Generative AI makes student homework look perfect, which conceals their real grammatical errors.	3 (30%)	6 (60%)	1 (10%)	0 (0%)	0 (0%)	4.20	0.63
3.	I feel that I am assessing the capability of AI tools (like ChatGPT/Gemini) rather than the students' own writing skills.	5 (50%)	3 (30%)	2 (20%)	0 (0%)	0 (0%)	4.30	0.82
4.	The use of AI in unsupervised take-home assignments leads to artificially high grades in the Writing 3 course.	3 (30%)	3 (30%)	2 (20%)	2 (20%)	0 (0%)	3.70	1.16
5.	Due to AI interference, unsupervised take-home assignments have lost their effectiveness in identifying student errors.	2 (20%)	4 (40%)	2 (20%)	2 (20%)	0 (0%)	3.60	1.07

The table clearly demonstrates that Generative AI significantly hinders lecturers from evaluating students' authentic writing ability, creating a false impression of academic success through inflated performance indicators.

4.2.2. Difference in Student Performance

Table 3 presents the key findings regarding the difference in student performance across tasks. Item 10 receives the highest average score ($M = 4.20$, $SD = 0.92$), with 70% of the teachers agreeing or strongly agreeing that high grades on unsupervised home tasks no longer

reflect how well students perform in real exams. For Item 9, 80% of the participants agree that take-home essays from different students look very similar in structure and vocabulary, while their monitored in-class tests show clear individual differences ($M = 4.00$, $SD = 0.94$).

Similarly, a majority of the teachers (80%) state that in IELTS Task 2 (Units 5–8), take-home essays show complex arguments, while in-class drafts are brief and repetitive (Item 8, $M = 3.70$, $SD = 1.06$). In IELTS Task 1 (Units 1-4), 60% of the respondents observe that students use advanced trend vocabulary at home but struggle with basic data grouping in class (Item 7, $M = 3.60$, $SD = 0.84$). Finally, 60% of the teachers report a big difference in writing quality between students' take-home tasks and their monitored, handwritten in-class tests (Item 6, $M = 3.50$, $SD = 1.18$).

Table 3. The analysis of difference in student performance across tasks.

Item	Survey statement	SA (%)	A (%)	N (%)	D (%)	SD (%)	Mean	St. deviation
6.	I observe a big difference in writing quality between students' take-home tasks and their monitored, handwritten in-class tests.	2 (20%)	4 (40%)	1 (10%)	3 (30%)	0 (0%)	3.50	1.18
7.	In IELTS Task 1 (Units 1-4), students use advanced trend vocabulary at home but struggle with basic data grouping in class.	1 (10%)	5 (50%)	3 (30%)	1 (10%)	0 (0%)	3.60	0.84
8.	In IELTS Task 2 (Units 5-8), take-home essays show complex arguments, while in-class drafts are brief and repetitive.	1 (10%)	7 (70%)	1 (10%)	0 (0%)	1 (10%)	3.70	1.06
9.	Take-home essays from different students look very similar in structure and vocabulary, but their monitored in-class tests show clear individual differences.	3 (30%)	5 (50%)	1 (10%)	1 (10%)	0 (0%)	4.00	0.94
10.	High grades on unsupervised home tasks no longer reflect how well students perform in real exams.	5 (50%)	2 (20%)	3 (30%)	0 (0%)	0 (0%)	4.20	0.92

Table 3 clearly shows that there is a stark discrepancy between students' performance at home and in class, as unsupervised tasks often display artificial complexity while monitored tests reveal their actual limitations.

4.2.3. Issues Related to AI Policies and Detection

Table 4 presents the difficulties teachers face regarding AI policies and detection tools. Item 13 records the highest mean score ($M = 4.00$, $SD = 1.05$), with 70% of the participants agreeing that the lack of an official university policy on AI plagiarism makes it difficult to handle suspected cases and maintain academic integrity. Similarly, 70% of the teachers agreed or strongly agreed that current AI detection tools are unreliable and may incorrectly classify original student work as AI-generated text (Item 11, $M = 3.70$, $SD = 0.82$).

For Item 14, although 70% of the participants agreed or strongly agreed that they felt stressed when grading suspected AI-generated essays, responses were more varied ($M = 3.60$, $SD = 1.26$), suggesting different levels of grading stress among lecturers. Item 12 records a mean score of 3.80 ($SD = 0.92$), with 70% of the respondents agreeing or strongly agreeing that students can bypass AI detectors through paraphrasing and minor modifications. Overall, the findings indicate that inadequate AI policies and unreliable detection tools create significant challenges for teachers in evaluating student writing.

Table 4. Difficulties with AI policy and detection tools.

Item	Survey statement	SA (%)	A (%)	N (%)	D (%)	SD (%)	Mean	St. deviation
11.	Current online AI detection tools are unstable and frequently mistake original student work for AI-generated text.	1 (10%)	6 (60%)	2 (20%)	1 (10%)	0 (0%)	3.70	0.82
12.	Students can easily bypass AI detectors by making minor changes or paraphrasing the AI text.	2 (20%)	5 (50%)	2 (20%)	1 (10%)	0 (0%)	3.80	0.92
13.	The lack of an official university policy on AI plagiarism at Dong Nai University makes it difficult to handle suspected cases and protect academic honesty.	4 (40%)	3 (30%)	2 (20%)	1 (10%)	0 (0%)	4.00	1.05
14.	I feel stressed when grading suspected AI essays because there is no institutional guidance and detection tools are unreliable.	2 (20%)	5 (50%)	1 (10%)	1 (10%)	1 (10%)	3.60	1.26

4.3. Qualitative Interview Findings

To gain a deeper understanding of the survey results, the researcher selected six out of the ten survey respondents to participate in individual semi-structured interviews. It is important to analyse since the qualitative selection specifically focused on their specialized years of experience in teaching academic writing. This interview sub-sample effectively bridges two distinct instructional profiles based on their writing-specific tenure: developing writing lecturers (having 5 to 10 years of experience) and senior writing experts (having over 15 years of experience). This approach ensures that the qualitative findings capture deep, specialized insights into writing assessment from within the surveyed group.

4.3.1. The Impact of AI on the Quality of Students' Take-home Writing Assignments

Lecturer 1 provided a detailed explanation of this phenomenon: "Generative AI has created a surface-level 'perfection' in take-home assignments, which heavily masks students' actual language proficiency. On one hand, submissions look flawless with sophisticated vocabulary and seamless transitions. On the other hand, this superficial quality conceals their real grammatical errors and syntax weaknesses. It makes it nearly impossible for me to diagnose their true writing problems and provide the necessary pedagogical support until they write in class."

This response suggests that AI-generated writing may compromise one of the central functions of writing assessment – identifying learners' developmental needs. When instructors cannot observe authentic language production, opportunities for targeted feedback and individualized intervention become limited.

Similar concerns were expressed by Lecturer 2: "AI-generated perfection heavily hides students' real grammar mistakes. This makes it very difficult for lecturers to see and evaluate their true language weaknesses."

This finding indicates that AI may reduce the validity of take-home assignments as measures of language proficiency. If students rely heavily on AI-generated content, instructors may struggle to distinguish between students' actual competence and the linguistic output produced by the technology.

Lecturer 3 reinforced this perspective: "AI tools make homework assignments look flawlessly polished. This superficial perfection covers up students' actual grammatical flaws and prevents me from identifying their real writing weaknesses."

The statement highlights a growing assessment challenge in higher education. Although AI-enhanced assignments may appear to demonstrate improved writing quality, they do not necessarily reflect genuine language development. Consequently, instructors may overestimate students' proficiency levels and overlook areas requiring instructional support.

A more concise observation was provided by Lecturer 4: “They just take the most advantage of AI to complete their writing tasks.”

This response suggests that some students may use AI not merely as a support tool but as a substitute for independent writing effort. Such practices can weaken opportunities for skill development, particularly in the areas of grammar, sentence construction, and academic expression.

Likewise, Lecturer 5 noted: “[There is a clear] difficulty in seeing their true language weaknesses.”

This statement further confirms the recurring theme across interviews: Generative AI reduces the transparency of students’ actual writing performance. The inability to identify authentic weaknesses may hinder both formative assessment and the design of effective remedial instruction.

Interestingly, Lecturer 6 offered a more nuanced perspective: “Essays produced by AI are better in terms of surface grammatical features, academic words and phrases, and text organization. However, it is still an ongoing debate whether the essay content is of better quality or not.”

This observation introduces an important distinction between linguistic form and intellectual substance. While AI may improve the surface features of writing, it does not necessarily enhance critical thinking, originality, argumentation, or depth of analysis. This finding suggests that the perceived improvement in writing quality may be largely cosmetic rather than indicative of meaningful academic achievement.

4.3.2. Differences Between AI-Assisted Homework and In-Class Test Scores

Lecturer 1 described the phenomenon in particularly strong terms: “There is a stark and alarming contrast between the two. For take-home tasks, where students have unlimited access to AI tools, many consistently receive high marks like 8/10 or 9/10. However, during monitored, handwritten in-class tests, their scores often drop drastically to 4/10 or 5/10. This massive gap clearly indicates that the high grades on homework reflect the capability of AI tools rather than the students’ autonomous writing skills.”

This statement suggests that AI-assisted assignments may undermine the validity of take-home assessments as indicators of individual learning. The considerable gap between homework and in-class performance implies that students may be receiving credit for work that exceeds their actual writing abilities, thereby complicating instructors’ efforts to accurately measure learning outcomes.

Lecturer 2 reported a similar pattern: “There are big score drops during in-class tests. Students easily get 9/10 for AI-assisted tasks at home, but drop down to 4/10 or 5/10 when writing by hand in class.”

This observation reinforces concerns that access to AI tools may create an artificial enhancement of performance. The consistency of the reported score differences suggests that the issue is not isolated to individual students but may represent a broader shift in assessment outcomes associated with widespread AI use.

Likewise, Lecturer 3 emphasized the magnitude of the discrepancy: “I notice a massive drop in performance. Students can achieve an easy 9/10 for AI-supported homework, but their grades immediately plummet to 4/10 or 5/10 during paper-based tests in class.”

The lecturer’s use of the phrase “immediately plummet” highlights how quickly the apparent proficiency disappears when technological support is removed. This finding indicates that AI-generated assistance may not necessarily translate into improved independent writing skills, raising questions about whether students are learning from the technology or merely relying on it to complete assignments.

The same trend was observed by Lecturer 4: “Big scores drop during in-class tests, generally from 8,9 to 4, 5.” and Lecturer 5: “Big score drops during in-class tests, such as from 9/10 at home down to 4/10 or 5/10 in class.”

These responses further strengthen the pattern identified across interviews. The repeated references to nearly identical score reductions suggest a shared perception among lecturers that AI use is contributing to inflated homework results and widening the gap between assessed and actual performance.

However, an alternative perspective was provided by Lecturer 6: “For me, I can realize whether the homework essays are students’ or AI products. So, it does not affect the way I grade my students either at home or at school, or during test or exam. For me, assessment is a process in which students are assessed on many contributions, such as attendance, active participation, formative assessment and summative assessment, and how they use AI-assisted

tools responsibly. Therefore, there is not much difference regarding grades whether the essays are produced at home or at school. Experienced teachers can find out whether the written essay is produced by students themselves or AI and then give marks correspondingly.”

Unlike the other participants, this lecturer emphasized the importance of process-based and multidimensional assessment practices. Rather than relying heavily on a single written product, the lecturer evaluates students through multiple sources of evidence, including participation, formative activities, and responsible AI use. This approach appears to reduce the influence of AI-generated assignments on final grades and may represent a practical strategy for maintaining assessment fairness in AI-enhanced learning environments.

4.3.3. Challenges in IELTS Task 1 Writing Performance

The interview findings revealed that most participants perceive a strong relationship between extensive AI use in take-home assignments and students’ declining ability to independently analyze, organize, and interpret information during in-class writing tasks. Lecturers consistently reported that while students often produce highly polished homework containing advanced vocabulary and well-structured responses, many struggle to demonstrate fundamental analytical and organizational skills when required to write without AI assistance. These findings suggest that excessive dependence on Generative AI may hinder the development of critical thinking, data interpretation, and independent writing competencies that are essential for successful IELTS Task 1 performance.

Lecturer 1 provided a comprehensive description of these challenges: “When forced to write independently in class, students display severe difficulties in core analytical skills for IELTS Task 1. While their home assignments are packed with advanced trend vocabulary, their in-class performance reveals that they struggle with basic data grouping and logical comparison. They often look confused when analyzing a chart, fail to identify main trends, and frequently make basic grammatical mistakes in sentence structures. This proves they rely too heavily on AI to do the critical thinking and organizing for them at home.”

This response suggests that AI may be compensating for weaknesses in students’ analytical processes rather than helping them develop those skills. Although learners become familiar with sophisticated lexical items and reporting language through AI-generated texts, they may not fully understand how to interpret visual data, identify patterns, or organize information independently. Consequently, the apparent quality of take-home assignments may conceal significant deficiencies in higher-order cognitive skills.

Lecturer 2 expressed a similar concern: “Students rely too much on AI at home. As a result, they show clear difficulty and confusion when they have to analyze data and build sentence structures independently without AI support in class.”

This finding indicates that students may become dependent on AI-generated guidance for both content development and language production. When such support is removed, they encounter difficulties not only in organizing information but also in constructing grammatically accurate and coherent responses.

Likewise, Lecturer 3 highlighted the effects of over-reliance on AI on independent performance: “Due to over-reliance on AI at home, students struggle significantly with independent work. They become visibly confused when grouping chart data and lack the ability to organize paragraph structures without AI assistance.”

The lecturer’s observation points to a potential weakening of essential academic writing skills. Rather than learning how to synthesize information and develop logical text structures themselves, some students appear to depend on AI to perform these cognitive functions, resulting in reduced confidence and competence during supervised writing tasks.

This concern was further echoed by Lecturer 4: “Clear difficulty or confusion when analyzing data and building structures independently without AI support in class.” and indirectly reinforced by Lecturer 5: “Students relying too much on AI at home.”

These statements strengthen the emerging theme that AI dependence may limit opportunities for students to practice independent reasoning, data interpretation, and organizational skills. The consistency of these responses suggests that lecturers perceive the issue as widespread rather than isolated to individual learners.

However, Lecturer 6 offered a more balanced interpretation of the situation: “For students who are doing IELTS Task 1, the most difficult part among students is to read, understand, and interpret precisely the visual data, and summarize the data in a way that can make readers see the connectedness among the data reported. As inexperienced writing students, when writing Task 1 in class, they also encounter difficulty in terms of structuring their essays regarding introduction, overview, and body development. AI can help students

so much in easing these difficulties. By synthesizing the information from several sources through algorithm, students can produce perfect report essays. For me, if students know how to make wise use of AI for Task 1, it is a plus point, not a drawback. They still can write Task 1 and Task 2 successfully for test and actual exams without AI support. The most important thing is that teachers should train students how to use AI wisely and responsibly, as we cannot stop students from using AI these days. So, teach them to be masters of AI so that they still write essays and research effectively even when AI is not with them when they need.”

This lecturer viewed AI primarily as a supportive educational tool rather than a source of dependency. The response acknowledges that data interpretation and essay organization have traditionally been difficult areas for IELTS learners and suggests that AI can serve as a scaffold that facilitates learning when used appropriately. This perspective shifts the focus from restricting AI use to developing students’ AI literacy and teaching them how to engage critically and responsibly with AI-generated assistance.

4.3.4. Patterns of AI-Generated Writing and Detection Challenges

Lecturer 1 explained: “I have noticed highly repetitive patterns across different students’ essays, such as identical paragraph structures, predictable transition words, and shared AI templates. Regarding AI detectors, I find them highly unreliable and unstable. They generate frequent false positives while being easily bypassed by students who use minor paraphrasing tools or make slight manual edits to the AI text.”

This response suggests that AI-generated writing may contribute to the homogenization of student work, reducing stylistic diversity and individual expression. It also highlights concerns about the fairness of relying on AI detection tools in academic assessment.

Similar observations were made by Lecturer 2: “Many essays look copy-pasted or are clearly written from the same AI templates. Meanwhile, current AI detectors are highly unreliable because they give false reports and can be easily bypassed by students.” and Lecturer 3: “Multiple submissions look highly repetitive, sharing the exact same paragraph templates and robotic transitions. Furthermore, AI detection software is highly untrustworthy since it frequently flags original work and is simple for students to bypass.”

These comments indicate that instructors are increasingly encountering standardized writing patterns that may signal AI assistance, while simultaneously lacking confidence in technological methods designed to identify such use.

The issue of uniformity was further emphasized by Lecturer 4: “A large number of student submissions appear to be directly copy-pasted or constructed from the exact same pre-determined AI templates.” and Lecturer 5: “Many essays look highly uniform, as if they were heavily plagiarized or generated from identical artificial intelligence frameworks.”

Together, these responses suggest that AI-assisted writing may encourage dependence on common templates and reduce originality in students’ written work.

Lecturer 6 offered a somewhat different perspective: “There are many similarities you can realize whether your students use AI or not. For example, you can see the consistency from the beginning to an end of the essay, perfect grammar, condense information, error-free spelling and punctuation, salient hyphen, repetitive contents elsewhere, and academic words or phrases. Unlike AI writing, human writing can be affected by pauses, a few grammatical mistakes, inconsistency, misspelt words or phrases, or pronunciation, conversational, discourses and so on. With experience in teaching writing and reading and observing students throughout the whole semester, you can realize this difference.”

Rather than relying on detection software, this lecturer emphasized the value of professional judgment and long-term familiarity with students’ writing patterns. The response suggests that experienced instructors may be better positioned to identify AI-generated content through observation and continuous assessment.

4.3.5. Challenges in Addressing Suspected AI Plagiarism

Lecturer 1 described these challenges in detail: “The lack of an official institutional policy creates immense grading stress and ethical dilemmas. When I suspect an essay is AI-generated, I have no legal or official framework to penalize the student, especially if they strongly deny it. Proving AI plagiarism is incredibly difficult due to the unreliability of detection tools. Without clear university guidelines, I feel isolated and stressed, caught between protecting academic honesty and avoiding unfair accusations.”

This response highlights the uncertainty lecturers experience when evaluating potentially AI-generated submissions. The absence of formal regulations appears to weaken instructors’ confidence in making disciplinary decisions and increases concerns about fairness and

consistency in assessment practices.

Similar concerns were expressed by Lecturer 2: “I face high grading stress due to the lack of official university rules. It is extremely difficult to penalize suspected cheating when students strongly deny using AI.” and Lecturer 3: “It causes a lot of anxiety because we have no official institutional regulations to back us up. It is almost impossible to enforce penalties when a student flatly denies using AI tools.”

These comments suggest that lecturers feel unsupported when handling suspected cases of AI misuse. Without clear procedures, proving misconduct becomes difficult, particularly when students challenge accusations.

The same issue was reinforced by Lecturer 4: “I experience significant grading stress due to the absence of official institutional rules, which makes it extremely difficult to penalize suspected cheating, especially when students flatly deny their reliance on AI tools.” and Lecturer 5: “I feel very stressed during the evaluation process because our university has not yet established an official AI policy. Consequently, enforcing penalties becomes highly challenging whenever students deny using these generative tools.”

Together, these responses indicate a strong consensus that the lack of institutional guidance creates anxiety, inconsistency, and uncertainty in the assessment of AI-assisted work.

In contrast, Lecturer 6 emphasized the importance of continuous assessment and professional expertise: “As said above, I rely on many methods to assess students’ essay performance. Classroom observation, students working in pairs, groups, or alone can help you realize whether they actually commit plagiarism. Once again, assessment is a process; therefore, it is not a one-shot assessment. Teacher, with patience, good training and experience in AI technology can realize whether students’ essays are theirs or assisted by AI technology. While we wait for a transparent university policy concerning plagiarism, we should equip ourselves with knowledge of AI to help our students use AI responsibly.”

Rather than focusing on punitive measures, this lecturer advocated for process-based assessment and greater teacher competence in AI literacy. The response suggests that instructor expertise and ongoing observation may help address some challenges while institutions work toward developing formal policies.

4.3.6. Strategies for Managing Generative AI in Writing Assessment

The interview findings revealed a strong consensus that assessment practices must be adapted to address the growing influence of Generative AI on student writing. Participants recommended increasing the emphasis on supervised assessment, redesigning writing tasks, and incorporating verification methods that allow instructors to confirm students’ authorship and understanding of their work.

Lecturer 1 emphasized the importance of adjusting grading practices: “Shift the grading weight: Reduce the percentage of grades allocated to take-home assignments and increase the weight of monitored, handwritten in-class writing tasks.”

This recommendation suggests that greater reliance on supervised writing activities may provide a more accurate measure of students’ actual language abilities and reduce the impact of AI-assisted work on final grades.

Similarly, Lecturer 2 proposed several complementary strategies: “Suggest three main strategies: Changing exam formats to focus more on monitored, in-class writing. Modifying essay prompts to limit the accuracy of generic AI responses. Introducing quick oral defenses (viva) where students must explain their own writing.”

The response highlights the need for assessment methods that evaluate not only the final written product but also students’ understanding of the ideas and arguments presented in their work.

Lecturer 3 expressed a similar view: “Increase the weight of supervised, in-class handwritten essays. Create highly customized or localized writing prompts to restrict AI capability. Use short face-to-face interviews to double-check students’ authorship of their work.”

This finding suggests that authentic, context-specific tasks may be more resistant to generic AI-generated responses and better able to reveal students’ independent writing skills.

The importance of assessment redesign was also emphasized by Lecturer 4: “Changing exam formats (e.g., more in-class writing); modifying essay prompts to limit AI response accuracy.” and Lecturer 5: “Changing exam formats.”

These responses further support the idea that traditional take-home writing assignments may require modification in order to maintain assessment validity.

Lecturer 6 expanded the discussion by focusing on oral verification, assessment transparency, and teacher preparedness: “First, we can supplement oral defenses as you suggested. Through oral defenses, teachers can find out whether the essays are their products or the products of AI. To some certain extent, the way students explain or defend their essays orally can demonstrate their real ability of writing... secondly, I think we should have a concrete rubric for writing in which each writing criteria is defined and explaining clearly to students... Last but not least, teachers of writing should arm themselves with the knowledge of many sorts of AI software industries and pilot themselves with these technologies so that we can see the strengths and weaknesses of each AI.”

This perspective highlights the importance of combining assessment innovation with teacher professional development. Familiarity with AI tools may enable instructors to design more effective tasks and make more informed judgments about student work.

By combining the quantitative survey percentages with the qualitative interview themes, this study highlights three main findings. First, unsupervised take-home assignments may no longer provide a fully accurate indication of students’ actual writing ability because AI-driven perfection hides their grammatical and structural weaknesses. This is statistically supported by the 60% of lecturers who agreed that AI leads to artificially high grades, and the majority of interview participants reported a clear difference between take-home assignments and supervised classroom tests. Homework scores were often reported to be around 8-9/10, while paper-based test scores frequently dropped to 4-5/10, suggesting that many students experienced difficulties performing writing tasks independently without AI support. Second, current online AI detectors are considered too unstable and prone to false positives to be reliable. This technical challenge is compounded by an institutional policy vacuum at Dong Nai University, which increases grading stress and creates ethical challenges for instructors. Finally, to address these issues, the faculty recommends a structural adjustment to the Writing 3 course. The consensus favors increasing the grading weight of monitored, handwritten in-class tasks, designing localized or customized writing prompts, and introducing short oral defenses to verify true student authorship.

5. Discussion

The study results show that Generative AI tools like ChatGPT and Gemini create serious challenges for English lecturers in the Writing 3 course at Dong Nai University.

First, regarding challenges in assessing real student ability, there is a strong agreement among lecturers ($M = 4.30$ for Item 3) that they are grading AI capabilities rather than students’ actual skills. The interviews confirm that AI tools create a “surface-level perfection” that hides students’ real grammar and vocabulary errors. This matches Nguyen and Barrot (2024), who found that AI-generated texts often do not contain the natural errors of L2 learners, making the teacher’s intuition less effective in evaluation. This diagnostic challenge also aligns with Dau and Lam (2026), who showed that separating AI text from real student work is now a major practical obstacle in modern language classrooms. Furthermore, for IELTS Task 1, lecturers noted that students face confusion and difficulty when analyzing data or grouping trends in class without AI support. These findings align with Chen and Gong (2025) regarding how AI masks true language proficiency, and strongly support Warner (2018) in that over-reliance on technology prevents learners from developing independent writing skills and autonomous analytical competence. This operational difficulty also supports the argument of He (2024) that relying too much on AI writing tools makes students dependent, leading to weaker independent thinking and a lack of self-discipline.

Second, a clear performance gap exists between take-home tasks and monitored, handwritten in-class tests ($M = 3.50$ for Item 6). Students easily get high grades (8/10 or 9/10) at home, but their scores drop drastically to 4/10 or 5/10 in exam conditions. This grade drop happens because students cannot use technology in class. This result is consistent with Cotton et al. (2024), who argued that traditional take-home assessments are heavily challenged because universities face immense difficulties in detecting and preventing AI-driven academic dishonesty outside the controlled classroom environment. However, a counter-perspective from Lecturer 6 suggests that if teachers use a process-based continuous assessment (including attendance, participation, and oral checks), the negative impact of AI on grades can be minimized. This recommendation matches the findings of Ta et al. (2025) in Vietnam, which show that structured guidance from teachers is essential for students to use AI effectively and ethically.

Finally, the lack of an official university policy on AI plagiarism ($M = 4.00$ for Item 13)

and the unreliability of online AI detectors ($M = 3.70$ for Item 11) create high grading stress ($M = 3.60$ for Item 14) for instructors. Teachers feel anxious because they cannot penalize cheating when students deny using AI. This anxiety is validated by Garg et al. (2024), who confirmed that tools like ChatGPT and Gemini are not yet reliable for academic writing tasks, which underscores the clear need for close teacher oversight to maintain evaluation quality (Madhavi, 2025). This situation is associated with a high degree of similarity among student essays, where submissions often display comparable structures, vocabulary choices, and organizational patterns ($M = 4.00$ for Item 9). Similar concerns regarding the standardized nature and homogenization of AI-generated texts have been noted by Sun (2026).

The lecturers' recommendations have several important practical implications for the Faculty of Foreign Languages. A key recommendation is the restructuring of assessment practices by reducing the grading weight assigned to take-home assignments and increasing the emphasis on supervised, paper-based in-class writing tasks, including both continuous assessments and formal examinations. Another recurring suggestion involves redesigning writing prompts to incorporate localized contexts, personal experiences, and customized topics that are less vulnerable to generic AI-generated responses. Participants also advocated the use of complementary verification methods, such as brief oral defenses or follow-up interviews, which would enable instructors to confirm students' authorship and assess their understanding of the ideas presented in their written work.

This study has two limitations. First, the sample size was relatively small, with only 10 survey participants and 6 interviewees from one provincial university. However, the sample reflected the limited number of lecturers responsible for teaching academic writing courses within the research setting. Therefore, the findings should be interpreted within the specific context of Dong Nai University and may not be generalized to other institutions. Second, the study focused only on lecturers' perspectives. Future research could examine students' perceptions and actual AI-use practices to provide a more comprehensive understanding of the impact of Generative AI on academic writing assessment.

6. Conclusions

In conclusion, this mixed-methods descriptive case study investigated the challenges faced by English lecturers in evaluating student essays in the GenAI era, focusing on the Writing 3 course at Dong Nai University.

The findings reveal that unsupervised take-home assignments may no longer provide a fully accurate indication of students' actual writing ability. AI tools create a "surface-level perfection" that hides students' real grammar and vocabulary errors. The majority of interview participants reported that homework scores often ranged from 8-9/10, whereas in-class test scores frequently dropped to 4-5/10. Additionally, the lack of an official university policy on AI plagiarism combined with unstable technical detection tools places a heavy emotional and ethical burden on instructors, who face high grading stress and encounter highly repetitive essay structures from students.

To maintain academic honesty and ensure real language development, the study recommends a necessary shift in the Writing 3 assessment framework. Faculty members should reduce the grading weight of take-home tasks and focus more on supervised, handwritten in-class writing. Furthermore, teachers need to design customized, localized writing prompts and adopt interactive methods like brief oral defenses to verify true student authorship. Ultimately, while waiting for official university regulations, writing lecturers must actively equip themselves with AI knowledge to guide second-year English majors toward using these advanced technologies responsibly and ethically.

Funding: This research received no external funding.

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

References

- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. <https://doi.org/10.1191/1478088706qp063oa>
- Bui, T. T. U., & Tong, T. V. A. (2025). The impact of AI writing tools on academic integrity: Unveiling English-majored students' perceptions and practical solutions. *AsiaCALL Online Journal*, 16(1), 83-110.
- 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, 38. <https://doi.org/10.1186/s41239-023-00408-3>

- Chen, C., & Gong, Y. (2025). The Role of AI-Assisted Learning in Academic Writing: A Mixed-Methods Study on Chinese as a Second Language Students. *Education Sciences*, 15(2), 141. <https://doi.org/10.3390/educsci15020141>
- Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61(2), 228-239. <https://doi.org/10.1080/14703297.2023.2190148>
- Crosthwaite, P., Hutchison, S., Wijayatilake, C., & De Souza, N. (2017). *Mindset for IELTS Level 1: Student's book*. Cambridge University Press.
- Dau, T. T. L., & Lam, T. T. (2026). Assessing L2 Student Writing in the AI Era: A Systematic Review on Challenges and Best Practices. *International Journal of TESOL & Education*, 6(1), 72-93. <https://doi.org/10.54855/ijte.26615>
- Deep, P. D., & Chen, Y. (2025). The Role of AI in Academic Writing: Impacts on Writing Skills, Critical Thinking, and Integrity in Higher Education. *Societies*, 15(9), 247. <https://doi.org/10.3390/soc15090247>
- Garg, S., Ahmad, A., & Madsen, D. Ø. (2024). Academic writing in the age of AI: Comparing the reliability of ChatGPT and Bard with Scopus and Web of Science. *Journal of Innovation & Knowledge*, 9(4), 100563. <https://doi.org/10.1016/j.jik.2024.100563>
- He, M. (2024). The impact and challenges of AI on students' writing skills – Take ChatGPT for example. *SHS Web of Conferences*, 220, 04004. <https://doi.org/10.1051/shsconf/202522004004>
- Lo, C. K. (2023). What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. *Education Sciences*, 13(4), 410. <https://doi.org/10.3390/educsci13040410>
- Madhavi, V. (2025). Academic Writing in the Age of AI: Opportunities, Challenges, and Best Practices. *Trends in Scholarly Publishing*, 4(1), 8-14. <https://doi.org/10.21124/tsp.2025.08.14>
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (pp. 13-103). New York, NY: American Council on education and Macmillan.
- Nguyen, L., & Barrot, J. S. (2024). Detecting and assessing AI-generated and human-produced texts: The case of second language writing teachers. *Assessing Writing*, 62, 100899. <https://doi.org/10.1016/j.asw.2024.100899>
- Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*, 3, 121-154. <https://doi.org/10.1016/j.iotcps.2023.04.003>
- Sun, Y. (2026). Resisting Homogenization in EAL Writing Education: Translingual Practices in the Rise of Generative AI. *TESOL Quarterly*. <https://doi.org/10.1002/tesq.70063>
- Ta, T. T. H., Bo, T. L., Khuong, Q. N., & Nghiem, H. P. (2025). Impacts of artificial intelligence student guide on academic writing among third-year students at the Faculty of International Education, VNU University of Languages and International Studies. *VNU Journal of Foreign Studies*, 41(4), 40-53. <https://doi.org/10.63023/2525-2445/jfs.ulis.5489>
- Vangibhurathachhi, S. K. (2025). The impact of generative AI on educational assessment. *International Journal of Innovative Research in Engineering & Multidisciplinary Physical Sciences*, 13(4). <https://doi.org/10.37082/IJIRMP.v13.i4.232678>
- Warner, J. (2018). *Why they can't write: Killing the five-paragraph essay and other necessities*. Johns Hopkins University Press.
- Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Education and Information Technologies*, 28, 13943-13967. <https://doi.org/10.1007/s10639-023-11742-4>