

Evaluating L2 Learners' Experience with GenAI-Powered Academic Reading Tool in Higher Education: A Small-Scale Exploratory Study

Haoming Lin¹, Wei Wei^{1*}, Handan Lu²

¹ Faculty of Applied Sciences, Macao Polytechnic University, Macao, China

² The Assessment and Qualifications Alliance (AQA), Surrey, UK

*Corresponding author's email: weiweitestng@hotmail.com

 <https://orcid.org/0000-0002-1919-3178>

®Copyright (c) 2025 Haoming Lin, Wei Wei, Handan Lu

Received: 22/01/2025

Revision: 28/07/2025

Accepted: 30/07/2025

Online: 28/10/2025

ABSTRACT

Keywords:

Generative AI,
Academic Reading,
Learner Experience

Generative artificial intelligence (GenAI)- powered tools have been widely used to enhance language learning skills. However, little is known about their role in shaping learners' experiences during academic reading. To address this gap, our study engaged twelve postgraduate students in evaluating reading reports generated by Fullpicture, an academic reading tool powered by ChatGPT, over seven weeks. Students reflected on their experiences using the GenAI tool for academic reading and rated its effectiveness on a 6-point Likert scale. Based on the three-level model of reading comprehension, learners' reported benefits of the tool include the literal level, which encompasses the facilitation of quick comprehension and the reduction of cognitive load; the inferential level, which includes the interpretation of the author's tone and purpose and the reinforcement of personal beliefs; and the critical level, which involves the promotion of reflective reading practices and the generation of new perspectives for academic writing. The perceived challenges can also be categorized into three levels: selective summarization and misrepresentation of details at the literal level; overgeneralization of interpretations and irrelevance to personal reading and writing goals at the inferential level; and a lack of practical suggestions and weak evidence to support critiques at the critical level. The analysis of questionnaire results also revealed the supporting role of GenAI-powered tools for L2 learners in evaluating arguments, evidence quality, or the generalizability of research; however, their limited effectiveness was also identified, especially concerning the adequacy of discussion and theoretical frameworks. The study suggested that learners should collaborate with AI instead of fully relying on it to support academic reading.

Introduction

Academic reading posed high requirements for L2 learners, as it usually involves frequent use of cognitive resources to evaluate different ideas before making well-informed decisions (Goldman, 2024; Lin et al., 2025; Liu et al., 2021), and, along with this, linguistic barriers and

cross-discipline knowledge requirements also hindered L2 learners' sustained academic development (Green, 2009). In recent years, the emergence of Generative Artificial Intelligence (GenAI)-powered tools has attracted academic interest in supporting L2 reading comprehension by reducing cognitive load (Feng, 2024; Srinivasan & Murthy, 2021). However, there remains a gap regarding whether these tools can help L2 learners to overcome the challenges in academic reading, therefore making it particularly crucial to investigate their experiences and perceptions of GenAI assistance.

A review of existing studies identified three primary concerns. First, although previous research has explored L2 learners' experiences with GenAI-powered tools in general language learning (Jeon, 2024), writing (Kim et al., 2024), speaking (Zou et al., 2024), there has been limited attention to tracing students' experiences specifically in GenAI-assisted academic reading contexts. Second, most studies on GenAI-assisted reading tools focus on short-term interventions, such as one-week or four-week durations, to compare the effects of AI integration versus non-AI scenarios. However, as noted by previous researchers (Jeon, 2024; Lin et al., 2024; G. L. Liu et al., 2024; Monib et al., 2025), it is equally important to first understand the benefits and challenges that AI brings to learning in order to avoid potential pitfalls when using GenAI. Third, despite the increasing adoption of GenAI tools, ethical concerns, such as plagiarism and the potential hindrance of higher-order thinking due to the direct provision of answers, have sparked heated debates. These concerns might prevent learners from expressing their personal experiences of utilizing GenAI in academic contexts (Wang, 2024). This silence also poses a threat to educators and policymakers in listening to learners' voices and designing curricula that appropriately balance the reasonable use of GenAI tools with meeting actual learner needs. This study aimed to address these gaps by adopting a mixed-methods approach. First, we applied the three-level reading comprehension framework and conducted a qualitative analysis to explore postgraduate learners' reported benefits and challenges of GenAI-powered academic reading. Second, we conducted a quantitative analysis based on a survey investigating students' perceptions of which dimensions of academic reading were most beneficial to them.

Literature review

The three-level model of reading comprehension

Reading comprehension is a complex cognitive process in which readers decipher meaning from written work (Hudson et al., 2008; Snow, 2002). To achieve effective reading, it usually involves not only reader itself and also include interactions with the text and the reading context (Rouet et al., 2017). Prior researchers have emphasized the hierarchical nature of reading comprehension and have deconstructed it into three levels (Basaraba et al., 2013; de-la-Peña & Luque-Rojas, 2021; Herber, 1978; Saadatnia et al., 2016): literal comprehension, which is the basic level of reading and often involves understanding, extracting, and summarizing explicit information in a text; inferential or interpretative comprehension, which requires readers to "read between the lines," including the interpretation of implicit information, constructing connections among multiple ideas in the text, and making comparisons with prior knowledge (Alptekin, 2006; Basaraba et al., 2013); and critical or evaluative comprehension, which

demands the highest cognitive resources, as it requires readers to apply higher-order thinking after a thorough evaluation of the main arguments of the text and to make critical judgments about the reliability and quality of the information (Basaraba et al., 2013; Wallace & Wray, 2021).

After taking a quick glance at the theoretical framework of reading comprehension, in the next section we will briefly review the challenges faced by L2 learners, who often differ from first language learners.

Challenges in L2 learners' reading comprehension

Although researchers, like Jeon and Yamashita (2014) and Pretorius and Spaull (2016), have pointed out the importance of reading for L2 learners, a review of the existing literature reveals several factors that hinder their reading development. These challenges can be understood through the lens of reading comprehension model, which divides reading into three levels: (1) literal, which includes limited language proficiency and linguistic barriers (Alptekin, 2006; Alptekin & ErÇetin, 2011), (2) inferential, which involves inactivation of context or background knowledge, and cultural and educational differences (Elbro & and Buch-Iversen, 2013). and (3) critical comprehension, which entails high cognitive demands, posing challenges such as evaluating arguments and assessing authorial intent (Wang & Gierl, 2011). Across all three comprehension levels, learners' insufficient use of metacognitive reading strategies, defined as readers' awareness and control of their own reading processes such as planning, monitoring, and evaluating (Sheorey & Mokhtari, 2001; Taki, 2016), can further impede effective comprehension (Johnson et al., 2010; Ku & Ho, 2010).

First, regarding literal comprehension, limited language proficiency poses an obvious challenge for L2 learners when they encounter linguistic issues, which may hinder their basic-level comprehension prevent them from progressing to the next level of comprehension.. For instance, Yüzlü and Dikilitaş (2022) found that some students could not fully comprehend the reading materials because of limited English proficiency even when multimodal texts were utilized to support the translanguaging classroom.

Second, for inferential comprehension, failure to activate background knowledge and cultural value pose additional difficulties for L2 learners. For example, in their case study investigating young learners' reading of multimodal texts about global warming, Fazio et al. (2022) found that learners faced challenges in interpreting content about coral reef deterioration because they lacked the relevant background knowledge to support inference-making. L2 learners also frequently encounter difficulties in activating prior knowledge during academic reading, especially when processing unfamiliar or culturally different texts. Similarly, Luk and Hui (2017), whose study focused on adolescent readers, indicated that the Confucian cultural background significantly impacted Hong Kong students' reading of English advertisements, as they often relied on established facts and were limited in exploring alternative interpretations. This suggested that cultural values emphasizing obedience to authority may hinder learners' deeper understanding of texts, since they avoid contradicting higher-level authorities.

Third, regarding critical comprehension, the major challenge for L2 learners lies in the relatively high cognitive demands required during critical reading. For example, Kiili et al.

(2018) found that Finnish primary school students felt overwhelmed when asked to evaluate a health-related text from multiple perspectives in digital format. This task imposes an even greater burden on L2 learners, as they must first overcome language barriers before engaging in higher-order thinking processes. Furthermore, the insufficient use of metacognitive reading strategies can negatively affect the development of critical thinking. This is illustrated by Ku and Ho's (2010) study, which showed that low-performing readers in critical reading tasks rarely employed metacognitive strategies, resulting in limited critical comprehension.

AI and reading

Although L2 learners face great challenges in reading, the recent advancements in GenAI such as ChatGPT, Deepseek, and other GenAI-powered tools have brought new possibilities for supporting learners in overcoming these barriers and achieving deeper levels of reading comprehension.

First, regarding literal comprehension, GenAI-powered tools can offer summarization functions or provide feedback on learners' vocabulary questions, which help L2 learners to understand the explicit meaning of the text more effectively. For example, Yang et al. (2021) employed BERT, which served as an advanced text summarization tool, to extract key concepts from learning materials and grade students' markings. They suggested that the AI-assisted marking function effectively helped students identify and focus on key information and therefore enhanced their literal understanding of the text. Toyokawa et al. (2023) performed a case study using the LEAF system with two students in inclusive education settings and indicated that AI-assisted active reading tasks, by analyzing learning logs and providing personalized feedback, helped students improve their literal understanding by detecting learning difficulties and enhancing their focus on explicit information in texts. Rees and Lew (2024) conducted an experimental study with 43 L2 English students using AI-generated definitions from OpenAI GPT-3 in a lexically oriented reading task and found that the AI-assisted definitions effectively supported students' literal understanding by helping them grasp the meaning of unfamiliar words in context.

Second, in terms of inferential comprehension, GenAI can scaffold learners' understanding by activating and integrating prior knowledge, offering contextual as well as and tailoring explanations to scaffold them overcoming cultural or educational gaps. M. Liu et al. (2024) carried out a user study with 59 undergraduate students utilizing the CoAsker tool, which leverages AI-driven question cocreation, and found that AI-assisted question generation improved students' inferential comprehension by supporting reflective revision of questions to enhance inferential skills and by facilitating repeated activation of their knowledge through cocreated question banks. Cheng et al. (2024) implemented a study on Chinese kindergarteners from underprivileged backgrounds to assess the effects of chatbot-led dialogic reading on story comprehension. Using AI to deliver inferential questions targeting causal reasoning, predictions, and themes, combined with tailored feedback mechanisms, the study revealed that AI significantly enhanced inferential understanding and providing adaptive scaffolding.

Third, for critical comprehension, GenAI tools can guide learners in evaluating arguments, synthesizing diverse perspectives, and applying metacognitive reading strategies through

personalized feedback and interactive prompts. Peng et al. (2022) investigated how CReBot, an AI-driven interactive tool, supports routine paper readers and novices in critically engaging with scientific papers. Through interactive prompts and tailored hints, CReBot facilitated deeper critical analysis by encouraging users to evaluate arguments, identify weaknesses, and reframe the content in meaningful ways, demonstrating its potential to enhance critical understanding in academic reading. Cheung et al. (2024) studied Hong Kong junior secondary students to explore their critical understanding of climate change in a ChatGPT scenario. The study found that AI facilitated students' evaluation skills by supporting their metacognitive skills, enabling them to reflect on the reliability of AI-generated content and critically assess the tentative and empirical nature of scientific claims.

However, the limitations inherent in GenAI cannot be overlooked. While previous studies have briefly touched upon the challenges of AI-assisted tools in reading, such as the high cognitive load for novice readers (Peng et al., 2022), and the biased content devoid of contextual understanding (Cheung et al., 2024; M. Liu et al., 2024; Rees & Lew, 2024), existing research on AI-assisted academic reading remains limited in several important ways. First, current studies rarely consider the benefits and challenges of GenAI tools systematically and comprehensively, often discussing them only as secondary findings rather than dedicated research objectives. Second, there is a notable absence of theoretical frameworks explicitly employed to evaluate the strengths and weaknesses of AI-assisted reading, resulting in fragmented and context-specific understandings of GenAI effectiveness. Third, empirical research explicitly targeting postgraduate students' perceptions and experiences, who typically engage in more sophisticated and cognitively demanding academic reading tasks, is particularly scarce.

Research Questions

To address these gaps, this study proposed two research questions:

1. What are the benefits and challenges of integrating GenAI tools in supporting academic reading?
2. Which dimensions of academic reading do postgraduate students find most and least supported by GenAI tools?

Methods

Research Design

This study adopted a mixed-methods approach, specifically an exploratory sequential design (Creswell & Plano Clark, 2018), to gain a comprehensive understanding of how AI-assisted tools support postgraduate students' academic reading. In this design, qualitative data were first collected through reflection journals to explore participants' subjective experiences and insights. These findings then informed the development and interpretation of a subsequent quantitative questionnaire, which measured participants' perceptions across key dimensions.

Participants and context

The participants in this study consisted of twelve second-year postgraduate students majoring in Translation and Interpretation at a university in South China, which is renowned for its focus in foreign language studies. Their ages ranged from 21 to 24 years, with a mean age of 22.63 years. Given the widespread use of machine-assisted translation in both teaching and professional translation practice, the participants were already familiar with using translation tools such as DeepL and Google Translate. They also reported personal use of generative AI chatbots like ChatGPT and Ernie Bot to support their translation work. However, their prior experience was primarily limited to tools designed for translation or general-purpose AI assistance, rather than academic reading assistants like Fullpicture. Therefore, to ensure a fair starting point, all participants were introduced to the AI-powered academic reading tool Fullpicture before the experiment, during which they also signed consent forms. Following this introduction, they used the tool over a two-week period to familiarize themselves with its features and functionalities. Although they had not previously used academic reading tools of this kind, they had been trained to engage in critical reading for academic purposes for over five years, a skill essential to their postgraduate academic work.

Instruments

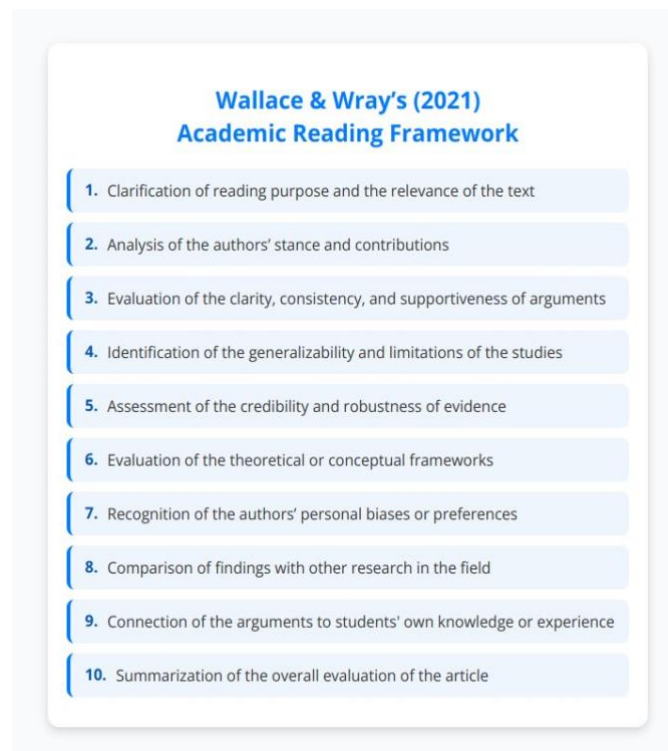
In order to obtain both qualitative and quantitative data regarding postgraduate students' perceptions and experiences with GenAI-assisted academic reading, reflection journals and questionnaires were employed as research instruments.

To address RQ1, we employed reflection journals, which allowed students to express their authentic and deeper feelings when interacting with AI-generated reading reports. Specifically, participants reflected on their perceptions of the AI-generated reports in relation to their academic reading after completing five research articles and the corresponding AI-generated reports. They were also encouraged to discuss their opinions on the benefits and challenges of AI, particularly from the perspective of the 10 dimensions of academic reading (see Figure 1) proposed by Wallace and Wray (2021).

To answer RQ2, we used a questionnaire, which aimed to evaluate participants' perceptions of key academic reading dimensions in a quantitative way. Adapted from the academic reading framework of Wallace and Wray (2021), the questionnaire focused on evaluating how participants perceived the usefulness of the AI-generated reading reports. It consisted of 10 items, including one item for each of the 10 dimensions (e.g., "The AI-supported reading report helped me clarify my reading purpose and how the text is relevant to the purpose"). The questionnaire used a six-point Likert scale (1 = strongly disagree to 6 = strongly agree) to measure the extent of participants' agreement with each statement. This structure ensured that the questionnaire comprehensively addressed both specific aspects of the AI reports and participants' general evaluation of their utility. To ensure seamless integration with the reflection journal, the questionnaire was placed directly below the journal template, allowing participants to complete it immediately after reflecting on their experiences.

Figure 1.

The ten dimensions of Wallace and Wray's (2021) academic reading framework

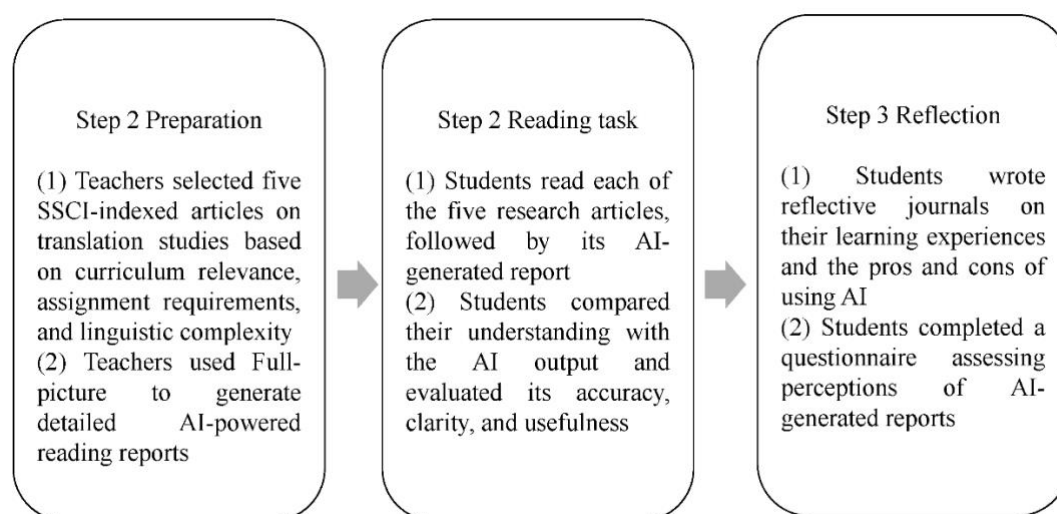


Data collection

The data collection process spanned approximately seven weeks and involved the following three detailed steps (see Figure 2):

Figure 2.

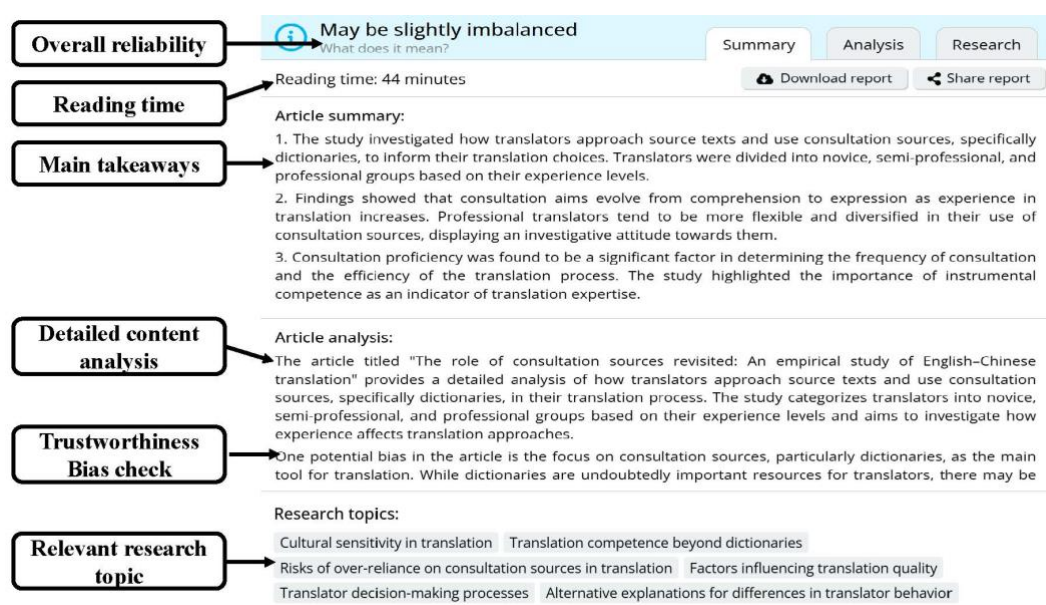
The flowchart of the data collection



In the first step (lasted about 1 week), the teacher selected five research articles published in SSCI-indexed journals, all focusing on topics related to translation studies. The selection was guided by three criteria: (1) the articles' relevance to the teaching curriculum of the postgraduate program, (2) their potential for integration into future assignments requiring critical evaluation and summary, and (3) the complexity of their linguistic features and their compatibility with the academic genre. These five articles were then processed using the AI tool Full-Picture (see Figure 3), which was powered by ChatGPT-3, which generated detailed reading reports including the information of overall reliability, estimated reading time of the report, main takeaways, detailed content analysis, trustworthiness and bias check, and relevant research topics. To align with the reading comprehension framework, we mapped these features to the three comprehension levels: literal (main takeaways, reading time), inferential (content analysis, related topics), and critical (bias check, reliability evaluation).

Figure 3.

Fullpicture - The ChatGPT-powered academic reading tool



In the second step (which lasted about 5 weeks), the students were instructed to read each of the five research articles one at a time. Each student was instructed to read the AI-generated report after finishing the article. The process of reading both the article and the AI-generated report was repeated five times.

The final step lasted about one week, and students were required to document their positive and negative experiences in their reflection journals when using the reports generated by Fullpicture during their academic reading, and then they completed the questionnaire.

Data analysis

We utilized the coding book adapted from the study by Zhang et al. (2024) and employed thematic analysis to examine qualitative data in reflective journals. In detail, we classified students' perspectives as either benefits or challenges, and next we, following the reading

comprehension model (Basaraba et al., 2013; Saadatnia et al., 2016), divided the benefits and challenges of students into three levels (Basaraba et al., 2013; Saadatnia et al., 2016), where the literal level is related to understanding explicit information, the inferential level is related to the personal goals of readers, and the critical level involves in-depth evaluation or critical insights of the text. The coding process consisted of three stages. First, the two coders familiarized themselves with the reflection journal data through multiple readings. Second, we coded the data following a predefined framework, which included levels, sub-categories, and representative quotes. Third, the coders compared their results, discussed discrepancies, and resolved differences through consensus. To minimize potential bias, both coders were aware of their personal views toward the AI tool and took deliberate steps to prevent these from influencing the analysis. These perspectives were openly discussed during the discussion stage to ensure objectivity. The final coding achieved an interrater agreement rate of 85.50%, ensuring the process was rigorous, consistent, and aligned with the theoretical framework.

Table 1.

The coding book

	Literal	Inferential	Critical
Benefits	Facilitating quick comprehension (Student C: <i>The report concisely summarized the source, saving me time compared to reading it myself.</i>)	Supporting interpretation of author's tone and purpose (Student G: <i>The report helped me identify the author's tone, such as their supportive attitude toward multimodal technology.</i>)	Promoting reflective reading practice (Student F: <i>The report made me reevaluate the contribution and limitations of the article, aiding my critical reading.</i>)
	Reducing cognitive load (Student E: <i>ChatGPT helps me to maintain focus on the article, and reduce my workload.</i>)	Reinforcing personal beliefs (Student A: <i>The findings mentioned in ChatGPT's summary helped reinforce my views on learner self-assessment, making me more confident that my research direction is meaningful.</i>)	Generating new perspectives for academic writing (Student I: <i>ChatGPT suggested potential research directions in the report, which inspired me to expand the content of my paper.</i>)
Challenges	Selective summarization (Student B: <i>The report barely mentioned the author's conclusion or reproduced the entire reasoning process.</i>)	Overgeneralization of interpretations (Student H: <i>ChatGPT's analysis is sometimes too broad, such as concerning the interpretation of the research context and participants</i>)	Lack of practical suggestions (Student F: <i>ChatGPT's critiques pointed out problems but didn't provide actionable suggestions, such as how to improve the experimental design.</i>)
	Misrepresentation of details (Student J: <i>The analysis sometimes misunderstood the article, such as claiming that the research didn't consider the effect of stress on interpreting, when it actually did.</i>)	Irrelevance to personal reading and writing goals (Student K: <i>Some suggestions seemed vague and didn't align well with my writing needs.</i>)	Weak evidence for critiques (Student K: <i>When pointing out the limitations of the article, the report didn't provide specific examples to support its claims, making it feel vague.</i>)

The questionnaire demonstrated good reliability and validity, with Cronbach's alpha of 0.727,

indicating good internal consistency. The Bartlett's test of sphericity was significant ($p < 0.001$), confirming that the data showed acceptable validity. For content validity, the items were carefully adapted from Wallace and Wray (2021) academic reading and reviewed by two experts (PhD holders in educational technology) in the field of academic reading, who ensured the clarity, relevance, and alignment of the items with the study's objectives. Subsequently, the descriptive data from the questionnaire were calculated, including means and standard deviations for each of the 10 items, as well as the overall mean score, to summarize participants' perceptions of the AI-generated reports.

Research Findings

Benefits of the GenAI-powered academic reading tool

We identified three levels, namely literal, inferential, and critical levels (see Table 1), of benefits reported by students during their utilization of the academic reading tool. When it comes to literal comprehension, the benefits include the facilitation of quick comprehension and the alleviation of cognitive load. In terms of inferential comprehension, students reported advantages such as the enhancement of the interpretation of the author's tone and purpose, and the reinforcement of readers' personal beliefs. As for critical comprehension, the perceived benefits involve the stimulation of reflective reading practices and the inspiration of new perspectives for academic writing.

First, regarding the benefits at the literal level, students indicated that the tool can provide concise summaries of the article to facilitate their quick comprehension of the main ideas. An example can be seen from Student C, who said, "*The report concisely summarized the source, saving me time compared to reading it myself.*" In addition, students also illustrated that the tool has helped them to maintain understanding of the key points of an article and reduce cognitive load because of the information simplification function, as one student reflected, "*ChatGPT helps me to maintain focus on the article and reduce my workload*" (Student E).

Second, for the inferential level, students expressed their favoring of the tool in supporting them to interpret the implicit information of the article, such as the author's tone and purpose. For example, one student noted, "*The report helped me identify the author's tone, such as their supportive attitude toward multimodal technology*" (Student G). Moreover, students believed that the tool has reinforced their personal views and beliefs about the article and helped them stay on the right track, as one individual remarked, "*The findings mentioned in ChatGPT's summary helped reinforce my views on learner self-assessment, making me more confident that my research direction is meaningful*" (Student A).

Third, at the critical level, the tool promoted reflective reading practices by encouraging students to critically evaluate the strengths and weaknesses of the texts they engaged with. One participant noted, "*The report made me reevaluate the contribution and limitations of the article, aiding my critical reading*" (Student F). Moreover, the tool generated new perspectives for academic writing by inspiring students to think creatively and critically about potential research directions or ways to expand their own work. As one student commented, "*ChatGPT suggested potential research directions in the report, which inspired me to expand the content of my paper*"

(Student I).

Challenges of the GenAI-powered academic reading tool

While the tool provided notable benefits, participants also reported challenges across the literal, inferential, and critical levels (see Table 1). In navigating the literal dimension, challenges included selective summarization and misrepresentation of details. Moving to the inferential level, students raised concerns about overgeneralization of interpretations and irrelevance to personal reading and writing goals. As for the critical comprehension, challenges involved the lack of actionable suggestions and weak evidence to support critiques.

First, with reference to the literal level, students expressed frustration with the tool's tendency of selective summarization, where the tool omitted important information, leading to incomplete representations of the text. For example, a participant remarked, *"The report barely mentioned the author's conclusion or reproduced the entire reasoning process"* (Student B). Another challenge was misrepresentation of details, as the tool occasionally misunderstood specific parts of the text. One student noted, *"The analysis sometimes misunderstood the article, such as claiming that the research didn't consider the effect of stress on interpreting, when it actually did"* (Student J).

Second, as for the inferential level, one issue was evident in cases of overgeneralization, where the tool failed to provide nuanced or context-specific analyses. One participant observed, *"ChatGPT's analysis is sometimes too broad, such as concerning the interpretation of the research context and participants"* (Student H). Furthermore, the tool was criticized for irrelevance to personal reading and writing goals, as some of its suggestions did not align with students' academic needs. A student reflected, *"Some suggestions seemed vague and didn't align well with my writing needs"* (Student D).

Third, at the critical level, challenges were most pronounced in the tool's inability to provide actionable suggestions, as students highlighted the absence of practical advice for improving research methods or addressing specific problems. For instance, a student explained, *"ChatGPT's critiques pointed out problems but didn't provide actionable suggestions, such as how to improve the experimental design"* (Student F). Another concern was weak evidence to support critiques, where the tool's feedback often lacked detailed examples or explanations. One participant commented, *"When pointing out the limitations of the article, the report didn't provide specific examples to support its claims, making it feel vague"* (Student D). These challenges suggest that the tool requires further refinement to better support students' academic engagement, especially at higher cognitive levels.

Questionnaire results

The questionnaire results provide valuable insights into which dimensions of academic reading postgraduate students perceive as most and least supported by the GenAI-powered academic reading tool. Among the ten dimensions adapted from Wallace and Wray's (2021) framework, the tool demonstrated strong support for evaluating the credibility, sufficiency, and robustness of sources and evidence ($M = 4.17$, $SD = 1.19$), as well as assessing the clarity, consistency, and supportiveness of arguments ($M = 4.08$, $SD = 1.24$). These findings suggest that students found the tool particularly effective in facilitating evaluation of core components of academic

texts, such as the validity of evidence and the logical coherence of claims. Similarly, the tool showed moderate support for evaluating the generalizability and limitations of studies ($M = 3.92$, $SD = 0.79$) and for providing overall evaluation ($M = 3.83$, $SD = 0.83$), and for analyzing stances as well as research contributions ($M = 3.67$, $SD = 1.07$). This suggested that while the tool is helpful for key evaluative tasks, its performance begins to taper off when addressing more interpretive and integrative aspects of academic reading.

However, the results also showed that participants reported the perceived inadequacy of the tool's ability to support other key dimensions of academic reading. For instance, lower mean scores were observed for clarifying reading purposes and understanding how the text aligns with those purposes ($M = 3.35$, $SD = 0.90$) and identifying authors' personal preferences and biases ($M = 3.33$, $SD = 1.07$), suggesting that students struggled to use the tool effectively for these interpretive tasks. Even lower ratings were given for evaluating the relevance of authors' arguments or claims in relation to students' own knowledge and experience ($M = 2.58$, $SD = 1.00$), as well as comparing and contrasting the findings of studies with others' work in the field ($M = 2.00$, $SD = 1.13$). The tool's lowest-rated dimension was understanding and evaluating theoretical or conceptual frameworks and how well they are adopted in studies ($M = 1.75$, $SD = 0.45$). These findings indicate that the tool offers limited support for more nuanced and inferential dimensions of academic reading, such as drawing connections between research findings, personal knowledge, and other works, or critically assessing the adoption of theoretical frameworks. Overall, the data reveal a clear divide between the dimensions of academic reading that students found well-supported and those where gaps remain.

Table 2.

The questionnaire results

Questionnaire item	Mean	SD
5. ChatGPT-generated reading reports help me evaluate whether the sources and evidence in the studies are credible, sufficient, and robust.	4.17	1.19
3. ChatGPT-generated reading reports help me evaluate whether the arguments in the studies are clear, consistent, and supportive.	4.08	1.24
4. ChatGPT-generated reading reports evaluate the generalizability and limitations of the studies.	3.92	0.79
10. Overall, the ChatGPT-generated reading reports are very helpful.	3.83	0.83
2. ChatGPT-generated reading reports help me analyze the authors' stance and the contributions of the studies.	3.67	1.07
1. ChatGPT-generated reading reports help me clarify my reading purpose and how the text is relevant to the purpose.	3.35	0.90
7. ChatGPT-generated reading reports help me identify authors' personal preferences and biases in the study.	3.33	1.07
9. ChatGPT-generated reading reports help me evaluate the relevance of the authors' arguments / claims with my own knowledge/experience	2.58	1.00
8. ChatGPT-generated reading reports help me compare and contrast the findings in the studies with others' work in the field.	2.00	1.13
6. ChatGPT-generated reading reports help me understand and evaluate what the theoretical or conceptual framework is and how well it is adopted in the study.	1.75	0.45

To clarify the reasons behind both lower and higher ratings, questionnaire results were compared with qualitative feedback from reflective journals. This comparison revealed that

higher ratings were often associated with dimensions involving well-structured analytical tasks, such as evaluating the credibility of sources or assessing the clarity and coherence of arguments. These aspects of academic reading tend to follow established evaluative criteria and rhetorical conventions, making them more accessible to the tool's generative capabilities. And students appreciated the tool's ability to identify logical inconsistencies and assess evidential support. For instance, one student noted, *"It helped me see whether the claims were backed up by evidence"* (Student Q), while another remarked, *"It pointed out logical inconsistencies that I hadn't noticed"* (Student R).

In contrast, lower ratings were linked to more interpretive and cognitively demanding dimensions. The lowest-rated item, understanding and evaluating theoretical or conceptual frameworks, reflects the tool's limited capacity to engage with abstract or discipline-specific constructs. Several students expressed dissatisfaction with the lack of theoretical insight, as in *"The summary failed to mention the theoretical model the article was based on, which is crucial for my own literature review"* (Student K). Similarly, dimensions such as comparing and contrasting findings or relating arguments to personal knowledge received low scores, likely because these tasks require contextual integration and critical synthesis, which go beyond the tool's general summarization capabilities. One student remarked, *"The tool doesn't help me see how this article fits into the broader academic conversation, it just summarizes it in isolation"* (Student M).

Discussion

This exploratory study aimed to evaluate the experiences of L2 learners using a Generative AI-powered academic reading tool, focusing on the benefits and challenges across three levels of comprehension. The findings highlight that the tool provided significant benefits, such as enabling quick comprehension and reducing cognitive load at the literal level, enhancing the interpretation of tone and purpose while reinforcing personal beliefs at the inferential level, and fostering reflective reading practices alongside inspiring new ideas for academic writing at the critical level. It should be noted that several challenges were reported, including selective summarization and misrepresentation of details at the literal level; overgeneralization of interpretations and irrelevance to personal reading and writing goals at the inferential level; and lack of practical suggestions and weak supporting evidence for critiques at the critical level. The questionnaire results provided further insights into the affordances and limitations of GenAI tools in academic reading.

RQ1: What are the benefits and challenges of integrating GenAI tools in supporting academic reading?

We identified several benefits that aligned with existing studies and further extended the literature by uncovering some novel insights. At the literal level of benefits, students reported that the content summarization affordance enhanced their quick comprehension of the material. This finding is consistent with several previous studies (Yang et al., 2021; Rees & Lew, 2024; Monib et al., 2025), which highlighted AI's affordance to support literal comprehension by offering summarization and vocabulary assistance. Additionally, the current study found that

GenAI-powered tools can help students focus on key points within large volumes of information without feeling overwhelmed. This finding aligns with Wang (2024), who also observed that the cognitive load reduction affordance of GenAI may be particularly beneficial for L2 learners as they strive to balance linguistic and academic demands.

Regarding the inferential level, this study supported prior research that highlighted GenAI's capacity to scaffold understanding by activating background knowledge and tailoring explanations (Cheng et al., 2024; M. Liu et al., 2024). However, a unique contribution of this study lies in the observation that the tool validated students' personal beliefs, reinforcing their confidence in academic perspectives. This affective dimension has been less emphasized in these studies and suggested that GenAI tools may play an important motivational role for L2 learners (Jeon, 2024; Zou et al., 2024).

Similarly, at the critical level, the findings aligned with Peng et al.'s (2022) and Cheung et al.'s (2024) studies, which demonstrated how AI tools can support critical comprehension by fostering reflective practices and encouraging evaluation of arguments. Moreover, we found that the tool can inspire new perspectives for academic writing, which has long been regarded as a complex and creative process. These findings suggest that GenAI tools can serve not only as assistive tools for improving evaluative skills but also as catalysts for creativity (Wang, 2024).

We also identified several challenges in this study that were consistent with previous research. In terms of challenges at the literal level, students reported feeling frustrated with selective summarization and the misrepresentation of details, which aligned with the findings of Yang et al. (2021), who also reported inaccuracies in AI-generated summaries. While previous work (Jeon, 2024) attributed these challenges to technical limitations, the current study suggests that they may also be due to the complexity of academic texts (Conrad, 1996), which are beyond the comprehension capacity of the GenAI tool.

Regarding challenges at the inferential level, we first identified the overgeneralization of interpretations, which aligns with the viewpoints of previous researchers such as M. Liu et al. (2024) and Cheung et al. (2024), who indicated that AI tools often generate texts that are too broad and lack contextual understanding. In addition, participants also expressed dissatisfaction with the tool, as it cannot provide fully personalized feedback aligned with their reading and writing goals. These findings suggest that the integration of GenAI tools should place greater emphasis on enhancing customization and promoting user-centered design to meet users' specific needs (Kim et al., 2024; Kohnke, 2024).

When it comes to challenges at the critical level, the present study indicated that the tool sometimes could not provide actionable suggestions, which is consistent with the findings of Peng et al. (2022), who also observed that AI often fails to offer truly feasible instructions to students. In addition, our findings demonstrated that the tool was incapable of providing sufficient evidence to support its critical views on the article, which echoes the viewpoints of Darwin et al. (2024), who revealed that GenAI tools often evaluate texts without offering clear explanations for the reasons behind such critical assessments.

RQ2: Which dimensions of academic reading do postgraduate students find most and least supported by GenAI tools?

The results of the survey showed that the tool supported certain dimensions of academic reading more effectively than others. First, students perceived that the GenAI-powered tool performed well in assisting them with evaluating the credibility of evidence, the clarity of arguments, and the generalizability of findings. This aligned with the view of Wallace and Wray (2021), who stated that these are foundational skills in academic reading, and therefore, students would put more effort into these dimensions. Oates and Johnson (2025) further revealed that GenAI tools are capable of assisting learners in developing their evaluation skills. Second, students rated certain dimensions lower, such as the understanding of theoretical frameworks, comparisons of the study with other researchers' work, and the relevance of arguments to user experience. This was consistent with the findings of G. L. Liu et al. (2024), who highlighted the limitations of GenAI tools in facilitating contextual and conceptual understanding. Similarly, Alqahtani et al. (2023) pointed out that GenAI tools perform less effectively when it comes to understanding personal experiences and linking them with other scholars' perspectives.

Implications and limitations

The present study reveals two implications and identifies two limitations. First, students need to form a collaborative relationship with GenAI-powered tools instead of relying entirely on them when these tools are applied in a university reading context. Second, when developing GenAI reading tools, developers should consider the hierarchical levels of reading comprehension and students' varying levels of AI literacy. As for the limitations, first, the sample size was relatively small, constrained to the context of language education, and relied on self-reported data. Second, it should be noted that some students in the study might have shown a preference for using GenAI tools due to personal interest, which could introduce bias.

References

- Alptekin, C. (2006). Cultural familiarity in inferential and literal comprehension in L2 reading. *System*, 34(4), 494-508. <https://doi.org/10.1016/j.system.2006.05.003>
- Alptekin, C. E. M., & ErÇetin, G. (2011). Effects of Working Memory Capacity and Content Familiarity on Literal and Inferential Comprehension in L2 Reading. *TESOL Quarterly*, 45(2), 235-266. <https://doi.org/10.5054/tq.2011.247705>
- Alqahtani, T., Badreldin, H. A., Alrashed, M., Alshaya, A. I., Alghamdi, S. S., bin Saleh, K., Alowais, S. A., Alshaya, O. A., Rahman, I., Al Yami, M. S., & Albekairy, A. M. (2023). The emergent role of artificial intelligence, natural learning processing, and large language models in higher education and research. *Research in Social and Administrative Pharmacy*, 19(8), 1236-1242. <https://doi.org/10.1016/j.sapharm.2023.05.016>
- Basaraba, D., Yovanoff, P., Alonzo, J., & Tindal, G. (2013). Examining the structure of reading comprehension: do literal, inferential, and evaluative comprehension truly exist? *Reading and Writing*, 26(3), 349-379. <https://doi.org/10.1007/s11145-012-9372-9>

- Carnine, D. W., Silbert, J., Kame'enui, E. J., Slocum, T. A., & Travers, P. A. (2015). *Direct instruction reading* (6th ed.). Pearson.
- Cheng, X., Yin, L., Lin, C., Shi, Z., Zheng, H., Zhu, L., Liu, X., Chen, K., & Dong, R. (2024). Chatbot dialogic reading boosts comprehension for Chinese kindergarteners with higher language skills. *Journal of Experimental Child Psychology*, 240, 105842. <https://doi.org/10.1016/j.jecp.2023.105842>
- Cheung, K. K. C., Pun, J. K. H., & Li, W. (2024). Students' Holistic Reading of Socio-Scientific Texts on Climate Change in a ChatGPT Scenario. *Research in Science Education*. <https://doi.org/10.1007/s11165-024-10177-2>
- Conrad, S. M. (1996). Investigating academic texts with corpus-based techniques: An example from biology. *Linguistics and Education*, 8(3), 299-326. [https://doi.org/10.1016/S0898-5898\(96\)90025-X](https://doi.org/10.1016/S0898-5898(96)90025-X)
- Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. Sage publications.
- Darwin, Rusdin, D., Mukminatien, N., Suryati, N., Laksmi, E. D., & Marzuki. (2024). Critical thinking in the AI era: An exploration of EFL students' perceptions, benefits, and limitations. *Cogent Education*, 11(1), 2290342. <https://doi.org/10.1080/2331186X.2023.2290342>
- de-la-Peña, C., & Luque-Rojas, M. J. (2021). Levels of Reading Comprehension in Higher Education: Systematic Review and Meta-Analysis. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.712901>
- Elbro, C., & and Buch-Iversen, I. (2013). Activation of Background Knowledge for Inference Making: Effects on Reading Comprehension. *Scientific Studies of Reading*, 17(6), 435-452. <https://doi.org/10.1080/10888438.2013.774005>
- Fazio, X., Gallagher, T. L., & DeKlerk, C. (2022). Exploring Adolescents' Critical Reading of Socioscientific Topics Using Multimodal Texts. *International Journal of Science and Mathematics Education*, 20(1), 93-116. <https://doi.org/10.1007/s10763-022-10280-8>
- Feng, L. (2024). Investigating the Effects of Artificial Intelligence-Assisted Language Learning Strategies on Cognitive Load and Learning Outcomes: A Comparative Study. *Journal of Educational Computing Research*, 62(8), 1961-1994. <https://doi.org/10.1177/07356331241268349>
- Goldman, S. R. (2024). Reading is complex: Implications for research and practice. *Educational Psychologist*, 59(4), 315-325. <https://doi.org/10.1080/00461520.2024.2418062>
- Green, D. A. (2009). New academics' perceptions of the language of teaching and learning: identifying and overcoming linguistic barriers. *International Journal for Academic Development*, 14(1), 33-45. <https://doi.org/10.1080/13601440802659254>
- Herber, H. L. (1978). *Teaching Reading in Content Areas*. Prentice-Hall. <https://books.google.com/books?id=RWFBezXMPVIC>

- Hudson, R. F., Pullen, P. C., Lane, H. B., & Torgesen, J. K. (2008). The Complex Nature of Reading Fluency: A Multidimensional View. *Reading & Writing Quarterly*, 25(1), 4-32. <https://doi.org/10.1080/10573560802491208>
- Jeon, E. H., & Yamashita, J. (2014). L2 Reading Comprehension and Its Correlates: A Meta-Analysis. *Language Learning*, 64(1), 160-212. <https://doi.org/10.1111/lang.12034>
- Jeon, J. (2024). Exploring AI chatbot affordances in the EFL classroom: young learners' experiences and perspectives. *Computer Assisted Language Learning*, 37(1-2), 1-26. <https://doi.org/10.1080/09588221.2021.2021241>
- Johnson, T. E., Archibald, T. N., & Tenenbaum, G. (2010). Individual and team annotation effects on students' reading comprehension, critical thinking, and meta-cognitive skills. *Computers in Human Behavior*, 26(6), 1496-1507. <https://doi.org/10.1016/j.chb.2010.05.014>
- Kiili, C., Leu, D. J., Marttunen, M., Hautala, J., & Leppänen, P. H. T. (2018). Exploring early adolescents' evaluation of academic and commercial online resources related to health. *Reading and Writing*, 31(3), 533-557. <https://doi.org/10.1007/s11145-017-9797-2>
- Kim, J., Yu, S., Detrick, R., & Li, N. (2024). Exploring students' perspectives on Generative AI-assisted academic writing. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-024-12878-7>
- Kohnke, L. (2024). Exploring EAP students' perceptions of GenAI and traditional grammar-checking tools for language learning. *Computers and Education: Artificial Intelligence*, 7, 100279. <https://doi.org/10.1016/j.caeai.2024.100279>
- Ku, K. Y. L., & Ho, I. T. (2010). Metacognitive strategies that enhance critical thinking. *Metacognition and Learning*, 5(3), 251-267. <https://doi.org/10.1007/s11409-010-9060-6>
- Lin, H., Chen, Z., Wei, W., & Lu, H. (2025). GenAI tools in academic reading: A study on AI-assisted metacognitive strategies and emotional reactions. In E. C. K. Cheng (Ed.), *Innovating education with AI: AETS 2024. Lecture Notes in Educational Technology* (pp. 98–112). Springer, Singapore. https://doi.org/10.1007/978-981-96-4952-5_7
- Lin, H., Xiong, Z., Tang, H., Jiang, S., Wei, W., & Fang, K. (2024). Implementing generative AI agent game to support reading of classical Chinese literature: A needs analysis. In *2024 4th International Conference on Educational Technology (ICET)* (pp. 86–91). IEEE. <https://doi.org/10.1109/ICET62460.2024.10868296>
- Liu, G. L., Darwin, R., & Ma, C. (2024). Exploring AI-mediated informal digital learning of English (AI-IDLE): a mixed-method investigation of Chinese EFL learners' AI adoption and experiences. *Computer Assisted Language Learning*, 1-29. <https://doi.org/10.1080/09588221.2024.2310288>
- Liu, M., Zhang, J., Nyagoga, L. M., & Liu, L. (2024). Student-AI Question Cocreation for Enhancing Reading Comprehension. *IEEE Transactions on Learning Technologies*, 17, 815-826. <https://doi.org/10.1109/TLT.2023.3333439>
- Liu, X., Gu, M. M., & Jin, T. (2021). Strategy use in collaborative academic reading:

- Understanding how undergraduate students co-construct comprehension of academic texts. *Language Teaching Research*, 28(4), 1391-1411. <https://doi.org/10.1177/13621688211025688>
- Luk, J., & Hui, D. (2017). Examining multiple readings of popular culture by ESL students in Hong Kong. *Language, Culture and Curriculum*, 30(2), 212-230. <https://doi.org/10.1080/07908318.2016.1241258>
- Monib, W. K., Qazi, A., & Mahmud, M. M. (2025). Exploring learners' experiences and perceptions of ChatGPT as a learning tool in higher education. *Education and Information Technologies*, 30(1), 917-939. <https://doi.org/10.1007/s10639-024-13065-4>
- Oates, A., & Johnson, D. (2025). ChatGPT in the Classroom: Evaluating its Role in Fostering Critical Evaluation Skills. *International Journal of Artificial Intelligence in Education*. <https://doi.org/10.1007/s40593-024-00452-8>
- Peng, Z., Liu, Y., Zhou, H., Xu, Z., & Ma, X. (2022). CReBot: Exploring interactive question prompts for critical paper reading. *International Journal of Human-Computer Studies*, 167, 102898. <https://doi.org/10.1016/j.ijhcs.2022.102898>
- Pretorius, E. J., & Spaull, N. (2016). Exploring relationships between oral reading fluency and reading comprehension amongst English second language readers in South Africa. *Reading and Writing*, 29(7), 1449-1471. <https://doi.org/10.1007/s11145-016-9645-9>
- Rees, G. P., & Lew, R. (2024). The Effectiveness of OpenAI GPT-Generated Definitions Versus Definitions from an English Learners' Dictionary in a Lexically Orientated Reading Task. *International Journal of Lexicography*, 37(1), 50-74. <https://doi.org/10.1093/ijl/ecad030>
- Rouet, J.-F., Britt, M. A., & Durik, A. M. (2017). RESOLV: Readers' Representation of Reading Contexts and Tasks. *Educational Psychologist*, 52(3), 200-215. <https://doi.org/10.1080/00461520.2017.1329015>
- Saadatnia, M., Ketabi, S., & Tavakoli, M. (2016). EFL Learners' Levels of Comprehension Across Text Structures: A Comparison of Literal and Inferential Comprehension of Descriptive and Enumerative Expository Texts. *Journal of Psycholinguistic Research*, 45(6), 1499-1513. <https://doi.org/10.1007/s10936-016-9414-6>
- Sheorey, R., & Mokhtari, K. (2001). Differences in the metacognitive awareness of reading strategies among native and non-native readers. *System*, 29(4), 431-449. [https://doi.org/10.1016/S0346-251X\(01\)00039-2](https://doi.org/10.1016/S0346-251X(01)00039-2)
- Snow, C. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Rand Corporation.
- Srinivasan, V., & Murthy, H. (2021). Improving reading and comprehension in K-12: Evidence from a large-scale AI technology intervention in India. *Computers and Education: Artificial Intelligence*, 2, 100019. <https://doi.org/10.1016/j.caeai.2021.100019>
- Taki, S. (2016). Metacognitive online reading strategy use: Readers' perceptions in L1 and L2. *Journal of Research in Reading*, 39(4), 409-427. <https://doi.org/10.1111/1467-9817.12048>

- Toyokawa, Y., Horikoshi, I., Majumdar, R., & Ogata, H. (2023). Challenges and opportunities of AI in inclusive education: a case study of data-enhanced active reading in Japan. *Smart Learning Environments*, 10(1), 67. <https://doi.org/10.1186/s40561-023-00286-2>
- Wallace, M., & Wray, A. (2021). *Critical Reading and Writing for Postgraduates*. SAGE. <https://books.google.com/books?id=3UbYzQEACAAJ>
- Wang, C., & Gierl, M. J. (2011). Using the Attribute Hierarchy Method to Make Diagnostic Inferences about Examinees' Cognitive Skills in Critical Reading. *Journal of Educational Measurement*, 48(2), 165-187. <https://doi.org/10.1111/j.1745-3984.2011.00142.x>
- Wang, C. (2024). Exploring Students' Generative AI-Assisted Writing Processes: Perceptions and Experiences from Native and Nonnative English Speakers. *Technology, Knowledge and Learning*. <https://doi.org/10.1007/s10758-024-09744-3>
- Yang, A. C. M., Chen, I. Y. L., Flanagan, B., & Ogata, H. (2021). From Human Grading to Machine Grading: Automatic Diagnosis of e-Book Text Marking Skills in Precision Education. *Educational Technology & Society*, 24(1), 164–175. <https://www.jstor.org/stable/26977865>
- Yüzlü, M. Y., & Dikilitaş, K. (2022). Translanguaging as a way to fostering EFL learners' criticality in a hybrid course design. *System*, 110, 102926. <https://doi.org/10.1016/j.system.2022.102926>
- Zhang, D., Wu, J. G., & Fu, Z. (2024). From shy to fly: Facilitating EFL learners' willingness to communicate with an AI chatbot and an intelligent tutoring system. *System*, 127, 103501. <https://doi.org/10.1016/j.system.2024.103501>
- Zou, B., Liviero, S., Ma, Q., Zhang, W., Du, Y., & Xing, P. (2024). Exploring EFL learners' perceived promise and limitations of using an artificial intelligence speech evaluation system for speaking practice. *System*, 126, 103497. <https://doi.org/10.1016/j.system.2024.103497>

Biodata

Haoming Lin is a PhD candidate at the Faculty of Applied Sciences, Macao Polytechnic University. His research interests include AI in education and computer-assisted language learning.

Dr. Wei Wei is an associate professor at the Faculty of Applied Sciences, Macao Polytechnic University. He obtained his PhD from the School of Education, University of Leeds, UK. He also serves as the coordinator of the PhD program in Educational Technology and Innovation. His research interests include computer-assisted language learning, learning-oriented assessment, and language assessment. His address is Room P325, Pearl Jubilee Building, Macao Polytechnic University Taipa, Macao SAR, 999078. Tel: 8399 8625. His email address is weiweitesting@hotmail.com.

Dr. Handan Lu obtained her PhD from School of Education, University of Leeds, UK. She currently works as a researcher with The Assessment and Qualifications Alliance (AQA) in UK.

Her research areas include: reading and language comprehension, second language acquisition from psycholinguistic perspective, digital literacy and on-screen assessment.