



Reinventing assessments with ChatGPT and other online tools: Opportunities for GenAI-empowered assessment practices

Dennis Foung^{a,*}, Linda Lin^b, Julia Chen^c

^a School of Journalism, Writing and Media, The University of British Columbia, Canada

^b English Language Centre, The Hong Kong Polytechnic University, Hong Kong

^c Educational Development Centre, The Hong Kong Polytechnic University, Hong Kong

ARTICLE INFO

Keywords:

GenAI
Online tools
Language learning
Assessments

ABSTRACT

The recent emergence of generative artificial intelligence (GenAI) tools, such as ChatGPT, has brought profound changes to higher education. While many studies have examined the potential use of ChatGPT in teaching and learning, few have explored the opportunities to develop assessments that facilitate the use of multiple technological innovations (i.e. traditional AI and GenAI tools). We conducted qualitative research to address this gap. The assessments of an elective English course in Hong Kong were re-designed to incorporate GenAI and other tools. Students were asked to employ and reflect on their use of these tools for their writing assessments. We analyzed the written reflections of 74 students and conducted focus group interviews with 28 students. The results suggest that the students possess an acumen for choosing the appropriate online tools for specific purposes. When they can choose freely, they develop skills that allow them to evaluate and select between traditional AI and GenAI tools when appropriate. Some students mentioned concerns with the different features of the free and premium versions. The results of this study call for (1) assessment practices that allow the flexibility to use different AI tools and (2) the equitable use of various AI tools.

1. Introduction

1.1. Prevalence of GenAI in higher education

The rapid rise of generative artificial intelligence (GenAI) since the end of 2022 has significantly impacted the education sector. Of the myriad GenAI tools, ChatGPT is by far the most popular. A global survey conducted in five countries (i.e., Brazil, India, Japan, the UK, and the US) to gather insights from over 1600 university educators and students showed that 74% of students had used ChatGPT when completing assessments (Ibrahim et al., 2023). In a similar study, out of nearly 6000 Swedish university students surveyed, an overwhelming 95% reported familiarity with ChatGPT (Malmström et al., 2023). Fewer students reported using Bing AI, CoPilot, OpenAI Playground, and Bard AI. Three groups of students reported a particularly strong inclination to use GenAI: those with special education needs (Malmström et al., 2023), those from lower socio-economic backgrounds, and non-native English speakers (Ibrahim et al., 2023). These students believe that using such tools supports their learning and allows them to compete with other students.

Similarly, educators have reported the presence of GenAI in teaching. A poll of 1070 educators at Educause Community Groups member organisations (Muscanell & Robert, 2023) revealed that 54% of the respondents have been impacted by GenAI in the following sectors: undergraduate teaching (37%), graduate teaching and faculty development (30%), integration in lesson activities (24%), use in course assignments (22%) and writing reports, manuscripts and proposals (21%). This reflects that GenAI can be used to cater to diverse educational needs. Though this technology has become popular in recent years, the research is still nascent: more should be done to understand how GenAI tools can help facilitate the education process.

In higher education, traditional AI (i.e. not GenAI) has long played an important role in facilitating the learning of undergraduate and postgraduate students. One simple definition of AI is computers performing “near or human-like functions” (Chen et al., 2020). Learning management systems can give students human-like feedback (e.g. feedback input by teachers) based on their answers to quiz questions. Other examples include adaptive learning systems that artificially adapt content based on students’ performance (Sahin et al., 2022).

Due to recent technological advancements and the availability of

* Corresponding author.

E-mail address: dennis.foung@gmail.com (D. Foung).

language data, GenAI has become the dominant form of AI. In addition to the capabilities of traditional AI, GenAI tools can produce content, including text and images. Instead of giving students limited feedback in response to their answers, GenAI acts more ‘human-like’ by producing full sentences and paraphrasing ideas based on the examination of large language models (Barrett & Pack, 2023). Therefore, GenAI tools help students earlier in the learning process, long before they engage with an intelligent tutor or input a piece of writing to check for errors. The feedback provided by GenAI tools has been shown to be beneficial to students (see Yilmaz & Yilmaz, 2023a, b).

Significantly, these examples are from higher education in general; they are not necessarily rooted in the language learning context. These diverse examples show how the use of traditional AI and GenAI tools can facilitate students’ learning. However, little research has explored how using both traditional AI and GenAI can contribute to students’ learning.

1.2. AI tools in language education

In the field of language education, AI tools can support language learning by (1) helping students with assignments, (2) building their proficiency, and (3) helping teachers facilitate their learning. First, GenAI can be used early in the writing process as a writing assistant (Imran & Almusharraf, 2023) or partner (Gimpel et al., 2023) that helps students generate ideas. For example, Quillbot can help students paraphrase using a deep-learning-based paraphrasing system (<https://quillbot.com/>). Another important GenAI tool that students can use to facilitate the planning stage is ChatGPT, a chatbot that uses large language models. ChatGPT is prevalent across disciplines including STEM, medicine, and education (Imran & Almusharraf, 2023).

Later in the writing process, other GenAI tools can give content-specific grammar and vocabulary suggestions and facilitate reflection on the writing process (Su et al., 2023). For example, Grammarly helps learners detect errors in their writing (Godwin-Jones, 2022), while Just The Word (<http://www.just-the-word.com/>) provides suggestions on word collocation. It is also important to note that many traditional AI tools (e.g. Grammarly, 2023) have begun to introduce GenAI features. This makes the boundaries between traditional AI and GenAI tools blurry, though it is apparent that both remain common in the context of higher education because they facilitate language learning and provide resources for English as a foreign language (EFL) students (Al-Imam et al., 2023; Friginal, 2018; Yoon, 2016).

Second, in addition to completing assignments, GenAI tools help students enhance the pace of language learning. Some of these tools can simulate real-life communication to help students understand words in context, as well as provide dictionary definitions, translations and sample sentences (Kohnke et al., 2023). In addition, they can help students acquire targeted vocabulary words based on a theme (Sarrion, 2023) and adjust the complexity of a dialogue to suit students of different levels (Kohnke et al., 2023). GenAI tools can also support teachers by analyzing student writing, determining proficiency levels, and using established assessment criteria (e.g., IELTS) to provide context-specific feedback (Koraishi, 2023). They can engage students in extracurricular study and practice (Mohamed, 2023).

However, despite the strengths of GenAI tools for language learning, they have also brought challenges, including a decrease in students’ critical thinking and writing skills (Chan, 2023; Kumar, 2023; Mizumoto, 2023). Many concerns relate to their ability to “perform a wide range of language tasks and generate human-like texts” (Chan, 2023, p. 3) which could encourage students to rely on them or use them excessively, leading to academic integrity issues. A further concern is the limitations of the texts created by GenAI tools, which include inappropriate referencing, little in-depth discussion, and insufficient consideration of context (Chan, 2023; Kumar, 2023; Warschauer et al., 2023). Using such texts may affect students’ learning outcomes and impair the quality of education. Addressing this issue will require universities to enforce regulations on GenAI use (Chan, 2023) while encouraging

learners to actively explore, select from, and coordinate the resources available to them (Mizumoto, 2023).

1.3. GenAI tools in language education

GenAI tools have transformed the landscape of language education, particularly for EFL students. This is primarily because of the powerful role they can play as writing assistants (Warschauer et al., 2023). Their assistance takes the form of the editing and translation functions of GenAI tools such as ChatGPT, which can help students with word choice, sentence structure, and writing style. The improvements to these features help learners produce high-quality assessments, papers, and even journal articles. GenAI tools have created a level playing field for EFL students (Lim et al., 2023).

This is evidenced by Jacob et al.’s (2023) case study, which explores how an EFL PhD student at a university in the United States effectively leveraged ChatGPT for academic writing. By engaging constructively with ChatGPT throughout the academic writing process, the student was not only able to successfully produce highly demanding academic texts (e.g. grant proposals) but also maintain her authentic voice and agency. The results of this study corroborate with those of several other investigations (e.g. Chan & Hu, 2023; Chan & Lee, 2023; Tlili et al., 2023) into how students use GenAI in academic writing. Similar to research focused on other fields within higher education, research into GenAI use in language education has primarily focused on the use of GenAI. They often do not consider the fact that traditional AI and GenAI tools are often used simultaneously when students work on assignments.

Despite the enthusiasm for GenAI writing, concerns have been raised about the challenges GenAI technologies pose for language education. The main concerns center on academic integrity, fairness in education, and author voice (Chan & Hu, 2023; Dignum, 2023; Michel-Villarreal et al., 2023). These concerns have prompted the development of the GenAI Acceptance Scale (Yilmaz et al., 2023c) and AI Assessment Scale (Furze et al., 2024; Perkins et al., 2024). These scales identify how GenAI technologies can be used in assessments. Based on a common understanding, traditional modes of assessment should be redesigned to promote the appropriate use of GenAI tools (Perkins & Roe, 2023). To avoid the possibility of overreliance on GenAI, the redesigned assessments should also encourage the use of various online language tools in the completion process (Mizumoto, 2023).

1.4. Research gap and goals

Despite extensive research on how to redesign assessments in the GenAI era (e.g., Chen, 2023; Dignum, 2023; Kumar, 2023), empirical evidence on their effectiveness is rather limited. In particular, there is a need for studies on how to redesign assessments to accommodate both traditional AI and GenAI tools. To bridge this gap, the current study examines how students constructively use various AI tools when completing a writing assessment.

2. Material and methods

This study adopted a qualitative design to answer the research question about how students use GenAI to complete a writing assessment. The qualitative approach is rooted in the postmodernist paradigm and examines the experiences and perceptions of learners (de Costa et al., 2017, p. 526). These are subjective in nature and do not have a single “truth.” Therefore, the positivist and/or quantitative approach to hypothesis testing, with (in)dependent variables and statistical analyses, is not applicable to the current study. We are not arguing that the qualitative approach is superior, but we do believe that understanding the experiences and perceptions of learners can contribute to the field of research on the use of AI tools for writing assessments.

In practice, we retrieved and analyzed written self-reflections completed by students taking the re-designed course. They were

supplemented with focus group interviews conducted with a sample of students. All the participants were informed about the study objectives and completed the research consent form before data collection began.

2.1. Research context

The current study was conducted with students enrolled in a communication course offered by the English language centre at a major university in Hong Kong. This course is only for Senior Year Admitted (SYA) students. SYA students have graduated from sub-degree programs offered at various institutions in Hong Kong, such as community colleges and vocational training schools. They are admitted to complete a government-funded “senior year” at a more prestigious university. However, many of these students have come from secondary schools where Chinese is the medium of instruction and thus have limited English language skills. This is a major reason that they cannot be admitted to the university as direct entrants. In addition, their low language proficiency makes it more difficult for them to read course materials and complete written and spoken assessments. In some cases, it can pose challenges to understanding lectures (Bobe & Cooper, 2019). To address this issue, the university offers several general education courses specifically designed to enhance the spoken and written English communication skills of SYA students.

The course was offered long before GenAI tools were available. However, after these tools became accessible in Hong Kong and the university announced an open and forward-looking GenAI policy, faculty were encouraged to change their assessments to promote the appropriate and constructive use of these technologies. Responding to the call for change, instructors revised their assessments based on the suggestions of Chen (2023) and Rudolph et al. (2023), who suggested four features.

- (1) Multimodal: analysis of images and other visual media
- (2) Context-specific: connections to what has been taught in class or included in the textbooks; analysis of longer texts that do not fit into a prompt
- (3) Learner-specific: inclusion of self-reflection, personal experiences, and perspectives
- (4) Integration of multiple sources: a combination of the other three elements

Guided by these four principles, the newly designed assessments removed any components for which students could obtain answers from GenAI tools directly (e.g. blog posts), added context specifications (e.g. inclusion of information from lectures and textbook), and increased learner-specific requirements (e.g. integration of personal developmental experiences). The instructions encouraged students to leverage various online language tools, including GenAI, constructively.

As previously mentioned, AI tools can help “level the playing field” for students with low language proficiency; however, students also need to understand the affordances and constraints of this new technology so they can use it appropriately. The revised assessments have been implemented for two semesters, so it is possible to evaluate how students perceive using AI tools for language learning assessments.

2.2. Participants

All participants were enrolled in a course using the new assessment scheme in the summer semester of the 2022–23 academic year. They were in the SYA program (see Section 2.1). Since GenAI tools only became available in late 2022 and college transfer students do not have a significant advantage over other students in terms of understanding GenAI, we hypothesized that their background would not affect the validity of the study.

The course assessments were redesigned to encourage the constructive use of GenAI. This study explores the effectiveness of the redesigned

written assessment. Two types of data were gathered: students’ self-reflections about the assessment (74 pieces in total) and four focus group interviews. The focus groups ranged from five to nine students with a total of 28 participants. While we did not have access to the students’ demographic information, most students taking this course come from hard science disciplines, including engineering, computing, and biochemistry.

In a qualitative study, it is important to address the roles of various participants. The second author, who conducted the focus group interviews, was one of the students’ instructors. However, she only delivered lectures (i.e. in a lecture hall seating 70+ students) and had limited interactions with the participants. Student–teacher interactions mainly took place during tutorial sessions, led by another instructor, who was not involved in the research process. The other two authors did not interact with the research participants.

2.3. Procedures

At the beginning of the semester, students were encouraged to use a range of online language tools, including traditional AI and GenAI tools, to complete their assessments. They were also introduced to two specific AI tools with in-class demonstrations to maximize their benefits: WeCheck!, an in-house grammar and style checker, and Just The Word, a widely utilized word collocation platform (Mizumoto, 2023). More importantly, they were provided with explicit instructions on how to use GenAI tools, such as ChatGPT, constructively and appropriately. The students then completed one writing assessment and one speaking assessment.

As part of the second assessment (Critical Response Writing, which made up 45% of their overall course grades), students were required to complete a reflection on their use of GenAI tools. These reflections served as a data source for this study. Only reflections from students who had given their consent to participate in the current study were retrieved.

At the end of the semester, students were invited to participate in focus group interviews, which were used to complement the self-reflections. The two types of data collected addressed the primary issues under investigation: *how and why* students utilized GenAI tools. Insights into their usage were primarily derived from the reflection data, while their reasons for doing so were discerned from both the reflection and interview data.

2.4. Instruments

This research adopted a qualitative approach, using data from self-reflections and focus group interviews. The overarching goal was to understand how students use GenAI tools in practice, so an open-ended approach to data collection was adopted. The instrument for the self-reflections focused on individual students’ experiences, while the focus group interviews focused on their collective understanding of the issue.

The self-reflection data were produced in response to the following general prompt: “Explain how you have effectively used these online tools [traditional AI and GenAI tools] to complete your assessment.” This approach allowed students to elaborate based on their own experiences, maximized flexibility, and facilitated a wide range of responses.

The objective of the focus group interview was to reach a collective understanding of how students use GenAI tools. As it was a semi-structured interview, there were four guiding questions used with each group, focusing on the following aspects: (1) whether students used any GenAI tools; (2) which tools they used; (3) why they chose to use (or not use) these tools; (4) which other online language learning tools they used. Based on the respondents’ answers, the interviewer asked follow-up questions. All students in the focus group were encouraged to express their views and experiences with using AI tools in the course.

2.5. Data analysis

We performed a thematic analysis of the data collected through the self-reflections and focus group interviews. While the second author conducted the interviews, a research assistant (not one of the authors) was trained to conduct the thematic analysis (Braun & Clarke, 2006). First, the research assistant transcribed the interviews verbatim and read through the transcripts to identify themes. Then, the second author checked the transcripts.

Next, the research assistant adopted an inductive approach to coding the themes that emerged from the data. After this, the assistant grouped these themes based on the tools students mentioned (e.g., “ChatGPT”). Within these categories, the assistant further grouped and categorized the codes that emerged from the data. After that, the assistant merged and split themes when necessary. At this stage, the assistant also identified any themes common to both the self-reflections and focus group interviews. To ensure the validity of the coding/merging process, the first and second authors verified the themes. There were minor discrepancies, which were discussed and resolved.

3. Results and discussion

3.1. Overview of results

This study was designed to understand how students leverage GenAI tools to complete assessments as a foundation for re-formulating assessment policies in higher education. After the meticulous process of coding and verification (based on 456 coded items), a total of four main themes and 12 sub-themes emerged. The main themes were as follows: perception and use of tools, limitations of GenAI tools, versions of GenAI tools, and critical evaluation of tools. Table 1 displays the themes, the corresponding number of coded items, and some representative quotes from the data. While certain themes appeared more often than others, this study adopted an interpretivist approach and focused more on the meaning and significance of the themes than the number of occurrences.

Table 1
Themes and corresponding quotes.

Theme	Number of items	Representative quotes
Perception and use of GenAI tools	n = 321	<p>“I was amazed at how convenient it was to use ChatGPT to brainstorm ideas and search for information. It helped me to quickly understand the problem and identify key points that I might have missed.” (Reflection Group 3, Student 8)</p> <p>“You just put some word[s], but ChatGPT know [s] what you mean. Actually, just put the assignment outline, the background, the guidelines and then let ChatGPT generate an outline.” (Student 4, Interview 4)</p>
Critical evaluation of GenAI tools	n = 66	<p>“I think that WeCheck! is [more] useful rather than Grammarly.” (Student 2, Interview 1)</p> <p>“WeCheck! will be more suitable for Asian students to correct the ‘Chinglish’ and language styling.” (Reflection Group 4, Student 17)</p>
Limitations of GenAI tools	n = 53	<p>“... also sometimes provide the wrong sources.” (Reflection Group 3, Student 3)</p> <p>“For learning, I think [in] some of the cases, you can ask questions, but you can’t assume everything it [ChatGPT] answers right.” (Student 1, Interview 1)</p>
Versions of GenAI tools	n = 16	<p>“I only used 3.5 because I’m a poor guy.” (Student 2, Interview 3) [Note: ChatGPT-4 requires a paid subscription plan]</p> <p>“Grammarly Premium would advise me on a better sentence structure.” (Reflection Group 4, Student 17)</p>

3.2. Use of GenAI tools

The data show that students used a variety of online language tools, including GenAI tools, at different stages of the writing process. During the planning stage, students used GenAI tools (e.g. ChatGPT) for planning, brainstorming, and summarizing. They served as a “writing partner” for students (Gimpel et al., 2023). In the development stage, students needed help from GenAI tools with paragraph development (e.g., composing sentences and organizing ideas). This corroborates with the observation of Kohnke et al. (2023). After drafting their assessments, the students used a range of online tools, including ChatGPT, WeCheck!, and Grammarly, to check their grammar and improve their writing style. Some also used Just the Word, Thesaurus, and QuillBot (a paraphrasing platform) to polish their writing: For example, one student stated that Just the Word helped “make the essay more advanced” (Reflection Group 1, Student 8). However, the first three tools were used most often. They levelled the playing field and motivated students to engage in independent language learning (Al-Imam et al., 2023).

Despite the students’ active engagement with GenAI tools, many were aware of their limitations, including false information, bias, and problematic citations. More impressively, some participants identified limitations related to the course guidelines. For example, one student noted that ChatGPT generated information that differed from the assessment criteria (Student 3, Interview 2). While previous studies (e.g., Dignum, 2023) have reported on the false information generated by GenAI, there has been little examination of context-specific concerns, such as information that is true but irrelevant to the assignment.

3.3. Making informed decisions about online language tools

As previously specified, the participants were strongly encouraged to leverage various language tools, including traditional AI and GenAI, while completing their assessments. This policy thus permitted them to engage in the critical evaluation of different tools in terms of their affordances and constraints, which is evident in their choice to use particular tools at different writing stages. For example, many students utilized ChatGPT for “brainstorming ideas” (Students 1, 3, and 4, Interview 1) and “generating an outline” (Student 4, Interview 4). They then turned to Grammarly or WeCheck! to “check grammar” (Reflection Group 1, Student 15) and refine their “word use and sentence structure” (Reflection Group 1, Student 17). Some also explored the differences between WeCheck! And Grammarly, both of which are grammar and language style checkers. This is evidenced by this comment:

“WeCheck! will be more suitable for Asian students to correct Chinglish [Chinese English] and language styling.” (Reflection Group 4, Student 17)

Such comments indicate that the participants are aware of the benefits and limitations of the online language tools they use. They expressed appreciation for how WeCheck! empowers them to accept or reject the platform’s recommendations (Student 5, Interview 1). Meanwhile, they found value in Grammarly’s immediate and actionable suggestions (Student 2, Interview 3). These insights reveal the students’ critical evaluation of each tool, which enabled them to make informed decisions about which tool to use for a specific purpose.

These findings suggest that understanding the affordances of various online language tools and making informed decisions is an essential aspect of modern, technology-assisted writing. They are corroborated by Mizumoto (2023), who considers the selection process an important step in adopting a range of online tools. Unlike many prior studies that focus on a specific online language tool, the current study prompted students to adopt a diverse range of tools. This assessment policy was designed in accordance with the “balanced approach to AI adoption” recommended by Chan (2023, p. 13). While the students were introduced to an in-house tool, WeCheck!, they were also encouraged to use GenAI tools (e.g., ChatGPT) and traditional AI tools (e.g., Just the Word, QuillBot,

and Grammarly). This balanced approach empowers students to experiment with different tools and determine which one to select based on the purpose and stage of the writing. It facilitates the development of AI literacy and prepares students for the future workplace (Cardon et al., 2023).

The redesigned assessment also enhanced students' ability to evaluate the performance of various online language tools. While GenAI tools play an increasingly important role as writing assistants – particularly among those with low language proficiency, such as the participants in the current study – many students were also aware of the constraints of these technologies. For example, some commented on the “wrong information” (Student 2, Interview 1; Reflection Group 3, Student 3) and “fake citations” provided by ChatGPT (Student 6, Interview 2). Others noted the mistakes in the references on the official website of the university library (Reflection Group 2, Student 3).

Understanding the constraints of various online language tools is one of the core elements of AI literacy, which is closely interrelated with language skills and communication competency. The interdependent relationship suggests that fostering AI literacy should be a major goal of higher education. The centrality of AI literacy has been widely discussed in prior studies (e.g., Cardon et al., 2023; Jacob et al., 2023; Zhang & Hyland, 2023). The current study provides empirical support for their suggestions.

3.4. Training and support for GenAI assessments

While learners endeavor to develop AI literacy, the support offered by higher education institutions cannot be ignored, especially when students are allowed to use GenAI for their assessments. A lack of support may result in the inappropriate and even unethical use of GenAI. In some cases, it can cause a technology gap, where some learners can take full advantage of online language tools but others cannot. Addressing this issue depends on two main factors: a clear and comprehensive policy on GenAI use for assessments and sufficient training for learners (Chan, 2023; Perkins & Roe, 2023). This is evident in the findings of this study, in which students were provided with detailed instructions on how they could use GenAI to complete the assessment. They were also given in-class demonstrations on how to leverage several other online language tools. These detailed instructions and training sessions were widely appreciated by the participants (e.g., Reflection Group 1, Student 5; Reflection Group 1, Student 10) because they helped them understand the language tools, which boosted their confidence and resulted in constructive use (Reflection Group 1, Student 11; Reflection Group 2, Student 16).

Following the implementation of the GenAI policies for the redesigned assessments, this study gathered data exploring how the students used GenAI tools. The purpose of this was to allow instructors to formulate more effective assessment strategies. The process formed a positive feedback loop, where students gained awareness of the affordances and constraints of various tools and instructors learned how to promote the constructive use of GenAI. The results reinforce the utility of various online language tools, which can benefit both students and instructors.

3.5. Equity considerations

Despite the efforts of policymakers and educators to reduce the technology gap among students, a digital divide may remain, even in metropolitan cities such as Hong Kong. This is evidenced by the findings of this study. All of the participants were able to access WeCheck! and the basic versions of Grammarly and ChatGPT (3.5 at the time of the study). The participants appreciated WeCheck! not only because the platform “provides more guidance than I had anticipated” but also because it was “free of charge” (Reflection Group 3, Student 12). This suggests that cost is a relevant factor for these students when evaluating AI tools.

Some students also identified differences between the free and paid versions of AI tools. Many of them used the basic versions of ChatGPT or Grammarly as a “free user” (Student 6, Interview 2) but knew that there were premium versions. Although some students considered the paid versions inexpensive in terms of the cost “per day” (Student 2, Interview 3), others commented that it was not an option for them because they perceived themselves as “poor” (Student 2, Interview 3). This poses an equity issue because the premium version of Grammarly offers more features than the free version, including article summaries (Student 6, Interview 2). More importantly, the free version does not provide specific suggestions on “how to modify” writing (Reflection Group 4, Student 4).

These observations imply that students from disadvantaged communities or low-income families may not obtain as much support as their counterparts who can pay for the premium versions of AI tools. Chan (2023) related this to the governance dimension of institutions and suggested that higher education institutions implement policies to mitigate the equity issues caused by the digital divide.

3.6. Limitations

The findings of this study have provided insights into how students utilized various online language tools when writing. While we adopted strategies to ensure the validity and reliability of the results, some limitations remain. While data were gathered from two sources (i.e. self-reflections and focus group interviews) and the study included 74 students from different disciplines, they were all SYA students taking the same course with similar levels of language proficiency. This could affect the generalizability of the results. For example, students with higher language proficiency may have different needs and adopt different strategies when using online language tools. Second, as a qualitative study, the interviewer and coder could have impacted the results of the study. Their teaching and/or experience with using GenAI could have affected how they solicited answers from the interviewees or coded responses.

4. Conclusion and implications

The results and discussion suggest that online language tools, including those powered by GenAI, provide a level playing field for students with low language proficiency in higher education. This conclusion has significant pedagogical and policy implications.

1. At the pedagogical level, the “balanced approach” proposed by Chan (2023) and supported empirically by the current study, should be promoted. Instructors can strive to promote the use of various online language tools for writing. This can help to prevent overreliance on GenAI tools. Some possible ways to achieve this goal are to “re-imagine” and “re-design” assessments (Dignum, 2023; Kumar, 2023) and expect students to critically analyze online language tools and leverage them in the writing process (Mizumoto, 2023)
2. At the policy level, it is essential to ensure the equitable use of tools. One potential solution is to ensure that all students have access to premium versions of both traditional AI and GenAI tools. This can be organized via the university network. While there may be usage and funding considerations, institutions can explore purchasing premium subscriptions for selected tools. Students can either access the university network on campus or log into the university's virtual network off campus. To ensure that students use these tools effectively, universities can also arrange training workshops and provide instructions. This can help disadvantaged communities to have equal access to online language tools.
3. Another important consideration at the policy level is to integrate AI literacy into the curricula of higher education. This includes critically evaluating AI, using it appropriately, and engaging with ethical issues (Concannon et al., 2023). Previous studies have indicated

concerns about the decrease in critical thinking skills with the availability of GenAI tools (Kumar, 2023; Mizumoto, 2023). The results of the current study, however, suggest that students can derive critical thinking skills through the evaluation of online language tools and the information powered by GenAI. These findings point to the need to promote AI literacy, as it is essential for the future of the AI-driven workplace.

GenAI is an emerging and rapidly advancing field. In the face of students' growing enthusiasm for these new technologies, policymakers and faculty in higher education should encourage students to use a range of online language tools responsibly and effectively. Institutions should also develop AI literacy among students and faculty members. To achieve these goals, one key strategy is to re-formulate assessment policies and practices so that assessments become opportunities for students to practice using AI to further their learning.

Disclosure statement

The authors report there are no competing interests to declare.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

5. Statement on open data and ethics

The datasets used in this study are not publicly available. The authors do not have the authority to share the interview transcripts/recording but other materials are available from the corresponding author upon request.

This study was approved by The Hong Kong Polytechnic University, and all procedures were conducted in accordance with applicable laws and institutional guidelines. Informed consent was obtained from all participants, and their privacy rights were strictly observed.

CRedit authorship contribution statement

Dennis Foung: Writing – review & editing, Writing – original draft, Project administration, Methodology. **Linda Lin:** Resources, Data curation. **Julia Chen:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Al-Imam, A., Al-Hadithi, N., Alissa, F., & Michalak, M. (2023). Generative artificial intelligence in academic medical writing. *Medical Journal of Babylon*, 20(3), 654–656.
- Bobe, B. J., & Cooper, B. J. (2019). The effect of language proficiency on approaches to learning and satisfaction of undergraduate accounting students. *Accounting Education*, 28(2), 149–171.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
- Cardon, P., Fleischmann, C., Aritz, J., Logemann, M., & Heidewald, J. (2023). The challenges and opportunities of AI-assisted writing: Developing AI literacy for the AI age. *Business and Professional Communication Quarterly*, 86(3), 257–295.
- 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(1). Article: 38.
- Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2305.00290>
- Chan, C. K. Y., & Lee, K. K. (2023). The AI generation gap: Are Gen Z students more interested in adopting generative AI such as ChatGPT in teaching and learning than their Gen X and Millennial Generation teachers? *Smart Learning Environments*, 10. <https://doi.org/10.1186/s40561-023-00269-3>. Article: 60 (2023).
- Chen, J. (2023). Four directions for assessment redesign in the age of generative AI. *Higher Education*. <https://www.timeshighereducation.com/campus/four-directions-assessment-redesign-age-generative-ai>.
- Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access*, 8, 75264–75278. <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9069875>.
- Concannon, F., Costello, E., Farrell, O., Farrelly, T., & Wolf, L. G. (2023). There's an AI for that: Rhetoric, reality, and reflections on EdTech in the dawn of GenAI. *Irish Journal of Technology Enhanced Learning*, 7(1). <https://doi.org/10.22554/ijtel.v7i1.116>
- De Costa, P. I., Valmori, L., & Choi, I. (2017). Qualitative research methods. In S. Loewen, & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 522–540). Routledge. <https://doi.org/10.4324/9781315676968-29>.
- Dignum, V. (2023). Responsible artificial intelligence: Recommendations and lessons learned. In *Responsible AI in Africa: Challenges and opportunities* (pp. 195–214). Cham: Springer International Publishing.
- Friginal, E. (2018). *Corpus linguistics for English teachers: Tools, online resources, and classroom activities*. Routledge.
- Furze, L., Perkins, M., & MacVaugh, J. (2024). The AI Assessment Scale (AIAS) in action: A pilot implementation of GenAI supported assessment. *arXiv*. <https://doi.org/10.48550/arXiv.2403.14692>
- Godwin-Jones, R. (2022). Partnering with AI: Intelligent writing assistance and instructed language learning. *Language, Learning and Technology*, 26(2), 5–24, 10125/73474.
- Grammarly. (2023). *Enterprise-Grade Generative AI is now available for Grammarly business customers*. <https://www.grammarly.com/business/learn/enterprise-grade-generative-ai/>.
- Ibrahim, H., Liu, F., Asim, R., Battu, B., Benabderrahmane, S., Alhafni, B., Adnan, W., Alhanai, T., AlShebli, B., Baghdadi, R., Bélanger, J. J., Beretta, E., Celik, K., Chaqfeh, M., Daqaq, M. F., Bernoussi, Z. E., Fougine, D., Garcia de Soto, B., Gandolfi, A., ... Zaki, Y. (2023). Perception, performance, and detectability of conversational artificial intelligence across 32 university courses. *Scientific Reports*, 13(1). <https://doi.org/10.1038/s41598-023-38964-3>. Article 1.
- Imran, M., & Almusharraf, N. (2023). Analyzing the role of ChatGPT as a writing assistant at higher education level: A systematic review of the literature. *Contemporary Educational Technology*, 15(4), Article ep464. <https://doi.org/10.30935/cedtech/13605>
- Jacob, S., Tate, T., & Warschauer, M. (2023). Emergent AI-assisted discourse: Case study of a second language writer authoring with ChatGPT. *arXiv preprint* <https://doi.org/10.48550/arXiv.2310.10903>.
- Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, 54(2), 537–550. <https://doi.org/10.1177/00336882231162868>
- Koraishi, O. (2023). Teaching English in the age of AI: Embracing ChatGPT to optimize EFL materials and assessment. *Language Education and Technology*, 3(1). <https://languedutech.com/letjournal/index.php/let/article/view/48>.
- Kumar, A. H. (2023). Analysis of ChatGPT tool to assess the potential of its utility for academic writing in the biomedical domain. *Biology, Engineering, Medicine and Science Reports*, 9(1), 24–30. <https://doi.org/10.5530/bems.9.1.5>
- Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., & Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *International Journal of Management in Education*, 21(2), Article 100790.
- Malmström, H., Stöhr, C., & Ou, A. W. (2023). Chatbots and other AI for learning: A survey of use and views among university students in Sweden. *Chalmers Studies in Communication and Learning in Higher Education*, 2023(1). <https://doi.org/10.17196/cls.cslche/2023/01>
- Michel-Villarreal, R., Vilalta-Perdomo, E., Salinas-Navarro, D. E., Thierry-Aguilera, R., & Gerardou, F. S. (2023). Challenges and opportunities of generative AI for higher education as explained by ChatGPT. *Education Sciences*, 13(9), Article 856.
- Mizumoto, A. (2023). Data-driven learning meets generative AI: Introducing the framework of metacognitive resource use. *Applied Corpus Linguistics*, 3(3), Article 100074.
- Mohamed, A. M. (2023). Exploring the potential of an AI-based chatbot (ChatGPT) in enhancing English as a foreign language (EFL) teaching. *Perceptions of EFL faculty members. Education and information technologies*. Advance Online Publication. <https://doi.org/10.1007/s10639-023-11917-z>
- Muscanel, N., & Robert, J. (2023). EDUCAUSE QuickPoll results: Did ChatGPT write this report? EDUCAUSE. <https://er.education.edu/articles/2023/2/educause-quickpoll-results-did-chatgpt-write-this-report>.
- Perkins, M., & Roe, J. (2023). Decoding academic integrity policies: A corpus linguistics investigation of AI and other technological threats. *Higher Education Policy*, Advance Online Publication <https://doi.org/10.1057/s41307-023-00323-2>.
- Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024). The Artificial Intelligence Assessment Scale (AIAS): A framework for ethical integration of Generative AI in educational assessment. *Journal of University Teaching and Learning Practice*, 21(06). <https://doi.org/10.53761/q3azde36>. Advance Online Publication.
- Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching*, 6(1), 1–22.
- Sarrion, E. (2023). Using ChatGPT to learn a language. In E. Sarrion (Ed.), *Exploring the power of ChatGPT: Applications, techniques, and implications* (pp. 117–128). Apress: Media. https://doi.org/10.1007/978-1-4842-9529-8_13.

- Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing*, 57, Article 100752. <https://doi.org/10.1016/j.asw.2023.100752>
- Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10(1), 15.
- Warschauer, M., Tseng, W., Yim, S., Webster, T., Jacob, S., Du, Q., & Tate, T. (2023). The affordances and contradictions of AI-generated text for second language writers. *Journal of Second Language Writing*, 62, Article 101071. <https://doi.org/10.1016/j.jslw.2023.101071>
- Yilmaz, R., & Yilmaz, F. G. K. (2023a). The effect of generative artificial intelligence (AI)-based tool use on students' computational thinking skills, programming self-efficacy and motivation. *Computers and Education: Artificial Intelligence*, 4, Article 100147. <https://doi.org/10.1016/j.caeai.2023.100147>
- Yilmaz, R., & Yilmaz, F. G. K. (2023b). Augmented intelligence in programming learning: Examining student views on the use of ChatGPT for programming learning. *Computers in Human Behavior: Artificial Humans*, 1(2), Article 100005. <https://doi.org/10.1016/j.chbah.2023.100005>
- Yilmaz, F. G. K., Yilmaz, R., & Ceylan, M. (2023c). Generative artificial intelligence acceptance scale: A validity and reliability study. *International Journal of Human-Computer Interaction*. <https://doi.org/10.1080/10447318.2023.2288730>. Advance Online Publication.
- Yoon, C. (2016). Individual differences in online reference resource consultation: Case studies of Korean ESL graduate writers. *Journal of Second Language Writing*, 32, 67–80. <https://doi.org/10.1016/j.jslw.2016.04.002>
- Zhang, Z., & Hyland, K. (2023). The role of digital literacy in student engagement with automated writing evaluation (AWE) feedback on second language writing. *Computer Assisted Language Learning*. <https://doi.org/10.1080/09588221.2023.2256815>. Advance Online Publication.