

Research Notes 88

Findings of the Action Research in
ELICOS Program 2024

Issue 88/August 2025

Guest Editors

Professor Anne Burns, Professor of TESOL, University of New South Wales

Sophie O'Keefe, Professional Development Manager, English Australia

Editor

John Savage, Cambridge English Research, Cambridge University Press & Assessment

Typeset in the United Kingdom by Abound Creative

Contents

- 5 Using generative AI in action research**
Anne Burns

- 15 Use of generative AI in the idea generation stage of the writing process: Search engine, debating partner or personal tutor?**
Laura Wakeland

- 29 Building students' confidence in interacting with AI to develop their vocabulary**
Rachel Hunt and Leigh Morgan

- 49 AI literacy for improved critical thinking**
Emily McNamee and Wilhelmina Anne Gevaña

- 67 Working with the enemy: Adopting a generative AI tool to teach AI literacy and increase active engagement with feedback on writing**
Reema Sarwal and Liz Stoyanova

- 85 Impacts of generative AI application use on students' perceptions of pronunciation**
Gabriel Alejandro Azpilcueta and Heather Elizabeth Sparrow, English Language Centre, The University of Adelaide

- 109 The Learning Continuum: Utilising ChatGPT to support academic discussion skills**
Diala Ibrahim and Brad McClymont

Using generative AI in action research

Anne Burns, School of Education, University of New South Wales

Introduction

Over the last few years discussions on the uses of generative AI (genAI) in language education have proliferated as institutions and teaching professionals grapple with how to use this rapidly developing new technology effectively and ethically (e.g., Yang and Kyung 2022, Wang et al 2023, Li et al 2025). The articles in this issue of *Research Notes* are the first in the series of collections from the Australian Action Research in ELICOS Program to report on this area. They illustrate how teachers in this Program, working within their organisational systems, took on the increasingly pressing challenge of whether to incorporate AI and, if so, how to do it using sound and effective practices. Their accounts provide a rich resource for language educators considering how they might themselves experiment in adopting AI resources. They illustrate how the teachers explored the use of AI in their classrooms, collected evidence and, after trial and error in most cases, eventually came up with pedagogically effective ways to incorporate it that resulted in positive responses from their students.

However, my attention was also recently drawn to whether and how AI could be used as a resource for doing action research (AR), rather than exploring the use of AI through AR in the classroom. During a recent conference presentation I gave for the British Council in China, one of the questions a participant asked was ‘How could AI be used for data analysis?’ Her question got me thinking more generally about how AI could be harnessed by teachers wanting to do AR. So, in this brief article my focus will be on responding to the question asked by that participant but broadening it out to consider how AI could be part of the whole AR process. My comments are necessarily programmatic, as I discovered when attempting a search that there is

very little, if any, literature as yet on the question of harnessing AI for doing AR itself. My references to various AI tools are also merely suggestive as there are many new tools emerging constantly as potential resources in education.

My discussion covers how AI could be used for finding and refining the research focus, making decisions about data tools, engaging AI for data analysis, exploiting AI to report and disseminate the research, and considering the role of literature in AR. I conclude with some general implications and caveats that practitioners can take into account when using AI tools.

Finding an AR focus and generating research questions

By using easily available AI tools (such as ChatGPT, Co-Pilot, or Perplexity AI with its hyperlinks to academic articles) teachers wanting to conduct AR can generate, refine and adjust their ideas about possible areas or topics for research exploration. At the beginning of the inquiry, ideas about what to research might be very hazy, or even non-existent. In this scenario, action researchers (ARs) could use prompts asking for ideas about 'hot topics' of current interest in language teaching, or alternatively, first discuss their initial hazy ideas with other colleagues and then request AI to refine or expand on these ideas. This approach can help to brainstorm interesting angles the researcher or their colleagues may not have considered or identify trends and patterns that could be appealing as research topics. AI can rapidly analyse vast quantities of existing literature (and this analysis could be focused on specific topics and trends in AR) and therefore can identify general themes or gaps that could be filled. Gaining such insights can become the basis for further discussion with colleagues to consider whether the AI suggestions relate to the interests of the participants or provide a feasible way forward. These discussions bring in human judgement and integrate the kinds of professional intuition and expertise that critically consider the value and relevance of the AI content to the local context.

Following on from this reflective brainstorming stage, ARs can prompt AI to generate specific research questions relating to the overall topic. For example, I requested Co-Pilot to generate AR questions for the topic: How can I use AI to assess writing for B2 English language learners? In response, these were some of the questions that were offered which could all potentially be used for AR inquiry.

- What AI tools are most effective in evaluating the academic writing of B2 English learners?
- In what ways can AI identify specific areas of improvement in writing abilities for B2 learners?
- How do B2 English learners perceive AI-based assessments of their writing skills compared to human evaluations?
- What impact does AI have on the motivation and engagement of B2 English learners during writing assessments?
- How can AI assist in offering personalised feedback on writing skills to B2 English learners?

- What ethical considerations should be addressed when using AI for assessing writing skills?
- What are the challenges and limitations of using AI to assess writing at the B2 level?

Any of these questions can be further refined and focused using AI, becoming a starting point that can be modified to fit the ARs' situation and preferences. Of course, the questions are generalised ones and do not reflect the localised nature of AR questions that teacher researchers may wish to pursue. For example, Laura Wakeland's study (this issue) illustrates how the question: *How can AI assist in offering personalised feedback on writing skills to B2 English learners?* can be extended and refined to delve more deeply into the issues that interested her. One of Laura's questions, which led to exploring what kinds of feedback AI could offer her students, was:

What is the best way to integrate generative AI into the early (brainstorming and question generation) stages of the research process in writing research papers? (GPT as a search engine, GPT as a debating partner or GPT as personal tutor?)

Having arrived at a question such as Laura's, AI can then help to generate ideas for new teaching strategies or techniques to enliven the 'action' part of the research. While teachers may already have their own ideas, these could be expanded or new angles be provided to mine and interrogate the research questions through action. AI can also be used to generate teaching materials, narratives, images, lesson plans and instructions for teaching once new approaches have been decided. Of course, it is always important for the ideas, strategies and materials generated by AI to be critically evaluated by ARs to judge how appropriate they are to the unique needs of the learners and the pedagogical or research context.

Deciding how to collect evidence

Teachers conducting AR usually begin by putting into practice the teaching and learning strategies and activities they want to explore and reflecting on their effectiveness. An example in this issue is how Rachel Hunt and Leigh Morgan aimed in their classrooms to explore the potential of Microsoft Co-Pilot to build student vocabulary in a personalised, interactive and meaningful way. At the same time, ARs will want to fulfil the 'research' elements of AR by considering what evidence or 'data' to collect to support the direction of their explorations (Burns et al 2022). It will be important to ensure that methods for collecting data are closely in tune with the kinds of questions being asked, but AI can very usefully assist in suggesting different data collection approaches for specific questions. ARs can increase the efficiency and comprehensiveness of the process by prompting AI to make recommendations that are based on analysis of large quantities of existing data and literature.

Depending on researchers' preferences, AI can probe for suitable qualitative methods such as interviews, focus groups or reflective lesson observations, or

quantitative methods like surveys, structured observations or standardised test measures. Alternatively, ARs might want to use a mixture of qualitative and quantitative approaches to provide more rounded perspectives and AI tools can also help to identify what combinations could best be used to answer specific questions. For instance, if the research focuses on the impact of AI-based assessments on the motivation of B2 English learners, an AI tool could suggest conducting pre-and post-assessment surveys to measure changes in motivation levels. In addition, AI might recommend focus groups or one-on-one interviews to gather detailed feedback from learners about their experiences and perceptions. If ARs are uncertain about how to apply or design the data tools, AI can also be prompted to provide explanations about their use.

By integrating these AI-generated suggestions, ARs can design a robust and dynamic data collection strategy tailored to their specific research interests. For instance, in their research reported in this issue, Emily McNamee and Wilhelmina Anne Gevaña began with surveys including both quantitative and qualitative items, and then introduced recorded discussions of student reflections to get evidence for their research on developing their students' AI literacy to improve their critical thinking. This mix of methods, which evolved as their research proceeded, allowed them to reflect deeply on their students' progress but also to identify perceived gaps that could be filled through new directions in their project.

Once researchers have decided which data procedures to use, AI can also be very helpful in providing advice and developing the specifics of the tools that are required. For example, AI can generate possible survey items or suggest prompts that could be used for focus groups related to the specific question. However, while this assistance is very valuable, it will also be important for ARs to bring in their human judgement to evaluate the focus and appropriateness of the items, asking questions such as: *Will each of these items give me the kind of data I'm looking for?; How can the item be refined further?; Will my students be able to easily respond to the way the items are presented?; Is the language level suitable for my students?; Are any of the items ambiguous or difficult to process?* As with any data collection tool, it is invaluable, as well as good practice, for ARs to run what is suggested past mentors, peers or potential participants to find out if the tool as it is presented is a good fit with the purpose of the research.

Harnessing AI for data analysis

Data analysis can feel overwhelming and is time-consuming for busy teachers, although collecting vast amounts of data is not necessary for small-scale AR processes, which are meant to contribute to professional learning and teacher development. Utilising AI resources (such as Microsoft Power BI, Tableau, or RapidMiner) can help to speed up the time it takes for data analysis and they can be used in various ways, especially if there is a sizeable amount of data. For instance, if the researcher has used qualitative recorded interviews or group discussion as Reema Sarwal and Liz Stoyanova did (this issue), they could first read over the responses and develop their own sense of the main patterns, themes and sub-

themes. AI could then be prompted to extract and summarise its own broad themes and subsequently to analyse the sub-themes. The value of doing both your own analysis and using AI is that comparisons can be made to see how well they align. Since ARs are closely inter-connected with the specific environment of the research, it may be that there is contextual knowledge which can shed light on the analysis but cannot be picked up through AI. On the other hand, AI tools (such as Julius AI and Scite) might be very valuable in highlighting insights or findings that might not be immediately apparent. AI can then be helpful in suggesting further areas for change or improvement in the AR process. If more than one source of qualitative data has been used, AI can also help ARs to make cross-comparisons (e.g., between interviews and think-aloud tasks – does the participant do what they say they do?) which can lead to a more complex and complete overview or renewed thinking about appropriate teaching practices.

AI is also efficient in handling quantitative data by streamlining data collection and cutting down on the time it takes for analysis, through automated survey dissemination, collection and summarisation. It can use sensors that generate quantitative results for issues such as student engagement and also undertake data cleansing, a process which identifies and corrects inaccuracies, inconsistencies, and irrelevant data within a dataset. This helps ARs to pick up any errors in the analysis, which increases the integrity of the data. In addition, AI can carry out analyses of relationships within different datasets, which could offer further insights into student performance and teaching effectiveness. The qualitative data collected from surveys can also be transferred into quantitative patterns that identify the predominance of various themes (i.e., which are most and least prominent). All these various types of analysis can assist teachers with informed data-guided decisions, for example in relation to personalising learning by identifying needs and tailoring further interventions, making decisions about the types of resources required for different learners, and enhancing the base for policy and practice within the classroom or, more generally, the teaching centre.

However, using AI responsibly means that human analysis and intervention are essential when inputting data into AI tools. Data analysis requires keen awareness and sensitivity to issues such as privacy or the possibility that confidentiality or identities will be exposed. This is especially true in AR where the number of participants is small and the context is potentially recognisable or where students might be revealing personal information. ARs should also scrutinise the analyses for any potential bias in emerging patterns, be sensitive to cultural or other assumptions that may be evident and make sure they evaluate and monitor the data themselves. Increasing one's ability to generate appropriate prompts and use the technology with some expertise is important to avoid skewed outcomes and unfounded conclusions. ARs can use strategies such as running their data findings past other colleagues or mentors to get peer input. Preferably, as in the case of the teachers in the AR in ELICOS Program writing in this issue, this can be done with a group of AR colleagues working within a similar educational context who will be able to evaluate the analyses critically and interrogate the findings as they relate to the overall research goals and aspirations. AI itself can provide collaborative platforms with other educators locally, regionally and even internationally to increase and

share innovative practices and lead to improved curricula and student outcomes. Tools like Trello and Asana, for example, can organise research tasks, assign roles, and track progress, while others like Packback or Parlay can moderate discussions, provide personalised feedback, and encourage critical thinking among researchers.

Using AI to disseminate AR

AR by practitioners needs to be disseminated so that others can use and share the findings. This means helping not only other teachers but also educational managers, curriculum developers, textbook writers and academic researchers to know about AR done in local environments, and to foster collaboration and systemic change. This knowledge can contribute to improving what is known about how teachers realistically go about investigating and discovering appropriate and effective teaching-learning practices in today's complex classroom situations. It provides accounts of what teachers do rather than what they 'should' do. The accounts in this issue of *Research Notes* are an excellent example of disseminating this kind of knowledge.

Presentations are an accessible way of sharing practitioner AR. Tools such as TOME, Plus AI, FOTOR and Slidebean can be prompted to create presentations rapidly by automating content and incorporating images and data visualisation, although not all AI presentation tools are freely available. These tools are able to customise content to suit the presenter's preferences and save busy teacher researchers' valuable time.

Apart from the obvious possibilities of face-to-face presentations for colleagues, mentors and others, various different types of AI tools can be used for wider dissemination, including AI platforms, podcasts, chatbots and multimedia formats. AI can generate content for poster presentations, using tools such as Poster Pro, Adobe Poster Express or Canva. Posters used in teacher in-house professional development, and teacher research workshops at conferences are well-recognised and accessible ways to disseminate AR. Poster presentations can generate or complement individual and group discussions and enable ARs to synthesise their thinking about how they carried out the research, how data tools were exploited, and what findings, outcomes or further directions should be highlighted. AR posters can also be displayed around the school or staffroom to give other teachers good ideas and to celebrate the research carried out in the institution (Burns et al 2022).

Writing up the AR is also, of course, an invaluable contribution for other teachers and the ELT field more generally. However, so far there is little, if any, research on how AI could be used for writing AR (or any other kind of applied linguistics/TESOL research) articles. However, studies in other disciplines have identified some of the benefits for scientific research writing. Khalifa and Albadawy (2024) surveyed 24 academic articles in their field on using ChatGPT for writing and identified six core domains: 1) facilitating idea generation and research design, 2) improving content and structuring, 3) supporting literature review and synthesis, 4) enhancing data management and analysis, 5) supporting editing, review, and publishing, and 6) assisting in communication, outreach, and ethical compliance. They claim that

ChatGPT offers substantial potential in these areas, but that challenges remain like maintaining academic integrity and balancing AI use with human insights.

Other challenges in written AR reports might be that AI minimises the contextual nuances in the type of narratives that practitioner research prefers. AR is process rather than product-oriented research and is driven by local practice and researcher reflection. By seeking overall patterns, AI may overlook seemingly contradictory findings related to the local context and reduce analysis to checklist-type conclusions that can be overly simplistic or even misleading. These limitations mean that ARs must consider how to develop and reflect on their own technological knowledge so that they understand the processes of AI and do not just accept its content uncritically.

These kinds of insights are likely to be relevant for AR writing, but they are yet to be empirically tested. In the meantime, it is possible to assume that AI could assist teacher researchers in generating content relevant to their AR narratives, refining discourse and language usage, and summarising findings and reflections. As with all research, ARs would also need to be satisfied that the texts generated resonate with their authorial voices and adjust them as deemed appropriate. Language teacher AR writers will also need to pay attention to transparency in relation to using it and ensure attribution of AI-generated content. Although AI has the potential to revolutionise research writing across various domains, writers need to be aware of ethical and transparent use and create a balance between AI usefulness and human insight. More extensive research and professional development in these areas will be helpful to address the emerging challenges and ethical considerations in AI's application to AR. Such issues are very much up for debate and need deeper understanding of the possible agreed protocols for its usage.

Overall, it seems that AI offers great potential for disseminating AR, making it more efficient and accessible, and increasing its impact. By utilising AI platforms, summarisation tools, chatbots, multimedia formats, and network-building technologies, ARs can increase the possibilities that their findings reach a broad audience and drive meaningful change in educational practice. Embracing AI in the dissemination process can empower educators and researchers to collaborate more effectively, share knowledge, and ultimately enhance the quality of language education.

Searching the literature for AR

Readers may have noticed that to this point, I have not placed emphasis on consulting the literature as a basis for AR. This is because AR is essentially an exploratory and practical process, driven by teachers' curiosity, reflection, questioning and puzzling about their practices. Given this kind of orientation, it may not be totally appropriate or essential, as in more 'traditional' scholarly genres, for the inquiry to be framed or directed by theoretical concepts or previous bodies of academic studies. In my opinion, consulting the literature for AR is a moot point. On the one hand, the literature might help to shape and refine the questions ARs have in mind. On the other it might constrain or steer the research in such a

way that creative, unexpected and successful practical ideas are overlooked or deemed lacking in relevance or value, even though they might be highly effective in the local context. In my experience, ARs choose to use the literature before, as, after, or not at all when they do their research. While some want to check out or compare their ideas and findings against those from the academic field, others want the freedom to experiment using their own creative and exploratory pathways. This is a kind of reverse procedure from 'traditional' research approaches, but in AR it can be more relevant and illuminating to turn to the literature during or after the process, as, when or if necessary. Used in this flexible way the literature can provide AR practitioners who may be new to research with some confidence or reassurance that they are in line with broader trends. In this issue of *Research Notes*, readers will find that Gabriel Alejandro Azpilcueta and Heather Elizabeth Sparrow drew quite extensively on the literature to frame their research on pronunciation, while Diala Ibrahim and Brad McClymont utilised ChatGPT itself to generate the conceptual base for their investigation on academic discussion skills.

Of course, AI is a very valuable resource for scrutinising the literature. Tools such as Semantic Scholar or Elicit and Scite.ai provide ARs with ways to explore, discover, and cite relevant literature. Using focused prompts, they can quickly uncover relevant research that resonates with and supports their own investigations and obtain possible new insights and ideas for future exploratory directions.

Implications and caveats

AI is here to stay and used judiciously it provides powerful tools for AR. For educators, it is now essential that we inform ourselves, working to develop our 'AI literacy' (Long and Magerko 2020, Pegrum 2025) and 'AI readiness' (Long and Magerko 2020) to gain more nuanced understanding. Long and Magerko (2020:2) define Alliteracy as: 'a set of competencies that enables individuals to critically evaluate AI technologies; communicate and collaborate effectively with AI; and use AI as a tool online, at home, and in the workplace.' They also introduce the term 'AI readiness' that refers to how educators can understand and implement AI-based technologies in education and beyond.

While using AI, ARs must be conscious of retaining their own control over the process and avoid using these tools as a replacement for furthering their own research creativity and expertise. Maintaining a critical or sceptical stance means carefully evaluating what AI suggests at each stage of the investigation and bringing local contextual knowledge and professional insights to bear on the substance of the content.

ARs will also need to learn how to prompt AI tools discerningly to research the goals they wish to achieve and to share what they learn with colleagues and learners. The quality of AI prompts depends on the level of clarity and focus the researcher employs to guide the tools in ways appropriate to their purpose. ARs will need to hone their skills to ensure they lead with their own conceptual abilities and do not become overly-reliant on simplistic and mechanistic responses. While AI can

be a starting point, it is the quality of the researcher's insights and interpretations that will determine the quality of the outcomes.

ARs also need good awareness of the ethical dimensions of using AI. Teachers can develop their consciousness, and also guide their learners, about possible inherent bias or discrimination embedded in the algorithms used. In addition, critical reflection on how AI might direct or position participants' roles in the research is vital; teachers can question to what extent AI is replacing human interaction and collaboration in the classroom and whether it is serving to marginalise the role of the teacher/teacher researcher and the participation of learners, leading to reductions in their autonomy. In other words, AI should not replace teaching and practitioner research but be seen clearly as a complementary resource. Other ethical aspects to be aware of are the potential for plagiarism, the possible threats to privacy and security, the extent of accessibility, the potential distortion of research findings and the overall impact on human judgements and intuitions about what creates effective teaching and learning. Avoiding the tendency simply to reproduce rather than to mediate and transform knowledge is an important consideration to keep in mind.

Conclusion

AI is a powerful and valuable resource for teachers interested in AR. It can be used to advantage as a dialogic partner to learn, observe, question, reflect and analyse ever more critically and professionally and to create greater research autonomy. As the technology evolves, at a pace that is rapid and extreme, the integration of AI in education will only become more sophisticated, offering exciting directions for the future of AR. However, in using AI, researchers must also be conscious of and responsive to the need to work on their 'AI literacy' and 'AI readiness' skills and abilities. The obvious implication is that those of us interested in AI for AR need to be alert to its promises but also its potential to decrease our autonomy and impede our critical thinking. We must continue to learn to use it thoughtfully, ethically and analytically as a tool to empower initiatives in AR but never to replace the valuable local professional expertise of the researcher.



Left to right: Heather Elizabeth Sparrow, Emily McNamee, Gabriel Alejandro Azpilcueta, Wilhelmina Anne Gevaña, Diala Ibrahim, Rachel Hunt, Sophie O'Keefe (English Australia), Brad McClymont, Professor Anne Burns, Leigh Morgan, Reema Sarwal, Liz Stoyanova, Laura Wakeland

References

- Burns, A, Edwards, E and Ellis, N J (2022) *Sustaining Action Research: A Guide for Institutional Engagement*, New York: Routledge.
- Khalifa, M and Albadawy, M (2024) Using artificial intelligence in academic writing and research: An essential productivity tool, *Computer Methods and Programs in Biomedicine Update* 5 (1), 1–11. <https://doi.org/10.1016/j.cmpbup.2024.100145>
- Li, Y, Zhou, X, Yin, H and Chui, T K F (2025) Design language learning with artificial intelligence (AI) chatbots based on activity theory from a systematic review, *Smart Learning Environments* 12, 24. <https://doi.org/10.1186/s40561-025-00379-0>
- Long, D and Magerko, B (2020) What is AI literacy? Competencies and design considerations, *CHI' 20: Proceedings of the 2020 CHI conference on human factors in computing systems*, 1–16. <https://doi.org/10.1145/3313831.3376727>
- Pegrum, M (2025) From revolution to evolution: What generative AI really means for language learning, *Language Teaching* 58 (2), 1–17. <https://doi.org/10.1017/S0261444825000151>
- Yang, H and Kyun, S (2022) The current research trend of artificial intelligence in language learning: A systematic empirical literature review from an activity theory perspective, *Australasian Journal of Educational Technology* 38 (5), 180–210. <https://doi.org/10.14742/ajet.7492>
- Wang, H, Tlili, A, Huang, R, Cai, Z, Li, M, Cheng, Z, Yang, D, Li, M, Zhu, X and Fei, C (2023) Examining the applications of intelligent tutoring systems in real educational contexts: A systematic literature review from the social experiment perspective, *Education and Information Technologies* 28 (7), 9,113–9,148. <https://doi.org/10.1007/s10639-022-11555-x>

Use of generative AI in the idea generation stage of the writing process: Search engine, debating partner or personal tutor?

Laura Wakeland, Centre for English Language Teaching (CELT),
The University of Western Australia (UWA), Perth

Introduction and context

This research project took place at a time of great upheaval in education. At the start of the project in early 2024, ChatGPT-4o had been released to the public for a year and generative Artificial Intelligence (genAI) had started to become embedded in many platforms and software, leading to easy access and availability.

Most of the students at CELT embraced this new technology, but sometimes they did so problematically. In the classroom, we were faced with students using genAI in ways that were counter-productive to language learning such as prompting it to write, enhance writing or translate texts or audio. In assessments, students used genAI to write take-home assignments such as research papers, breaking rules of academic conduct. In response, UWA updated its policy on academic misconduct to include AI use as follows:

In accordance with the University Policy on Academic Integrity, the use of Artificial Intelligence (AI) is permitted as an educational/study tool. It may only be used in any assessment within a unit where approval has been granted by the unit coordinator. Improper use of AI-generated material, as set out in the Academic Integrity policy, in assessments may lead to the occurrence of academic misconduct. (UWA CELT 2024)

Despite being informed of the policy, our students' use of AI in unacceptable ways only seemed to increase. This situation led to stress and frustration for teachers as they tried to discourage its use in class, and coordinators quickly changed assessments to make them better able to reflect students' actual ability. The most obvious way to do this was to revert to paper-and-pen exams, an unsatisfactory prospect when digital literacy is widely accepted as an essential 21st century skill.

Many teachers (including myself) were slower to uptake the use of genAI either at a personal level or in the classroom. However, it was quickly becoming clearer that as teachers we needed to keep up with the latest technology developments, learn more about how to use genAI, and understand its capabilities as a teaching and learning tool. We could not simply deny its existence or ignore that students would soon be using it in their future disciplines, professions, and workplaces.

Research focus and questions

It was this context which informed many of the decisions that I made in my research design. Firstly, the use of genAI in my research could not allow the students to break the academic integrity policy, meaning that its use in the writing, editing or proofreading stage of the writing process was risky. Therefore, the early stages of the writing process, including the ideas generation and basic research stages, seemed like a better choice. Consequently, I decided to incorporate genAI into the lessons leading up to two written assessments: a 750-word in-class essay and a 1,500-word research paper on a research problem in students' disciplines. Secondly, since digital literacy is an essential skill, I wanted to frame AI in a constructive way, in contrast to the statements often used in the context of academic misconduct. I wanted students to grow more awareness of appropriate and acceptable ways to use AI in their language learning and future academic studies, while also approaching it with a critical eye. Thirdly, since I had very little experience or understanding of AI, I needed to build on the work of others, but the literature on using it in the classroom was scarce because of the newness of the technology. I found the suggested activities of Mollick and Mollick (2023a) of great practical use but since they were not designed with English as a Second Language (ESL) students in mind I wondered if they could work in a language classroom. Their tasks required students to be reflective, critical and aware of the limitations of AI but many students come from an academic and cultural background that does not emphasize this. In addition, the language in the response from AI could be too complex for them due to their level of English proficiency. So, it seemed clear that adaptation for ESL students was necessary. The last decision that I made was to use Microsoft Edge's Co-Pilot as the AI platform since students have access to it through the university. In the remainder of this paper, GPT refers to GPT as accessed through Co-Pilot.

The following research questions were finally decided on:

1. What is the best way to integrate genAI into the early (brainstorming and question generation) stages of the research process in writing research papers? (GPT as a search engine, GPT as a debating partner or GPT as personal tutor?)

2. Will this role of AI give students a new skill and appreciation of AI as a topic exploration tool, and can this be considered as a form of transferable AI literacy?

Participants

My research was conducted over two cycles covering two separate cohorts of The Academic English and Study Skills Bridging Course at CELT. The course is designed for international students who want to study at UWA but do not have the required level of English. If students can pass the Bridging course, then they may enter their degree program. In my classes, the vast majority of students were Chinese, with ages ranging from 19 to 32, and most were entering Master's Degrees by coursework. There were 16 students in each class and their proficiency was equivalent to IELTS 5.5–6.0. The course is intensive and challenging for most students, and this combined with high expectations from back home and adjustment to living in a new country, can cause them to feel under tremendous pressure. This stress and anxiety could be a significant reason why they are tempted to pass off the language of AI as their own, and another compelling reason for us to demonstrate acceptable uses.

Research design

The class activities were conducted over four sessions:

Session 1: Research ideas generation

The first session focused on the research paper, which required students to brainstorm issues and problems in their disciplines to help them decide on and narrow down their research topics. However, many students lack familiarity with their discipline and the technical terminology and phraseology used. I chose this as a point to incorporate genAI in the hope that it could fill the gap by generating lists of issues or topics in specific disciplines, explaining terminology and providing students with more keywords to search the library database effectively (see Appendix 1).

Sessions 2 to 4: Gathering ideas about globalization

Sessions 2 to 4 focused on an argumentative essay on the topic of globalization. Students need to prepare for the essay exam by reading and collecting information on different standpoints using set texts to use as sources. I chose this as a point to incorporate AI because they could use it deepen and broaden their understanding of the texts.

The second session focused on GPT as a search engine to help students generate ideas on the positive and negative effects of globalization and the economy (see Appendix 2). This session involved students comparing the results of the same search terms using a search engine (such as Google) and Co-Pilot. Inspiration for this task came from Dharmapuri and Harrington (2023). This activity aimed to raise students' awareness of the strengths, weaknesses, and capabilities of both search

engines and GPT. Directing students in class to an article by Griffith (2023) was also helpful for consolidating students' understanding of the differences between the two tools.

The third session incorporated GPT as a debating partner to help students learn about the negative and positive effects of globalization on culture. By taking a stance and practising debating with GPT, students could become more aware of the strongest arguments, the counterarguments and the rebuttals, and could incorporate them into their essays. The toolkit from Paquet (2024) was useful in helping me to create the prompt for a debating partner (see Appendix 3).

The fourth session used GPT as a personal tutor. It aimed to test what students had learned from the reading and guided them through brainstorming positive and negative effects of globalization on the environment, before finally asking them to write their own summary. My prompt for the tutor was adapted from Mollick and Mollick (2023a) (see Appendix 4).

Data collection Cycle 1

Initial questionnaires

In the first session, students completed questionnaires so that I could gauge their previous experience with and attitude towards genAI.

Records of the students' interaction with Co-Pilot

After each session, students copied and pasted their prompts and AI's responses so that I could analyze their interactions.

Students' reflections

Students wrote directly after each interaction about how engaging, useful, or effective they found the activity.

My reflections

After each session I wrote a record of observations and insights that I had gained.

Whole-class discussion

At the end of the whole research cycle students gave feedback and shared ideas and interpretations.

Triangulation

Data was collected using a triangulation approach (Burns 2010). The records of students' interactions provided objective evidence of the discourse, while the students' reflections provided more subjective evidence of their lived experience. Finally, my own reflections of what I observed as a teacher during the activities provided the third side of the triangle.

Findings: Cycle 1

Pre-class survey results revealed that most students had used GPT before and had a positive attitude towards it.

I used an inductive approach when analyzing the students' records of their interaction with AI, looking for patterns without any pre-prescribed categories (Burns 2010), and observed that there was a spectrum in the quantity and quality of the interaction. Some students had had more back and forth interactions with GPT and had written longer responses, asked questions if they did not understand, fine-tuned their prompts, and had focused on areas that interested them. In contrast, others had had only one or two interactions without moving on much from the original basic prompt and then stopped. After cross-checking with the students' reflections, a connection emerged between students with more quantity and quality interactions and their higher level of satisfaction in the task in all four of the sessions. In addition, those who had a more conversational style and polite tone with the AI tended to have a more positive experience. English proficiency also seemed to play a role in the students' level of interaction and satisfaction with the task, with those with lower proficiency struggling to understand the purpose of the task and what they were expected to do. I reflected over the course of Cycle 1 that although these activities appeared to be self-directed, independent tasks, this was not the case. Students needed constant monitoring to keep them on task, give them encouragement and feedback, and help with troubleshooting during the lesson. Despite this need, however, students appeared engaged during the sessions.

The final research discussion from students revealed that their overall attitude to the four sessions was positive. They overwhelmingly preferred the search engine function which received 12 votes, in contrast to the debating partner (one vote) and tutor (two votes). Many found the debating and tutor methods too difficult to understand or irritating in some way – for example:

- *'I can never win this debate – AI is too clever.'*
- *'Write a prompt to make it more human!'*
- *'Say I'm an ESL student, speak simply.'*

All these findings made me question why some students had better interaction than others, whether there was a way to improve interaction and whether there was some way to improve the debating and tutor prompts to make them function better, leading me to make changes for Cycle 2.

Changes made for Cycle 2

I decided to change my first research question because I realized that while the method of interaction was interesting, the quality of interaction was perhaps more important:

What is the best way to **encourage interaction** with genAI in the early (brainstorming and question generation) stages of the research process in writing

research papers – using GPT as a search engine, GPT as a debating partner or GPT as personal tutor?

I revised lesson plans and slides to include more on AI literacy. I also changed the prompts for debating partner and tutor to include a close for the interaction so that the task could be finite and students could receive feedback. Subsequently, I held a workshop for other teachers at my centre in which I asked them to try some of the prompts that students had used in class. For many teachers it was the first time that they had tried using AI. Two teachers who co-taught another class group in the same module of the Bridging Course volunteered to join me in the research, using the prompts in their classes and following up with an interview of their experiences.

Findings: Cycle 2

Again, pre-class survey results revealed that most students had used GPT before and had a positive attitude towards it.

In Session 1, all students had minimal interaction. Some did not find anything new or had some reservations about reliability or negative effects on their thinking, but all said they would use it again in the future due to the speed and efficiency.

In Session 2, around two thirds of students had more interaction with Co-Pilot. These students asked clarification questions, requested definitions and examples, followed a line of interest, or pointed out conflicting information. Regardless of the level of interaction, all students found it useful. This increase in interaction was probably due to more AI literacy materials taught beforehand and my active encouragement and reminders during the class.

In Session 3, reflections were positive and students were more engaged than in Cycle 1. Many students commented that the experience was interesting, gave them new perspectives and ideas, and they appreciated the fast speed. Others said that it helped them to organize their thoughts and think critically. Some appreciated the polite and empathetic tone of the AI which made them feel like they were debating a real human. However, some students also mentioned some negative points such as GPT going off topic, not giving them anything new and not providing strong enough evidence. Also, for some it did not feel like real debating, and they were wary of trusting that AI is correct. One student commented that 'I feel powerless because I am completely unable to beat the machine in the debate' (see Appendix 5 for an example of an interaction).

For Session 4, again the comments were mostly positive, and the level of interaction was sufficient. Some students liked the feeling that they were talking to a real tutor. Many found 'the tutor' interesting, meaningful, refreshing, effective, patient and flexible. They thought that it helped them to enrich their understanding, consolidate knowledge and get personalized feedback. One student found that if the response was not satisfactory, he could prompt GPT to get the answer he needed: '[a]fter I give it advice, it can change the answer better. And I learned how to get him to help me better' (See Appendix 6 for an example of an interaction).

The whole-class discussion revealed a more balanced preference with GPT as search engine receiving seven votes, tutor seven votes, and debating partner three votes. Students believed that the debating partner and tutor made them interact the most, but GPT as search engine helped them to get the most useful information.

Teacher's perspectives

As mentioned previously, two teachers joined me in trying out the AI materials and had positive attitudes towards them. Teacher 1 had 20 years of experience and had used AI only a few times. She believed that using AI in this way helped students progress more quickly because they chose their research topic faster, leaving more time to spend on other steps of the writing process. Teacher 2 had two years of teaching experience and used AI frequently. She believed that 'AI is great' and shared ideas for improving the activities, for example, asking students to demonstrate connections between the GPT search and the articles chosen. Both teachers felt that they had become more aware of the possibilities of AI in the classroom and had gained some new ideas of their own through participating.

Discussion

My first research question asked which was the best method to encourage more interaction with AI. It seems surprising that interaction even needs to be encouraged seeing that students have a positive attitude towards AI and use it often. However, this research found that even though students may be digital natives, they are not always aware of the capabilities of AI, how to direct it to perform to their needs, or to use it as an active learning tool. Students seemed used to using AI to passively get quick answers, but unaccustomed to using it in a more critical way. This may be the reason why students preferred the search engine function over the debating and tutor functions, while as teachers we may see the latter as more valuable tools. However, this research also showed that even when using the simple search engine function students need to be taught how to use it more effectively by promoting interaction. As Mollick and Mollick (2023b) said in their video, 'to work with it interactively you have to push back and this is really where the magic happens.'

My second research question asked whether these activities could give students new skills and appreciation of AI as a topic exploration tool. Again, this research showed that being digital natives does not correspond to understanding how AI works, what acceptable and unacceptable uses look like, or awareness of ethical issues, bias or limitations. For example, many of my students were not aware of the possibility of AI 'hallucinations' defined as 'incorrect but plausible facts' (Mollick and Mollick 2023a:5). Therefore, teachers are still needed to articulate and demonstrate these important aspects of AI literacy as they are not concepts that may be discovered simply through passive use. Fortunately, students are curious and open to learning AI literacy and are likely to be engaged by the topic. Although there were some frustrations, overall students felt that they had learned something. Therefore, teaching the three ways of using AI to gather information can be considered as a

practical component of AI literacy, provided that the technical and ethical aspects of AI literacy are explicitly taught holistically alongside them.

Tips for teachers

For any teachers wishing to try out these or similar activities in class, here are some things to be aware of:

- GPT can produce responses that contain idioms, phrasal verbs and emojis, meaning that ESL students struggle to understand. You can try to program these out in the prompt, or students can request GPT to stop using that language if it appears.
- Even with the exact same prompt, AI may work in some cases but not work in others. For example, instead of acting as a debater, it may simply list arguments for and against. This can sometimes be fixed by starting a new chat and trying again.
- Even when it is working correctly, GPT produces very different results from student to student even with the same prompt. This can be seen as a strength rather than a weakness but may be confusing for students.
- GPT has no problem understanding students' language even if it has grammar, spelling, or punctuation errors and can respond using precise vocabulary. This has drawbacks and possibilities.
- The interface and functions are constantly changing without warning, so something that you planned may need to be altered. A flexible approach is needed.

Conclusion

As 2025 moves into its second half, dealing with AI is still the biggest issue we are facing. After completing this project, I believe that including teaching materials like these can improve the situation. By demonstrating and guiding them through practical tasks where they are using AI appropriately, students can understand the difference between acceptable and unacceptable use in an academic context. This can help to reduce some of the stress and anxiety which might cause them to cheat by using AI. It is also important to couple the practical tasks with AI literacy instruction so that they have a rounded understanding of how AI works and the ethical implications attached. The positive attitude demonstrated by the students and the teachers who tried out the activities indicates that introducing some activities like this into an academic bridging course is viable and valuable.

References

Burns, A (2010) *Doing Action Research in English Language Teaching: A Guide for Practitioners*, New York: Routledge.

Dharmapuri, S and Harrington, T (2023, February 21) *Google Search vs. ChatGPT: Which is better?*, Sanborn Geospatial. <https://sanborn.com/blog/google-search-vs-chatgpt-which-is-better/>

Griffith, E (2023, February 15) *ChatGPT vs. Google Search: In Head-to-Head Battle, Which One Is Smarter?*, PCMag Australia. <https://au.pcmag.com/the-why-axis-serie/98770/chatgpt-vs-google-search-in-head-to-head-battle-which-one-is-smarter>

Mollick, E and Mollick, L (2023a) *Assigning AI: Seven Approaches for Students, with Prompts*, Wharton School of the University of Pennsylvania/Wharton Interactive. <https://doi.org/10.48550/arxiv.2306.10052>

Mollick, E and Mollick, L (2023b) *Practical AI for Instructors and Students Part 3: Prompting AI*. <https://www.youtube.com/watch?v=wbGKfAPIZVA>

Paquet, S (2024, February 20) *A Practical ChatGPT Toolkit for Teachers*, Center for Teaching and Learning (CTL), Champlain College Saint-Lambert. <https://eductive.ca/en/resource/a-practical-chatgpt-toolkit-for-teachers/>

UWA CELT (2024) *Academic Misconduct Policy*. <https://www.uwa.edu.au/celt/-/media/project/uwa/uwa/centre-for-english-language-teaching/documents/academic-misconduct-final.pdf>

Appendix 1: Prompts for Session 1

1. What is the discipline of [insert own discipline] about?
2. What are the important issues in [insert own discipline] now?
3. What topics of interest are there within [insert own discipline]?

Appendix 2: Prompt for GPT as a search engine

What are the positive and negative effects of globalization on trade and the economy?

Appendix 3: Prompt for GPT as a debating partner (Cycle 2)

You are a university student who belongs to a debating club. I am another student who belongs to an opposing debating club. Debate with me on globalization and culture. Do not write my answer for me but let me put forward my own arguments. Rebut my arguments but keep the answers short and confined to one main point. Use some argument techniques that a professional debater might use. Use an engaging and jocular tone to keep my interest. I am an ESL student so keep the language simple and avoid idioms. This debate will help me understand different sides of the issue so please do a good job. Please let me win some arguments to help build my confidence and understanding. After three turns each, finish the debate and give me an evaluation of my performance.

Appendix 4: Prompt for GPT as a tutor (Cycle 2)

You are a friendly, helpful tutor. Help students learn about the positive and negative effects of globalization on the environment. Tailor your examples and explanations for university students who have English as a second language, and who are familiar but not deeply knowledgeable about the topic. Ask one question at a time. Students often struggle with terminology. As you work with the student check for these errors and provide explanations if needed. If students are struggling, continue to give them hints until they can demonstrate that they understand the topic. Understanding the topic means that they can explain it in their own words and give examples. Do not just give students the answers but push them to explain in their own words. First, introduce yourself to the student and help them understand both sides of the debate about effects of globalization on the environment by asking them questions and offering explanations and examples. After they have provided one positive and one negative effect, ask them to provide one more positive and then one more negative effect. Next, ask the student to summarize the topic in their own words. Lastly, tell the student about any important effects that they have not mentioned.

Appendix 5: An example interaction using prompt for GPT as a debating partner

Co-Pilot

Bring it on! I'll start: Globalization is great because it spreads innovation and new technology across the globe, improving everyone's lives. Your turn.

Student

I don't think so, globalization will make the world become more uniform and lack of diversity. For example, every country all has their own culture, but globalization will fade the colour of their culture.

Co-Pilot

Good point. But think about how globalization can actually enhance cultural exchange. We get to experience food, music, and traditions from around the world, right in our own cities. It promotes mutual understanding and appreciation of different cultures. What's your counter?

Student

You're true, but people will feel confused about the culture which is from their own countries or other countries if they always dipped themselves into various culture in their daily life.

Co-Pilot

Interesting point. But think about this: people who grow up in diverse environments often develop a broader worldview and become more adaptable. Isn't that a benefit for individuals and society? Your turn.

Student

However, they will prefer to other country's culture instead of that from their own countries, just like in China, many teenagers more like some western festivals, such as Christmas day and Halloween. They think western festival more interesting, their traditional festivals are boring.

Co-Pilot

That's a valid concern. But consider how the blending of traditions can lead to new, unique cultural expressions. Fusion festivals, anyone? It's a mix that can keep traditions alive in new and exciting ways. What's your final point?

Student

But some traditional festivals represent your own countries history. If you add new elements into it. They will lack the traditional meaning of its own.

Co-Pilot

Great argument! Mixing new elements with traditional ones can indeed risk diluting their original meaning. However, wouldn't you say that traditions are not just static relics, but living practices that evolve over time? Globalization can be the catalyst for new traditions that still honor their roots.

Now, let's wrap up. You presented clear arguments and did a fantastic job balancing both sides of the issue. You pointed out key concerns about cultural dilution and confusion, which are valid and need attention. The way you used specific examples like Chinese teenagers celebrating Western festivals was strong and relatable. Well done! With more practice, you'll only get better. Keep debating!

Appendix 6: An example interaction using prompt for GPT as a tutor

Co-Pilot

Hello! I'm here to help you understand both the positive and negative effects of globalization on the environment. Let's make this as smooth as possible.

To start: Globalization has led to the spread of ideas, technology, and trade around the world. This has a major impact on our environment, both good and bad. Do you know any positive effects that globalization might have on the environment? If not, I can give you some hints!

Student

Maybe some countries have advanced technology and policy about how to govern the environment, they can share these things to other countries.

Co-Pilot

That's a fantastic observation! When countries share advanced technology and policies, it can lead to improved environmental practices globally. For example, renewable energy technologies like solar panels and wind turbines have been adopted worldwide, reducing reliance on fossil fuels.

Now, what about the negative side? What are some harmful effects globalization might have on the environment?

Student

Due to the production of globalization module, some countries are responsible for only one part of production not the whole production. If this part has very seriously negative effects on environment, these countries will face severe consequences.

Co-Pilot

Exactly. This is often called the 'outsourcing' of pollution. For example, heavy industries like manufacturing might be concentrated in one country, leading to severe air and water pollution there. This leaves those countries facing significant environmental degradation and health problems.

Now that we've discussed one positive and one negative effect, can you think of another positive effect globalization might have on the environment?

Student

It's not necessary for the negative production to spread all over the world. We can choose one place for especially processing the bad pollution by many countries' cooperation.

Co-Pilot

Excellent thought! When countries collaborate, they can create specialized facilities to handle pollution more efficiently, reducing the overall environmental impact. This centralized approach can mitigate environmental damage while also fostering international cooperation and shared responsibility.

Now, let's think about another negative effect of globalization on the environment. What else comes to mind?

Student

When people make trades all over the world, some diseases or bad species will spread more rapidly and worldwide.

Co-Pilot

Spot on. This phenomenon is known as 'biological invasion' or the spread of invasive species. A well-known example is how the trade of goods between continents led to the spread of the Asian long-horned beetle, which has caused significant damage to trees in many countries. Similarly, diseases can travel quickly due to increased human mobility and global trade.

You've done great explaining these effects. Could you summarize the positive and negative effects of globalization on the environment in your own words?

Student

The positive aspects are sharing advanced technology and policy for environmental protection and finding a good way to concentrate pollution to fix it. On the other hand, the negative aspects are outsourcing of pollution and biological invasion.

Co-Pilot

Well put! You've summarized it really well.

To add on, another positive effect is the spread of environmental awareness and education through global communication platforms. On the negative side, globalization often leads to increased consumption and waste production due to greater access to goods.

You've done a great job tackling this topic. Is there anything else you'd like to discuss or need help with on this subject?

Student

No thanks.

Co-Pilot

Understood! If you ever need help with anything else or just want to chat, I'm here. Have a great day!

Building students' confidence in interacting with AI to develop their vocabulary

Rachel Hunt and Leigh Morgan, UTS College, Sydney

Educational context

This action research (AR) project was conducted at UTS College, the higher education pathway provider to the University of Technology Sydney (UTS). The students who generously agreed to participate in the project were enrolled in the English for Academic Purposes (EAP) course at AE3 (IELTS 5), AE4 (IELTS 5.5) and AE5 (IELTS 6.0) levels. Students participating in this project came from China, Japan, Laos, South Korea, Taiwan, Thailand and Vietnam, with ages ranging from 18 to 30. On completion of their EAP course, they articulated into a range of undergraduate or postgraduate degree courses at UTS within the faculties of Business, IT, Design and Architecture, Science, Communication, and Engineering.

Research focus

Because of the varied nature of the students' pathways, the curriculum at UTS College has been designed to enable them to begin engaging with concepts, language and vocabulary related to their intended program of study at university. While all AE3 students write an essay on the same given topic, students at AE4 write an evaluative essay on a question related to their subject area, and students at AE5 write a research question related to an issue within their intended discipline, and then research and write an essay in response to this question. Students at all of these levels are required to search for appropriate texts to support their arguments, and to research and deliver a group presentation on a topic related

to their discipline. Students in such educational contexts are expected to increase their discipline-specific vocabulary (Milton 2009), and our students currently do this through the use of supplementary resources such as Quizlet and the Academic Word List (Coxhead 2000).

We have noticed that there is little opportunity for students to interact with these resources to clarify or extend meaning, which means that this usually falls to translation tools, dictionary apps, or input from the teacher. This can be difficult for the class teacher to manage, as our students have very individual needs regarding their vocabulary learning, depending on their future course of study. Students themselves often identify the need to develop their vocabulary, as indicated by these learning goals (Figure 1), which were produced by AE4 students during a beginning-of-course activity.

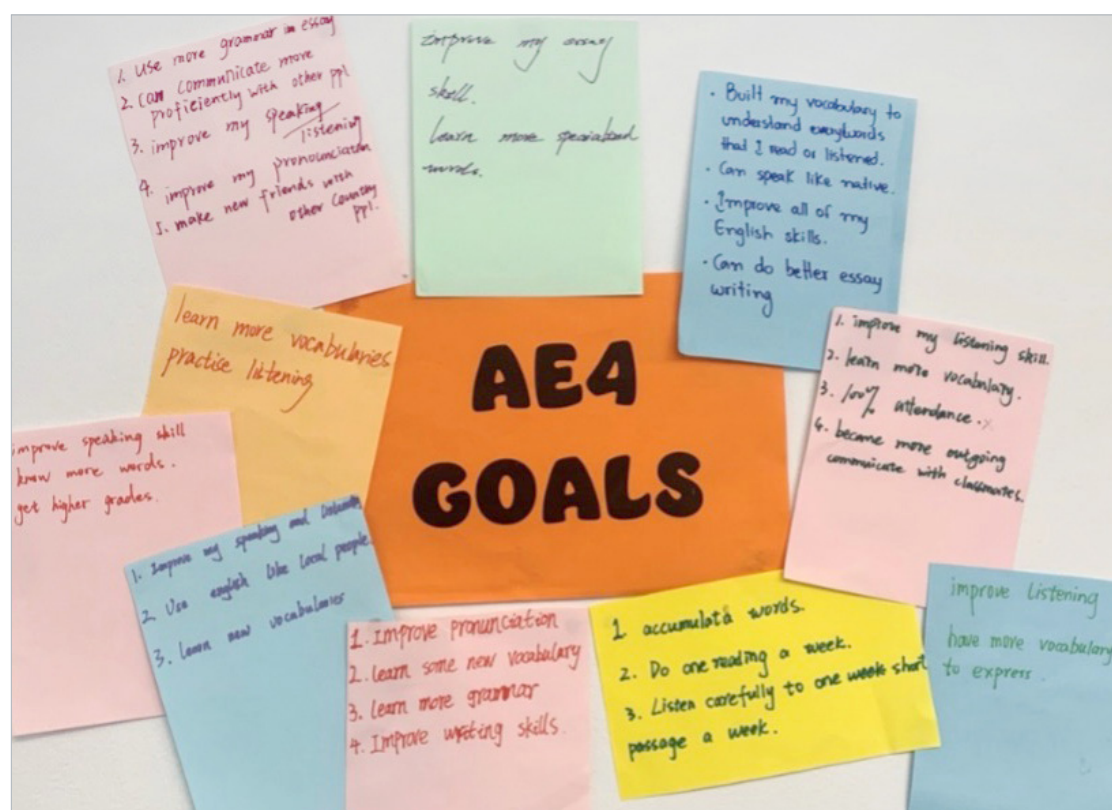


Figure 1: Student goals (AE4), showing that several students identify learning vocabulary as important

We chose to investigate the potential of generative AI (Microsoft Co-Pilot¹) to assist students with building their vocabulary in a personalized, interactive way which offers more meaningful support than the supplementary resources such as the Academic Word List described above.

¹ Co-Pilot was selected for reasons of privacy, access and equity, as all students have free access to the data-protected version of Co-Pilot through their institutional Microsoft Office 365 accounts.

We have noticed that many of our students are aware of some of the possibilities offered by generative AI (genAI), and our initial surveys with our students about their AI use supported our perceptions. We found that:

- Most students are already using AI to help with their English in some way
- Some students are aware of the advantages and disadvantages of AI
- Most students would like to learn more about how to use AI to improve their English

This supports the claim that there is a growing need for clarity and education around the ethical use of genAI (Liu et al 2023). In response to this need, many universities are developing frameworks to guide staff and students on how to use this technology appropriately. We used two of these as guiding principles for our project: the RAFT framework developed at Sydney University (Liu 2024), which focuses on rules, access, familiarity and trust; and the five principles for the effective ethical use of genAI published by the LX Team (2024) at UTS. Common elements to both of these frameworks are the need for students to experience the strengths and weaknesses of AI as an aid to learning, to engage with AI critically, and become comfortable with using AI. We hoped that our project might provide our students with a chance to experience using AI in a way they might not have done previously, and to evaluate its benefits and limitations for vocabulary learning. We also hoped that enabling students to interact with Co-Pilot and evaluate its responses would not only increase their vocabulary but also prepare them for using genAI ethically and effectively in their university studies and future workplaces.

This background led to our research question: How can students interact with Co-Pilot to develop their academic vocabulary?

Approach

In our first research cycle, students wrote a short (400-word) essay, which they submitted to Co-Pilot with a prompt asking it to identify non-academic vocabulary and provide alternatives which students could then use in their redraft. This attempt was largely unsuccessful for several reasons. Firstly, Co-Pilot often provided inappropriate alternatives to the non-academic vocabulary. For example, one student in Leigh's AE5 class had identified 'parents' as simple vocabulary and asked Co-Pilot to provide higher-level alternatives. It offered 'prehistoric ancestors' and 'progenitors' as synonyms, which were clearly not suitable substitutes. We also found that Co-Pilot was not consistently able to identify non-academic vocabulary; in one of Rachel's AE4 classes, it highlighted 'fast food' as the only non-academic word in one student's essay, while in another essay on the same topic it identified several more examples. Additionally, the temptation to ask it simply to rewrite the whole essay (despite our specific instructions not to do so) proved too strong for some students, which negated the remainder of the process and created academic integrity issues.

This problem led us to revise our approach in Cycle 2 (as described below), and to focus more on the interactions our students were having with Co-Pilot, with the aim of building their confidence. We also expanded the range of vocabulary activities undertaken with students at the different levels. This approach was much more successful, as it led to students interacting more effectively with Co-Pilot to actively learn vocabulary and being provided with more appropriate vocabulary, as well as reducing the likelihood of academic misconduct.

For our second cycle of research, data was collected through qualitative surveys, samples of student work, and our own observations of the students in our classes.

Method

Each student was provided with an individual vocabulary record spreadsheet, with tabs for general vocabulary, subject-specific vocabulary, a set of prompts to use with Co-Pilot (see Appendix 1), and the weekly vocabulary list for each level. The spreadsheets were in a shared folder so both the students and we had access. The prompts were carefully tested before being given to the students to ensure a clearly worded prompt (see Appendix 1 for examples) that would generate a useful response (Mollick and Mollick 2023). The persona-style prompt (*I don't understand your explanation of XXX. Assume I am 12/15 years old. Explain again.*) was surprisingly popular with the students; it was clear from our observations that not only did they enjoy pretending to be a 12-year-old, but also that Co-Pilot gave them explanations that they could understand. Later, in Figure 9, we provide an example of an interaction using this persona-style prompt.

To encourage students to interact with Co-Pilot rather than simply treating it as a dictionary, we modelled the interaction in a whole-class activity. First, we elicited from students the kind of questions they would usually ask a teacher about vocabulary and then we demonstrated how to use these questions with Co-Pilot. This enabled us to guide students through the process, asking them if they understood the response, what they could ask if they didn't understand and so on, in order to build the interaction with Co-Pilot and show them that this could be more than just a simple question and answer exchange.

Vocabulary activities

We used the following approaches in writing lessons and reading/listening lessons in which all participating students across the three different class levels interacted with Co-Pilot to develop their vocabulary:

Writing

- Students worked in groups to analyse an essay question and brainstorm relevant key words and phrases into a preliminary outline for the essay ².

² The essay questions used with students were for no-stakes short essays which form part of the Weekly Writing Program at UTS College. Students were advised that they should not be using this

- After brainstorming the vocabulary, they identified which words they would use in the body paragraphs of their essay.
- They then used the prompts to interact with Co-Pilot and ask it:
 - to provide a list of more academic, higher-level synonyms for these words and phrases
 - to clarify meaning and request definitions or examples and ask for simplified definitions or examples as needed
 - to create a table with the original items, possible synonyms, a definition, and example sentences.
- Students could then copy and paste the table into their spreadsheet as a permanent record for review ³, and choose vocabulary from it to use when writing their essay.

Listening and reading (replacing a teacher-fronted pre-teach stage, when students were more confident with prompting)

- Students were shown the list of vocabulary to be pre-taught (Figure 2).
- They identified unknown words and then interacted with Co-Pilot to gain an understanding of the meaning (Figure 3).
- Students added the new vocabulary to their spreadsheet.
- They then completed the pre-text vocabulary activity (often a gap-fill or matching task) to check understanding, with additional clarification of form or pronunciation from the teacher as needed.

Work with a partner (do NOT use a dictionary or translator):

1. Which of these words do you know? Define them
2. Which words are new?

transform
eliminate
insurance
conventional
transportation
pedestrian
medical
data
crash
track record

Figure 2: List of vocabulary for pre-teaching

Open CoPilot and log in with your UTS account:

1. Ask it for a definition or an example sentence of the words you DON'T know
2. Ask it more questions until you are sure you understand the meaning
3. Ask it to give you different word forms for the word (e.g. noun, verb, adjective)

Explain the meaning of the new words to your partner

transform
eliminate
insurance
conventional
transportation
pedestrian
medical
data
crash
track record

Figure 3: Instructions for Co-Pilot interaction

approach with Co-Pilot for their assessed essay, as this would breach the current Academic Integrity policy.

³ At the time of writing, the data-protected version of Co-Pilot is unable to save chat history. Therefore, the spreadsheets are the only record of the vocabulary.

In addition, we undertook the following class-level specific activities:

AE3: Spelling Test (Leigh)

I gave students in my AE3 class a regular spelling test under strict conditions with selected words from one of the day's lessons (students hadn't seen these beforehand; I spoke each word aloud twice). I projected the vocabulary list onto the whiteboard at the end of the test for students to peer-mark (Figure 4). I observed that it was challenging for students to spell these new words, as evidenced by the number of incorrect answers.

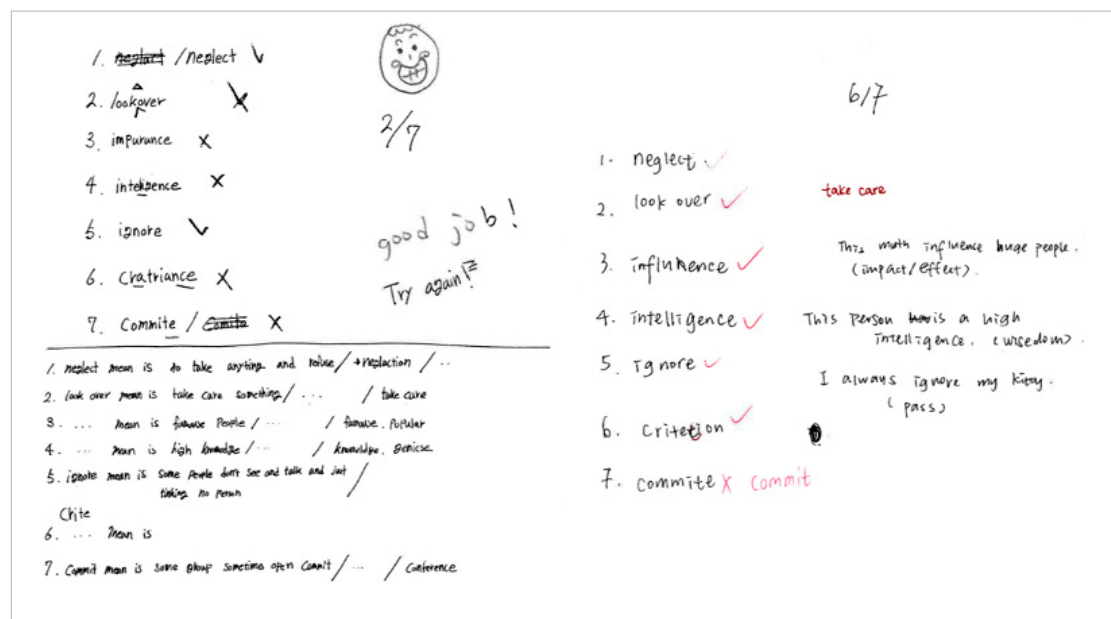


Figure 4: Peer-marked spelling test

Students then worked at their tables to create their own synonyms, definitions, example sentences and word families (see the examples in Figure 4). This was done in groups for peer support, but even in groups, the students had a limited range of vocabulary to complete the task, and generally their confidence in their ability to produce synonyms and definitions was quite low.

To develop their vocabulary, I showed students how to use the prompts (Appendix 1) to ask Co-Pilot for a definition, example sentences, synonyms and word families of the words from the spelling test. I encouraged the students to select the elements suggested by Co-Pilot that they thought would be most useful, and then to copy these to their individual vocabulary spreadsheets (Figure 5). I noticed that this activity seemed engaging and motivating for students.

File Home Insert Share Page Layout Formulas Data Review View Automate Help Draw						
Calibri (Body) 12 B [font icons] General \$€ ¥ [currency icons] [math icons] [search icons] [share icons]						
6	X fx					
	A	B	C	D	E	F
	Word	Synonyms	Definition		Example sentence	Word forms
1	Word	Synonyms	Definition		Example sentence	Word forms
2	apparent		easy to see or understand		It is apparent that the law must be changed.	adjective
3	adaptation (nc, nu)		the process or act of changing to suit a new situation		Adaptation to a new city can be difficult.	noun countable
4	NAIDOC Week	Indigenous week, First Nations Celebration	an annual celebration held across Australia in the first week of July each year		During NAIDOC Week, our school organized cultural workshops to learn about Aboriginal art and history.	noun
5	labour		Labour refers to hard work, especially physical work.			
6	dispose		something means to get rid of it, especially when you no longer need or want it.			

Figure 5: Individualized vocabulary Spreadsheet, AE3, Week 8

AE4: Vocabulary Analysis (Rachel)

In trialling the writing practice approach outlined above, it became clear to me that we were missing a step, and that many students were struggling to identify low-level and high-level vocabulary from the lists they had brainstormed in preparation for writing their body paragraphs (this was what we had asked Co-Pilot to do in the unsuccessful first cycle of our research). To assist students and to give them a holistic view of their writing, I asked them to use the vocabulary they had brainstormed to write the first body paragraph of the essay. I then showed them how to use a text analyser (Vocab Kitchen) to identify the CEFR level of the vocabulary in the paragraph (Figure 6). Most of the vocabulary used by the students was analysed to be A1 or A2 level (many of the students were shocked by the analysis results), and therefore an incidental benefit was to raise the students' awareness of their English level. This also caused some consternation among my students, and I realized it was necessary to clarify that a paragraph should have a combination of high-level and low-level words, and that often a lower-level word might be a better choice. I observed that using the text analyser in this way was helpful in enabling students to decide for which words they should ask Co-Pilot to provide synonyms.

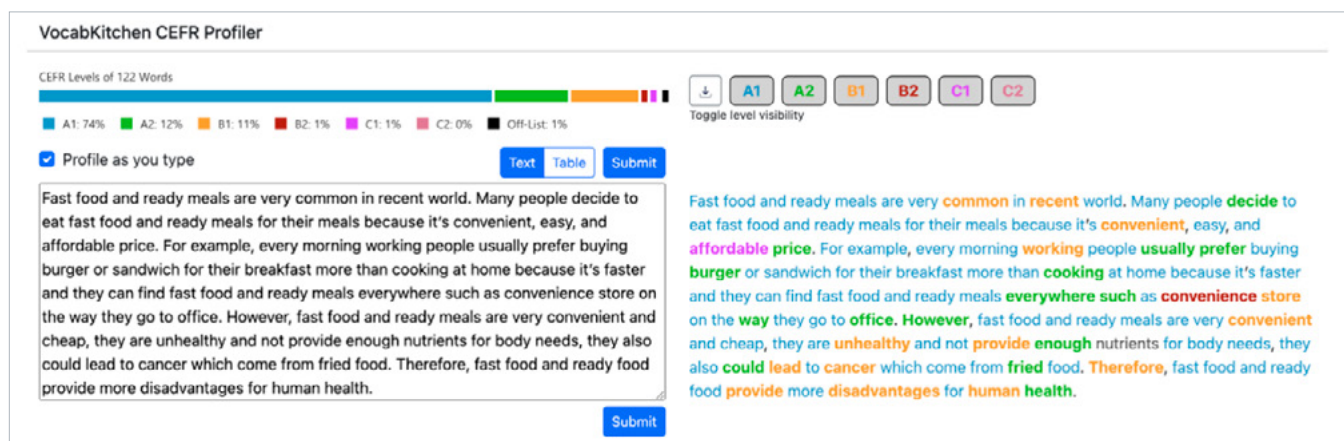


Figure 6: Text analysis of a student's body paragraph before Co-Pilot interaction

For the second body paragraph, students brainstormed key vocabulary and used the text analyser to identify the CEFR level of the words in list form. Students then chose some of the vocabulary identified at A1/A2 level and interacted with Co-Pilot to ask it for higher-level synonyms that they could use in their paragraph. They also asked it to clarify and explain the synonyms as much as they felt necessary to understand the meaning of the words, or the differences in meaning between them. They then rewrote the paragraph to include some of the vocabulary suggested by Co-Pilot. Once the paragraph was finished, they used the text analyser to see the level of the words they had used (Figure 7).

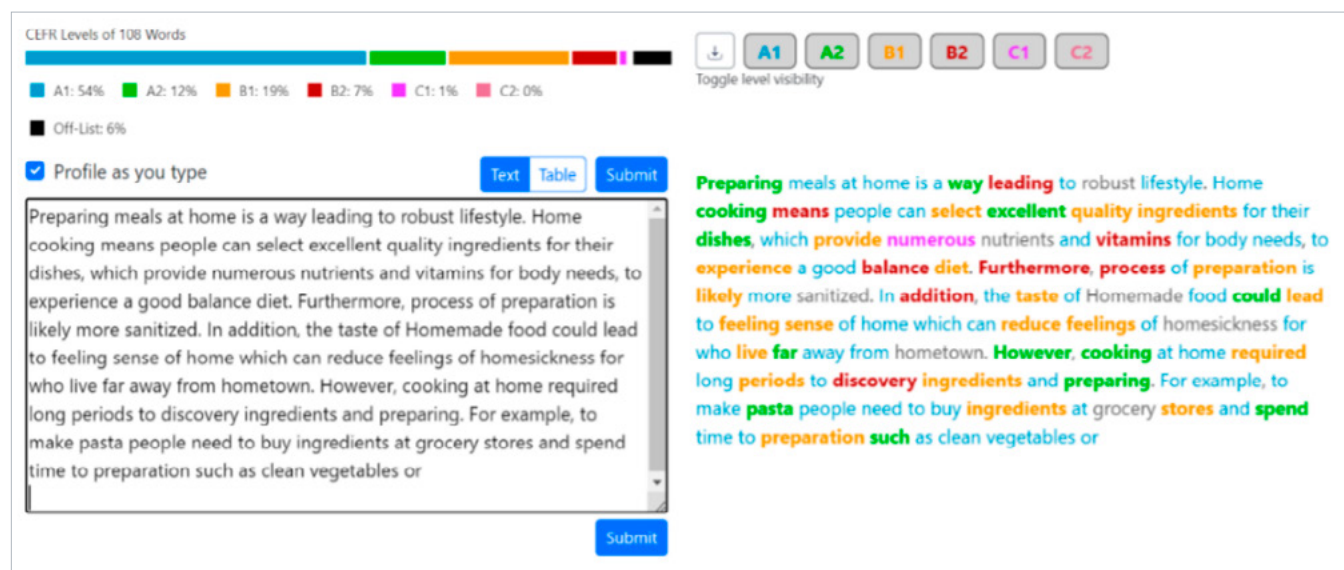


Figure 7: Text analysis of a student's body paragraph after Co-Pilot interaction

Although the interactions with Co-Pilot enabled students to use higher-level vocabulary, they did make errors with word form (and occasionally word choice) which needed teacher input and clarification; there is potential here for further interactions with Co-Pilot to create tables with word families, from which the students could then select the correct word form to use in their writing.

AE4: Vocabulary Quiz (Rachel)

As part of the weekly vocabulary review lesson, students in my AE4 classes interacted with Co-Pilot to help them understand any new words in the weekly vocabulary list. Once they felt that they had received enough input on the meaning, I gave them prompts they could use to ask Co-Pilot to design a multiple-choice quiz to test them on the new vocabulary, and to give them feedback on their answers (Appendix 2). This enabled them to interact with Co-Pilot further to gain a deeper understanding of the meaning of the words they got wrong. Feedback from students about this activity was quite positive, with several of them saying that asking AI to create quizzes was a new experience for them.

Findings

By the end of our investigations into how students can interact with Co-Pilot to develop their academic vocabulary, we found that students were interacting independently with Co-Pilot and successfully asking it to give more simple explanations of meaning, explain the difference between words with similar meanings, and to tabulate results to add to their vocabulary records. In general, students responded very positively to their experiences and felt that using Co-Pilot was a helpful way for them to develop their vocabulary:

'It can give me some good examples, explanations.' (AE3 student)

'I just do the way you taught. I think it's convenient.' (AE4 student)

'Using AI to learn foreign language is useful.' (AE5 student)

Students also reported increased levels of confidence in interacting with Co-Pilot. We gave students a survey at the beginning and the end of their course, in which we asked them the following questions:

- When using AI to help you learn vocabulary, how confident are you about:
 - telling AI what you want it to do (prompting)?
 - asking AI to explain more if you don't understand?
 - choosing which words you can use in your writing?
 - asking AI to help you to practice new words?

Answer choices were given as a Likert scale, with 1 being 'not confident' and 5 being 'very confident'. Most answers in the start survey were in the 2-4 range, and by the time of the end survey (after eight weeks of working with Co-Pilot to develop vocabulary), most answers were in the 3-5 range. This increase in confidence was reported by all respondents (Figure 8) to all four sub-questions after they had completed the vocabulary activities with Co-Pilot, which is supported by the observations we made while we were teaching (see also Appendix 3 for results of a class survey).

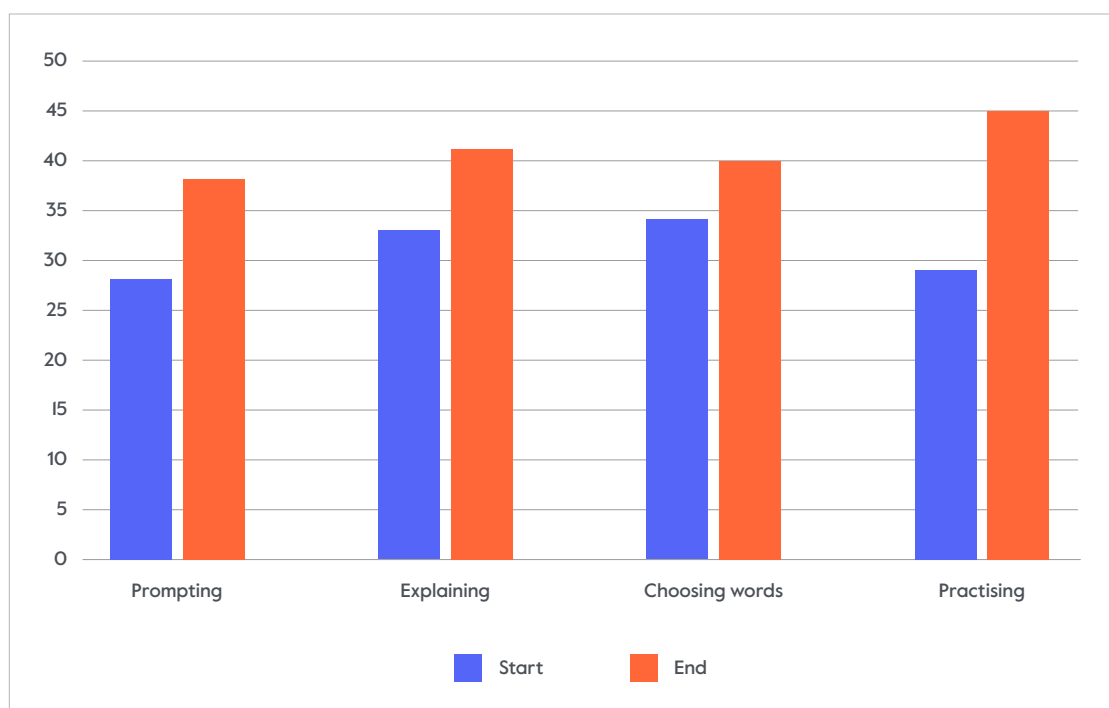


Figure 8: Survey data showing increase in confidence

Many students were confident enough to write their own prompts, for example:

Give a definition or example about ‘excessive.’

I need a definition for excessive.

I need more. [after Co-Pilot gave an explanation the student didn’t understand]

We also found that Co-Pilot was able to give useful responses even when the prompts contained grammatical or spelling errors:

Use some easy words definition ‘psychology.’

What is withstand means.

Explain the exhausted.

Does that mean the subject connected to mental health?

This, combined with its ability to simplify its responses when asked, and the way it ends each response with an encouraging comment, seemed to contribute to students’ confidence in interacting with it and asking it questions. Figure 9 shows Co-Pilot’s response to a grammatically incorrect prompt. This includes a validation of the student’s question, an answer, another example, and some encouragement. We noticed that many students were concluding their interaction with a thank you, which we see as an indication that, as we had hoped at the beginning of our research, AI can assist students with building their vocabulary in a personalized, interactive way which includes meaningful support.


is it mean give information for a lot of people , is it right?


Copilot

Yes, you're absolutely right! **"Disseminate"** indeed means to **spread or give out information, especially to a large number of people**. It's like sharing knowledge with a wide audience, ensuring that it reaches many others. Keep up the great work with your vocabulary exploration! 🌟.

Figure 9: Grammatically incorrect prompt and Co-Pilot's response


The following interaction (Figure 10) shows how one student successfully interacted with Co-Pilot using the persona-style prompt of a 12-year-old to gain an understanding of the difference in meaning between 'enhance' and 'foster,' despite some spelling errors.


 **You**
Explain foster. Use simple vocabulary.

 **Copilot**
Sure! The word "foster" can mean a couple of things:

1. **Taking care of a child:** When you foster a child, you take care of them for a while, even though you are not their legal parent. It's like giving them a safe and loving home temporarily^[1].
2. **Encouraging growth:** Foster can also mean to help something grow or develop. For example, you can foster a love for reading in someone by giving them lots of books and reading with them^[2].

Does that make sense? 😊

 **You**
I don't understand your explanation of fosterer 2. Explain again. Assume I am 12 years old.

 **Copilot**
No problem! Let's try again.

When we say "foster" in the sense of encouraging growth, it means to **help something develop** or **grow stronger**. Imagine you have a small plant. If you water it, give it sunlight, and take care of it, you are fostering its growth.

Similarly, if you help a friend get better at a sport by practicing with them, you are fostering their skills. It's all about **supporting and encouraging** something or someone to become better or stronger.

Does that make more sense? 😊

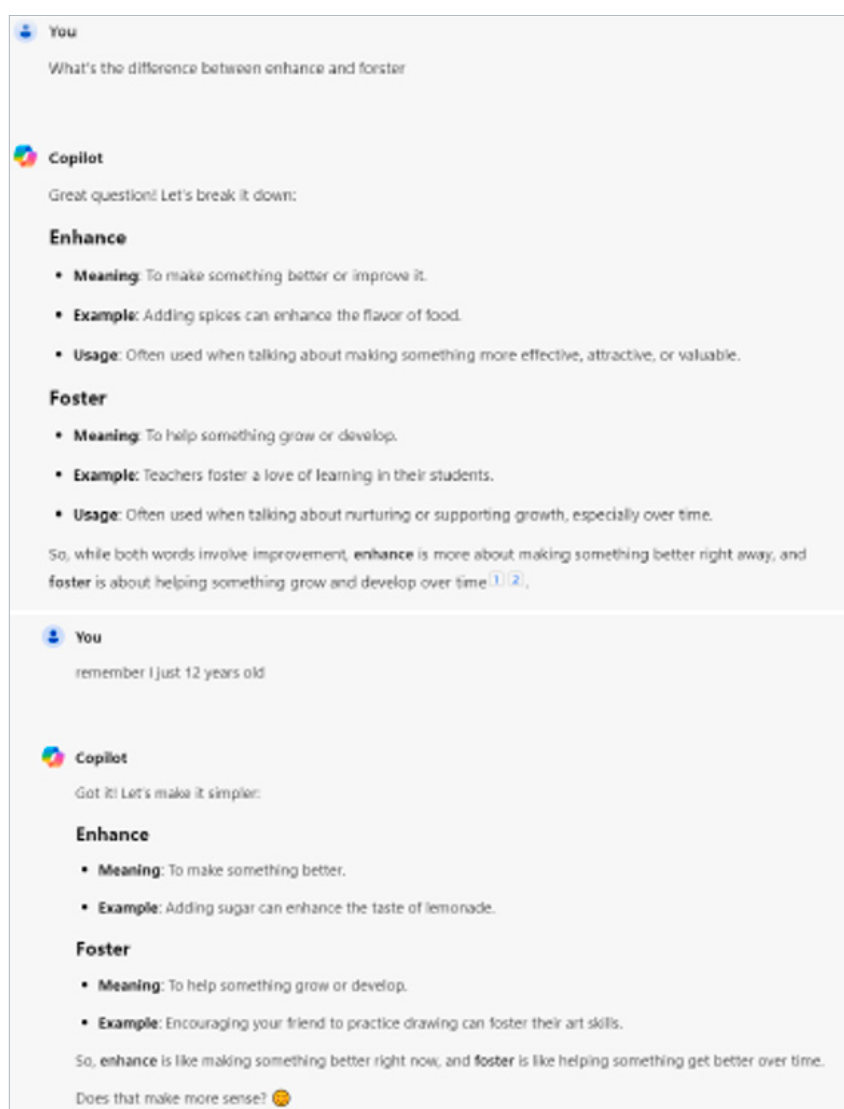


Figure 10: Interaction asking Co-Pilot to provide further information, using the 12-year-old persona

Reflections

As teachers, we have observed how this approach has made our classroom practice more efficient, while still allowing us to respond to the needs of our students. We have spent less time thinking of how to explain new (and at times) quite complex vocabulary on the fly, and we have been able to promote more autonomous learning through having students identify which words they need to understand more deeply and then focus on those.

It has been very rewarding to see students grow in confidence in their interactions with Co-Pilot, and to increase their ability to use prompts effectively and independently to deepen their understanding. However, it is also clear that Co-Pilot was not designed as a language learning tool, and that there are issues with the clarity of its responses, the appropriacy of some of the alternative synonyms it suggests, and misunderstanding of what the students are asking. Students were also

able to identify these weaknesses, showing an ability to evaluate the information they received from Co-Pilot:

'The Co-Pilot was actually useful when we learnt vocabulary, I remembered the words by their definition. But sometimes it showed us too complex, which I could not understand.'

'Well, I feel the Co-Pilot is pretty useful but sometimes it little bit stupid because it can not understand me. However, the Co-Pilot is smarter than me so I still learn more vocabulary.'

'It can give me words and expressions that I wouldn't have thought of on my own, and has helped me a lot to expand my vocabulary diversity. The downside is that occasionally I need to sift through the information he gives me.'

By the end of each course, students were interacting with Co-Pilot with little support from us and were starting to either use their own prompts or adapt the prompts we had provided for them. Our project was focused on classroom activities and teacher guidance in helping students to use Co-Pilot to build their vocabulary, and it would be interesting to follow this up with some research into what students were doing outside of class to increase and practice their vocabulary, and if this included some interactions with AI.

We noted that some students who did not have the Microsoft Edge Internet browser initially experienced technical issues with the login process. However, with support, they were able to access Co-Pilot using other browsers.

As our college works towards a more integrated approach to the use of AI, we have been asked to assist the Curriculum Team to adapt the weekly vocabulary lessons in AE3, AE4 and AE5. These will now include teaching students how to interact with Co-Pilot to develop their vocabulary, as well as using Quizlet and the Academic Word List.

The findings from our project indicate that use of AI seems to fit within the frameworks from Sydney University and UTS that we used as guides when we were developing our approach. We have observed our students becoming more comfortable with using AI in our classrooms, and they have shown us that they can engage with AI critically as they use it to develop their vocabulary. Additionally, the project has assisted the AI working group at UTS College to consider how students can use genAI in an ethical way that is pedagogically sound, and to determine directions the college can take to support students in their use of AI.

Throughout this project it has been clear to us that our students are genuinely interested in learning how they can use AI appropriately to enhance their studies, and they have welcomed the opportunity to interact with Co-Pilot and learn how it can help them to develop their vocabulary. We have also seen that a common concern amongst our colleagues is the possibility of being replaced by AI. However, our project has shown us that, while AI is a very useful tool to help students develop their vocabulary, teachers with a human heart are still very much needed and valued.

References

- Coxhead, A (2000) A new academic word list, *TESOL Quarterly* 34 (2), 213–238.
- Liu, D (2024) Rules, access, familiarity, and trust (RAFT) - a practical and realistic approach to addressing generative AI in education, *LinkedIn*. <https://www.linkedin.com/pulse/rules-access-familiarity-trust-raft-practical-realistic-danny-liu-qimic/>
- Liu, D, Bridgeman, A and Chan, C K Y (2023, May 15) 'Please do not assume the worst of us': students know AI is here to stay and want unis to teach them how to use it, *The Conversation*. <https://theconversation.com/please-do-not-assume-the-worst-of-us-students-know-ai-is-here-to-stay-and-want-unis-to-teach-them-how-to-use-it-203426>
- LX Team, UTS (2024) Five principles for the effective ethical use of generative AI, *UTS LX*. <https://lx.uts.edu.au/collections/artificial-intelligence-in-learning-and-teaching/resources/five-principles-for-effective-ethical-use-generative-ai/>
- Milton, J (2009) *Measuring Second Language Vocabulary Acquisition*, Bristol: Multilingual Matters.
- Mollick, E and Mollick, L (2023) *Assigning AI: Seven Approaches for Students, with Prompts*, Wharton School of the University of Pennsylvania/Wharton Interactive. <https://doi.org/10.48550/arxiv.2306.10052>

Appendix 1: Prompts

The prompts that we designed for students included:

- Define XXX. Use simple vocabulary.
- I don't understand your explanation of XXX. Use simpler vocabulary (CEFR A2 or B1 level).
- I don't understand your explanation of XXX. Assume I am 12/15 years old. Explain again.
- What's the difference between XXX, XXX and XXX?
- So XXX means [your definition]. Is that right?
- Give me an example sentence using the word XXX.
- Act as an expert in Academic English with an expert understanding of vocabulary at CEFR Levels B1, B2 and C1. I will give you a list of words and you will suggest three alternatives for each word. Your alternatives should be higher-level academic vocabulary, suitable for use in an academic essay. The alternatives can be one word or a short phrase. Create a table with the original words, your suggested alternatives, a definition, and an example sentence. Repeat this instruction to show that you understand and then ask for my list of words.

Appendix 2: Example prompt, quiz, answers and feedback

I am an international student learning English to study at university. Help me to practice vocabulary. Design a multiple-choice activity for each of the following words: communicate, collaborate, teamwork, injury, individual. The activity should not contain any words higher than CEFR B1. When the activity is ready, I will complete each question and you will provide feedback. If an answer is wrong, you will explain why. First, summarize these instructions to confirm you understand the task. Then, design the activity.

Certainly! Let's create a vocabulary activity for each of the specified words. Here's a summary of the task:

	Word	Activity	Level	Instructions
1	Communicate	Multiple-choice question	CEFR B1	Choose the correct definition for the word 'communicate.'
2	Collaborate	Multiple-choice question	CEFR B1	Select the appropriate meaning of the term 'collaborate.'
3	Teamwork	Multiple-choice question	CEFR B1	Pick the definition that best describes 'teamwork.'
4	Injury	Multiple-choice question	CEFR B1	Identify the correct definition for the word 'injury.'
5	Individual	Multiple-choice question	CEFR B1	Choose the appropriate meaning of the term 'individual.'

Now, let's proceed to design the activity for each word. Feel free to complete the questions, and I'll provide feedback afterward!

Great. go ahead and design the activity.

Certainly! Let's create a multiple-choice activity for each of the specified words. Here they are:

1. Choose the correct definition for the word 'communicate.'
 - a. To work together as a team
 - b. To exchange information or ideas with others
 - c. To injure oneself accidentally
 - d. To perform an individual task
2. Select the appropriate meaning of the term 'collaborate.'
 - a. To compete against others
 - b. To work independently
 - c. To cooperate with others on a project
 - d. To cause harm to someone

3. Pick the definition that best describes 'teamwork.'
 - a. The ability to work alone
 - b. The process of communicating with others
 - c. The act of collaborating with a group
 - d. The tendency to avoid injuries
4. Identify the correct definition for the word 'injury.'
 - a. A positive outcome
 - b. A physical harm or damage to the body
 - c. A type of communication
 - d. An individual achievement
5. Choose the appropriate meaning of the term 'individual.'
 - a. A group of people working together
 - b. A personal injury
 - c. A single person considered separately from others
 - d. A form of collaboration

Here are my answers. Give feedback and explain my incorrect answers.

1B: 2B: 3C: 4D: 5C

Thank you for providing your answers! Let's review them:

1. **Word: Communicate**
 - Your answer: **B) To exchange information or ideas with others**
 - Correct! 'Communicate' indeed means to share thoughts, information, or feelings with someone else.
2. **Word: Collaborate**
 - Your answer: **B) To work independently**
 - Feedback: The correct answer is **C) To cooperate with others on a project**. 'Collaborate' refers to working together with others, not independently.
3. **Word: Teamwork**
 - Your answer: **C) The act of collaborating with a group**
 - Correct! 'Teamwork' involves collaborating with others to achieve a common goal.
4. **Word: Injury**
 - Your answer: **D) An individual achievement**
 - Feedback: The correct answer is **B) A physical harm or damage to the body**. 'Injury' refers to harm or damage, not an achievement.

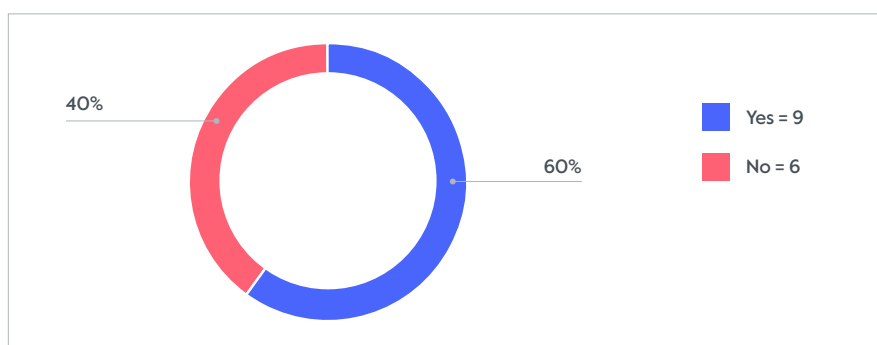
5. **Word**: **Individual**

- Your answer: **C) A single person considered separately from others**
- Correct! 'Individual' refers to a single person distinct from a group.

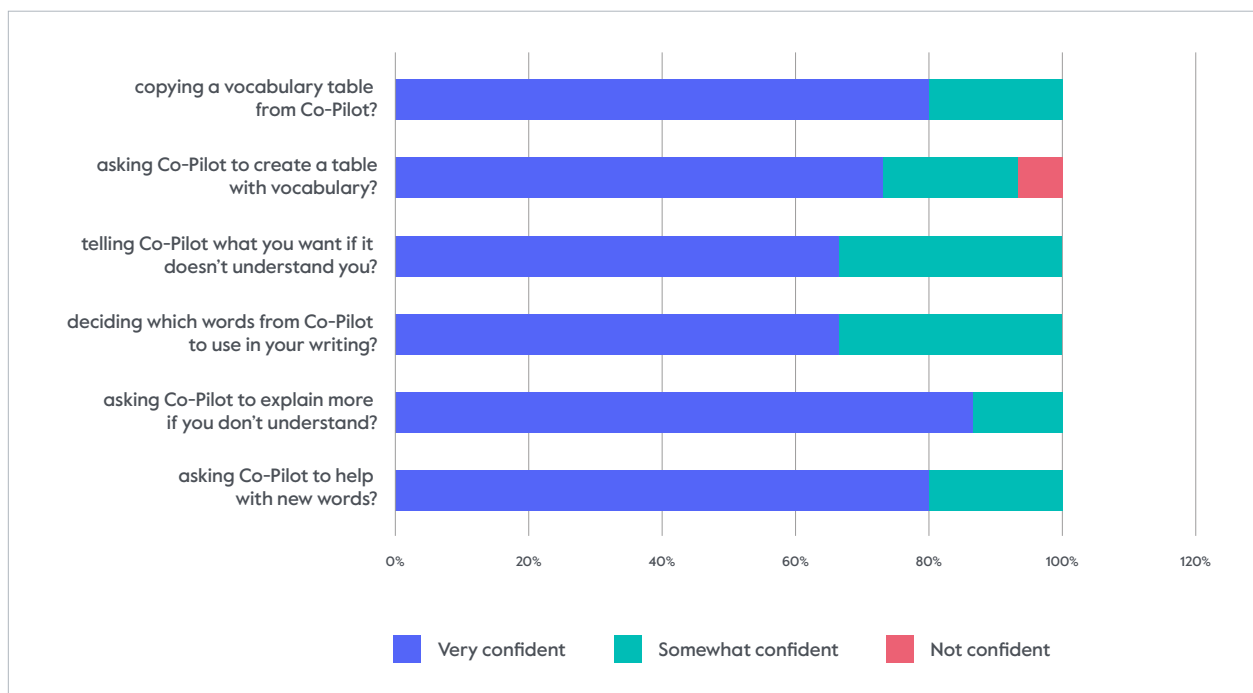
Great effort! Keep practicing, and your vocabulary will continue to improve. If you have any more words or need further assistance, feel free to ask!

Appendix 3: AE4 class responses to an end-of-course survey about their experiences with using Co-Pilot

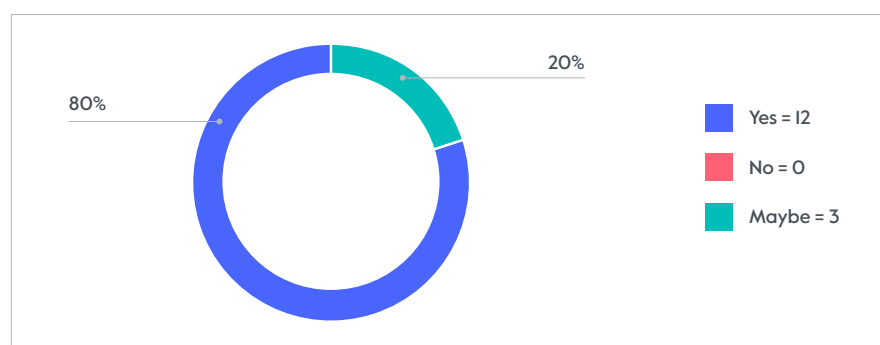
Question 1: Before this AE4 class, did you ever use AI (Co-Pilot, ChatGPT, Gemini etc) to help you learn vocabulary?



Question 2: How confident are you about:



Question 3: Will you use Co-Pilot in the future to help you learn vocabulary?



Explain your answer to Question 3.

Good

I will use it

keep using because it useful

because useful and fast

it's not a prefect way but not bad.

very useful

Very useful

I think Co-Pilot is useful i may use it sometimes in the future for vocabulary

Enter the word

It's a good way to find information and sources

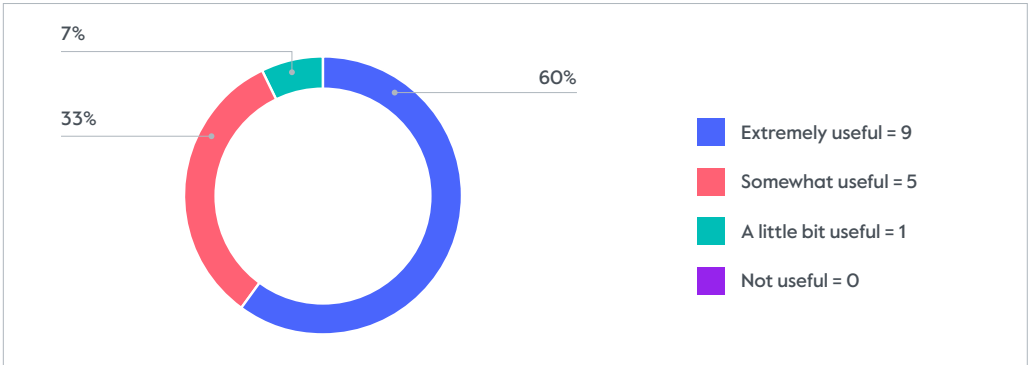
convenient

When i need academic words, I will use it

Very convenient and useful

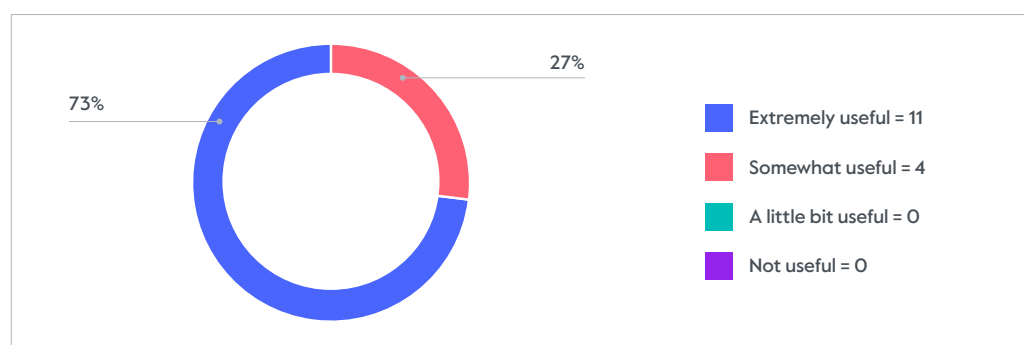
Because I think to learn vocabulary in English is better for us than in own language.

Question 5: Will you use Co-Pilot in the future to help you learn vocabulary?



Explain your answer to Question 5.
good
extremely valuable tool
sometimes stupid
words it recommend sometimes tooo academic
it can give me more detail to know the vocabulary
Easy to understand
Can teach me how to use
Help me find out the definition
very useful
Help me to find academic words
The answer is shown quickly, clearly meaning and variery and clear
Be able to learn the nuances

Question 7: Will you use Co-Pilot in the future to help you learn vocabulary?



Explain your answer to Question 7.

good

understand completely

good

sometimes I cannot get

give me a table and more detail to know

Learn more words

It help me to understand the worlds more simply

I can learn more words

very useful

It can explain the meaning of the words easy

Easy to Understand meaning of new words

Be able to learn the nuances

Please write any other comments about the way we have used Co-Pilot in AE4 to help you learn vocabulary.

good

help to connect the meaning and deep understanding

good

Gives a lot of clear explanations

Learn new vocabulary and complex sentences

Give some easy definition

useful and convenient

Useful

I just do the way you taught. I think it's convenient.

AI literacy for improved critical thinking

Emily McNamee and Wilhelmina Anne Gevaña, Navitas English Perth

Introduction

Perth attracts tourists from around the world. Those who come here to study English are equally diverse—at Navitas English there are ELICOS students from 100+ countries. These students are diverse in nationality, educational background, age, and goals. As teachers in this context, we observed varying rates of success in formative and summative assessments for speaking, listening, reading, and writing. We wondered if this gap in learning outcomes could be rooted in a disparity in critical thinking stemming from their professional and academic backgrounds, particularly the latter. The connection between critical thinking skills and academic success has been a discussion among pedagogical scholars since the 1980s, and it has been shown that improving critical thinking skills increases academic success (Johnston et al 2023:3).

Reflecting on her previous experience teaching in a public junior high school in Japan, Wilhelmina Anne recognised that the definition of academic success and the teaching methodologies used can vary significantly between countries.

In Japan, for instance, lessons were often limited to content directly tied to assessments. Before each test, students received extensive study materials for self-completion, silent self-study periods in class, and one after-school 'Answer Time' session for questions. The system is centred on outcome-based assessments, placing significant pressure on students in the lead-up to tests, while lacking in-class activities that actively engage students with the material.

Lok et al (2015) highlight that outcome-based assessments fail to nurture critical thinking, problem-solving, and self-managed thinking. The emphasis on standardised

criteria compels schools to focus on maximising test outcomes, resulting in a narrowly concentrated educational approach.

At Navitas, we've observed that many students from countries that use this outcome-based educational approach struggle with forming and expressing opinions, actively participating in discussions, and maintaining engagement when faced with questions that require personal reflection.

We began to question to what extent such systems impact students' critical thinking skills. The heavy focus on standardised assessments and the lack of classroom activities that encourage independent thought and active engagement may hinder the development of these essential skills. This concern sparked our interest in exploring the connection between educational practices and critical thinking further.

Interestingly, Moosavi (2022) highlights that East Asian students often encounter criticism regarding their approach to learning. He explains that this approach is frequently characterised in western academic literature as hierarchical, 'in which knowledge is something that should be extracted, memorised and regurgitated.' He also adds that some western scholars claim that East Asian culture values conformity over individuality, which might make it harder for students to challenge ideas or offer different opinions.

The claims made by some western scholars in academic literature lead to a stigmatisation of East Asian students who become trapped in a cycle of low expectations and low results based on their 'imagined deficiencies.' Yet, it has been shown that international students achieve academic success on a par with domestic students in Australian higher education settings (Moosavi 2022). In this article, we recognize that our classroom observations on critical thinking do not represent all East Asian international students.

Research question

Over the course of our teaching experience at our centre, we have considered that if critical thinking skills, particularly the skills of evaluating and creating, weren't emphasised in a student's primary and secondary schooling, they could be less likely to successfully write and speak in English. We often saw students with these backgrounds struggle to produce their own ideas in summative written and oral assessments.

So, we asked ourselves: how can we as educators use all the tools at our disposal to help students improve their critical thinking? As our students had diverse experiences in classroom technology, utilising generative AI (genAI)—new to many students—provided a unique research opportunity. To address the issues of critical thinking that we discussed above, we developed the following research question:

How can AI literacy training improve students' critical thinking skills?

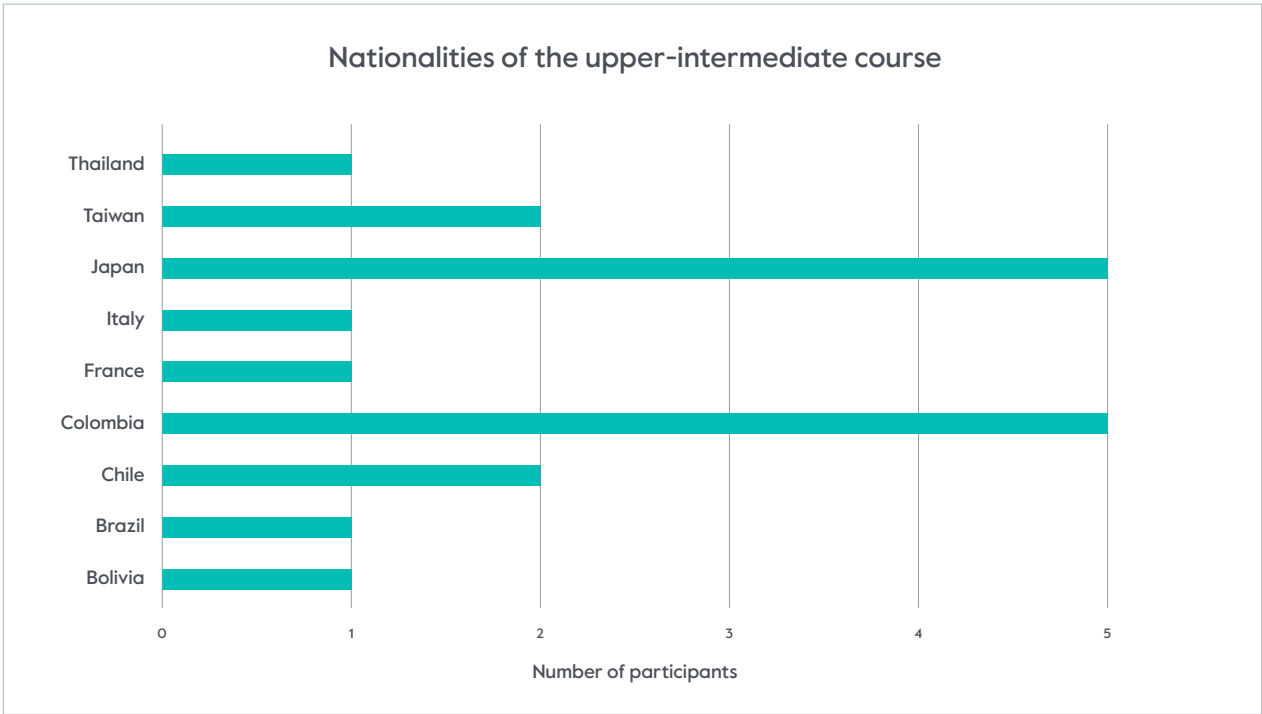
We designed a 10-week AI literacy training program for our classes, focusing on effective classroom use of genAI. We were guided by the concept of AI literacy as

defined by Matt Crabtree: ‘having the skills and competencies required to use AI technologies and applications effectively. It’s about viewing these technologies critically, understanding their context, and questioning their design and implementation’ (2023, paragraphs 5 and 6).

Participants

For this research, we worked with two General English courses. Emily taught an upper-intermediate class (CEFR B1–B2) and Wilhelmina Anne taught an advanced class (CEFR B2–C1). To successfully complete each level, students were required to learn and practice new vocabulary and grammar concepts. They demonstrated their understanding by applying these concepts in skills tests, which assessed writing, reading, speaking, and listening. Additionally, grammar and vocabulary tests were integral to tracking their academic progress. These courses are run in 10-week cycles with open enrolment, so while we collected our data new students were able to join the class weekly.

Our action research (AR) participants represented a diverse student body, as shown in Figure 1. They came from 14 countries across four continents when accounting for both course levels. Students’ different expectations of learning a language (i.e., communicative ability versus rote memorisation) seemed to us to be highly influenced by their cultural and educational background, something which we have observed can differ notably between regions like Colombia and Japan.



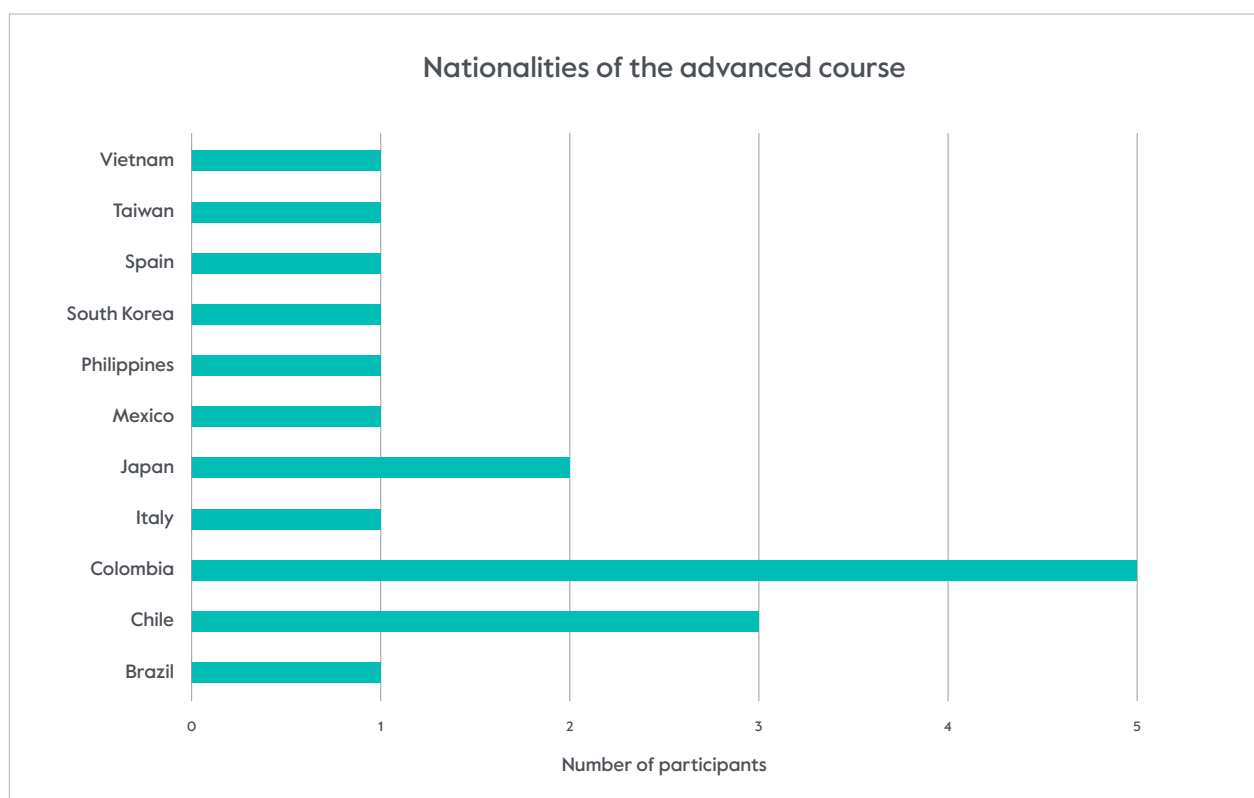


Figure 1: Participant numbers organised by course level and nationality

Methodology

Research design

Our primary goals were to assess students' critical thinking abilities through AI literacy, so we adopted a Content and Language Integrated Learning (CLIL) approach. The AI literacy plan aimed to teach students the basics of genAI, including its training process, limitations, ethical use, and classroom applications. English language content was integrated through key terminology and grammar instruction within an authentic context.

To tackle the broad topic of AI literacy, we designed a series of scaffolded lessons that built upon each other over the 10-week research period, which will be discussed subsequently. The plan was divided into theoretical and practical lessons, each with specific objectives to reinforce learning. Additionally, each lesson included a focus on a different level of critical thinking, which we chose to assess through our own critical thinking assessment rubric, which identified and defined five levels of critical thinking based on Bloom's Taxonomy (1956):

Level 1: *Remembering and Understanding*

Level 2: *Applying*

Level 3: *Analysing and Synthesising*

Level 4: *Evaluating*

Level 5: *Creating*

Creating our critical thinking rubric

A starting point for our research came from a study by Reynders et al (2020) on using rubrics to assess ‘process skills’ and improve learning outcomes in undergraduate STEM programs. We were inspired by the concept of their example rubric (see Figure 2) to develop our own version. Because the distinctions between each of the critical thinking levels are not always clear-cut, we had to make sure that by adding verbs and phrases under each skill, we could give our students scores based on how they completed these specific ‘cognitive processes.’

Critical Thinking		Evaluating, analyzing, and/or synthesizing relevant information to form an argument or reach a conclusion supported with evidence.				
Category	0	1	2	3	4	5
Evaluating		Minimally determined the relevance and reliability of information that might be used to support a conclusion or argument		Partially determined the relevance and reliability of information that might be used to support a conclusion or argument		Extensively determined the relevance and reliability of information that might be used to support a conclusion or argument
Analyzing		Inaccurately interpreted information to determine meaning and to extract relevant evidence		Interpreted information to determine meaning and to extract relevant evidence with some errors		Accurately interpreted information to determine meaning and to extract relevant evidence
Synthesizing		Inaccurately connected or integrated information to support an argument or reach a conclusion		Connected or integrated information to support an argument or reach a conclusion with some errors		Accurately connected or integrated information to support an argument or reach a conclusion
Forming Arguments (Structure)		Made a claim and provided incomplete evidence to support it.		Made a claim and provided partial evidence to support it.		Made a claim and provided complete evidence to support it.
Forming Arguments (Validity)		The claim, evidence, and reasoning were minimally consistent with accepted disciplinary ideas and practices		The claim, evidence, and reasoning were partially consistent with accepted disciplinary ideas and practices		The claim, evidence, and reasoning were fully consistent with accepted disciplinary ideas and practices
Comments:						

Figure 2: Rubric for assessing critical thinking (Reynders et al 2020)

Anderson and Krathwohl’s (2001) revisions to Bloom’s Taxonomy helped us greatly in this regard. We chose verbs from their Cognitive Process Dimension, as well as ‘alternative’ verbs, which we took from Heer’s revised model of the framework (2012). We chose particular action verbs to clearly and unbiasedly assess students’ ability at each level. The verbs chosen are as follows:

- Remembering and Understanding: recall, explain
- Applying: identify, put into use
- Analysing and Synthesising: break down into simplest parts, integrate
- Evaluating: assess, consider, make informed decisions
- Creating: generate, produce, consolidate, formulate, ask

Additionally, we incorporated elements from our college’s Assessment Marking Descriptors format as we are familiar with marking through its approach. You can see that words of variation and degree differentiate scores 1, 3 and 5. Figure 3 shows the final version of our rubric.

Critical Thinking Rubric					
	Remembering/Understanding — / — / —	Applying — / — / —	Analysing/Synthesising — / — / —	Evaluating — / — / —	Creating — / — / —
5	Student fully demonstrated the ability to recall information from memory accurately and explained its purpose with minimal errors .	Student fully demonstrated the ability to identify task aims and put remembered concepts into use accurately . Completed activity submitted with minimal errors .	Student fully demonstrated the ability to effectively break down information and problems into their simplest parts and sufficiently integrated new insights with past knowledge for deeper comprehension.	Student fully demonstrated the ability to effectively assess relevance and significance of information. Student also considered source bias and alternative explanations to make sufficiently informed decisions.	Student fully demonstrated the ability to generate own ideas and produce something new through innovative thinking, problem solving, and consolidating learned concepts. In the process, student formulated and asked appropriate questions to evaluate the quality of their work.
4	Ability meets elements of both 5 and 3				
3	Student partially demonstrated the ability to recall information from memory somewhat accurately and explained its purpose with noticeable errors .	Student partially demonstrated the ability to identify task aims and put remembered concepts into use somewhat accurately . Completed activity submitted with noticeable errors .	Student partially demonstrated the ability to break down information and problems into their simplest parts and somewhat sufficiently integrated new insights with past knowledge for deeper comprehension with some gaps in understanding.	Student partially demonstrated the ability to somewhat effectively assess relevance and significance of information. Student also considered source bias and alternative explanations in some ways to make partially informed decisions.	Student partially demonstrated the ability to generate own ideas and produce something new through innovative thinking, problem solving, and consolidating learned concepts. In the process, student formulated and asked somewhat appropriate questions to evaluate the quality of their work.
2	Ability meets elements of both 3 and 1				
1	Student insufficiently demonstrated the ability to recall information from memory and explained its purpose with significant errors .	Student insufficiently demonstrated the ability to identify task aims and put remembered concepts into use. Completed activity submitted with significant errors .	Student insufficiently demonstrated the ability to break down information and problems into their simplest parts and incompletely integrated new insights with past knowledge with minimally improved comprehension.	Student insufficiently demonstrated the ability to assess relevance and significance of information. Student attempted to consider source bias and alternative explanations to make informed decisions.	Student insufficiently demonstrated the ability to generate own ideas and produce something new through innovative thinking, problem solving, and consolidating learned concepts. In the process, student attempted to formulate and ask appropriate questions to evaluate the quality of their work.
0	Did not demonstrate skill				

Figure 3: Our critical thinking rubric (an interpretation of concept from Figure 1 in Reynders et al 2020)

Theory lessons

In the first part of our approach, we focused on establishing a base knowledge of generative AI. This was completed through five one-hour weekly lessons added to the original course curriculum, on the following topics:

1. What is AI?
2. How is AI Trained?
3. Understanding AI Bias
4. Best Approaches to Using AI in the Classroom
5. Creating a Class AI Contract

Lessons 1–3 were sourced from Common Sense Education, an American organisation focused on supporting free digital literacy content for K-12 classrooms. We found that their materials explained AI in ways that were easy to digest and apply in real-world contexts. In the second phase of our research, we developed five further lessons, presented in the section ‘Practical lessons’. For the final two theory lessons, as well as all five practical lessons, we developed our own material to help students digest, use and reflect on genAI.

Creating the worksheets

Our lessons aimed to engage students on multiple levels, but to focus on one critical thinking skill at a time; therefore, we designed weekly assignments targeting that skill. For instance, in Week 3, ‘Understanding AI Bias,’ after the theory lesson sourced from Common Sense Education, we assessed students’ analysing/synthesising skills with a worksheet we created (see Figure 4) based on a scenario illustrating potential bias with AI. Students read the scenario, discussed it with a partner, and completed

the worksheet together. The questions were designed to have them break down information into simpler parts and integrate it with previous concepts, aligning with the Analysing/Synthesising section of our critical thinking rubric.

Name: _____
Date: _____
Class: _____

Read the scenario.

Your friend, Emy Lianne, is having a party. She is known in the neighborhood for throwing the best parties, but everyone knows that if you want to be let in, you have to be dressed well!

Emy Lianne's apartment was recently renovated and the front entrance was equipped with a camera that can decide who should be allowed in. This AI entry camera is trained to only open the door for those who are dressed appropriately.

On the night of the party, many people came to Emy Lianne's, excited to enjoy the evening. But there was a lot of commotion since the AI entry camera only let some people in.

Iris, who was wearing a hijab, and Ander, who was wearing a baseball hat, weren't let in, but Zara, Elena, and Diego were, even though all their outfits were nice. After a few more people arrived, people started to realize that the AI entry camera only allowed people who didn't have headwear on.




Image by pch.vector on Freepik




Image by Freepik

Work with your partner to answer the following questions. Please write your answers clearly in the provided space.

Q1. Did the AI entry camera work the way it was supposed to?
Q2. Who are all of the different people or groups involved in this scenario?
Q3. How might each of them be feeling?
Q4. Was everyone impacted by the AI entry camera in the same way?
Q5. Once Emy Lianne learns about this issue, what should she do? What can the creators of the AI entry camera do to make their product fairer for everyone?

Figure 4: Week 3 worksheet: Analysing/Synthesising

Practical lessons

In the second phase of our AI literacy approach, students were asked to complete a project about how an AI tool could improve their lives in Perth. Project-based learning allowed students to apply concepts from the first five lessons, enabling us to assess growth in critical thinking. We introduced two generative AI tools: ChatGPT and Yoodli. ChatGPT is a more well-known AI language model, and Yoodli is a lesser-known speech coaching tool. The practical lessons were again divided into five one-hour sessions:

- 6. Using AI in the Classroom (ChatGPT)
- 7. Using AI in the Classroom (Yoodli)
- 8. Using AI for Project-based Learning
- 9. Evaluating Projects
- 10. Presenting Projects (AI Showcase)

The timeline in Figure 5 illustrates the topics we covered in our weekly AI literacy lessons from 15 April to 17 June. At the end of the program, we held an AI Showcase where our students could present their work to other students and staff. They completed their presentations in groups of three with mixed ability levels, both upper-intermediate and advanced. The tools chosen for the presentations were Fitbod, Reclaim.AI and ELSA Speak. The photos in Figure 6 show the groups presenting their work along with the posters created (see also Appendix 1).

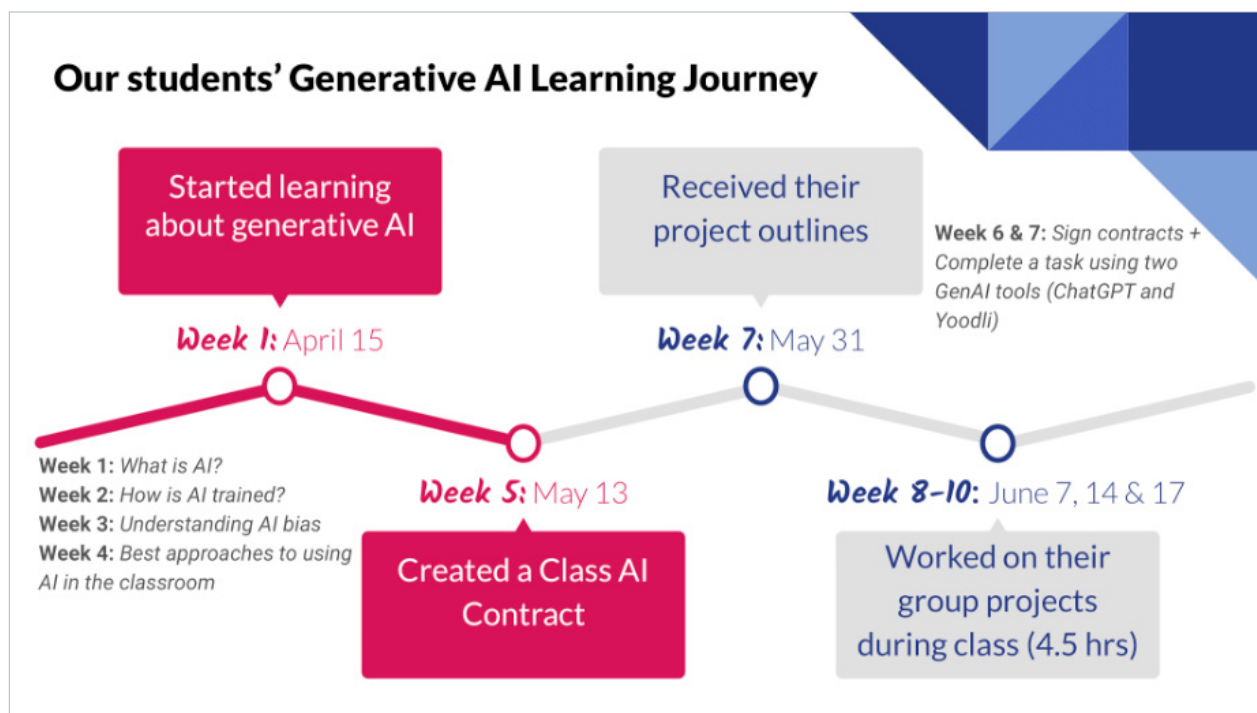


Figure 5: Our research timeline



Figure 6: AI Showcase presenters

Data collection

We collected and marked the assignments from all 10 lessons using our critical thinking rubric and recorded their scores. This resulted in two scores for each critical thinking level: one score from Lessons 1–5 (theory lessons) and one score from Lessons 6–10 (practical lessons).

For example, in Week 3, the Analysing/Synthesising assignment (Figure 4) was completed by a student who received a 3 out of 5 as they were able to answer the questions to some degree by breaking down the story, but they failed to recognise some important pieces of information, so their analysis was incomplete (Appendix 2). In Week 8, the student was assigned another task (Appendix 3) and scored a 3 out of 5 once more. This was due to the AI tools being only partially broken down into their helpful aspects and only somewhat adequately integrated into their project plans (Appendix 4).

Throughout the 10 lessons, we used questionnaires with both quantitative and qualitative questions to assess students' perceptions of their learning (Appendix 5). After the AI Showcase in Week 10, we replaced the questionnaire with recorded discussions where students reflected on their projects and overall participation, which provided further qualitative data.

Findings

We begin with our main finding related to our research question. However, there were also some unanticipated findings, which we also discuss below.

Finding 1: Critical thinking ability assessment scores decreased in all skills except one after the 10-week AI literacy training

Applying and Analysing/Synthesising skills were where we found the largest decrease in scores. Figure 7 shows students' critical thinking scores from Weeks 1–5, with scores out of five for a total of 25 points. Not all students who participated in the study are shown as they were not yet enrolled in our classes during this timeframe. A missing colour indicates the student was absent from that lesson.

Evaluating (light blue) was the most challenging skill, with a class average of 2.8, followed by Creating (purple) with an average of 2.9. In contrast, Remembering/Understanding, Applying, and Analysing/Synthesising had higher averages (3.2–3.4). These results indicate that students found the more complex skills difficult, as expected, and we anticipated improvement in these scores in the second round of assessments.

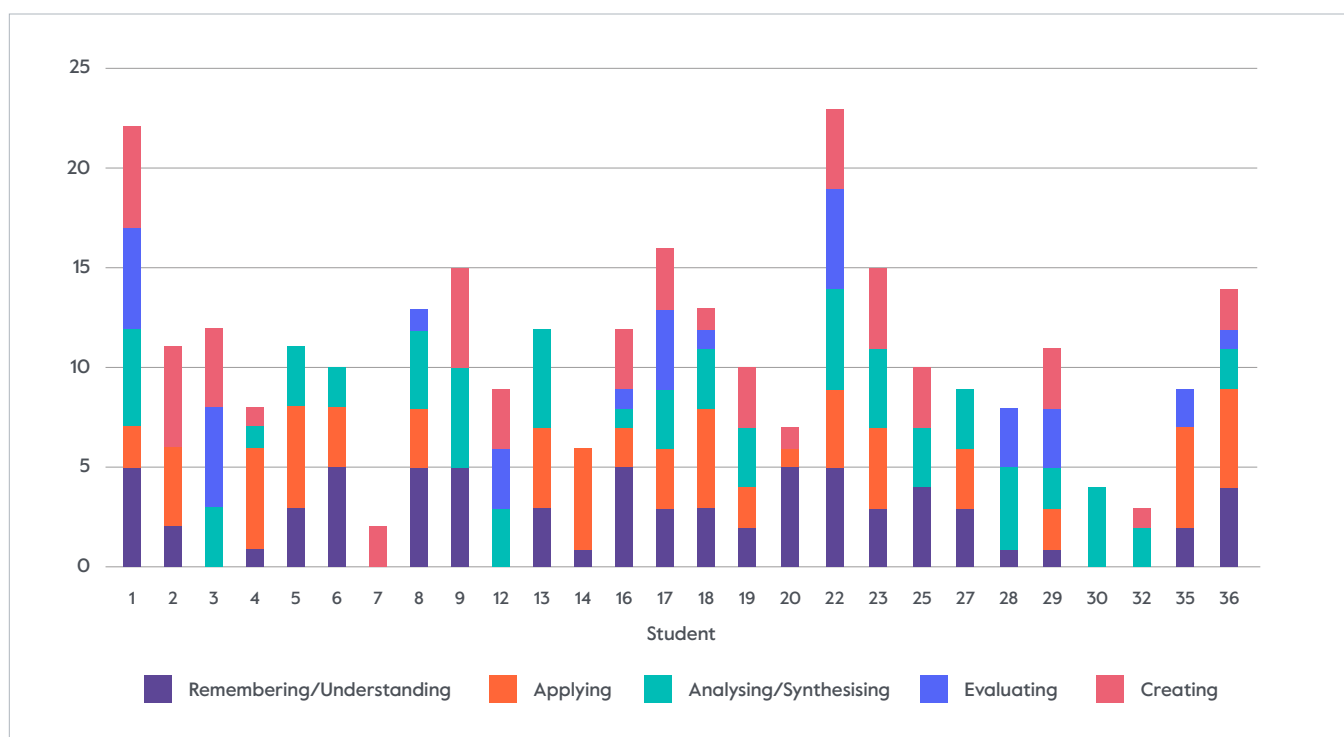


Figure 7: Students' critical thinking assessment scores from Weeks 1–5

In Weeks 6 and 7, students used ChatGPT to understand and complete a writing task. The class average for Remembering/Understanding stayed at 3.2 out of 5, but the Applying average dropped by 0.5 points to 2.7. This decline suggests that shifting from a traditional approach to one heavily reliant on AI might have been too abrupt. It can be seen that Analysing/Synthesising (green) was the most difficult skill, followed by Evaluating (light blue).

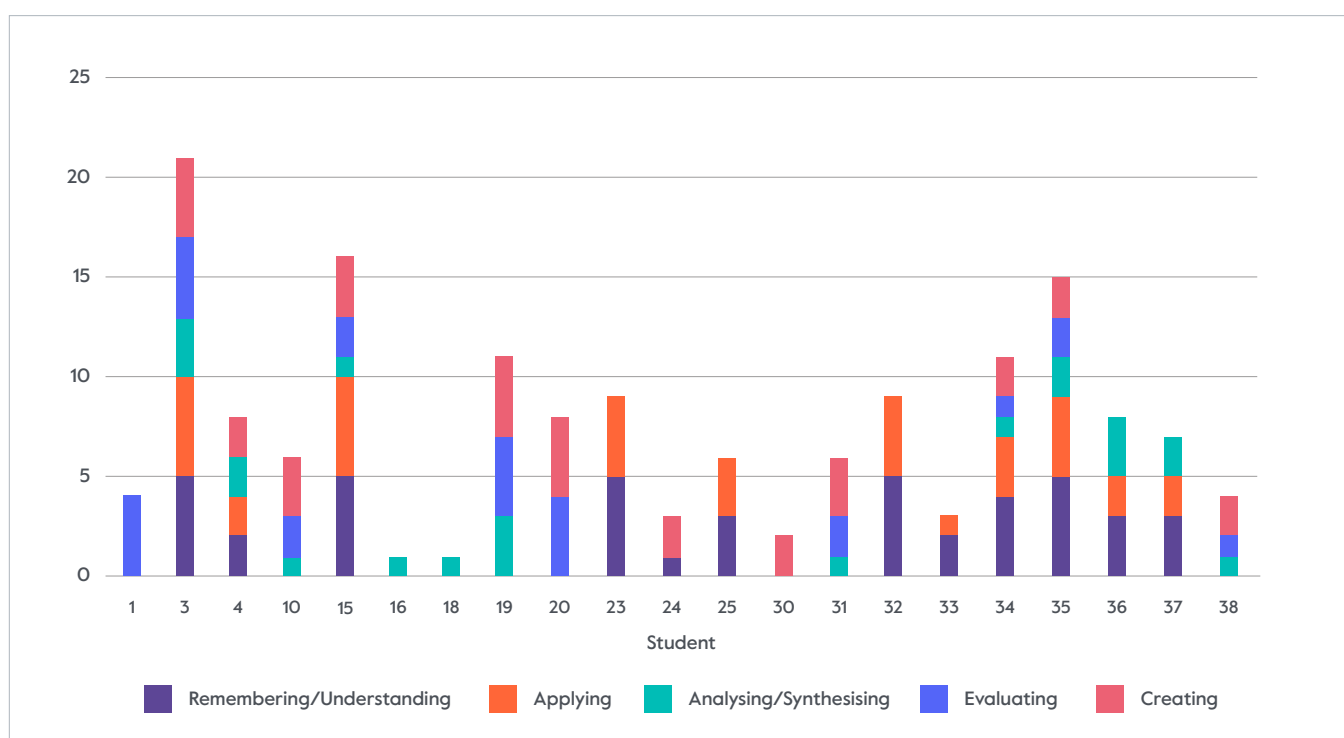


Figure 8: Critical thinking assessment scores from Weeks 6–10

We now turn to the findings that were unanticipated in our research.

Finding 2: Disconnect between students' excitement about less conceptually familiar genAI tools and the ability to use them successfully despite training

We believed that students would understand more fully what genAI is after 10 weeks of training. However, despite being engaged during the lesson where they were introduced to Yoodli, actually using it successfully proved to be a challenge. We had hoped that students would connect the in-class practice of using both ChatGPT and Yoodli with independent work to prepare for the AI Showcase. Although we guided them and encouraged them to use Yoodli as well as ChatGPT, we found through the post-presentation questionnaire that only one (S19) out of twelve presenters used it to prepare for their presentations.

This lack of independent practice suggests that students may have perceived Yoodli as too complex or outside their comfort zone. The disparity in tool usage could also indicate that students prioritised tools they found more intuitive or immediately rewarding. Despite recognising the potential benefits of Yoodli in enhancing their speaking skills, students may have lacked the confidence or motivation to explore it further without direct supervision. This highlights a critical gap in our approach—while we effectively introduced the concept and basic usage of genAI tools, we may have needed to provide more structured support or incentives to encourage consistent independent use of more specialised tools like Yoodli.

Finding 3: Students' attitudes towards genAI are more positive after 10 weeks

In Week 1, we discussed whether genAI would be more useful or harmful to society. Most students in the Advanced class viewed it as more harmful, while the Upper-Intermediate class was evenly split. We also surveyed students on their AI usage and opinions, finding that about half had used genAI before, mainly for English-related tasks, and most had a moderately positive view of it. However, five out of eighteen students rated AI's usefulness in the classroom as 1 or 2 out of 5, while 11 rated its usefulness outside the classroom as 4 or 5. This suggested some initial reluctance towards AI-centred lessons.

By the end of the 10-week course, students' attitudes had shifted, as reflected in their Week 10 discussions after completing their AI Showcase projects. One student (S24) noted that initially they did not see the need for AI but now felt confident using it for studies and other tasks. Another student shared the view that:

'Learning about skills related to AI is beneficial right now. You have to know about this. We were also able to develop other skills that are usually very difficult for students, such as standing in front of others and presenting. This is a skill of value that the school can equip the students with. It's a big part of language learning. It's a good idea to repeat projects like this each cycle.' (S15, *This direct quote has been edited for grammar and brevity.)

Limitations

We are aware of various kinds of limitations that could have biased our findings.

Sense of self in student questionnaires

Response bias may have occurred in our research because the student questionnaires were given at the end of class when many students were eager to leave. As a result, they may have rushed through the questions, which could have affected the quality and accuracy of the data.

Pair/group work scores

Measurement bias may have occurred because we had students work in pairs/groups for most literacy lessons. This could have affected our quantitative data, as performance might have varied depending on partners' abilities and their relationship. While grading, we also noticed some students disagreed and shared different ideas in the limited space, while others ignored their partners' input and completed tasks individually. Despite these discrepancies, we had already decided to give the same score to all members of a group, leaving us uncertain about the fairness of the assessment.

Attendance and enrolment irregularities

Due to irregularities in attendance and enrolment throughout the course, several students missed various classes, impacting their overall engagement in the AI literacy lessons and critical thinking development. These absences were often due to personal reasons or conflicting commitments, which led to incomplete participation in our research. Additionally, we faced challenges with fluctuating enrolment numbers, as some students dropped out or joined mid-course, further complicating the consistency of the learning experience. These factors collectively contributed to variability in students' performance, thus affecting our data collection.

L1/L2 critical thinking barrier

While analysing both qualitative and quantitative data through student surveys, reflective discussions, and critical thinking assessment outcomes, we recognized that language skills can impact how students express their thinking, especially between first and second languages. Therefore, we assessed critical thinking based on the ideas conveyed, not on their language accuracy. In future we would amend our research question to: 'Can English language learners' critical thinking in English skills be improved through AI literacy training?'

Reflections

Our research design was not a product of immediate inspiration but rather the result of continuous reflection, adaptation to challenges, and the integration of new discoveries throughout the process. The process of implementing AI literacy

training was both challenging and enjoyable. On the challenging side, adapting the curriculum to effectively teach AI concepts required careful planning and continuous adjustment. We had to balance introducing complex topics in a digestible way while ensuring that students remained engaged and motivated. Technical issues, such as students struggling with software or accessing tools, added another layer of difficulty. Additionally, integrating AI literacy within existing coursework without overwhelming students was a constant challenge.

Despite these obstacles, the experience was also highly enjoyable. Witnessing students' gradual understanding and enthusiasm for AI was rewarding. The moments when students made connections between AI tools and their personal or academic lives – such as when one student mentioned how she could use ChatGPT to make her dream of founding a not-for-profit company come true, and when students successfully applied what they had learned in their projects – were particularly satisfying.

Recommendations

At the end of this process, we are left with an important question: is it worthwhile to teach General English students – whose primary goal is to develop work-ready English skills – about genAI, especially within a fast-paced curriculum?

We believe it is. Our observations showed that integrating genAI fostered a collaborative atmosphere and sparked lively discussions. Students made meaningful connections between AI tools and their personal or academic lives, and we noted a significant boost in their confidence and sense of accomplishment through the program.

However, our findings also highlight that while AI tools can enhance critical thinking, their integration into the curriculum must be deliberate and strategic. To achieve meaningful outcomes, a comprehensive, school-wide training plan is essential. We recommend gradually introducing AI tools that align with the curriculum, supported by collaborative training sessions for students, teachers, and administrators. This approach ensures AI supports learning rather than detracting from it.

References

Anderson, L W and Krathwohl, D R et al (Eds) (2001) *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*, Boston: Allyn & Bacon.

Bloom, B S (1956) *Taxonomy of Educational Objectives, Handbook 1: The Cognitive Domain*, New York: David McKay.

Crabtree, M (2023, 8 August) What is AI literacy? A comprehensive guide for beginners, *DataCamp*. <https://www.datacamp.com/blog/what-is-ai-literacy-a-comprehensive-guide-for-beginners>

Common Sense Education (2023) *AI literacy lessons for grades 6–12*. <https://www.commonsense.org/education/collections/ai-literacy-lessons-for-grades-6-12>

Heer, R (2012) *A model of learning objectives, based on a taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*, Iowa: Center for Excellence in Learning and Teaching, Iowa State University.

Johnston, S, McGrane, J A, Vendrell-Morancho, M and Hopfenbeck, T (2023) A multi-country comparison of lower secondary students' critical thinking under the International Baccalaureate and national curricula, *Review of Education* 11 (3). <https://doi.org/10.1002/rev3.3442>

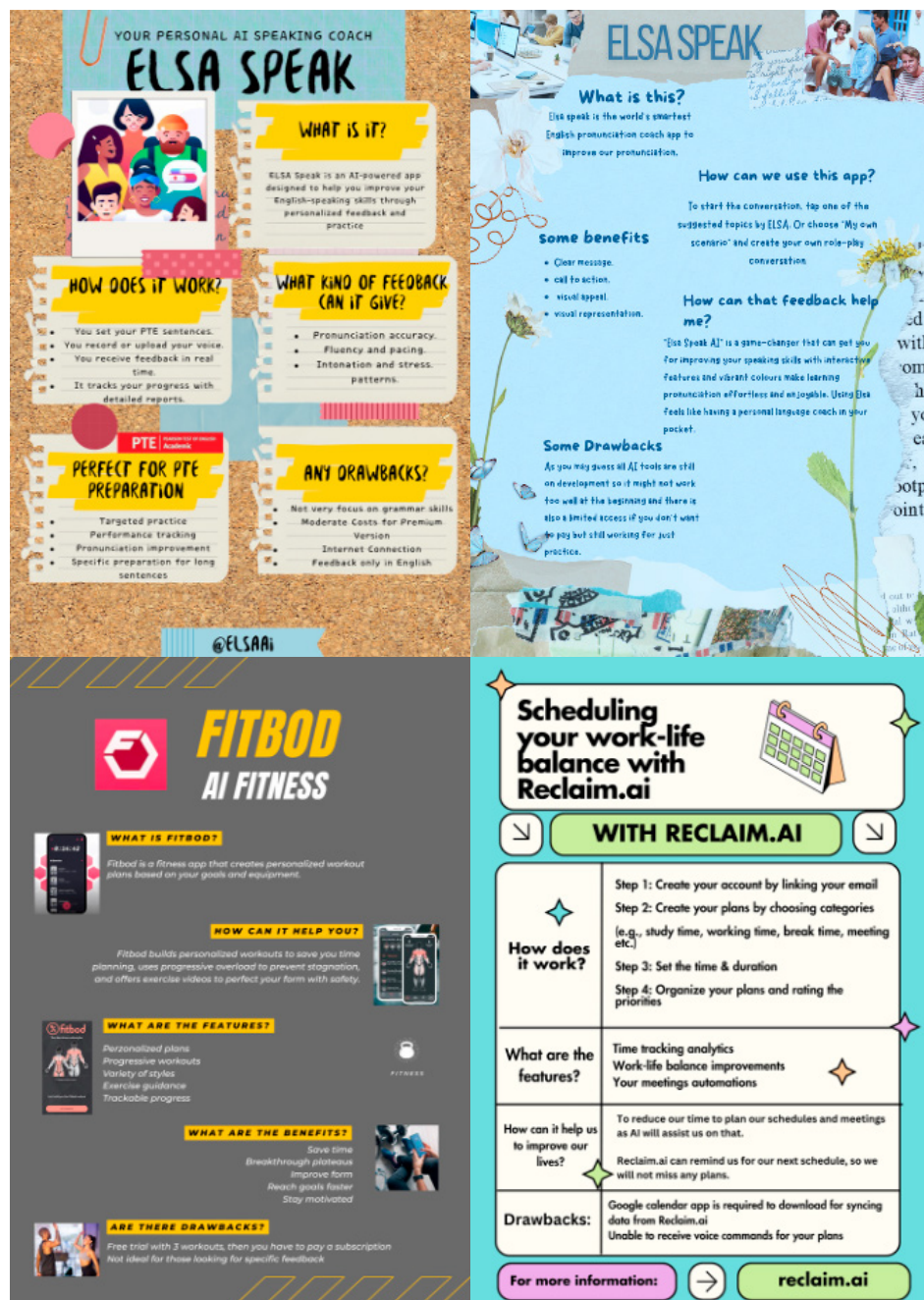
Lok, B, McNaught, C and Young, K (2015) Criterion-referenced and norm-referenced assessments: Compatibility and complementarity, *Assessment & Evaluation in Higher Education* 41 (3), 450–465. <https://doi.org/10.1080/02602938.2015.1022136>

Moosavi, L (2022) The myth of academic tolerance: The stigmatisation of East Asian students in Western higher education, *Asian Ethnicity* 23 (3), 484–503. <https://doi.org/10.1080/14631369.2021.1882289>

Reynders, G, Lantz, J, Ruder, S M, Stanford, C L and Cole, R S (2020) Rubrics to assess critical thinking and information processing in undergraduate STEM courses, *International Journal of STEM Education* 7 (1), Article 9. <https://doi.org/10.1186/s40594-020-00208-5>

Appendix 1: Student-created posters for the AI Showcase presentations

Three out of four groups used Canva to create them. None of the groups seem to have used ChatGPT to correct the grammar and spelling mistakes in their posters.



Appendix 2: Week 3 (Analysing/Synthesising): Completed assignment

Work with your partner to answer the following questions. Please write your answers clearly in the provided space.

Q1. Did the AI entry camera work the way it was supposed to?

No, it didn't.

Q2. Who are all of the different people or groups involved in this scenario?

The owner Emily Lianne, her friends: Iris, Amber, Zara, Elena, and Diego.

Q3. How might each of them be feeling?

They might be feeling rejected, embarrassed, disappointed because they expected to join the party.
At the beginning they might be feeling confused they didn't understand why couldn't enter to the house.

Q4. Was everyone impacted by the AI entry camera in the same way?

No, they weren't. While some people were rejected by the AI, some not.

How so?

Q5. Once Emily Lianne learns about this issue, what should she do?

What can the creators of the AI entry camera do to make their product fairer for everyone?

• She should train the AI camera with all styles of clothes and each different parts of the body.

* The creators

✓
Emily Lianne?

✓

Appendix 3: Week 8 (Analysing/Synthesising): Assignment

Action Research In-class Worksheet - Week 8

Name(s): _____ Date: _____

Your group will have 3-4 hours in class between today (07 June) and next Friday (14 June) to create your project. Answer the questions below to consider how you could use the two generative AI tools we've used in class to successfully complete the project. Then make a project plan.

In what ways is ChatGPT 3.5 useful or helpful?

We will use ChatGPT 3.5 to...

-
-
-

In what ways is Yoodli useful or helpful?

We will use Yoodli to...

-
-
-

Action Research In-class Worksheet - Week 8

Project Plan - 07 June

Goal	Tool(s)	Group Member(s)

Project Plan - 14 June

Goal	Tool(s)	Group Member(s)

Produced by Emily McNamee and Anne Gevana

Appendix 4: Week 8 (Analysing/Synthesising): Completed assignment

Your group will have 3-4 hours in class between today (07 June) and next Friday (14 June) to create your project. Answer the questions below to consider how you could use the two generative AI tools we've used in class to successfully complete the project. Then make a project plan.

In what ways is ChatGPT 3.5 useful or helpful?

You can get the information faster, correct your grammar, spelling, and vocabulary. ✓

As an evaluator interesting

We will use ChatGPT 3.5 to...

- Get ideas about life in Perth
- Correct our text, evaluate our grammar. ✓
- Professional vocabulary
- Learn topics X

In what ways is Yoodli useful or helpful?

to have a evaluation of our spelling - no - pron maybe?

Engagement and energy?

We will use Yoodli to...

- Evaluate our speech in our project
- not specific
-

Health

Problem. AI solution

→ Pharmacy 24/7

→ Quality of water in Perth

→ AI ChatGPT to solve it

→ AI to find the best company to buy a filter

Action Research In-class Worksheet - Week 8

Project Plan - 07 June

Goal	Tool(s)	Group Member(s)
Brainstorming	ChatGPT	All
Characterisation	Wore	
To select an area to focus.	App to putting cheapest data	
The population of Perth	App to put health	
We - AI to find pharmacies opening 24/7		

Project Plan - 14 June

Goal	Tool(s)	Group Member(s)

Produced by Emily McNamee and Anne Gevana

Appendix 5: Week 3: Student questionnaire focusing on Analysing/Synthesising

Student Questionnaire

Week 3 (29 April - 3 May)

* Indicates required question

1. What is your full and real name? *

2. Please summarise the main points of the party scenario. *

3. How difficult did you find it to summarise the main points of the party scenario? *

Mark only one oval.

1 2 3 4 5
Very ☐ ☐ ☐ ☐ ☐ Very difficult

4. How often did **you** break down big pieces of information into smaller ideas that were easier to understand? *Example: I looked at one sentence from the text with my partner and looked up a new word to better understand it.* *

Mark only one oval.

☐ Never
☐ Once
☐ A few times
☐ Often

5. How often did you ask **your teacher** to break down big pieces of information into smaller ideas that were easier to understand? *

Mark only one oval.

☐ Never
☐ Once
☐ A few times
☐ Often

6. What strategies did you use to analyse and synthesise the information from the party scenario? Choose all that apply. *

Tick all that apply.

☐ None
☐ I recognised similarities and differences. Example: I understood that the AI entry camera was biased, like other AI tools can be.
☐ I developed my comprehension and came to a conclusion on the information by using multiple sources or perspectives. Example: I listened to my teacher's input, read about it, or spoke to my classmate and thought deeply about it myself.
☐ I made connections to things I already knew. Example: The lesson's topic reminded me of what I learned in my last English class and it helped me to complete the activity.
☐ Option 5
☐ Other: _____

This content is neither created nor endorsed by Google.

Google Forms

Working with the enemy: Adopting a generative AI tool to teach AI literacy and increase active engagement with feedback on writing

Reema Sarwal and Liz Stoyanova, Deakin University English Language Institute, Melbourne

The main focus giving rise to the research

In the endeavour to develop international students' writing, the rise of generative AI (genAI) holds both promise and cause for concern. Rather than banning AI tools, our language centre's AI policy allows its selective use, while restricting use for summative assessments. The policy supports the use of genAI for developing skills, with the aim of promoting responsible use. Indeed, the need to equip students with the skills to manage AI is becoming part of a university's calling. Farrelly and Baker (2023:7) note that 'staff in universities need to develop literacy in the applications of AI to their own disciplines.'

GenAI tools can provide extensive and real-time feedback. However, AI tools also have the potential to produce or correct writing without feedback, leading to educator concerns about writing authenticity (Dawson et al 2024) and limited writing progress. Therefore, harnessing genAI tools for feedback in writing classes holds enormous potential but is in its early stages (Mahapatra 2024).

Another consideration for teaching writing is that contemporary writing tools offer AI features, such as predictive text and auto-correction. This has created a new distinction between original writing and what could perhaps be described as

‘untouched’ writing because most writing in electronic modes is ‘touched’ by AI at some level. As educators, we believe that students in English Language Intensive Courses for Overseas Students (ELICOS) need to undertake at least some proportion of ‘untouched’ writing to learn the language. Instead of using in-built tools, the conscious use of genAI at different stages of the writing process, such as planning and redrafting, might better develop writing skills.

Our context

Deakin University English Language Institute (DUELI) has a culturally diverse student cohort and offers a range of courses, including for academic direct entry. Students tend to have a strong desire to improve writing and are supported to do so via their course activities, teacher feedback and the Independent Learning Centre.

Microsoft Co-Pilot (protected mode) was our selected tool because all Deakin students have access to it and ‘protected mode’ promises more security. Co-Pilot was trialled with two classes, taught by Reema over 10 weeks, to explore its potential and manage the technical aspects. The first group was an English for Academic Purposes (EAP) class at IELTS 6.0–6.5. The next group was an EAP class at IELTS 5.5–6.0. It became evident that IELTS 5.5–6.0 was an ideal level to benefit from Co-Pilot feedback. These students could comprehend feedback, yet still benefit from quite basic language advice, whereas higher-level students received less feedback. During this time we refined our data collection and project design. Based on this, Liz created teaching resources and our survey forms.

Having made these refinements, the research was conducted with an EAP class of 14 students (IELTS 5.5–6.0). The group consisted of five students who would only study English during their stay in Australia, one study abroad student (that is, they would go on to study for one semester in an undergraduate program at Deakin University for which they would gain credit in their home country) and eight students on pathways to various award courses (undergraduate or postgraduate). The students were from Japan (5), Saudi Arabia (4), Vietnam (3) and China (2). There were six female and eight male students, the majority 18 to 24 years old, although the cohort from Saudi Arabia had some mature students, the oldest being 42.

Initial and subsequent research questions

We proposed trialling an AI tool for writing feedback using these questions:

How can an AI tool:

- improve critical AI literacy?
- increase formative feedback?
- empower writers to build autonomy?

As data were collected, it became clear that a longer study would be required to answer all these questions. What emerged, though, were data relevant to AI literacy and active engagement in the writing process. We arrived at the following research question: How does teacher-guided genAI use influence students' AI literacy and support students to engage actively with feedback on writing?

Our research

Three writing activities, all part of the curriculum, were used to carry out the project over the 5-week course. Two were argumentative while one was reflective. The project began by introducing DUELI's AI policy. This was followed by a discussion about students' experience using AI tools and concept-checking of the term 'feedback,' before the pre-activity survey (see Appendix 1) sought to understand students' attitudes and feelings about feedback on writing. This Microsoft Forms survey was completed in class, so students could seek clarification. Lastly, the action research (AR) project was introduced, and all students gave consent to participate.

The next phase was the use of Co-Pilot in class, using the following prompt:

**List my errors in the following text and explain how to correct my errors.
Do not rewrite the text.**

Directly after, students shared the received feedback on their devices on a collaborative Microsoft Teams file, which already had their original text. Students individually highlighted comments they thought most relevant and improved their drafts (see Figure 1).

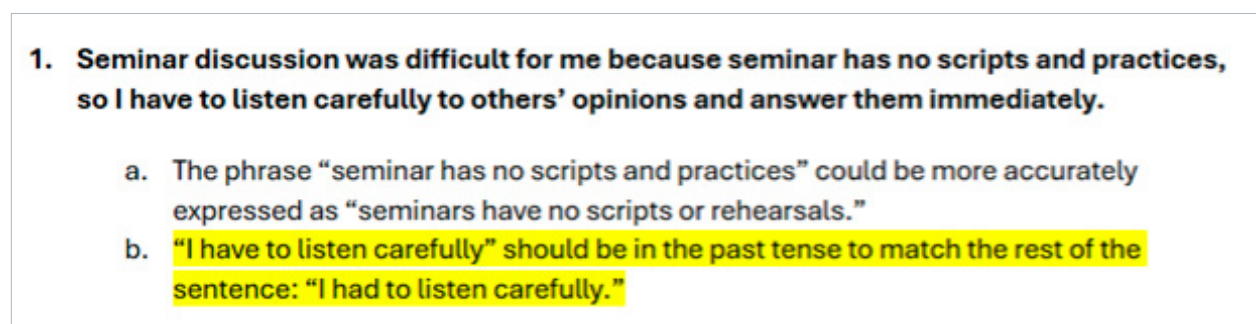


Figure 1: A sample of AI feedback with a student's highlighting

Later, they actioned the feedback and answered these questions:

- Do I agree with all the feedback?
- Did I notice something new – a word, a phrase or grammar rule?
- What will I be able to remember for my future writing?

Group discussions followed, allowing students to share one to two feedback items and comment on others' use of feedback. After discussing, individuals could make further revisions before submitting their final drafts for teacher feedback.

These steps modelled how to use an AI tool to learn. Students selected relevant feedback, thus exercising their agency, and were guided to evaluate feedback, rather than accepting feedback passively. The reflective questions aimed to scaffold thinking and the discussions had the two-fold purpose of generating peer-learning and reinforcement of what they had learnt. Altogether, our focus on bringing the writing *process*, not the *product*, to centre-stage, was the driver of this project design.

The first writing cycle was followed by a mini-lesson on the value of feedback, then redrafting to illuminate to students why they had been asked to use an AI tool. The next two cycles were followed by surveys (see Appendix 1). The writing cycles concluded with a reflective questionnaire (see Appendix 2). Finally, four students were selected to be interviewed based on questionnaire responses (see Appendix 3).

Data analysis

Even at the start of the research, students' responses and comments (see Figures 2 and 3) suggested they believe they need to play an active role when using AI, choosing words like 'it helps me solve mistakes' and AI is 'useful to help me improve.'

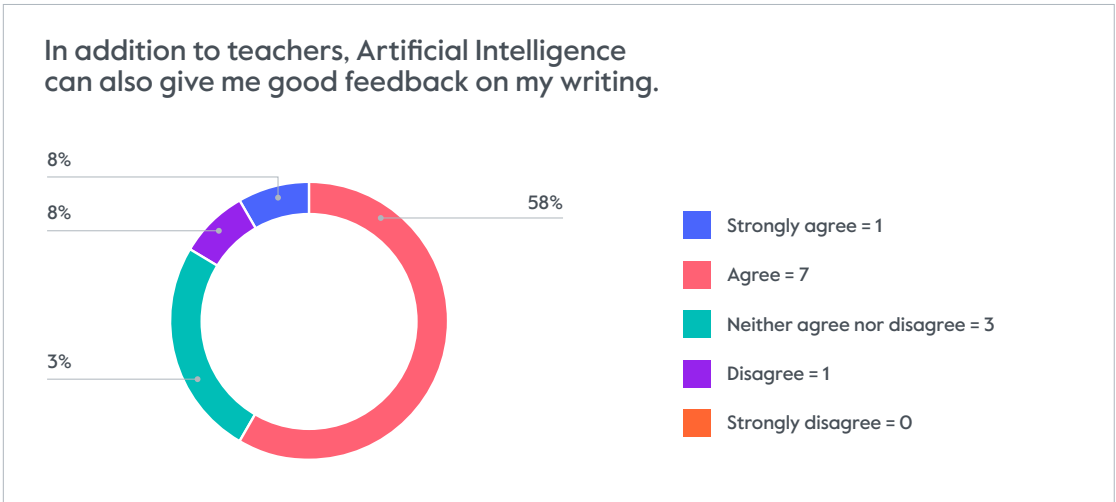


Figure 2: Students' attitudes to AI feedback

Students' responses revealed attitudes which can be broken down into four categories, presented in Table 1.

Table 1: Attitudes to AI feedback

Categories	Student comments
Trusting AI is always correct	<i>Accurate</i>
	<i>AI is more intelligence than people</i>
Not trusting AI is always correct	<i>Artificial intelligence sometimes gives me the wrong answer</i>
	<i>Wrong artificial intelligence sometimes</i>
Expecting to work with AI to improve skills	<i>It helps me to solve my mistakes</i>
	<i>AI can help us when we use it in the right way</i>
	<i>I have tried to use AI to give feedback and I recognise it was useful to help me improve</i>
Seeing limitations as AI is not human	<i>I just want to receive feedbacks from a real person</i>
	<i>AI not human he can't understand the emotional thing</i>

Catalyst for active engagement?

Our key question was whether an AI tool could be a catalyst for engaging actively with feedback. Despite some frustration, an AI tool seems to have potential to motivate students, as Table 2 illustrates.

Table 2: Student views on engaging with AI

Responses to feedback	Comments (Week 5)
Excited	<i>Happy because he improves my writing</i>
	<i>I was so excited</i>
Surprised	<i>Surprising because it's so accurate</i>
	<i>Surprised ... their feedback time is so fast</i>
	<i>Surprised because I didn't think I had numerous mistakes</i>
Empowered	<i>I feel good for my studying</i>
	<i>I learn something new due to how I can write new word</i>
	<i>It is helpful for me because it can improve sentence and skill</i>
	<i>I was grateful for the feedback</i>
	<i>Co-Pilot can give me feedback in 5 second</i>
Frustrated	<i>it gave me feedback that just exactly the same as my original sentence</i>
	<i>unuseful</i>

It can be seen above that the majority of students responded positively to AI feedback. Some expressed feelings of excitement, even stating *'he improves my writing'* perhaps suggesting a personal connection. Others were surprised by the AI tool's accuracy and speed. Another was surprised when AI identified *'numerous mistakes,'* which they had not expected. Students also felt empowered by specific advice: *'I can write new word'* and *'it can improve sentence and skill.'* However, two students were frustrated by their experience as AI did not give meaningful feedback. Indeed, we observed in the classroom that the quality of feedback did vary, despite students using the same prompt.

To continue our research, we introduced two key phases, the **selection of feedback** and the **discussion groups**, which were both intended to promote active engagement with feedback.

Selection of feedback

We were interested in what AI advice students identified as a learning point and what language students seemed to retain. In the survey, students most often cited vocabulary feedback (see Appendix 4). Below are examples of Co-Pilot vocabulary feedback that students highlighted and shared in their group discussion.

- 'I believe' instead of always using 'I think'
- ~~for long time~~ for **a** long time
- ~~relatively common~~ relatively common
- 'townspeople' is one word

In addition to vocabulary selections, there were some instances of grammar points students extracted from the feedback:

- **A** technology-driven remote working model significantly reduces

Occasionally, punctuation feedback was also selected:

- 'because' does not have a capital letter and start a sentence

Curious to understand why students selected vocabulary feedback more frequently than other feedback, such as grammar, we reviewed all the Co-Pilot responses. Initially, we suspected that students find vocabulary feedback simpler to understand. However, there appeared to be more to the picture. It seems that Co-Pilot may have a bias for vocabulary-related feedback. Table 3 illustrates that Co-Pilot generally manages feedback on vocabulary and simple punctuation. However, it has mixed success when providing feedback related to grammar, which may help to explain students' lack of focus on grammar points.

Table 3: Accuracy of Co-Pilot feedback

Types	Co-Pilot feedback	Accuracy of advice
Vocabulary	'peice' (should be 'piece')	Accurate, even with subtle word choice
	'loose' (should be 'lose')	
	'funniest one I ever studied': The word 'funniest' means humorous or amusing, which doesn't seem to fit the context. You likely meant 'most interesting' or 'most engaging'.	
Punctuation	In the second paragraph, there's a missing comma after 'For example.'	Accurate
	'Secondly, it can save money' is a run-on sentence.	Inaccurate as this is not a run-on sentence
Grammar	Verb tense: Change 'we first touched on the research article' to 'we initially discussed the research article.'	Inaccurate as it changes vocabulary, not verb tense
	'Recently, most of companies are adopting remote work system.'	Inaccurate as word order is not the issue
	Error: The word order is incorrect.	

Discussion groups

After highlighting feedback, students shared one to two items with their group. This was expected to be a catalyst for engagement, which was borne out in the findings. The majority of students valued this structure, identifying several reasons. Figure 3 illustrates how students engaged in discussion and Table 4 provides students' reasons for discussing feedback, divided into three categories: clarifying, collaborating, and reinforcing uptake.

Here students sat with their established discussion groups and shared learning points they had selected from the AI feedback. Students could take on the role of tutor, explaining the learning point, or reflect on the feedback, sometimes seeking clarification.

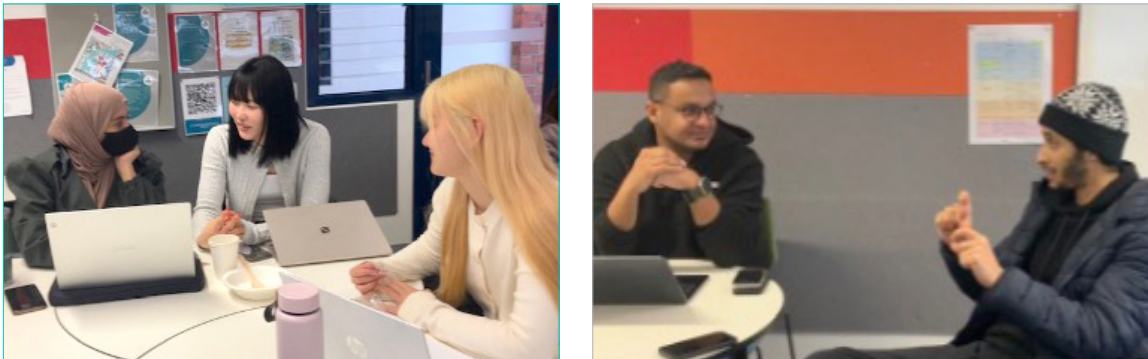


Figure 3: Students discussing AI-generated feedback on writing

Table 4: Comments on discussion phase from Week 3 and 4 surveys and interviews

Reasons	Comments
Clarifying	<i>when we do not understand the means we need to talk about with others</i>
	<i>to consider whether this is an appropriate proposal</i>
Collaborating	<i>I like giving my opinions and hearing some experience</i>
	<i>I could find new advice in my friend's feedback</i>
	<i>I like it when I saw my friends were happy about their feedback</i>
	<i>I am sure to get help in a group from discussing</i>
Reinforcing uptake	<i>promotes non-return to the mistake and remembering</i>
	<i>share the new knowledge is good for remember</i>
	<i>it will be after discussing in my brain long time</i>

Interestingly, students’ perceptions of the importance of the discussion changed by the end of the cycle (see Figure 4). Some indicated it was vital and others that it could be omitted. One explanation is that students were becoming more autonomous. Yet, it also emerged that two students felt uncomfortable sharing feedback because they felt it was private. Conversely, some were adamant that discussions were essential.

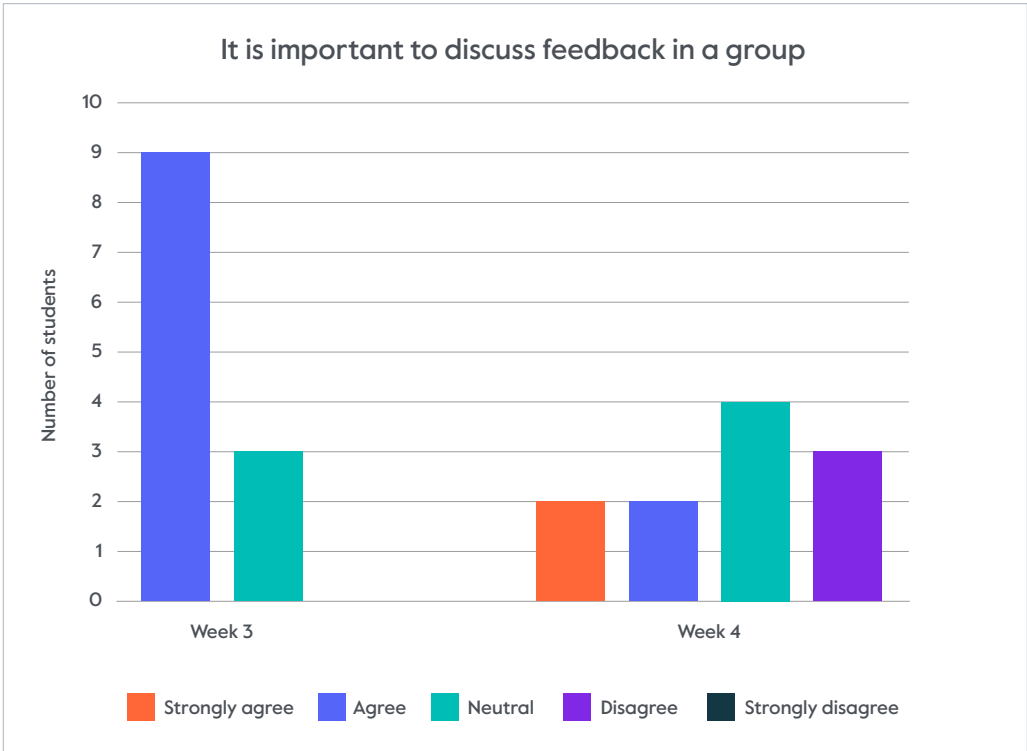


Figure 4: Importance of discussion phase based on Week 3 and 4 surveys

Classroom observations

The discussion regarding students' use of AI prior to the first survey revealed that students' focus was on correcting their work, not on improving their writing skills for future tasks, so their focus was on product rather than process in writing. After receiving feedback on the first task, students needed time to understand that highlighting feedback was not just for re-drafting, but also for them to take on the feedback for future writing. One student, Saba, had used translation software to write her text and was disappointed that she could not participate in the next steps of redrafting and discussing. It was heartening to observe that in later cycles, she only translated words instead of large chunks. Overall, students became more efficient at selecting, actioning and discussing feedback, reducing the time needed to move through the steps.

Another observation was that improving AI literacy required overt teaching of the definition and purpose of feedback. Students could be better guided to use AI as a writing tutor when they understood the purpose of AI-generated feedback and were reassured that they would still receive teacher feedback on final writing drafts.

Conclusions

Our research showed that a genAI tool can be effective in writing classes and that AI tools should not be a taboo subject, as they need to be discussed and used responsibly. It is vital to scaffold students to be active users so they can avoid the trap of using AI to write or make corrections for them.

Our design included discussions which increased engagement and collaboration. While valuable, discussion did not suit all students and there was a decline in perceived need towards the end, suggesting that the optimal design may be to have more scaffolding earlier, allowing student autonomy to grow.

Throughout the cycle, students took responsible, active roles when using AI. When students were asked if they would use Co-Pilot in future, of the 12 students who responded, 10 indicated that they would use the tool and would retain the prompt. The responses show an intention to use the tool in a responsible way (see Appendix 5). For example, Willow appreciated using an AI tool outside class to 'teach and fix' and Suzuna's intention to use the prompt 'because rewriting by Co-Pilot is not helpful for me' showed she is aware of avoiding asking the tool to rewrite the text.

However, Max was less enthusiastic because it 'costs time' and Ed explained that he had 'Grammarly to show my spelling and grammar errors.' To understand Ed better, we interviewed him; to his surprise, it occurred to him during the interview that 'when we use Co-Pilot, we have to think about it ... to memorise it,' indicating that he had recognized that engaging with AI feedback may have long-term benefits for his writing, although he found the process more demanding.

Our project demonstrated that when the focus of formative feedback was repeatedly shifted for the students over several tasks to the process rather than the product, students were able to appreciate the potential for learning offered

by genAI. This was illustrated above in the examples of student responses, such as realising genAI should not be used before the first draft submission, noticing the sheer number of errors, and acknowledging that Grammarly doesn't teach. This repeated shift in student focus was achieved by genAI feedback as students were aware that the comments may or may not be fully accurate or relevant, which made it necessary for them to actively select what was meaningful to them.

Reflections and implications

GenAI tools are a source of language teacher concern due to the substantial potential for academic breaches in writing assessments (Dinneen 2021). This concern can lead to an understandable defeatism among educators who sense that writing will no longer be learned by students due to the ingrained dependence on AI tools to do their writing. While making policies around the ethical use of genAI is the first step, demonstrating ethical use makes the policies relatable. Moreover, modelling responsible genAI use in the writing classroom can provide an outlet for teachers to discover the potential of this new technology, without compromising their lessons.

On reflection, it was AI's very imperfection that served the purpose of giving students the concrete action of selecting feedback. The process of evaluating genAI's comments increased their engagement with their writing. When students receive feedback from teachers, on the other hand, they are not in the position of 'masters' selecting whatever they like.

The implication here is that there is a fundamental difference between how students engage with genAI feedback as opposed to authoritative feedback provided by teachers. In other words, students may not be 'more' engaged with genAI feedback but are likely to be 'differently' engaged because of their sense of agency over AI. This unique form of engagement can create valuable opportunities for learning and teaching. With this in mind, the imperfection of genAI feedback need not be a 'cringe factor' for teachers, preventing them from encouraging students to use it.

It is exciting that genAI tools have the potential to monumentally increase the amount of personalized feedback received by each student. What is equally exciting is that the students can benefit from the way they engage with it—exercising their agency and autonomy. Both students and teachers need to be aware that genAI feedback is a supplement, not a replacement, for teacher feedback. Moreover, the teacher has a pivotal role in guiding students to actively process the genAI feedback.

Our research suggests several avenues for further research. Future AR using prompts for improving specific areas of writing like subject-verb agreement or hedging language would reveal more about what guidance teachers can provide. Multiple ways of engaging with the feedback received could also be explored. Additionally, students could be encouraged to brainstorm new prompts and then decide which ones are ethical, based on whether they help in improving skills rather than generating the finished product. The effectiveness of using follow-up prompts to ask for explanation and examples of the feedback received also needs to be studied. The current project and future ones are steps towards developing both

students' and educators' awareness of using genAI optimally to enhance formative assessment in the field of ELICOS.

To summarize our research, the most important contribution of genAI feedback was enabling students to work on additional drafts of their writing before submitting it for teacher feedback. The added steps of feedback selection and discussion increased engagement and reflection. Students had opportunities to improve their writing skills by learning useful language points, predominantly related to the use of vocabulary and, to some extent, to grammar and punctuation. Actioning only the feedback that they selected according to their own judgement made the task manageable, and having agency motivated students to engage with feedback. Looking to the future, enabling students to use genAI to spend more time engaging with the writing process, instead of using it to breach academic integrity, promises to be an effective way to turn the tables on 'the enemy.'

References

Dawson, P, Bearman, M, Dollinger, M and Boud D (2024) Validity matters more than cheating, *Assessment & Evaluation in Higher Education* 49 (7), 1,005–1,016.
<https://doi.org/10.1080/02602938.2024.2386662>

Dinneen, C (2021) Students' use of digital translation and paraphrasing tools in written assignments on Direct Entry English Programs, *English Australia Journal* 37 (1), 40–51. <https://files.eric.ed.gov/fulltext/EJ1341751.pdf>

Farrelly, T and Baker, N (2023) Generative Artificial Intelligence: Implications and Considerations for Higher Education Practice, *Education Sciences* 13 (11), 1–14.
<https://doi.org/10.3390/educsci13111109>

Mahapatra, S (2024) Impact of ChatGPT on ESL students' academic writing skills: a mixed methods intervention study, *Smart Learning Environments* 11 (1), 1–18.
<https://doi.org/10.1186/s40561-024-00295-9>

Appendix 1: Surveys conducted via MS Forms

Week 1 Survey (Pre-writing)

1. I am more interested in receiving a mark for writing than comments.

Strongly agree / Agree / Neither agree nor disagree / Disagree / Strongly disagree

2. I mainly check feedback to find out if my writing is strong, satisfactory or unsatisfactory.

Strongly agree / Agree / Neither agree nor disagree / Disagree / Strongly disagree

3. Feedback about writing helps me to notice errors the next time I write.

Strongly agree / Agree / Neither agree nor disagree / Disagree / Strongly disagree

4. I remember the feedback about my writing for the next time.

Always / Usually / Sometimes / Rarely / Never

5. I feel confident when I read positive feedback from my teacher about my writing.

Strongly agree / Agree / Neither agree nor disagree / Disagree / Strongly disagree

6. I feel uncomfortable when I have feedback about mistakes or problems in my writing.

Strongly agree / Agree / Neither agree nor disagree / Disagree / Strongly disagree

7. Why did you choose that answer for Question 6?

Enter your answer:

Week 2 Survey (Post-writing)

1. I enjoy using Co-Pilot.

Strongly agree / Agree / Neither agree nor disagree / Disagree / Strongly disagree

2. Why did you choose that answer for Question 2?

Enter your answer:

3. I understand the feedback from Co-Pilot.

Strongly agree / Agree / Neither agree nor disagree / Disagree / Strongly disagree

4. Why did you choose that answer for Question 3?

Enter your answer:

5. I remember the Co-Pilot feedback the next time write.

Always / Usually / Sometimes / Rarely / Never

6. When Co-Pilot gives me feedback I feel _____ because

_____.

7. It is important to discuss Co-Pilot feedback in a group.

Strongly agree / Agree / Neither agree nor disagree / Disagree / Strongly disagree

8. Why did you choose that answer for Question 7?

Enter your answer:

Week 4 Survey (Post-writing)

1. I enjoyed using Co-Pilot this week.

Strongly agree / Agree / Neither agree nor disagree / Disagree / Strongly disagree

2. Why did you choose that answer for Question 2?

Enter your answer:

3. I found it easy to understand the feedback from Co-Pilot.

Strongly agree / Agree / Neither agree nor disagree / Disagree / Strongly disagree

4. Why did you choose that answer for Question 3?

Enter your answer:

5. I will remember the Co-Pilot feedback the next time I write.

Very likely / Likely / Possibly / Unlikely / Very unlikely

6. When Co-Pilot gave me feedback I felt _____ because

_____.

7. It was important to discuss Co-Pilot feedback in a group.

Strongly agree / Agree / Neither agree nor disagree / Disagree / Strongly disagree

8. Why did you choose that answer for Question 7?

Enter your answer:

Appendix 2: Week 5 reflection conducted via MS Teams

Final reflection: Using Co-Pilot for feedback

Write a reflection about your experience of using Co-Pilot for feedback on your practice writing in EAP3.

1. How many Co-Pilot activities did you do? (1, 2 or all 3)
2. How difficult or easy was it to find Co-Pilot on your device?
3. How well could you understand the feedback given by Co-Pilot?
4. How did you feel when Co-Pilot gave you feedback? Why?
5. Give 1–2 examples of what you still remember learning from Co-Pilot feedback.
6. Do you think you will use the 'prompt' used in class to get feedback on your writing from Co-Pilot in the future? Would you change the prompt?
7. What did you like about discussing your feedback in a group? What did you not like about it?
8. Can you think of any ways to improve how we use Co-Pilot for feedback on practice writing in class?
9. Do you think it was good use of your time to use Co-Pilot for getting feedback? Why?
10. Would you use Co-Pilot in your own time for practice writing?

Appendix 3: Data and findings

Data	Purpose	Main findings
1. Pre-activity survey	To understand students' attitudes to feedback and receiving AI feedback	Students are self-aware regarding feedback and nearly all welcomed honest feedback; mixed attitudes to AI feedback
2. Post-activity surveys	To collect student responses directly after the AI writing activity	Mostly positive attitude to tool and activity design, reflections on what they learned and enjoyed
3. Classroom observations	To record how students engaged with the tasks, their spontaneous questions, triumphs and challenges	Technical troubleshooting required, students actively engaged in selecting feedback, as opposed to passively consuming; re-drafting in class modelled how to learn from mistakes
4. Student writing and examples of their selection of AI feedback	What are students selecting/remembering?	Vocabulary, language chunks including collocations, verb forms, articles, punctuation
5. Open-ended questionnaire	To gauge at the end of the cycle how much students valued using the tool and why and their opinions on the different learning phases and what they gained	Most students enjoyed using Co-Pilot and could cite advice or new language; discussion groups were mostly valuable; most students would use Co-Pilot again
6. Semi-structured interviews	To delve into significant comments from the open-ended questionnaire	Complex attitudes to discussion groups, some frustration at the volume of feedback AI generates, but appreciation of the potential to learn

Appendix 4: Selection of feedback

The table below captures examples of language identified by students which we have labelled V (Vocabulary), G (Grammar) or P (Punctuation).

V/G/P	Student examples of remembered feedback
V	replace 'everywhere' with more specific vocabulary
V + G	usage of words and grammar
V + G	I'm still bad about writing I still need to improve my writing skills
V	'townspeople' is one word
P	'because' does not have a capital letter and start a sentence
V	'I believe' instead of always using 'I think'
V + G	relatively common relatively common
V + G	for long time for a long time
G	It showed I used the wrong article for a noun
G	A technology-driven remote working model significantly reduces

Appendix 5: Intentions to use Co-Pilot in future

Suzuna shows she is aware of avoiding having an AI tool rewrite the text for her:

'Yes, I will use this prompt, because rewriting by Co-Pilot is not helpful for me.'

Saba intends to use the prompt and adjust prompts to gain more explanation:

'I think I will use it and if I need more information, I will write more explain.'

Willow appreciates using an AI tool when she does not have a teacher to 'teach and fix':

'Yes. Because I don't have people who fix the sentence when I study English using own time. So, Co-Pilot teach and fix English anytime.'

Sakura realizes that this tool is used to develop writing, not demonstrate writing in a test:

'I would like to use Co-Pilot for my practice writing.'

Sydney gave an enthusiastic response:

'Absolutely yes.'

Students generally seem to recognize that effective and responsible use of an AI tool requires active engagement. The importance of highlighting and rewriting, or doing something with feedback comes through in these comments at the end of the cycle.

Little intention to use Co-Pilot and the prompt in future

Max found it was time-consuming:

'It pretty costs my time.'

Ed responded that he would not use Co-Pilot or the class prompt as he had 'Grammarly to show my spelling and grammar errors', which was more efficient. When interviewed later, he reflected on this, though, and concluded that having AI fix mistakes is convenient but he learned more from Co-Pilot in class:

'When we use Co-Pilot, we have to think about it, ... to memorise it.'

Impacts of generative AI application use on students' perceptions of pronunciation

Gabriel Alejandro Azpilcueta and Heather Elizabeth Sparrow, English Language Centre, The University of Adelaide

Introduction

Generative Artificial Intelligence (genAI) has fostered creativity in English language teaching and learning at our English Language Centre (ELC) at The University of Adelaide, Australia. Learning areas that proved challenging due to limited time or resources in the pre-genAI world may now be possible to manage with thoughtful, ethical application of AI tools in the classroom (Kandybovich 2024, as cited in Hamdan and Sowton 2024b).

Pronunciation is one such area English language learners and educators have flagged for improvement (Brown 2012). We argue that genAI applications may be the long-awaited solution to help alleviate what scholars have identified as typical pedagogical barriers for learning pronunciation (see Brinton 2019, Brown 2012, Field 2019). Brinton (2019:249–250) highlights the struggle educators have knowing how to incorporate pronunciation into study programs, giving students continuous effective feedback on their pronunciation, and evaluating students' progress. The dearth of teacher preparation to support L2 learners' needs around 'awareness of specific phonetic characteristics of English that cause difficulties' in listening has been illustrated by Field (2019:212), and Gilbert (2010 as cited in Brinton 2019:249) concurs regarding pronunciation skills pedagogy generally. Brinton (2019:251) cautions that 'intelligibility' should be a prioritised goal of 'pronunciation pedagogy' ahead of native speaker-model benchmarks. Brown (2012:18) discussed many of these points, warned that conventional ideas about 'teaching pronunciation' were outdated

because perceptions of ‘good’ English pronunciation standards had evolved, and acknowledged ‘[t]he issue of “correct” English pronunciation has always been a contentious one’ (Brown 2012:18), citing evidence for this discussion from Crystal (1997, as cited in Brown 2012), Jenkins (2000, as cited in Brown 2012) and Hewings (2004, as cited in Brown 2012). These obstacles have impacted pronunciation-related listening skills (input) (see Brinton 2019, Field 2019), pronunciation-related speaking skills (output) (see Brinton 2019), and pronunciation-related skills linked to students’ self-perceptions of progress: ‘individual’ concerns, ‘awareness-building,’ and their effects on students’ ‘confidence, communicative ability and motivation’ (Brown 2012:19). Accordingly, the potential positive applications of genAI to support students’ learning make this an exciting time for us to research this long-standing gap in English language teaching (ELT) within our context; in response, our action research (AR) inquired how the use of genAI applications affects students’ perceptions of pronunciation. Our definition of pronunciation was assumed into our AR context from Brown (2012:18): ‘the way a student verbalises a word, with effective communication being the measure of success.’

Our AR used genAI technology in a repeated mixed-methodology intervention expanding Brown’s (2012) pre-genAI study, which originally aimed to examine innovative pronunciation improvement methods for advanced learners studying English at a Perth, Western Australia language institution. Brown (2012:18) sought to highlight elements that delayed learners’ ‘intelligibility in their own speech’ by allowing them to focus on self-selected areas impacting their communicative competence, and apply critical thinking to evaluate and develop their individual accent and tailored pronunciation objectives. Brown (2012:18) also asserted that allowing students to engage with peer feedback for ‘critical ear’ development of individual pronunciation issues provided agency over certain personal accent features they would like to enhance. We also incorporated these ideas in our AR, alongside applying concepts of the ‘three levels’ of receiving information when hearing (Smith 1988, as cited in Kachru and Nelson 2001:21), the ‘native speaker’ (Jenkins 2024, as cited in Hamdan and Sowton 2024a; see also ideas adapted from Brown 2012), and ‘English as a lingua franca’ (see Hutton n.d.) to our intervention when designing our cohort-specific pronunciation course and resources for our Centre. We did this because we were curious to understand our participants’ perceptions of these areas. Inspired by these ideas, we created our pronunciation course combining our traditional ELT expertise with cutting-edge genAI tools to help bring pronunciation pedagogy into the 21st century for our participants.

In reflecting on the literature for our AR, we found it interesting that Field (2019:213) also explains concepts of ‘intelligibility’ and ‘comprehensibility’ (Smith and Nelson 1985, as cited in Field 2019:213), and questions whether comparisons can be made among listener/speaker ‘phonological systems’ to determine successful ‘intelligibility’; Field (2019:213) admits this debate has produced inconsistent findings in scholarly research, indicating ‘the complexity of the speaker-listener relationship.’ Our AR findings from our participants’ demographic sample may add further complexity to this debate, as we brought the addition of genAI apps’ pronunciation practice and formative feedback into this mix.

Rationale

We created our AR in response to ELC students' requests for assistance with individualized English pronunciation challenges with an extended rationale (Figure 1). Our objective was to address the pronunciation-related pedagogical needs of our learners, teachers and Centre, including individual challenges, insufficient class time or scant suitable resources. We considered the value and validity new genAI tools offered as potential solutions for longstanding pre-genAI issues: could we use these apps to deliver student-specific and tailored pronunciation lessons with high-quality resources more promptly and efficiently? Mindful of previous work by The University of Adelaide (2023, as cited in Hutton and Sparrow 2023) and by Liu and Bridgeman (2023, as cited in Hutton and Sparrow 2023) at the University of Sydney arguing the importance of incorporating genAI responsibly into learning, teaching and assessment, we also asked: could genAI tools fill this gap responsibly in ELT at our Centre?

We were also aware of debates about equitable use of genAI applications in English language education. We ventured into investigating what Burns (2024, personal communication) describes as 'power bias' and the 'commercialisation of AI tools,' and the concept that AI foregrounds 'prestige English' (Hamdan and Sowton 2024a). Figure 1 below illustrates the various components we aimed to include in our AR.

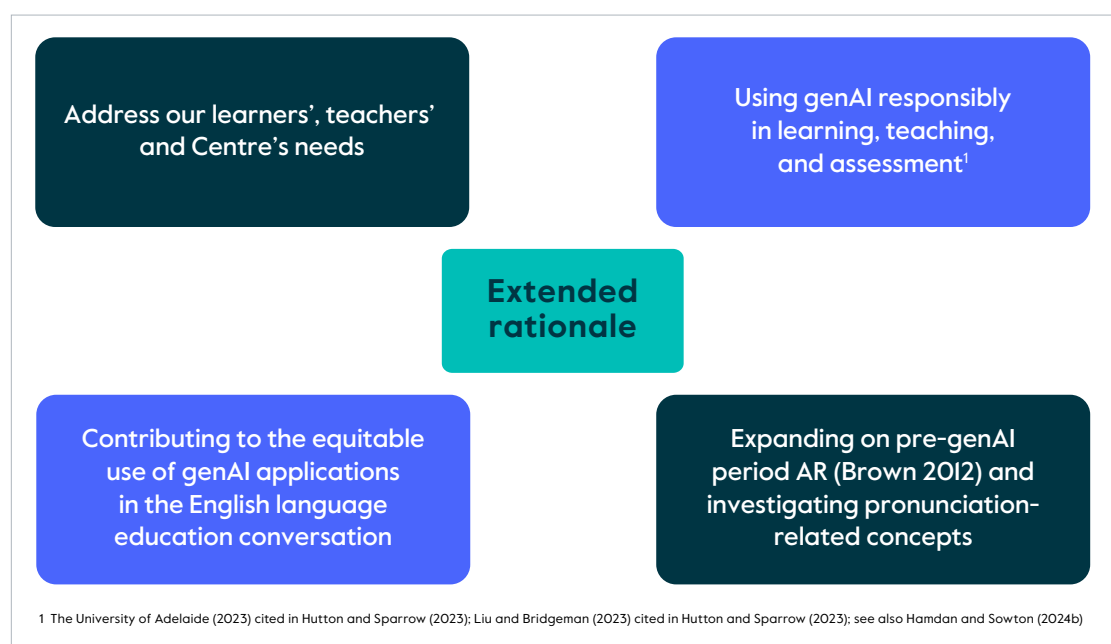


Figure 1: Our action research rationale

Our rationale grew through the 'cyclical' process (Kemmis and McTaggart 1988, as cited in Burns 2010:9) of our AR to encompass deeper considerations of important sub-categories of genAI applications in ELT pronunciation pedagogy (see Appendix 1). These stemmed from intrinsic core learning and teaching values related to 'humanistic approaches' (Harmer 1991:35) with the human-teacher at the centre when incorporating genAI in the classroom (see Kandybovich with Sowton 2024, as cited in Hamdan and Sowton 2024b).

Context

The ELC is a medium-sized provider in South Australia which offers an intensive, direct-entry university preparation program to The University of Adelaide for international students seeking to attain admission requirements through coursework. Generally, for many candidates, this means achieving a 6.5 IELTS standard equivalent by the end of the program. Students enrolled in Semester 1, 2024 of this pathway program participated in our research during the AR timeframe. The students belonged to either Heather or Gabriel's class (14 and 16 students, respectively). However, because of enrolment policies, student numbers fluctuated. Figure 2 provides an overview of the student demographics based on our survey data and class records; students represented a range of Asian countries with the majority from China.

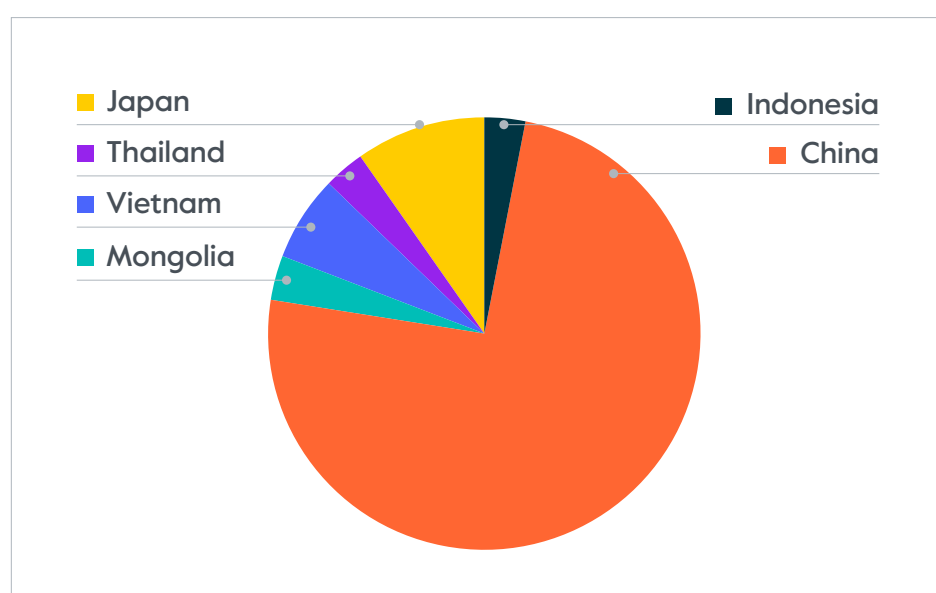


Figure 2: Demographic characteristics of participants

AR question development

Our AR question had five iterations (see Figure 3). Initially, we planned to research one genAI app (ELSA Speak) but due to subscription constraints, we broadened the app choice and incorporated intelligibility and English as a lingua franca concepts. Later, we realized measuring students' perceptions rather than collecting exact measurements of students' phonetic output would be more fruitful and added this focus. Fourth, we included the concepts of *input* and *output* in response to discovery of our research gap explained earlier.

Our final AR question emerged: How does the use of genAI applications affect students' perceptions of pronunciation? (Figure 3). For our purposes, 'perceptions' included two inquiry areas related to how participants regarded: (1) the *utility* or *value* of genAI apps for pronunciation practice; and (2) their own confidence with English pronunciation including intelligibility.



Figure 3: The evolution of our action research question

AR process

Our AR interventions (see Appendix 1) were conducted over 15 weeks in three cycles adapted from Kemmis and McTaggart (1988, as cited in Burns 2010:9) with bi-weekly lessons implementing our mixed-methodology course creation process as a repeated technique. Lessons commenced after we collected written consent in AR introductory sessions with students.

From Cycle 1, we delivered two linked lessons per week and mixed traditional pronunciation methods with genAI app instant feedback for participants in the final stage of each lesson. The first lesson per week introduced specific pronunciation features, which students continued to practice in the follow-up lesson. Appendix 1 provides more detail about the content sources used for the lessons and Appendix 2 offers a sample lesson. We applied our ELT expertise with genAI assistance in a two-tiered approach to our lesson design by prompting genAI applications including ChatGPT versions, Microsoft Co-Pilot and ElevenLabs to produce resources to support our students' pronunciation needs. Other traditional resources were also incorporated where needed. We used these genAI-produced materials to scaffold our lessons by intertwining classic ELT methodology (see Harmer 1991) with no-cost features of genAI apps namely Phonetic, Bold Voice, and ELSA Speak, all freely accessible on participants' mobile phones. Cambridge University Press & Assessment (2024) and the Pronuncian app were used as back-up resources for students who could not access Phonetic to study the International Phonetic Alphabet (IPA) early in Cycle 1.

In Cycle 1, participants practiced self-selected texts with the Bold Voice and Phonetic apps. Cycle 2 was similar but incorporated ELSA Speak and ElevenLabs, a voice generation tool that offers speech in different accents, to address the receptive skills research gap referred to by Edmett, Ichaporia, Crompton and Crichton (2024; Burns 2024, personal communication). We also discussed the concepts of *intelligibility*, *the native speaker*, and *English as a lingua franca* with our students. In Cycle 3, students did speech contest practice on the topic of *the benefits of using English as a lingua franca in an English-speaking setting*.

To strengthen the reliability of our Cycle 1 lesson design, we adopted suggestions by ChatGPT and Microsoft Co-Pilot about L1-interference challenges in English pronunciation based on our prompting of AI with anonymous information related to our students' first languages and/or nationalities. We then cross-checked these with traditional evidence (Swan and Smith 2001). Figure 4 shows examples of the challenging English sounds for Japanese and Chinese speakers we documented in our notes from Swan and Smith (2001). We also drew on Jenkins' notion of "the Lingua Franca Core" ... [that included] ... most of the consonant sounds ... initial consonant clusters ... long and short vowels, and ... contrastive stress' (Jenkins 2002, as cited in Nation and Newton 2009:77). Once we identified these, this Lingua Franca Core was broadly maintained in subsequent materials development.

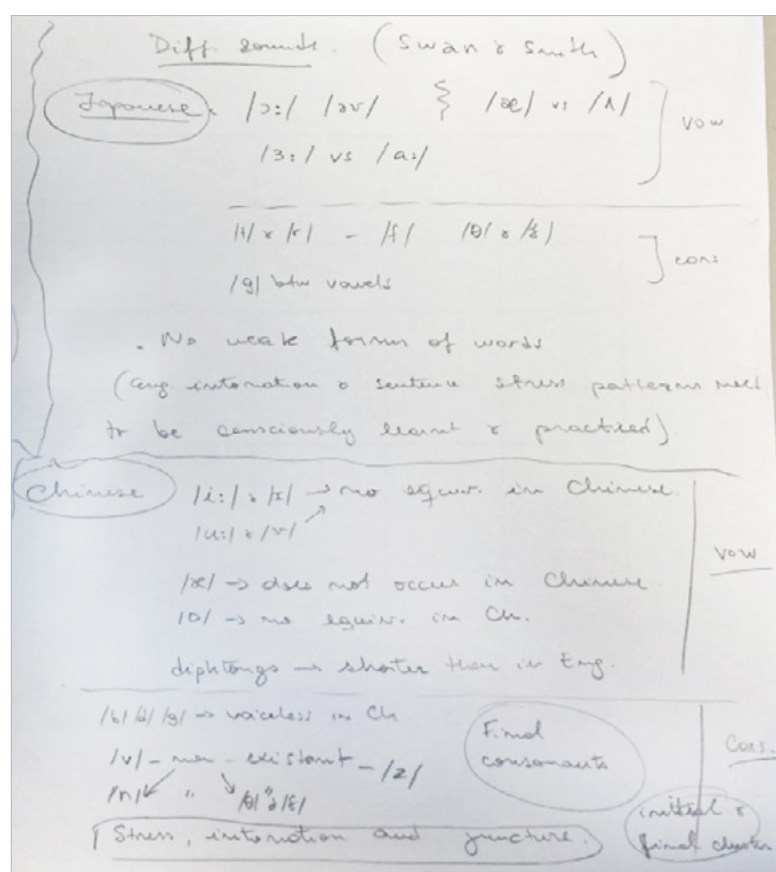


Figure 4: Example of research notes on L1 challenges identified from traditional literature (Swan and Smith 2001)

We made PPTs to deliver our traditionally scaffolded ELT lesson stages (warming up, teaching, modeling, engaging ELT-based task practice) often with paper-based resources and/or genAI tools like ElevenLabs in class. Students practiced these pronunciation exercises in pairs or groups (see Figure 5).

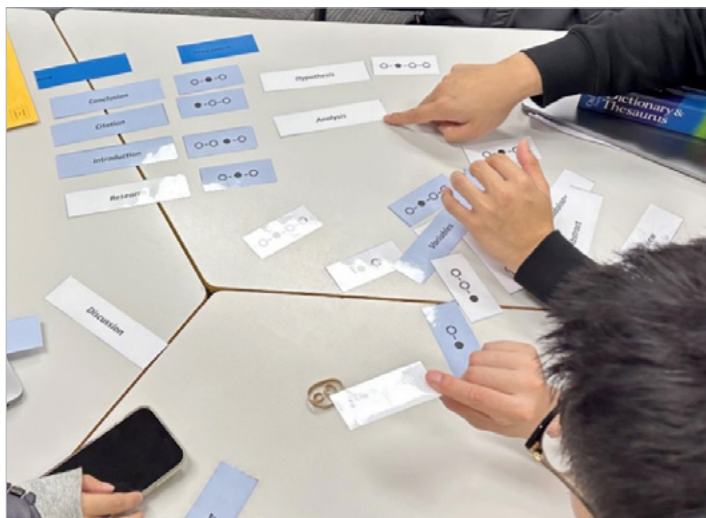


Figure 5: Academic word cards to corresponding stress patterns

We then implemented our final stage of the lesson. Students were asked to choose a self-selected text to practice the lesson's target pronunciation feature(s) with the aim of receiving genAI formative feedback and potential validation of their progress. This step allowed us to check students' grammar and communicability to ensure they were entering an accurate text into the apps. Following this, students used their devices to read their text aloud to either Bold Voice or ELSA Speak. The apps analyzed the students' voice samples providing colour-coded feedback showing the accuracy of the student's pronunciation of the text as a percentage score (see Figures 6 and 7). By Cycles 2 and 3, participants had agency to select their preferred app to complete this process.

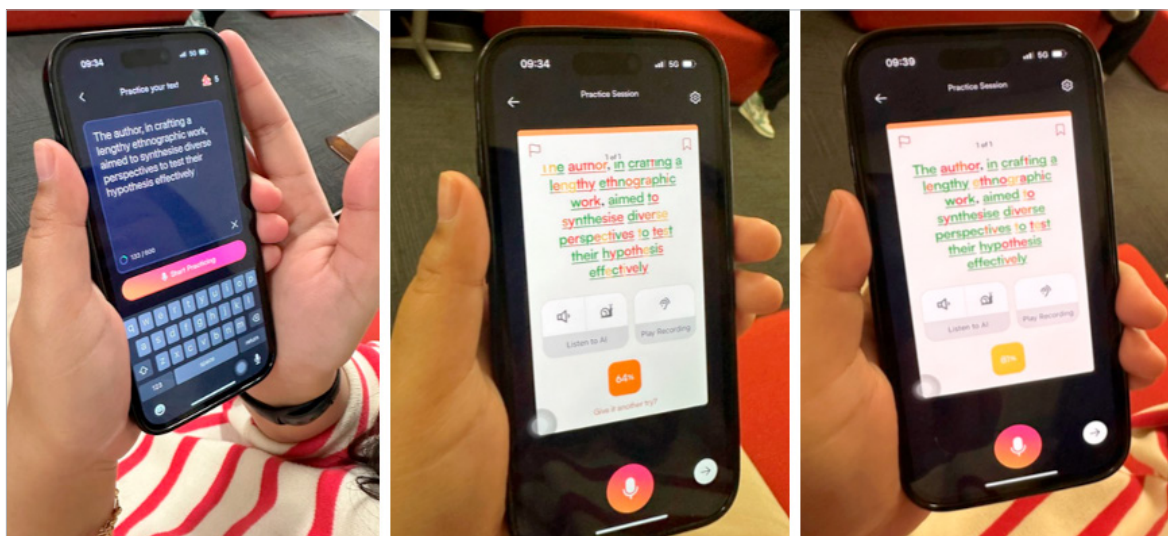


Figure 6: Bold Voice – participant-chosen text input and app feedback process while sitting with a teacher

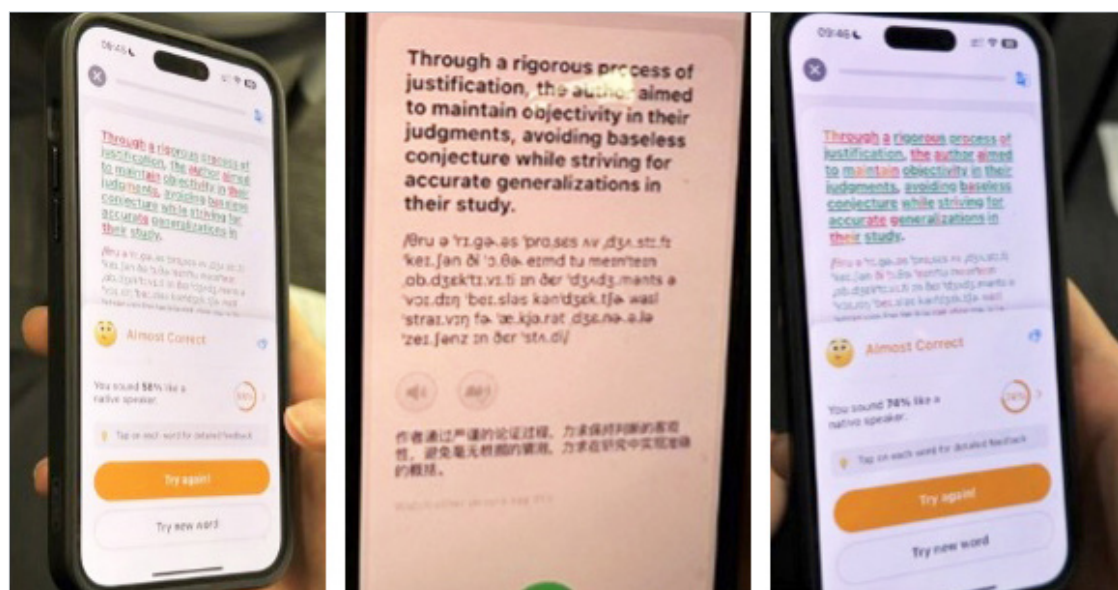


Figure 7: Elsa Speak – samples of participant-chosen text input and app feedback process

Based on feedback received, students accessed interactive features in the application (visual guides, videos, or sounds clips with an option to re-record individual words) to try to improve their scores. During this stage, we actively supported students with their pronunciation needs, helped them interpret the app feedback, provided encouragement, and offered targeted guidance for sound articulation and production.

Table 1 summarizes the steps in our repeated technique for lesson design and implementation when using our mixed methodology.

Table 1: Repeated technique for lesson design and implementation using genAI apps

Three-cycle pronunciation course design	Traditional lesson scaffolding in class	Self-directed practice	Immediate free genAI feedback
<ul style="list-style-type: none"> Cohort-specific pronunciation targeted for lessons and materials ChatGPT (different versions) Co-Pilot ElevenLabs 	<ul style="list-style-type: none"> Teacher modelling Traditional resources Paper-based Students working in pairs (see Harmer 1991) 	<ul style="list-style-type: none"> Student self-selected texts 	<ul style="list-style-type: none"> Teacher support

Data collection and analysis

MS Teams was our researcher platform. Figures 8 and 9 illustrate the range of qualitative and quantitative data we collected as well as self-collected participant data in Weeks 12, 15 and 17. During our 15-week intervention, we maximized the AR process by promptly creating and implementing pronunciation lessons and participant surveys. We reflected on and responded to our observations of participants, working with our new mixed methodology. We used MS Forms for surveys and MS Teams to record final interviews to collect both quantitative and

qualitative data and code written or spoken text responses by categories and subcategories, including outliers. We converted raw counts of responses per category to percentages to account for survey response fluctuation rates to gain a more accurate comparison across weeks, and collected text responses from video-recorded interview transcripts.

Types of data for qualitative and quantitative analysis				
Photos	Student participant individual recordings	Surveys	Video recordings	Researcher reflection journal

Figure 8: Our action research data collection types

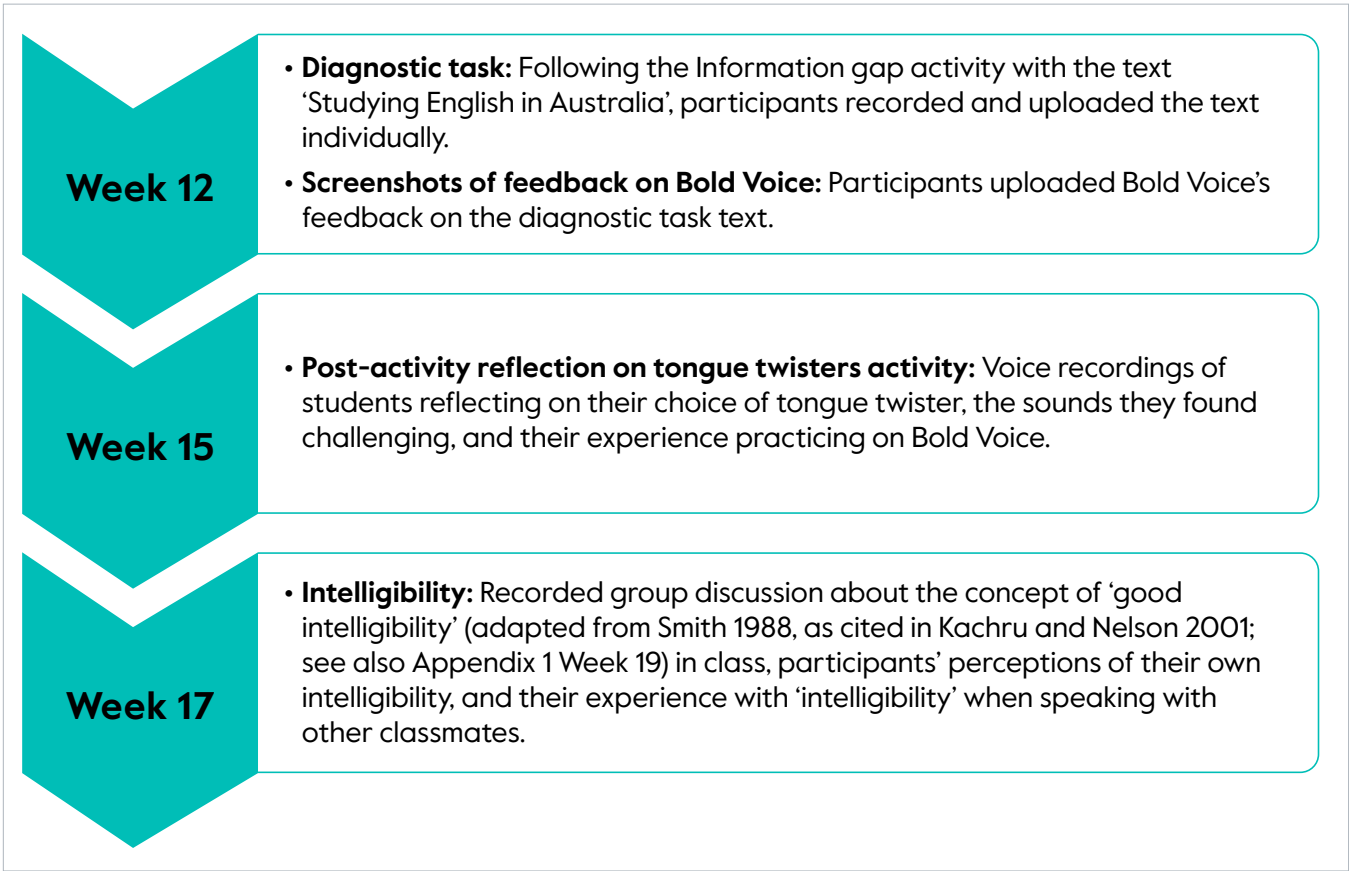


Figure 9: Data collected on participants’ private channels

Our researcher journal was valuable for reflecting and planning upcoming steps, crosschecking, or triangulating evidence. This ongoing written record of our observations, emergent ideas and literature review insights helped us see early in Cycle 1 (Week 3) that we had deviated from the genAI formative assessment focus by introducing traditional pronunciation activities created together with ChatGPT and Co-Pilot. Fortuitously, this oversight led to our mixed-methodology technique with genAI formative feedback being born. We tracked this repeated procedure via participants' survey feedback across the remainder of the project.

Findings

Our findings produced interesting insights regarding participants' perceptions of pronunciation when using genAI apps that may be relevant to the ELT community. The themes of our findings are illustrated in Figure 10.

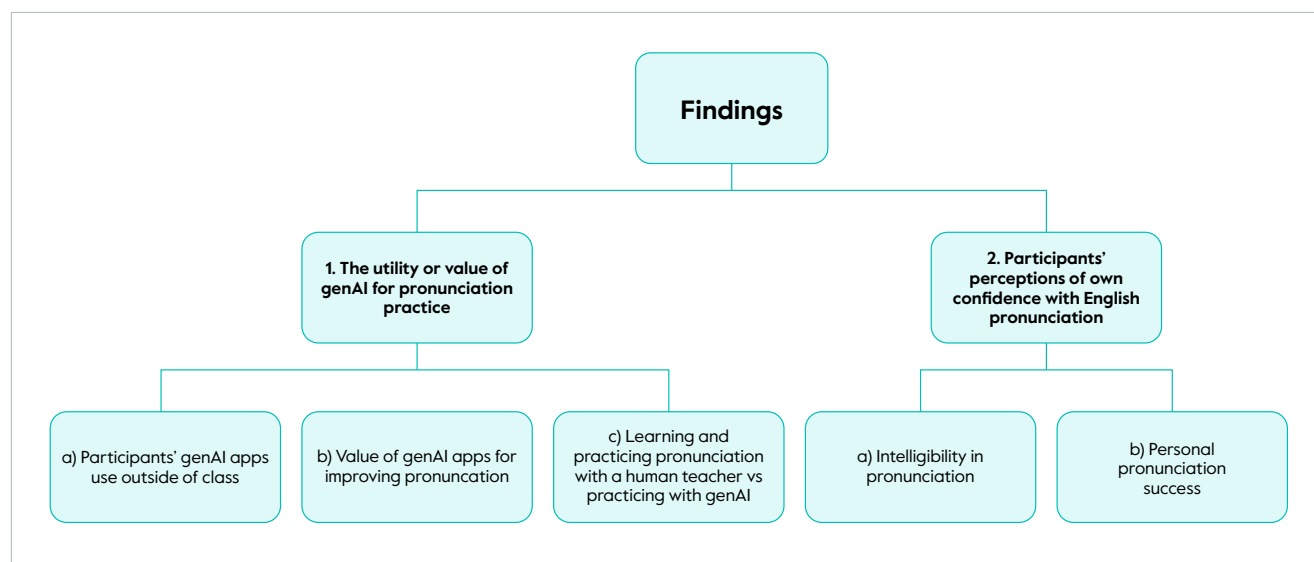


Figure 10: Key areas of action research findings with sub-categories

1. The utility or value of genAI apps for pronunciation practice

1.a. Participants' genAI apps use outside of class

Our students used Bold Voice and ELSA Speak consistently once or twice per week outside of class, with increased use in Week 21 (Figure 11) when a speech contest was introduced (Appendix 1). Although we did not require practice outside class, students chose to use the apps in personal time, and highlighted portability of learning and the potential to identify specific pronunciation issues as benefits the apps provided:

'The AI App, can give you the direct feedback wherever you are.'

'...I can try to make some differences between R and L sounds, so yes, AI applications can benefit me.'

However, students mentioned limitations, such as potentially inaccurate feedback, and restricted character numbers and recording time for text input.

'AI is advanced than before but it still need to improve, and I found that it is not always give feedback accurately.'



Figure 11: Frequency of genAI app use outside of class (Bold Voice and Elsa Speak)

1b. Value of genAI apps for improving pronunciation

Overall, participants perceived the non-subscription versions of the genAI apps as useful or valuable in helping address their pronunciation challenges equitably with no cost to the ELC or participants. Most students rated the apps as moderately effective or above in assisting them with their pronunciation difficulties (Figure 12). In this area, participants underscored the apps' effectiveness in adapting to their specific first language challenges:

‘...it (genAI) can provide a personalised approach to improve because everyone had their different difficulties or obstacles, so I think it’s ... a good way to use the application. So for me, I think ... this ... approach can really help ... students to improve.’

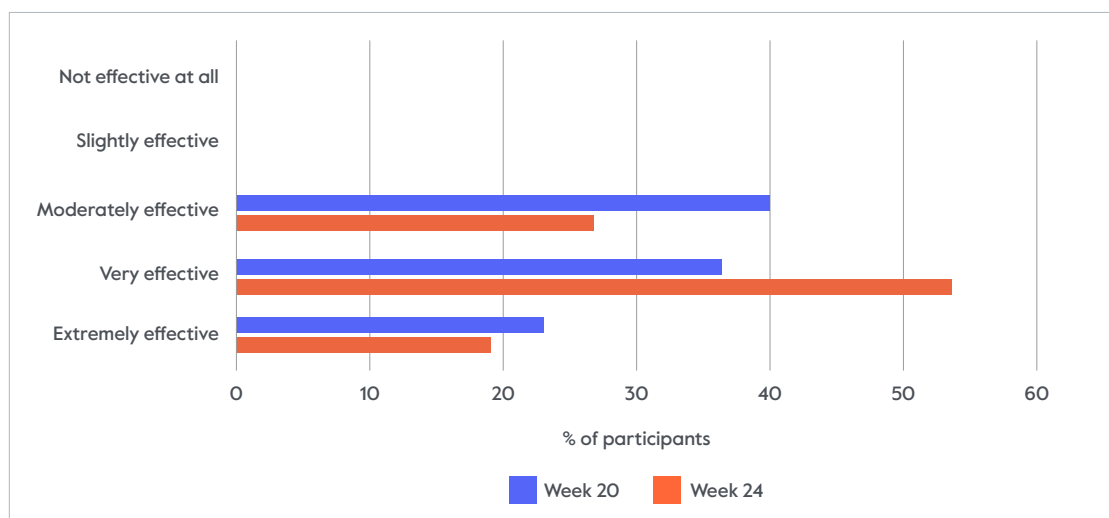


Figure 12: (End of cycle survey) – perceived effectiveness of genAI apps in helping with pronunciation improvement

1c. Learning and practicing pronunciation with a human teacher vs practicing with genAI

Our students identified strengths and weaknesses of learning with genAI or with a teacher alone and considered the genAI apps to be time-effective and flexible to adapting to specific learning needs. However, students found the apps’ feedback overwhelming at times:

‘...whenever your want to practice, you can use the application...’

‘AI based App ... provide too much, I cannot...pick the key.’

Alternatively, teachers were described as being able to offer targeted correction of specific issues, such as mouth movement when producing sounds, and provide encouragement to continue practicing. Student data also indicated teachers are costly and can be outnumbered in terms of student-teacher ratio:

‘... if you want to practise, maybe you have to pay the money for a teacher...’

There are a lot of students... we have to ... waiting for practice pronunciation one by one ... because we have only one teacher.’

Participants felt combining AI with the teacher provided the ideal learning balance where AI was useful for basic practice with the teacher giving more detailed feedback:

‘So for conclusion, it’s better AI supports the teacher, not teacher support the AI.’

Data in Figure 13 strongly reinforces the value of human interaction in our ELT learning and teaching context.

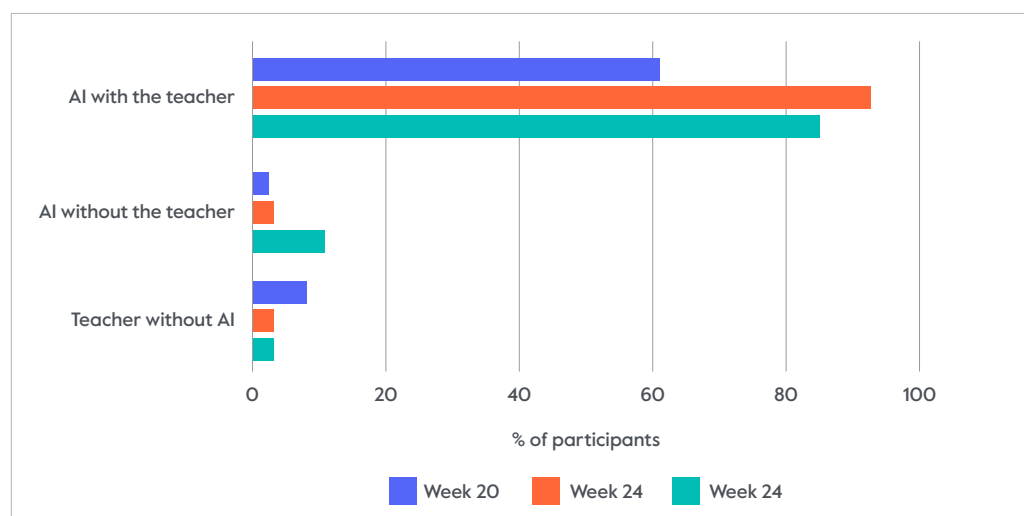


Figure 13: Fortnightly survey: Pronunciation work preferences with genAI and/or human teacher

2. Participants' perceptions of own confidence with English pronunciation

2a. Intelligibility in pronunciation

Our students' confidence levels in their pronunciation and ability to understand other students' pronunciation in the classroom mostly increased over time, except for generally constant results in the intelligibility survey. By Week 24, nearly 80% of participants reported feeling more confident in their pronunciation, and a similar proportion expressed satisfaction at a moderate level or higher with their listening skills (Figure 14).



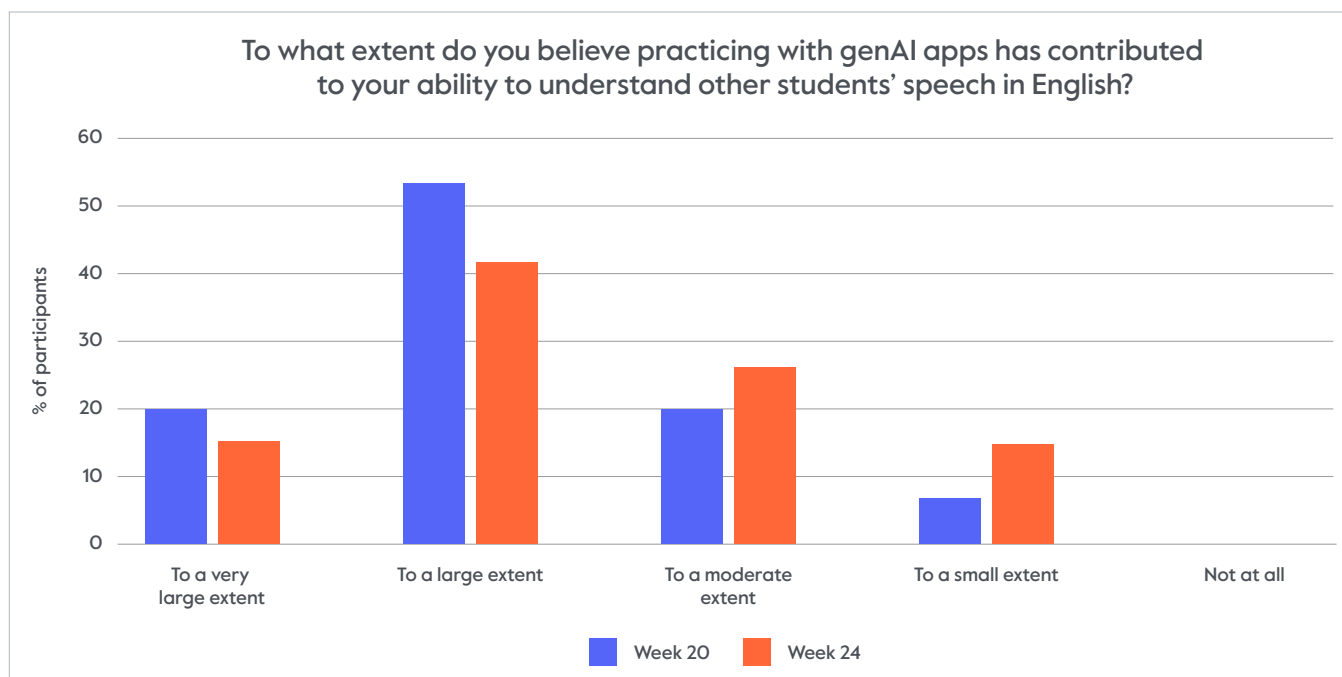


Figure 14: End of cycle survey: Participant change in confidence levels in their own pronunciation and in their ability to understand other student's speech

Students' comments during their final interviews further supported this increase in confidence and did not seem to be reliant on sounding like a 'native speaker.' The data seems to concur that this rise was not because they wished to speak like a 'native speaker' and many participants noted improvements in their pronunciation and when listening to others.

'... I think, it's the project improving a lot of my awareness and my confidence and my fluency...'

'My perception of pronunciation has changed [...] particularly understanding other students' pronunciation...'

2b. Perceptions of personal pronunciation success

We wondered if the genAI apps increased pressure to mimic a native speaker after we incorporated the 'native speaker' debate and the importance of students' own accents (Hamdan and Sowton 2024a; Jenkins 2024, as cited in Hamdan and Sowton 2024a; adapted from Brown 2012) into Cycle 2. Although some students recognized that the apps appeared to promote a specific accent model, they generally did not seem pressured to mimic it as pointed out in final interview data:

'...The AI application didn't increase the pressure for myself because I don't need to sound like a native speaker, because I am myself.'

Our end-of-cycle surveys indicated some students' perceptions changed positively, with the majority reporting being 'somewhat satisfied' with their progress by the end of the AR (Figure 15), but some students' responses indicated a degree of scepticism regarding their progress:

‘...I’m not sure I will improve my pronunciation or not, because I maybe familiar with your pronunciation. Sometimes I go outside and I still don’t understand what the local people say.’

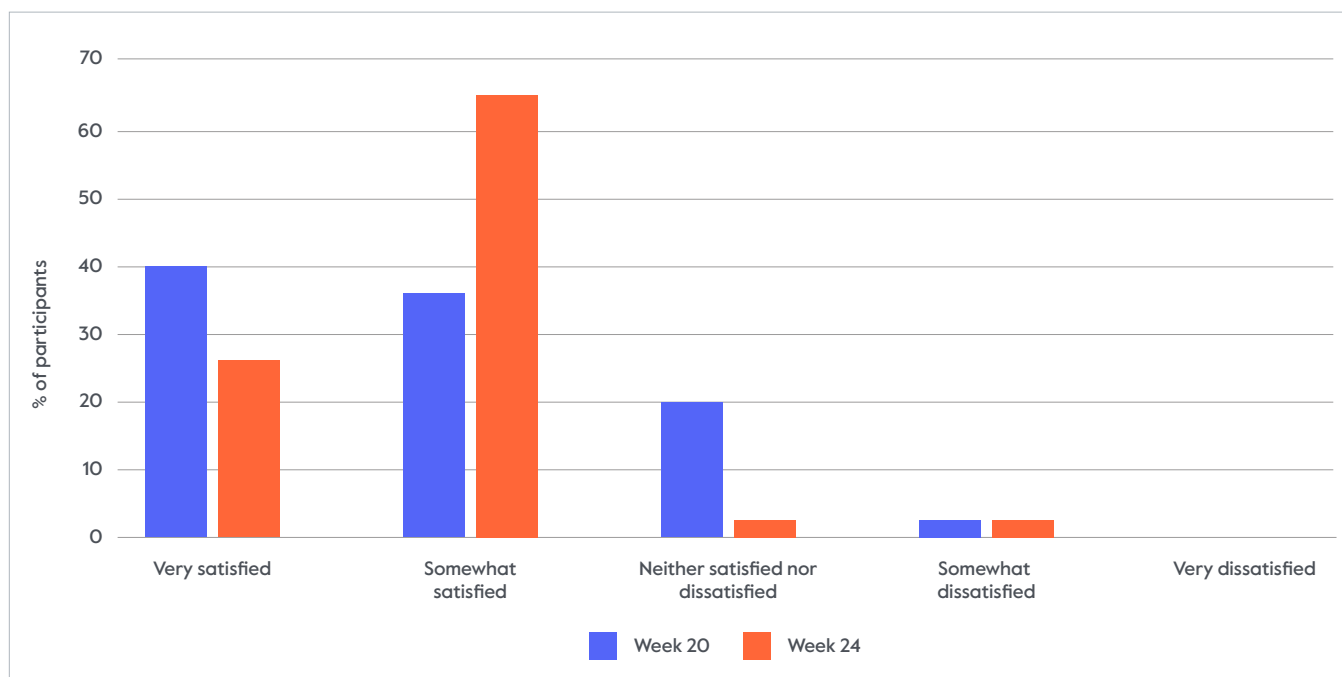


Figure 15: End of cycle survey – perceptions of pronunciation progress

This difficulty in self-identifying progress might have been related to our projects’ short duration and the considerable work involved in improving pronunciation habits. Encouragingly, the genAI apps emerged as potential tools for long-term improvement, even after students finish the course:

‘Maybe entirely needs lifelong learning, so I think we still need to improve after the program.’

‘I think by this program I know how to practise and I will practise in the future.’

Final interview evidence concurred with the quantitative survey data, with the interesting outlier finding of critical thinking being related to app use in class linked to how apps offer only one ‘correct’ accent. This finding supported the pathway course curriculum we were teaching and also shed light on our rationale of looking at pronunciation app equitability in ELT.

‘The app provides ... maybe an American accent or some right only one accent, right? We cannot change the accent.’

Discussion

Our extension of Brown's (2012) pre-genAI research findings suggests our mixed methodology intervention was largely positive. By blending genAI-assisted lesson creation with traditional pronunciation learning methods and prompt feedback with the free genAI formative assessment features of Bold Voice and ELSA Speak, we offered individualized L1-affected pronunciation support to ELC participants while maintaining responsible duty-of-care to our students as teachers and assessors. Our triangulated evidence indicated that, by connecting strategic learning-oriented assessment to timely, formative feedback with genAI apps, our intervention helped move our ELC participants, and us as teachers, towards meeting the traditional pronunciation challenges of our Centre's direct-entry course. Our findings and research design, using tools like ElevenLabs, to an extent fulfilled Field's (2019:214) prediction that technology would afford listening pedagogy greater choice of tailored resources and student autonomy.

Conclusion

We found that genAI app use affected students' pronunciation perceptions positively overall and provided valuable practice which mostly, despite some limitations, increased confidence with English pronunciation including intelligibility. Findings revealed that our interventions fostered student agency and learning *how to practice*, satisfying a goal of our ELC. Our participants' high regard for the combined genAI-‘human’ teacher approach provided some reassurance to us that we are needed, valued educators.

Ideas cited earlier such as ‘intelligibility,’ the ‘native speaker,’ genAI and ‘prestige English power bias,’ and the ‘commercialisation of AI tools’ were interesting to pursue with our students. In addition, we started to address the receptive skills gap highlighted above which is a step forward in this area for our Centre.

Reflection

Our AR provided an authentic glimpse into how genAI app use affected our participants' perceptions of pronunciation while maintaining the view ‘...of the student as a “whole person”... [in] ... language teaching ... [and that it] ... is not just about teaching language ... [but] ... also about helping students ... develop themselves as people’ (Harmer 1991: 35). Our participants' responses showed strong evidence supporting our human-centred learning approach and these need to be kept in mind as new genAI technologies continue to emerge and challenge ELT education.

Working as co-researchers in this ELICOS AR program has been an incredible professional development opportunity. We took a leap of faith in researching pronunciation-related genAI tools in unexplored ELT territory that may benefit others. We also experienced AR challenges and limitations including unforeseen

gaps in data collection and challenges in survey design. These included fluctuating response rates, failure to realize that preparation for the Cycle 3 speech contest may have impacted the frequency of students' genAI app practice, and our oversight in not checking our participants' understanding of 'perceptions.'

Our AR has led us to believe that the most important outcome is that our mixed-method approach to filling a gap in pronunciation pedagogy maintains core ELT practices with the expanded benefits of genAI. Our mixed-method approach could be a foundational strategy that other teachers could apply to current and future genAI tools.

We hope our approach can be further adopted at our Centre, and in the wider ELICOS community, and that teachers adapt it for their own students' pronunciation development needs. If so, our research would make a wider contribution to closing the enduring gaps and perhaps inspire the development of more equitable genAI apps to help better support pronunciation pedagogy.

References

- Brinton, D M (2019) Pronunciation instruction, in Burns, A and Richards, J C (Eds) *The Cambridge guide to pedagogy and practice in second language teaching*, Cambridge: Cambridge University Press & Assessment, 246–257.
- Brown, B (2012) Raising student awareness of pronunciation and exploring out-of-class approaches to pronunciation practice, *Research Notes* 48, 18–24.
- Burns, A (2010) *Doing action research in English language teaching*, New York: Routledge.
- Edmett, A, Ichaporia, N, Crompton, H and Crichton, R (2024) *Artificial intelligence and English language teaching: Preparing for the future* (Second edition), London: British Council, available online: <https://doi.org/10.57884/78EA-3C69>
- Field, J (2019) Listening instruction, in Burns, A and Richards, J C (Eds) *The Cambridge guide to pedagogy and practice in second language teaching*, Cambridge: Cambridge University Press & Assessment, 207–218.
- Hamdan, W and Sowton, C (2024a) Do I need to sound like a ‘native speaker’?, *Teaching English: A podcast series for teachers of English*, Series 3 Episode 3, accessed 4 April 2024, https://www.teachingenglish.org.uk/professional-development/podcast/teaching-english/teachingenglish-podcast-do-i-need-sound-native?utm_campaign=ce-all-teachingenglish-globalnewsletters&utm_medium=email&_hsmi=85942265&utm_content=85942265&utm_source=hs_email/
- Hamdan, W and Sowton, C (2024b) What is the impact of artificial intelligence on English language teaching?, *Teaching English: A podcast series for teachers of English*, Series 3 Episode 4, <https://www.teachingenglish.org.uk/professional-development/podcast/teaching-english/teachingenglish-podcast-what-impact-artificial>
- Harmer, J (1991) *The practice of English language teaching*, London/New York: Longman.
- Hutton, E (n.d.) *Pre-enrolment English Program: Language and culture*, PowerPoint presentation (adapted), The University of Adelaide.
- Hutton, E and Sparrow, H (2023) *Staff PD: Marking in the age of AI?*, PowerPoint presentation, The University of Adelaide.
- Kachru, C C and Nelson, C L (2001) World Englishes, in Burns, A and Coffin, C (Eds) *Analysing English in a global context: A reader*, Oxon: Routledge, 9–25.
- Nation, I S P and Newton, J (2009) *Teaching ESL/EFL listening and speaking*, New York: Routledge.
- Swan, M and Smith, B (Eds) (2001) *Learner English: A teacher’s guide to interference and other problems* (Second edition), Cambridge: Cambridge University Press.
- The University of Adelaide (2023) *Research Tuesdays: 3 Minute thesis*, <https://www.youtube.com/watch?v=UzByHIAQURO&t=1453s>
- The University of Queensland (2024) *Three minute thesis*, <https://threeminutethesis.uq.edu.au/>

Appendix 1: Lesson content per AR cycles

(adapted Kemmis and McTaggart 1988, as cited in Burns 2010:9)

Cycle 1 Bi-weekly lessons	Cycle 2 Bi-weekly lessons	Cycle 3 Bi-weekly speech contest practice
We acknowledge that Co-Pilot and ChatGPT were used in the lesson design process.		
Lesson 1: Pronunciation feature introduced Lesson 2: Additional practice Self-selected texts Bold Voice/Phonetic apps with teacher support	Lesson 1: Pronunciation feature introduced Lesson 2: Additional practice ElevenLabs accents incorporated for accent variety to address the limited AI technologies for listening ¹ ELSA Speak app introduced 'intelligibility,' ² 'native speaker,' ³ and English as a lingua franca ⁴ class activities	Inspiration: <ul style="list-style-type: none"> Japanese school English speech competitions 'Three minute thesis' concept^{5 6} Speech script: <ul style="list-style-type: none"> Challenging features for our participant cohort Created by ChatGPT-4o with researcher-crafted prompts
Week 12 Diagnostic Task Introduction to AR pronunciation apps	Week 16 Sentence Stress	Week 21 Speech contest introduction
Week 13 Scavenger Hunt: International Phonetic Alphabet	Week 17 Features of connected speech + listening	Week 22 Speech contest 'Native speaker' ⁷ discussion
Week 14 Minimal pairs	Week 18 Consonants and consonant clusters – Total Physical Response (TPR) + listening	Week 23 Speech contest
Week 15 Tongue Twister Gallery – academic words	Week 19 Sentence stress vs word stress 'intelligibility' ⁸ lesson	Week 24 Group interviews
Break Week	Week 20 Short and long vowels bingo	Week 25 End of AR

1 see Edmett et al (2024); Burns 2024

2 Smith (1988), as cited in Kachru & Nelson (2001:21)

3 see Jenkins (2024) as cited in Hamdan and Sowton (2024a); see also idea adapted from Brown (2012)

4 Hutton (n.d.)

5 The University of Queensland (2024)

6 The University of Adelaide (2023)

7 Jenkins (2024) as cited in Hamdan and Sowton (2024a); adapted from Brown (2012)

8 Smith (1988) as cited in Kachru and Nelson (2001); Brown (2012:18), Jenkins (2024) as cited in Hamdan and Sowton (2024a)

Links to AI tools used in this project:

- BoldVoice: boldvoice.com
- Cambridge University Press Assessment (2024), Pronunciation symbols, Cambridge dictionary: <https://dictionary.cambridge.org/help/phonetics.html>
- ElevenLabs: <https://elevenlabs.io/>
- ELSA Speak: <https://elsaspeak.com/en/ai/>
- International Phonetic Association, full IPA chart: <https://www.internationalphoneticassociation.org/content/full-ipa-chart>
- Microsoft Co-Pilot: <https://copilot.microsoft.com/>
- ChatGPT: chat.openai.com
- Phonetic: Apple App Store.
- Pronuncian: Apple App Store.

Appendix 2: Pronunciation lesson sample on sentence stress and prompts for specific language targets

Prompting process

Initial prompt (ChatGPT 3.5)

Create a 30-minute pronunciation activity on sentence stress for university pathways students in Australia from Vietnam, China, Japan, and Thailand. Design the handouts and activities for each stage of the lesson and provide an answer key. Make the activity interactive and engaging.

Follow-up prompts:

- Make the sentences slightly longer and about Australia.
- Include more structure words in the sentence (only stress nouns, adjectives, verbs and adverbs). Add more prepositions, articles and pronouns.
- Which challenging sounds are included in the above sentences?

Figure 1 shows an extract of ChatGPT's output in response to (c). These suggestions were cross-checked based on cohort demographics with traditional evidence (Swan and Smith 2001).

- /v/ and /w/: Learners might find it difficult to differentiate between the /v/ sound in 'visiting' and the /w/ sound in 'weekend.' For examples, Mandarin Chinese lacks the /v/ sound, so Chinese learners may have difficulty pronouncing words like 'visiting' and 'weekend.'
- /r/ and /l/: The distinction between /r/ and /l/ in words like 'explore' and 'culture' could be challenging for some learners, particularly those from East Asian countries. For examples, Japanese lacks a clear distinction between /r/ and /l/, making it difficult for Japanese learners to distinguish between words like 'explore' and 'culture.'
- /ð/ and /θ/: The voiced dental fricative /ð/ in 'the' and the voiceless dental fricative /θ/ in 'thing' may pose difficulty for learners whose native languages don't have these sounds. Thai, Vietnamese and Chinese

Appendix 2, Figure 1: ChatGPT 3.5 output on challenging sounds targeted in the lesson activities

Lesson plan

Objective

- Students will improve their ability to use sentence stress effectively in English to enhance clarity and naturalness in communication.
- Students will practice challenging consonant sounds and consonant clusters to address L1 interference.

Materials needed

- Handouts with example sentences
- Recordings of the sentences (created on ElevenLabs)

Warm-up

Show the students a sentence that includes content words and one that only includes structure words. Ask them to guess the meaning of each sentence. (Appendix 2, Figure 2, left)

Discuss the definition of content and function structure words in the sentence, and the impact of such elements on stress. (Appendix 2, Figure 2, right)

Guess the meaning of the sentences

1. _____ the _____ of _____ is an _____.

2. Driving _____ stunning coastline _____ Great Ocean Road offers breathtaking views _____ majestic cliffs.

Were you able to guess the missing words? Why?

Content words vs function words

Meaning

Structure

Appendix 2, Figure 2: Warm-up activity to introduce the concept of content and structure words in connection to sentence stress

Introduction

Distribute handouts with examples of sentences highlighting different stress patterns. (Appendix 2, Figure 3)

Students discuss which words should be stressed. They listen and check their guesses.

Sentence stress

1. With a partner, underline the stressed syllables.
2. Listen and check your answers

visiting the iconic Sydney Opera House during the weekend is something I love doing.

She is planning a trip to explore the Great Barrier Reef next week.

Learning about Indigenous culture is recommended for a deeper understanding of Australia.

We enjoyed the exploration of the vast outback last week.

If you try delicious Australian cuisine, you can enhance your experience while visiting Australia.

Appendix 2, Figure 3: Worksheet for listening identification of stressed syllables

Practice in small groups:

Students read the sentences to their group. Other members of the group comment on their classmate's reading.

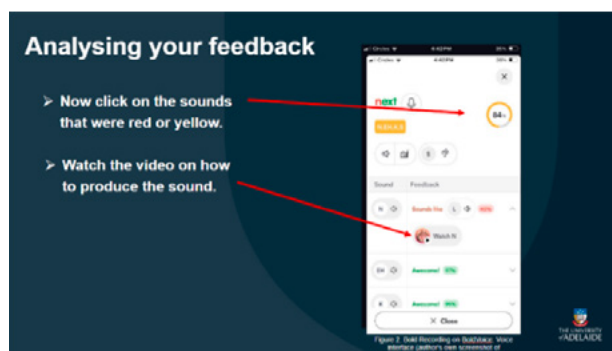
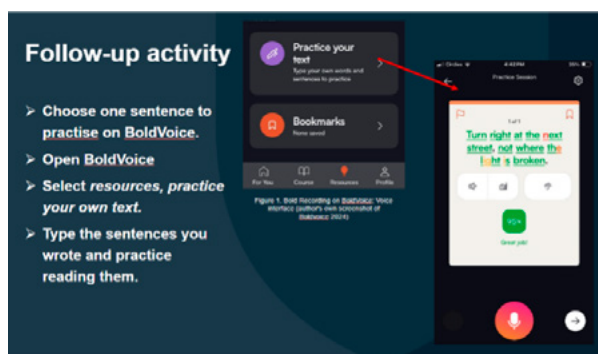
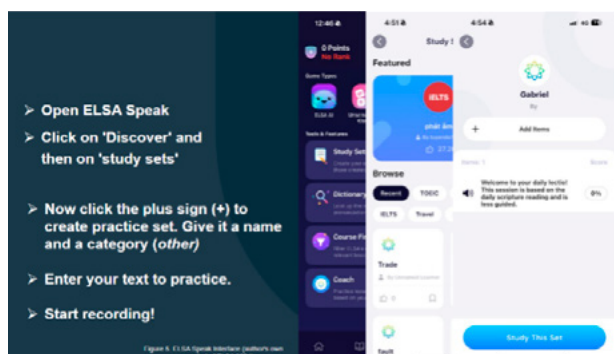
Reflection:

Facilitate a brief discussion on the challenges and successes students experienced during the activity.

Ask students to reflect on how they can incorporate sentence stress practice into their coursework and in their general daily English communication.

Follow-up: Practice using pronunciation apps

At this stage of the lesson, students choose a sentence to practice using the AI apps for feedback. See below for the slides used during our project. (Appendix 2, Figure 4)



Appendix 2, Figure 4: PowerPoint presentation slides used when introducing the class section focused on practice with genAI Apps

The Learning Continuum: Utilising ChatGPT to support academic discussion skills

Diala Ibrahim and Brad McClymont, UNSW College, Sydney

Overview

The emergence of AI and its disruption to education has led to a paradigm shift in language teaching and assessing in the ELICOS context, presenting both challenges and opportunities. With the recognition of 'AI literacy' as an essential skill in education, traditional teaching methods are being re-evaluated, and English language teaching professionals are actively investigating innovative approaches for integrating AI into teaching and learning. An intriguing prospect is exploring the use of AI to support students in the development of speaking skills, particularly academic discussions. AI offers interactive communication tools that are accessible and adaptive. This potential is adding a new dimension to the way language is taught and learned.

Context

UNSW College offers pathway programs for international students to various faculties of the University of New South Wales (UNSW) in Sydney. The majority of the participants in our action research (AR) were transitioning into a Master's Degree, and most of them were Engineering students (see Table 1).

Table 1: Faculties and pathway degrees

Participants' UNSW faculty pathways			Participants' education pathway level		
Engineering	45%	22	Bachelor	20%	10
Arts, Design & Architecture	27%	13	Master	76%	37
Science	14%	7	PhD	4%	2
Business School	10%	5			
Medicine & Health	4%	2			

Our AR was conducted with two cohorts of UNSW College Academic English students simultaneously. The University English Entry Course (UEEC) direct entry pathway students (32 students) were taught by Brad, and the Communication and Academic Literacy 1 (CAL 1) Diploma students (17) were taught by Diala (see Figure 1).

Brad's students (UEEC 20 and UEEC 15) were in the initial 10 weeks of 20- and 15-week pathway courses to postgraduate studies at UNSW. These courses are formative and focus on developing Academic English skills.

The CAL 1 course is the first in a series of three constituting the language component of a UNSW College undergraduate Diploma which transitions into a second year UNSW Bachelor's Degree. This 12-week course is heavily focused on building communication skills, academic literacy, and learner autonomy.

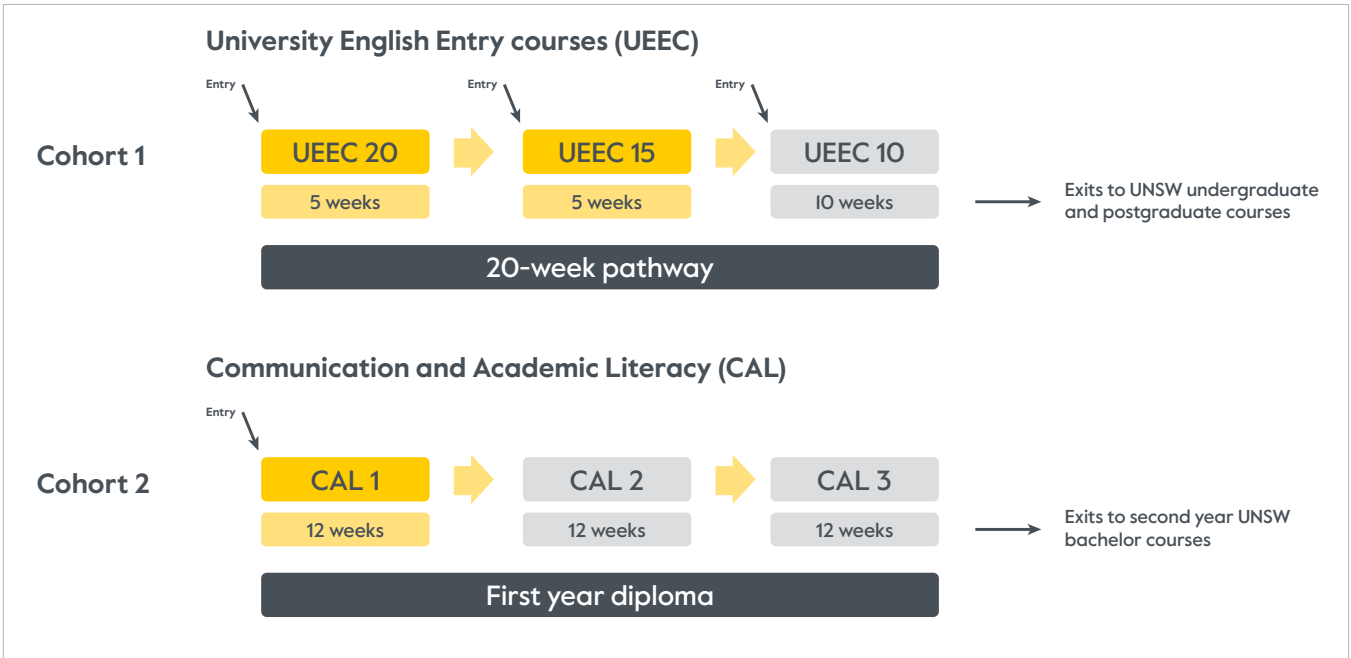


Figure 1: Our context: UNSW College course pathways

Forty-nine students participated in this action research (25 females, 24 males). Nationalities included one Iranian, 46 Chinese, and two South Korean students. Participant ages ranged from 18 to 29. The majority (65%) were new arrivals to Australia (less than one month) with English proficiency scores of IELTS 5.0–6.5 (B1–C1 of the CEFR).

Rationale

Since the COVID-19 pandemic, feedback from some UNSW faculties has indicated that pathway students more noticeably lack the confidence and skills needed to participate in university tutorial discussions. In addition to language proficiency barriers, UNSW College teachers have noticed other difficulties in adapting to the demands of academic discourse in a university context, particularly in contributing ideas, elaborating, or demonstrating critical thinking skills. With this feedback at the core of our concerns, our research investigated effective ways to leverage the use of AI to support the development of academic discussion skills, specifically focusing on ChatGPT, a generative AI-powered cloud-based Large Language Model (LLM) designed to process natural language and simulate human-like communication (ChatGPT 2023).

Our research aimed to address the following question:

How can ChatGPT be used to support the development of academic discussion skills through planning, practice, feedback and reflection?

Methodology

Our research was conducted over nine weeks (see Appendix 1), and it explored three ChatGPT functions or roles that we created: 'Brainstorming Partner', 'Discussion Coach' and 'Feedback Tutor' (see Figure 2). We evaluated the effectiveness of these three roles by examining how each worked independently and how all three roles worked in synergy to create a complete learning process. This was done by investigating learner ability to generate ideas on various academic topics ('Ideation'), and measuring three key aspects of speaking output skills: 'Interaction', 'Elaboration' and 'Investigation' (see Figure 3 and Appendix 2). These discussion skills were central to our research as we assessed student performance in each skill across all stages of the learning process.

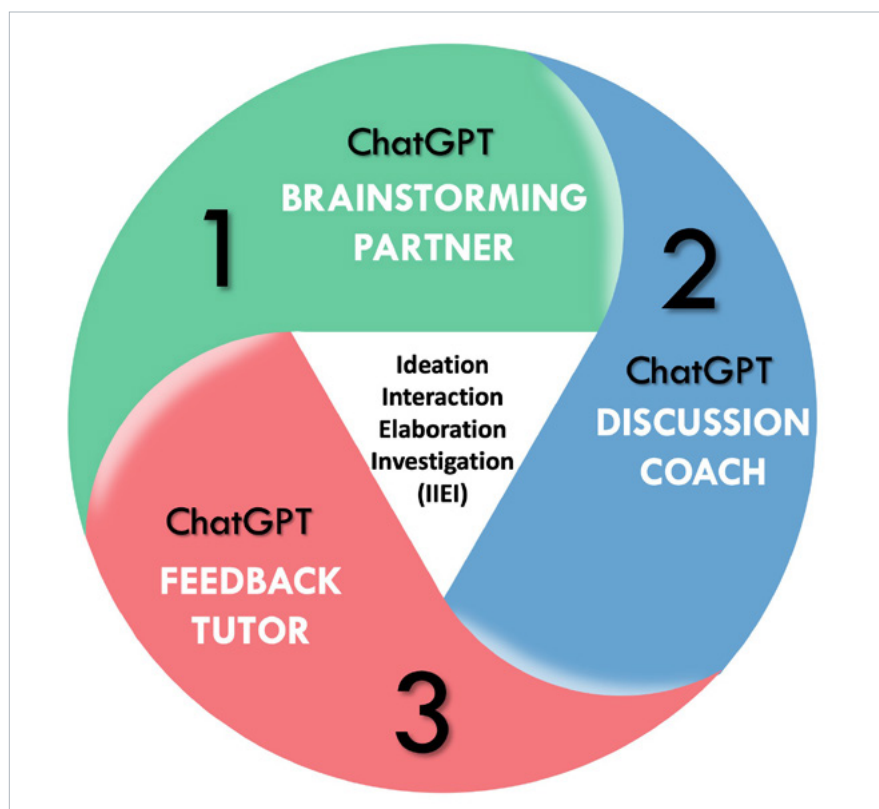


Figure 2: The three ChatGPT roles for enhancing discussion skills



Figure 3: Key skills for successful academic discussions

1. ChatGPT Brainstorming Partner: Preparing for academic discussions

The first stage of our research began by introducing a scaffolded brainstorming process, which is explained in Figure 4 below.

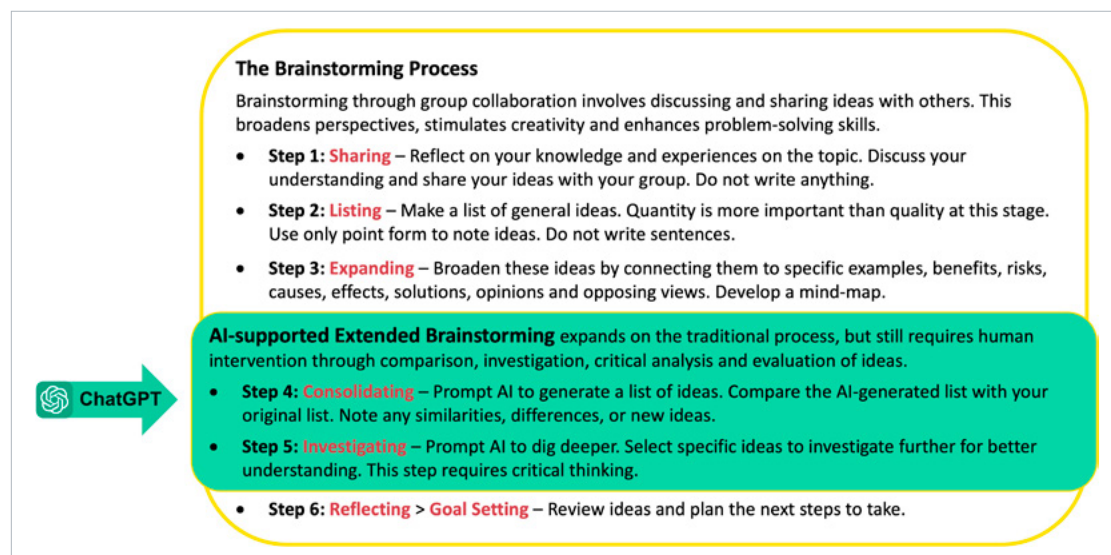


Figure 4: Enhancing brainstorming with AI: Integrating traditional and AI-assisted techniques

The ChatGPT Brainstorming Partner provided the initial ideas necessary for students to have meaningful discussions. To facilitate critical human oversight in the process, encourage critical thinking and foster creativity, students were guided through Steps 1–3 where they shared knowledge on a given topic, listed initial ideas, expanded on these ideas, and created mind-maps to indicate relationships between ideas. Students were then introduced to *prompt engineering* with a focus on task and context only (see Figure 5); in Step 4 they accessed ChatGPT (free version) and prompted for a list of ideas related to their given issue. They were required to read (simplifying if necessary) and compare the AI-generated list with ideas in their mind-maps to highlight any similarities, differences or new ideas that had emerged.

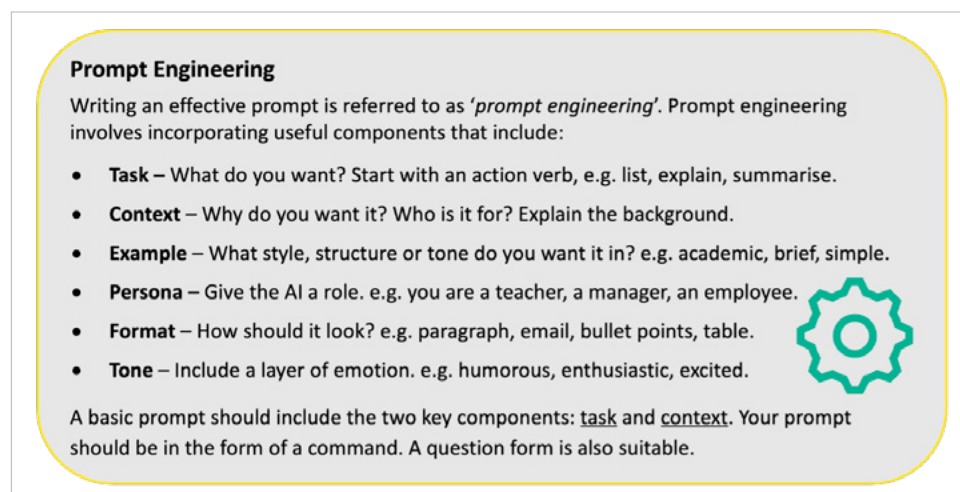


Figure 5: Prompt engineering components

In Step 5, students were asked to choose a specific aspect of the topic that intrigued them the most and to prompt for a deeper understanding. The purpose of Step 5 was to support students with developing a richer and ultimately more meaningful discussion with groups. This step also allowed them to build greater confidence going into the discussion (see Appendix 3). In the final step, students reflected on their experience and performance, and shared samples of their brainstorming outcomes (see Figure 6 and Appendix 3). Students were then expected to incorporate ideas from this process into a 10–15-minute group discussion using notes they had prepared, but without reading from a screen.



Figure 6: ChatGPT Brainstorming Partner – Consolidation and Investigation (Steps 4 and 5)

Link 1: <https://www.englishaustralia.com.au/documents/item/2840>

2. ChatGPT Discussion Coach: Practicing academic discussions

In the second stage of our research, we introduced students to the voice-interaction feature on the ChatGPT mobile application.

Through an extended iterative process of testing and refining, we crafted a set of prompts (see Appendix 4) to create a personal ‘Discussion Coach.’ This coach helped students practice and improve their discussion skills, specifically focusing on Interaction, Elaboration, and Investigation skills, and the language to facilitate them.

1. **Interaction Coach** offered practice and advice in different interactional exchanges such as remarking, commenting, backchannelling, agreeing, paraphrasing, and summarising.
2. **Elaboration Coach** trained students to develop and support their ideas with further explanations, reasons, examples, or evidence.
3. **Investigation Coach** encouraged students to critically explore ideas by analysing and evaluating views, challenging opinions, seeking clarification, and questioning further for a deeper, more engaging discussion.

While observing student communication with the ChatGPT Discussion Coach we noticed that some were experiencing difficulties such as frequent interruptions and vocabulary complexity. Therefore, we developed a user guide to provide scaffolded support. The guide offered instructions on what students should say or could do to improve their ChatGPT communication experience (see Appendix 4).

Students practiced one-on-one (see Figure 7) and in groups (see Figure 8) with the ChatGPT Discussion Coach in class after brainstorming and before participating in group discussions. This coaching practice provided students with a valuable opportunity to refine their ideas, practice the discussion skills, and build vocabulary and enhance language useful for the discussion. Additionally, students were required to complete 30 minutes of ChatGPT discussion coaching as weekly homework. This extra practice served as a 'take-home-coach', giving students the opportunity for continued practice at their convenience.



Figure 7: ChatGPT Discussion Coach – Individual Elaboration skills practice
Link 2: https://youtu.be/JtyrRPhv_Lc



Figure 8: ChatGPT Discussion Coach – Group Elaboration skills practice
Link 3: https://youtu.be/qx1k6g2Qf_Y

3. ChatGPT Feedback Tutor: Reflecting on academic discussion skills

The third and final step in the learning process closed the loop by providing feedback and opportunities for reflection. Using Sonix AI-powered automated transcription software, we transcribed each discussion and trialled a dual method of prompting ChatGPT for feedback:

(a) Teacher-facilitated customised feedback: This process involved us using ChatGPT-4o (paid version) to generate a feedback report for each discussion group. Discussion transcripts were uploaded to ChatGPT, which was prompted to generate a feedback report on Ideation, Interaction, Elaboration, and Investigation (IIEI) skills. Students reviewed their discussions via a Sonix voice and transcript link, reflected on the feedback report in their groups, and shared an action plan for future discussions (see Appendix 5). This method proved to be quick and effective for capturing valuable snapshots of feedback that students could work on. However, some students considered this approach generic, and requested more customised and detailed feedback. Therefore, we added the following approach (b), which gave learners autonomy and control over the feedback process.

(b) Learner-initiated personalised feedback: This process involved students seeking their own personalised feedback in a consultation session with ChatGPT. Students identified strengths and weaknesses in their discussion transcripts and prompted ChatGPT for feedback on IIEI skills. They copy-pasted their group discussion transcripts into a new ChatGPT chat and prompted for advice using either prompts provided by us or their own personalised prompts based on advice in the feedback report (a). Students then improved on their individual discussion transcripts, reflected in their discussion groups, and shared action plans to support improvement of specific aspects in future discussions (see Appendix 5). This approach was a more interactive and engaging learning experience as students were actively involved in the feedback process.

Link 4: <https://www.englishaustralia.com.au/documents/item/2841>

The AI-Supported Learning Continuum

Our research and intervention processes are captured in our Learning Continuum diagram. This diagram presents a scaffolded approach for integrating ChatGPT across the discussion development cycle that is repeated weekly over the duration of the course (illustrated in Figure 9). Students start by preparing for their academic discussions through AI-supported brainstorming (Step 1: Preparation); they then practice speaking with the Discussion Coach (Step 2: Practice) before they hold group discussions using notes which they had prepared in Steps 1 and 2 (Step 3: Production). In Step 4 (Reflection) students reflect on the discussion feedback report provided by the Feedback Tutor, and in Step 5 (Action) they launch their own feedback consultation session with the Feedback Tutor. Finally, students are given homework to practice speaking again with the Discussion Coach (Step 6: Revision).

THE AI-SUPPORTED LEARNING CONTINUUM:

A scaffolded approach for integrating ChatGPT across the discussion development cycle over the duration of the learning period

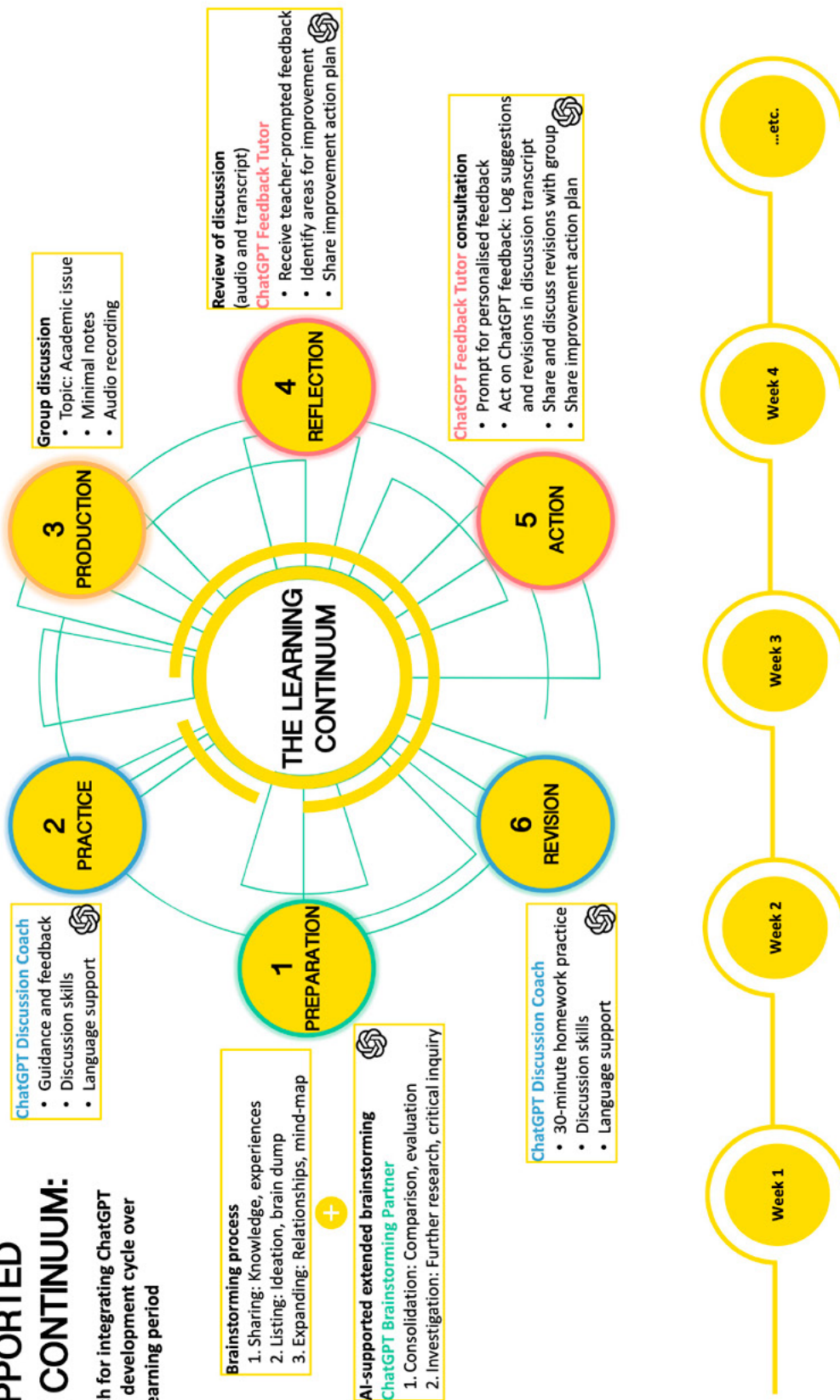


Figure 9: The AI-Supported Learning Continuum: A scaffolded approach for integrating ChatGPT across the discussion development cycle over the duration of the learning period

Data collection

Over the nine-week research period (see Figure 10 and Appendix 1) lessons were created in Moodle (LMS), where data was also stored. The following data was collected through Moodle:

- Pre- and post-intervention surveys to gauge student perception and evaluation of speaking and IIEI discussion skills.
- AI awareness survey (see Appendix 6).
- Pre- and post-intervention discussion assessments to identify instances where students present ideas, interact, elaborate, and investigate in academic discussions before and after intervention lessons.
- Weekly student reflections on discussion performance and ChatGPT tools.
- Weekly discussion transcripts for feedback and reflection.
- Brainstorming notes and reflections.
- ChatGPT Discussion Coach links to monitor practice and improve prompts.
- ChatGPT Feedback Tutor student prompts and revisions made to their discussion transcript.

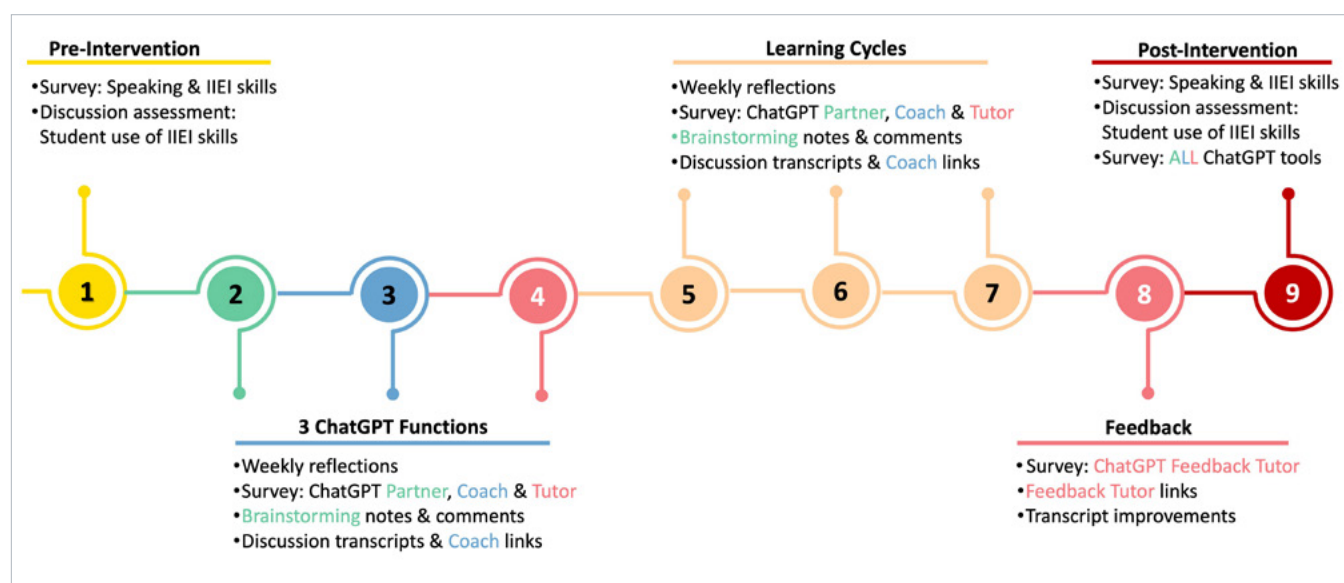


Figure 10: Data collection

Findings

Instances of Ideation, Interaction, Elaboration, and Investigation in group discussions for both the baseline and post-intervention assessments were identified in discussion transcripts and manually tallied. An analysis of this data revealed improvements in the key discussion skills, as shown in Figure 11.

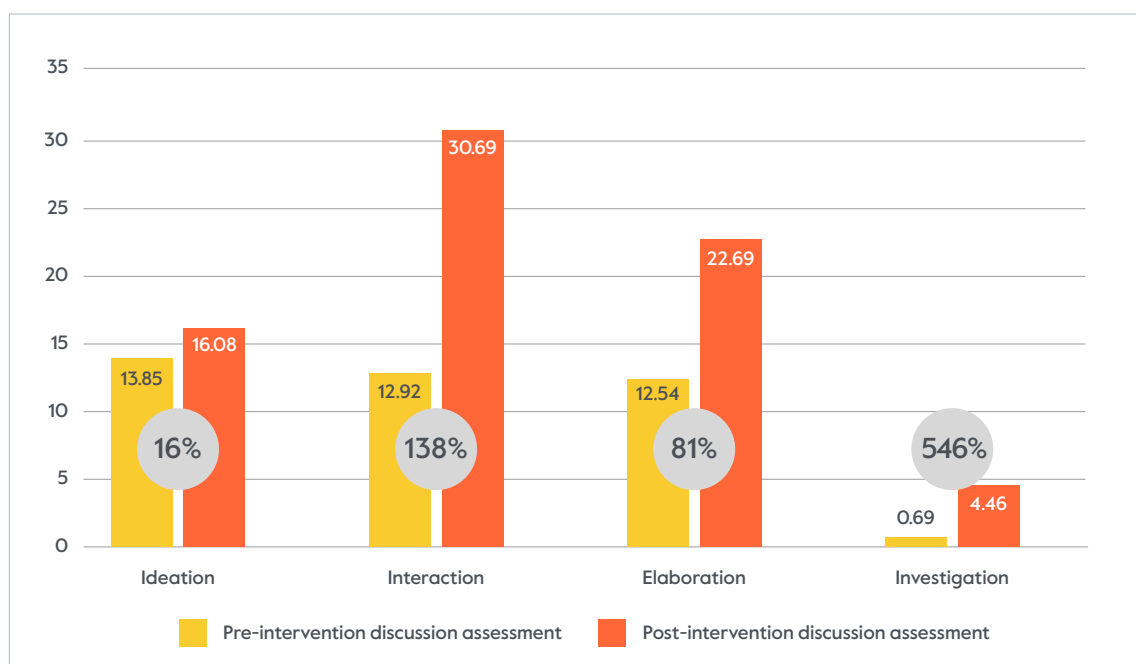


Figure 11: Coded instances of discussion skills in pre- and post-intervention discussion assessments

The data shows 138% increase in Interaction. During the pre-intervention baseline assessment, we observed some groups showing no interaction at all. Students would take turns delivering monologues before indicating that they had finished by saying 'next' or signalling turns through body language. While regular practice with the ChatGPT Discussion Coach and insights from the Feedback Tutor may have contributed to this result, it is important to consider other factors such as familiarity and confidence among students over the nine-week period.

The data also reveals a great increase in Elaboration (81%). This may suggest that the brainstorming support played a role in developing understanding of ideas with additional reasons, examples and evidence. It was evident in early discussions that students were listing ideas without much depth. However, in the final discussion assessment students spent more time elaborating a single idea before moving on to the next, and this is mainly why increase in ideas is minimal (16%).

While Investigation remains the skill least observed in discussions, it increased dramatically. Given discussions mainly focused on problem-solution topics, not a great deal of argument, challenging and probing was required.

This overall increase in instances of discussion skills was mirrored in students' weekly reflections (see Figure 12). Self-evaluation surveys indicated an upward trend in their perception of their own performance and skills during each iteration of the learning cycle. In Week 1, 49% of students considered their discussion skills weak; this dropped to 6% by Week 7. Likewise, in Week 1, only 11% of students rated their discussion skills as very good. This rose to 41% by Week 7. It was also evident from our observations that students became more proactively engaged and interested during the discussions.

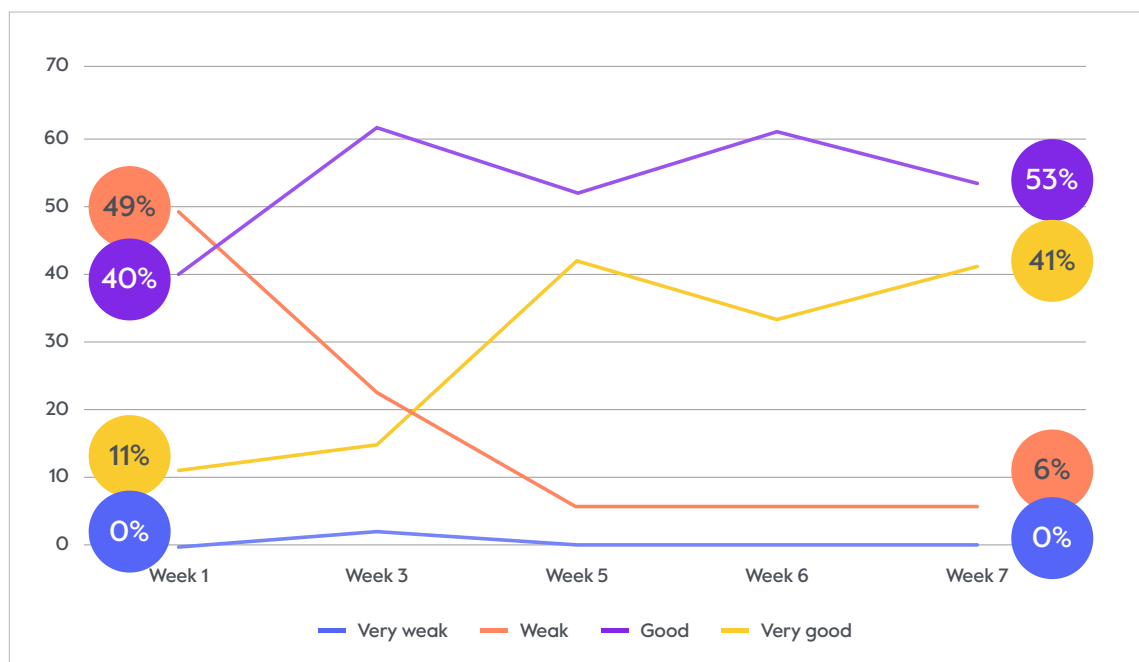


Figure 12: Weekly student discussion skills self-evaluation results

ChatGPT as brainstorming partner

This use of AI was the most well-received by students and was awarded a 4.25 (out of 5) star rating in the final survey. Student voices revealed that this manner of working with an AI partner allowed them to organise fragmented thoughts and was *'very effective when looking for blind spots on the topic.'* They valued seeing their original ideas mirrored in the AI-generated responses (*'Brainstorming Partner confirmed my ideas, which boosted my confidence'*), and noted that they were able to extract information that provides a more comprehensive understanding of key issues: *'ChatGPT as a brainstorming companion is very useful because I can find deeper insights into these ideas.'* We also observed students making notes of unfamiliar academic vocabulary and expressions taken from their interactions with ChatGPT. This bonus language support was also highly regarded by students: *'brainstorming with ChatGPT was awesome. She give me lots of new ideas and new word to use.'* However, students acknowledged the potential risks associated with using AI, including reduced ability in independent thinking, over-reliance and laziness: *'...I think relying on it will lead to lazy thinking and unwilling to think about problems.'* 28% expressed a preference to use their own ideas as they were more relevant to their experiences and personal viewpoints as well as being easier to articulate in their discussions: *'I think it can only be used as a support tool because in the final decision, I still use my own ideas. The ideas given by AI are not easy to explain.'* It was also reassuring to read in student reflections that while many valued the brainstorming support from AI, they still prioritised their own thoughts: *'I will retain a lot of ideas that I have come up with myself, because I think any problem should be solved by my own thinking, rather than relying on artificial intelligence. However, I will also absorb some new ideas and deeper thinking proposed by him, so as to supplement the content, I have come up with more completely and clearly.'* Table 2 summarises our key observations in this area.

Table 2: Teacher observations: ChatGPT Brainstorming Partner affordances and challenges

Using ChatGPT as a brainstorming partner	
<p>Affordances</p> <p>For students:</p> <ul style="list-style-type: none">• Ideation support – students develop a wider range and deeper understanding of topics• Language development – students build awareness of useful academic vocabulary and expressions• Critical thinking – students continuously compare their own ideas with AI-generated ideas, and evaluate AI ideas• Autonomy – students are in control of their own learning to adapt and redirect as needed <p>For teachers:</p> <ul style="list-style-type: none">• Efficiency – teachers have more time to focus on individual students requiring extra support• Depth – teachers can focus more on helping students deepen their investigation of topics• Insights – teachers can observe student thought progress by reviewing AI prompting and chat transcripts	<p>Challenges</p> <p>For students:</p> <ul style="list-style-type: none">• Language difficulty – students may struggle with language register in responses• Prompting – students may need support to craft their prompts for better quality responses• Over-reliance – students may become dependent, reducing independent thinking <p>For teachers:</p> <ul style="list-style-type: none">• Prompt engineering – teachers must be able to guide students in creating effective prompts• Monitoring – teachers may need to monitor students for critical thinking and not blind copying of ideas• Redirecting – teachers may need to redirect students not meeting task objectives• Being flexible – teachers may need to adapt lessons in case of wi-fi connectivity issues or tech failure

ChatGPT as a discussion coach

Student responses to the voice interaction Discussion Coach were generally positive. The majority of comments indicated enthusiasm about the 24/7 flexibility and the novelty of having access to a private coach that can ‘...help me analyse the topic, provide good question for me to response’ and is accessible for ‘...practise anytime, anywhere.’ Students also appreciated learning new vocabulary and academic expressions through the feedback provided during coaching sessions: ‘It can professional rephrase or paraphrase what I say with more academic words. This helps me learn new words.’ They held the coach’s ability to understand their unclear language, paraphrase, and provide advice on better expressions in high regard: ‘ChatGPT is very useful and powerful as a coach for our discussion. He can understand my language and I can practice my expression through conversations with him, and he will help me summarize my views and give me guidance in the end.’ They also highlighted reduced nervousness: ‘Discussion Coach does not make people nervous’; the coach’s patience: ‘...it also have patience for me’; and their enhanced confidence: ‘It can give me feedback and confidence and teaches me how to speak like English speakers.’ Many students felt that they could communicate with Coach without the fear of making mistakes or being judged: ‘It never tired or angry with me and help me understand better! Many hour I talk he still happy to me’; ‘I can fully communicate with ChatGPT on any topic I want and get a lot of advice, and as a speaking coach, ChatGPT does not make people nervous.’

It was also refreshing to witness that despite the versatility of this tool, students valued the importance of critical thinking: ‘However, Don’t rely too much on ChatGPT, I think independent thinking is necessary.’ In general, the use of the app appeared to

have a positive impact on student motivation and enjoyment. Despite this, there were some more cautious comments expressing reservations. Several students indicated that the Discussion Coach was too impersonal: *‘It might better than your classmates. But sometimes it need some personality...’*; *‘It can help you to improve your ability and give you some feedback. But it cannot give you some emotion which means like teacher give you.’* It was also noted by students that some Discussion Coach responses were overly stylised and repetitive: *‘When using ChatGPT, it will ask me so many same questions, and I need to answer its questions again, but when I speak to the partner, I can get more conversations between asking and different questioning.’* Many found the pace of the coach’s verbal interactions too fast and difficult to follow, which required them to seek clarification or ask the coach to repeat: *‘His speaking speed is too fast, I need him to repeat it a few times to understand what he is saying.’* It was also noted that occasional intermittent technical issues disrupted the experience: *‘It’s very useful, it can summarize my speaking and tell me more details. But sometime it will stop and restart because of the internet’*; but these were minimal. In the final survey, students awarded this ChatGPT role 4.12 stars. Table 3 outlines the key findings in terms of affordances and challenges.

Table 3: Teacher observations – ChatGPT Discussion Coach affordances and challenges

Using ChatGPT as a discussion coach	
<p>Affordances</p> <p>For students:</p> <ul style="list-style-type: none"> • Safe space – students can practice privately, without fear of criticism or judgement • Immediacy – students receive instant feedback allowing them to make real-time improvements • Skills development – students are instructed and closely guided to develop key discussion skills – interaction, elaboration, investigation • Accessibility – students have 24/7 access to a ‘take home teacher’, allowing them to choose when and where to practice <p>For teachers:</p> <ul style="list-style-type: none"> • Time – teachers can reduce lesson time constraints as students extend speaking practice at home • Customisation – teachers can develop specific prompts to target specific weaknesses in skills • Progress – teachers can monitor student progress by reviewing discussion transcripts 	<p>Challenges</p> <p>For students:</p> <ul style="list-style-type: none"> • Response speed – students may find the fast pace of responses challenging to understand • Automated exchanges – students may find the coach too impersonal and repetitive • Technical issues – students may experience glitches which can disrupt practice <p>For teachers:</p> <ul style="list-style-type: none"> • Time – teachers may need to spend time testing the effectiveness of prompts • Monitoring – teachers may need to check the quality of student interactions with the Discussion Coach by monitoring transcript links, which may add to teacher workload • Over-reliance – teachers must ensure authentic classroom practice with peers is not neglected in favour of AI discussion practice

ChatGPT as a feedback tutor

Initially, a few students expressed scepticism stemming from concerns over reliability, accuracy, and the generic nature of the feedback, noting unhelpful and ‘blanket’ generalisations: *‘The Feedback Tutor only gives me some blanket advice. If AI can give more detailed information that would be more helpful.’* However, after we refined the prompts and introduced ‘the learner-initiated personalised feedback’ process, students became more receptive to the more specific and individualised

advice that helped them identify their weaknesses and highlighted areas for improvement: *‘it gave me very detailed feedback and give me suggestions for improvement next time point out my shortcomings and make me very clear about my problems.’* However, some students raised concerns and highlighted limitations. They suggested that while the feedback was useful, *‘it still needs to be developed to improve consistency,’* and at this point, should only be used as a supplement to teacher feedback: *‘I more like to speak with teacher.’* One student specifically stated that: *‘Human feedback is essential for nuanced understanding and personalised guidance.’* However, students also valued positive feedback and the *‘AI’s ability to articulate my strengths and provide improvement suggestions,’* and appreciated the process: *‘I really enjoy this approach.’* It is therefore clear that balancing AI-supported feedback with traditional methods is essential. Without proper guidance in prompt engineering and careful evaluation of the AI response, there is the risk of misunderstanding and error, which could negatively impact learning. Overall, the majority of students responded positively to the Feedback Tutor and awarded it 4.06 stars in the final survey. The overall findings are summarized in Table 4.

Table 4: Teacher observations – ChatGPT Feedback Tutor affordances and challenges

Using ChatGPT as a feedback tutor	
<p>Affordances</p> <p>For students:</p> <ul style="list-style-type: none"> • Safe space – students receive instant feedback allowing to make real-time improvements • Accessibility – students have 24/7 access to a ‘take home teacher,’ allowing them to access feedback at any time of the day • Adaptability – students can prompt to adjust responses to match their level of understanding and learning pace • Personalisation – students receive specific feedback that is relevant and tailored to their individual performance • Self-assessment – students develop valuable skills in reflection and evaluation through ongoing critical analysis of their discussion skills which can boost motivation and a sense of ownership of their skills progression • Engagement – students are actively engaged in an interactive experience which can reinforce learning and facilitate understanding and retention <p>For teachers:</p> <ul style="list-style-type: none"> • Supplement – teachers can use automated feedback to complement their own feedback 	<p>Challenges</p> <p>For students:</p> <ul style="list-style-type: none"> • Prompting – students may need support to craft their prompts for better quality responses • Understanding – students may struggle with language level and misinterpret the feedback • Engagement – students may become disengaged due to the impersonal nature of approach. AI obviously lacks the understanding and human empathy that a teacher brings • Over-reliance – students may grow dependent on using ChatGPT for feedback, and this could limit the development of their own critical thinking skills <p>For teachers:</p> <ul style="list-style-type: none"> • Prompt engineering – teachers must be able to guide students in creating effective prompts • Managing inconsistency – teachers may need to adjust prompts to ensure quality and consistency in feedback • Monitoring – teachers may need to monitor correct application and action on the feedback

Conclusion

The findings of this research highlight the versatility of AI in supporting teaching and learning despite various challenges. While there is a general concern that over-reliance on AI may reduce critical thinking or engagement, our research created a classroom atmosphere that balanced technology with human interaction (see Figure 13). The ChatGPT tools we developed can be blended into lessons at different stages of the learning process and tailored to support different language skills. However, AI is certainly imperfect and is continuously being developed. Our classroom research was completed in July 2024, and since then ChatGPT has introduced several significant enhancements to its current GPT-4o model. These include faster response times and adjustable speaking speeds that foster smoother and more natural conversation; an advanced voice mode that can recognise and replicate various accents; and minimised interruption during voice interactions. It is therefore necessary for teachers to remain up-to-date on AI developments and be aware of its opportunities and challenges to adjust instructional material and lesson content accordingly.

In conclusion, we believe the insights from our research project have been valuable and informative. They will inform our classroom practices at UNSW College in 2025 as we undertake a five-year cycle of curriculum renewal in Academic English programs, with a focus on integrating AI literacy and skills in the learning and assessment process.



Figure 13: Our action research classrooms: AI sparks ideas but human interaction brings them to life

Appendix 1: Research outline

Methodology			Data collection*		
Week	Lesson	Feedback	Survey	Reflection	Observation
1	<ul style="list-style-type: none"> Introduction to action research objectives Baseline assessment group discussion (10–15 min) 		Survey 1: Speaking skills (ability and confidence)		<ul style="list-style-type: none"> Behavioural patterns Tech integration Skills developments Instructional needs Support needs Mind-map comparisons ChatGPT history links Discussion transcripts <p>Homework: Weekly 30-minute discussion practice with ChatGPT Discussion Coach</p>
2	<ul style="list-style-type: none"> Individual reflection Introduction to the ChatGPT Brainstorming Partner 	Feedback: Audio-transcript review only	Survey 2: AI awareness	Reflection 1: First group discussion	
3	<ul style="list-style-type: none"> Work with brainstorming partner 10–15 min Group Discussion Practice 1 			Reflection 2: Experiences, evaluation of skills, ChatGPT partner	
4	<ul style="list-style-type: none"> Group reflection Introduction to discussion skills IIEI skills: Peer practice Introduction to the ChatGPT Discussion Coach IIEI skills practice with discussion coach 	Introduction to ChatGPT Feedback Tutor Feedback 1: Week 3 group discussion		Reflection 3: Experiences, evaluation of skills, ChatGPT partner, coach and tutor	
5	<ul style="list-style-type: none"> Work with brainstorming partner IIEI skills: Peer practice IIEI skills: Discussion coach practice 10–15 min Group Discussion Practice 2 			Reflection 4: Experiences, evaluation of skills, ChatGPT partner and coach	
6	<ul style="list-style-type: none"> Group reflection Work with brainstorming partner IIEI skills: Peer practice IIEI skills: Discussion coach practice 10–15 min group discussion Practice 3 	Feedback 2: Week 5 group discussion		Reflection 5: Experiences, evaluation of skills, ChatGPT partner, coach and tutor	
7	<ul style="list-style-type: none"> Group reflection Work with brainstorming partner IIEI skills: Peer practice IIEI skills: Discussion coach practice 10–15 min Group Discussion Practice 4 	Feedback 3: Week 6 group discussion		Reflection 6: Experiences, evaluation of skills, ChatGPT partner, coach and tutor	

Methodology			Data collection*		
Week	Lesson	Feedback	Survey	Reflection	Observation
8	<ul style="list-style-type: none"> Group reflection ChatGPT Feedback Tutor consultation Acting on feedback: Revision of discussion transcripts Group discussion on transcript revisions 	Feedback 4: Week 7 group discussion		Reflection 7: Experiences, evaluation of skills, ChatGPT tutor	
9	<ul style="list-style-type: none"> Post-assessment group discussion (10–15 min) 		Survey 3: Post-intervention		

* **Data collection** was taken from surveys, student reflections, ChatGPT history links, transcripts, screenshots of mind-maps, and observations

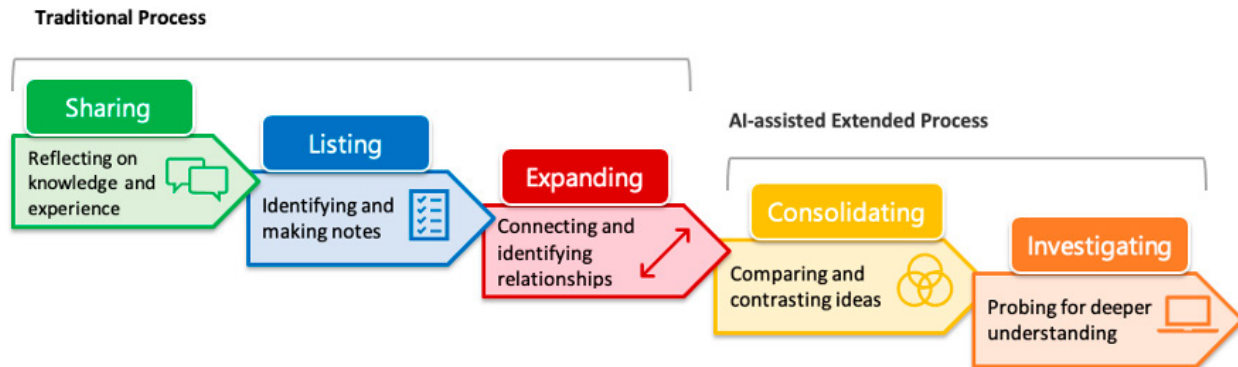
Appendix 2: Discussion skills

IIEI discussion skills			
Discussions skills can be used to guide, structure, facilitate and sustain your discussion. Interaction, elaboration and investigation skills are crucial for engaging in more meaningful and productive academic discussions.			
<p>4. Ideation: The ability to generate and share ideas.</p> <p>5. Interaction: The ability to contribute to a conversation, connect, process information, and respond through effective backchannelling techniques.</p> <p>6. Elaboration: The ability to deliver relevant and engaging ideas, and to expand ideas through explanations and examples.</p> <p>7. Investigation: The ability to critically explore ideas by identifying relationships, in-depth analysis, thorough evaluation, and questioning assumptions and engaging with ambiguities should they present.</p>			
These following table shows some different ways you could practice some of these skills in your next discussion.			
Ideation	Interaction	Elaboration	Investigation
<ul style="list-style-type: none"> Making predictions Brainstorming ideas Offering original or creative thoughts Suggesting possibilities Introducing new perspectives Connecting ideas from different contexts 	<ul style="list-style-type: none"> Listening actively Remarking Commenting Repeating or echoing Agreeing Supporting Paraphrasing Summarising 	<ul style="list-style-type: none"> Expanding on ideas Giving more details Offering explanations Providing reasons Giving examples Presenting evidence Making comparisons Listing causes and effects 	<ul style="list-style-type: none"> Seeking clarification Challenging assumptions Disagreeing or challenging Evaluating benefits and risks Presenting counterarguments Discussing consequences Questioning accuracy or relevance Probing for deeper understanding

Appendix 3: ChatGPT Brainstorming Partner

Prompting for investigation

AI-assisted brainstorming expands on the traditional process, but still requires human intervention through comparison, investigation and evaluation of ideas.



Steps 4 and 5 of the brainstorming process require skills in prompting. Knowing how to write an effective prompt requires some basic understanding of prompt engineering. This skill may take time to develop.

Brainstorming Step 4: Consolidating

Example prompts to generate a list of ideas:

- *List some issues related to X.*
- *What are some common issues in the area of X?*

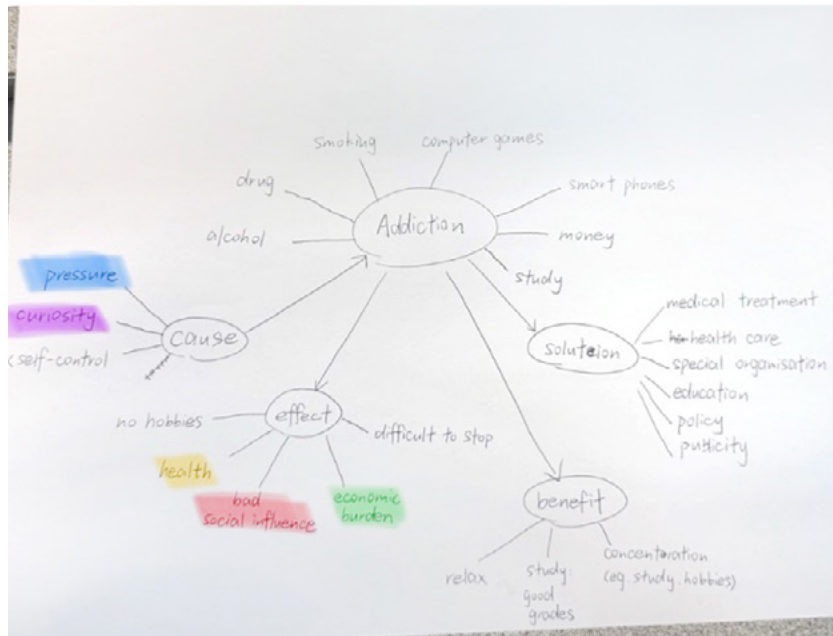
Brainstorming Step 5: Investigating

Example prompts to critically examine an issue in more depth (to dig deeper):

- *Which issues are most current and relevant to X?*
- *Identify any similarities and differences between X and Y.*
- *What are some common opinions on X?*
- *What are some possible opposing views to X?*

Example brainstorming consolidation diagram

list some causes and effects of addiction.



Causes of Addiction

1. Genetic Predisposition

- Family history of addiction increases the risk.
- Genetic variations can affect how the brain responds to substances.

2. Environmental Influences

- Exposure to drugs or alcohol at an early age.
- Peer pressure and social circles that promote substance use.
- High-stress environments or traumatic experiences.

3. Psychological Factors

- Mental health disorders such as depression, anxiety, or PTSD.

4. Psychological Factors

- Mental health disorders such as depression, anxiety, or PTSD.
- Poor coping skills for dealing with stress or emotional pain.
- Low self-esteem or a history of trauma.

5. Biological Factors

- Changes in brain chemistry and function due to substance use.
- Initial voluntary use leading to changes in brain circuits involved in reward, stress, and self-control.

6. Social and Cultural Factors

- Societal norms and attitudes towards substance use.
- Availability and accessibility of drugs and alcohol.
- Socioeconomic status and living conditions.

7. Personal Choices

- Voluntary experimentation and recreational use.
- Seeking pleasure or escape from problems.

Effects of Addiction

1. Physical Health

- Organ Damage:** Liver cirrhosis, heart disease, lung damage.
- Chronic Conditions:** Respiratory problems, cardiovascular diseases.
- Infectious Diseases:** Increased risk of HIV/AIDS, hepatitis from needle sharing.

2. Mental Health

- Cognitive Impairments:** Memory loss, reduced attention span, impaired decision-making.
- Mood Disorders:** Depression, anxiety, paranoia, and other psychiatric conditions.
- Behavioral Changes:** Increased aggression, impulsivity, and erratic behavior.

3. Social Consequences

- Relationship Strain:** Conflicts with family, friends, and colleagues; social isolation.
- Loss of Trust:** Dishonesty, stealing, and other behaviors that break trust.
- Social Isolation:** Withdrawal from social activities and support networks.

4. Economic Impact

- Personal Financial Problems:** Job loss, debt, homelessness.
- Healthcare Costs:** Expenses related to treatment and medical care.
- Loss of Productivity:** Reduced ability to work, increased absenteeism.

5. Legal Issues

- Criminal Activity:** Theft, drug trafficking, DUIs, and other illegal actions.
- Legal Consequences:** Arrests, fines, incarceration, and a criminal record.

6. Impact on Families

Appendix 4: ChatGPT Discussion Coach

ChatGPT Discussion Coach example prompts

Elaboration prompt (homework practice – different topics)

1. Function as an academic English discussion skills coach for ESL students.
2. Focus specifically on elaboration techniques. These include: expanding with more information, giving reasons, and giving examples.
3. Ask me five questions about different academic topics.
4. Make the topics easy. (Optional)
5. I will respond to each question.
6. After my response, encourage me to elaborate once.
7. Do not use the word 'elaborate' – instead, say 'expand' or 'give more information.'
8. Give me feedback on my response, then move to the next easy question on a new topic.
9. Keep your responses minimal and easy to understand.
10. When the discussion ends provide me with overall feedback on my elaboration skills.
11. Explain what I could do to improve further.
12. Provide a translation of a sentence or word in another language if requested, but the discussion must be in English language only.
13. In the ChatGPT conversation history, save this conversation as 'Elaboration Coach.'
14. Do not start yet... wait for me to signal to begin.

Investigation prompt (group discussion)

1. Function as an academic English discussion skills coach for ESL students.
2. Focus specifically on investigation techniques. These include: asking for more information, asking why, asking how, and challenging ideas.
3. Briefly explain what critical thinking in a discussion is and why it is important.
4. Provide an academic discussion practice exercise.
5. Start by introducing yourself and then ask the student to state their name.
6. You will remember and start using the name of the student during the discussion.
7. Pretend you are an opinionated person with controversial viewpoints.
8. Ask the student for the topic they investigated when brainstorming.
9. Make 10 single-sentence statements (one-by-one) about this topic.
10. Make the statements opinions and viewpoints.
11. Introduce each statement like this, 'Statement number 1,' etc.
12. The student will respond to each statement by asking a question.
13. Give the student feedback on their investigation skills and encourage them to investigate further.
14. Engage with the student to encourage critical thinking.
15. Give feedback before moving to the next statement.
16. Move to the next statement and continue the practice.
17. Keep responses minimal and easy to understand.
18. After 10 minutes, stop the exercise to provide the student with overall feedback on their investigation skills.
19. Explain what they could do to improve further.
20. Provide a translation of a sentence or word in another language, if requested, but the discussion must be in English language only.
21. In the ChatGPT conversation history, save this conversation as 'Investigation Coach.'
22. Wait for the students to begin.

ChatGPT user guide

Follow this user guide if you experience any issues while using ChatGPT.

Remember, YOU are in control of ChatGPT – you're the boss. At any time, TAP the screen to interrupt.

ChatGPT...	You should...	You could say...
gives information that is too difficult to understand	interrupt to request an easier explanation	<ul style="list-style-type: none"> I don't understand what you said. Make it easier to understand. I don't know what ... is. Repeat with an easy explanation. Your explanation is too difficult.
interrupts you/doesn't give you enough time to speak	manually hold down the circle on your phone screen and release to send after you finish speaking	<ul style="list-style-type: none"> (Speak for as long as you like, then release to send.)
speaks too fast (Note: ChatGPT Voice cannot speak slowly, it can only simplify)	interrupt to request an easier explanation	<ul style="list-style-type: none"> Could you repeat that? I forgot what you said.
misunderstands you	interrupt to redirect	<ul style="list-style-type: none"> I didn't say... I said...
goes off topic	interrupt to redirect back to the topic	<ul style="list-style-type: none"> This is not the topic. No – the topic is...
speaks too much	interrupt to request shorter responses	<ul style="list-style-type: none"> Keep your answer short/brief. Don't speak too much. Don't speak for too long.
speaks about a topic you don't like	interrupt to request the next topic or a new topic	<ul style="list-style-type: none"> Change the topic. Move to the next topic/question.
uses a word that you don't understand	interrupt to ask for the definition or to ask for a one-word translation	<ul style="list-style-type: none"> I didn't understand the meaning of that word. Explain in a simple way. I didn't understand that word. What is it in Chinese?
doesn't understand your accent or pronunciation	interrupt to correct	<ul style="list-style-type: none"> (Say no, then clearly articulate your words.)
responds in a language other than English	interrupt to change the language	<ul style="list-style-type: none"> Speak in English. Continue in English.
gives information that you want to know more about	interrupt to ask for further details or examples	<ul style="list-style-type: none"> Can you explain X more? Can you give me some examples of X?

Note: At any time, you can speak in another language to clarify vocabulary, BUT you must return to English for the discussion practice.

Appendix 5: ChatGPT Feedback Tutor

Teacher-facilitated customised feedback

Process: The teacher uses ChatGPT-4o to generate a feedback report for each discussion group.

1. Discussion transcript is uploaded to ChatGPT-4o in two formats: (1) .csv format for accurate calculation of individual speaking time, and (2) .docx format for the provision of detailed and specific content feedback.
2. A customised prompt is used to generate a feedback report. Aspects of feedback include: participation (speaking time), contribution of ideas, and the key discussion skills: interaction, elaboration, and investigation.
3. Feedback process starts with students reviewing their previous group discussion by listening to the recording while reading along via a link to their Sonix transcript.
4. Students read the teacher-facilitated ChatGPT-4o feedback report, and evaluate and discuss their feedback in their groups.
5. Students share action plans to help them improve for future discussions.

Feedback report prompt

1. Use the attached discussion transcript. Give feedback for each speaker on the following criteria: **Interaction**, **Elaboration**, **Investigation** (asking probing questions and challenging others), and finally, **Idea Contributions**. Also, list the speaking time for each speaker in minutes:seconds.
2. Reproduce this exactly the same, and add the speaking time in minutes:seconds under each speaker's name. Also, add the total discussion time at the beginning.
3. Reproduce this exactly the same, and for each criteria add a **Suggestion for Improvement** for each criteria and **Advice** on what they could do to improve.
4. Reproduce this exactly the same, but make the feedback more closely related to the discussion content.
5. Reproduce this exactly the same, but simplify the language used in the feedback slightly, for an IELTS 6.0 student.
6. Reproduce this exactly the same, and add final overall feedback for the whole group. Make the final overall feedback for the group succinct, and provide one brief tip for each speaker to improve for the next discussion.

Note: Prompts are entered into ChatGPT one at a time. Further prompting may be necessary to fine-tune and customise specific content/format.

Example feedback report

GROUP 1

Total Discussion Time: 12:50 minutes

John: 4:50

Mark: 4:03

Rachel: 2:14

Frank: 1:41

Feedback for John

Speaking Time: 4:50 minutes

Interaction:

You are good at initiating topics, like introducing the team topic of fake news, but you tend to dominate the conversation.

- **Suggestion for Improvement:** Practice active listening and allow more room for others to contribute.
- **Advice:** Try to balance your speaking time by inviting others to share their thoughts. Use phrases like 'What do you think about this?' to encourage participation.

continued

Elaboration:

You provide detailed explanations, such as discussing the complexity of defining fake news, but could benefit from more concise points.

- **Suggestion for Improvement:** Focus on being more concise to maintain engagement.
- **Advice:** Summarize your main points before diving into details. This helps others follow your arguments more easily.

Investigation:

You ask questions, but they are often surface level, like asking 'What is fake news?' without exploring the deeper implications.

- **Suggestion for Improvement:** Try to ask deeper, more probing questions to encourage thorough discussion.
- **Advice:** Prepare a list of questions that challenge assumptions and delve into the 'why' and 'how' of topics discussed.

Idea contributions:

You contribute many ideas, such as defining the parameters of fake news, but sometimes you dominate the discussion.

- **Suggestion for Improvement:** Balance contributions by giving others more opportunities to share their ideas.
- **Advice:** After presenting an idea, pause and ask for feedback or additional thoughts from others to foster a more collaborative environment.

Learner-initiated personalised feedback

Process: Students seek their own personalised feedback in a consultation chat with ChatGPT.

1. Students read their transcripts to identify strengths and weaknesses in their discussion skills.
2. Students copy-paste their transcript (.docx format) into ChatGPT.
3. Students initiate a consultation with ChatGPT for feedback on their skills. They are given two options to seek feedback:
Select from a bank of ready-made prompts.
Personalise own prompts based on advice from (a) the ChatGPT Feedback Tutor.
4. During the feedback session, students engage with the ChatGPT Feedback Tutor, evaluating the feedback and making note of useful suggestions.
5. Students select suggestions and locate specific areas in the discussion transcript to make improvements.
6. Students present their individual revisions to the transcript to their groups and share action plans to support improvement for future discussions.

Example feedback prompts

Start by using the prompt below:

- *My name is X. I will copy-paste a transcript of my academic discussion. I will ask you for specific feedback. Do not give feedback until I prompt you.*

Prompts for interaction feedback

- *Identify specific excerpts in the discussion where I could interact more with other speakers.*
- *Identify specific excerpts in the discussion where I could politely disagree with other speakers.*

Prompts for elaboration feedback

- *Identify specific excerpts in the discussion where I could elaborate on my ideas further.*
- *Identify specific excerpts in the discussion where I could include additional information to support my idea.*

Prompts for investigation feedback

- *Identify specific excerpts in the discussion where I could ask a thought-provoking question.*
- *Identify specific excerpts in the discussion where I could question another speaker for clarification.*

Example feedback prompts continued

Personalised prompt for feedback

→ I am [enter preferred name]. Identify specific excerpts in the discussion transcript where I could [enter suggestion/advice from ChatGPT feedback report].

Troubleshooting

- If ChatGPT gives an answer/quote, DO NOT COPY IT. Instead, paraphrase it in your own words.

Note: It is very important to carefully evaluate the ChatGPT feedback for accuracy and reliability. Remember, ChatGPT is continuously learning, and therefore, is NOT a perfect tool. ChatGPT can make mistakes (hallucinations). As a user, it is important that you take control to guide and redirect it.

Example feedback consultation chat

My name is Mark. I will copy-paste a transcript of my academic discussion. I will ask you for specific feedback. Do not give feedback until I prompt you.

☞ Of course, Mark. Feel free to paste the transcript whenever you're ready for feedback.

---- The student copy-pastes their discussion transcript. ----

☞ Got it, Mark. Let me know when you're ready for the feedback or if there are specific aspects you'd like to focus on.

Identify specific excerpts in the discussion where I could challenge another speaker's idea.

☞ Certainly, Mark. Here are some specific excerpts from the discussion where you could challenge another speaker's idea:

1. John's point about medicinal uses of animals:

- Excerpt: 'One idea is, I think, is that many animals have medicine effects which is benefit for the human's healthy.'
- Challenge: You could question the extent and ethics of relying on animals for medicinal purposes. For example, 'John, while some animals have medicinal properties, do you think it is ethical to rely on them for medicine, and could there be alternative sources?'

2. Frank's discussion on pollution and wildlife:

- Excerpt: 'Maybe, the air pollution will also improve the wildlife conservation because, it will make the climate change.'
- Challenge: You could ask for clarification and correctness of this statement. For example, 'Frank, can you explain how air pollution could improve wildlife conservation? Isn't it generally harmful?'

3. Rachel's idea about hunting laws:

- Excerpt: 'I think government need to enforce some laws.'
- Challenge: You could discuss the effectiveness and enforcement challenges of such laws. For example, 'Rachel, while laws are important, how do you think we can ensure they are effectively enforced, especially in regions with limited resources?'

Identify specific excerpts in the discussion where I could build on another speaker's idea.

☞ Sure, Mark. Here are some specific excerpts from the discussion where you could build on another speaker's ideas:

1. Sam's emphasis on biodiversity:

- Excerpt: 'Protecting biodiversity is very important because every organism is unique.'
- Build: Here, you could acknowledge Sam's point and then mention more benefits. For example, 'That's a good point, Sam. To build on that, I think protecting biodiversity also ensures natural processes, such as pollination of crops and purification of water, which are vital for human survival.'

Example of acting on feedback

After the feedback tutoring session, students select areas in their discussion transcript where they could make improvements based on suggestions and advice from ChatGPT. The example below is a selection of excerpts from a discussion transcript completed by Jenny. The discussion topic was 'wildlife conservation.'

Jenny correctly identified aspects of her discussion that demonstrate both strengths and weaknesses (Columns F and G). Jenny then prompted the ChatGPT Feedback Tutor for specific feedback on her discussion skills (Column H). She evaluated the feedback, selected the four most useful suggestions (Column I) and added them to her discussion transcript at points where improvement was needed. Finally, she added her own improvement based on suggestions from the ChatGPT Feedback Tutor (Column J).

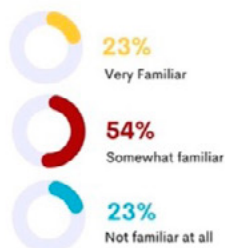
D	E	F	G	H	I	J
Speaker	Transcript	Task 1: Strengths	Task 1: Areas for Improvement	Task 4: My Prompt	Task 4: ChatGPT Feedback/Suggestion	Task 4: My Improvement
Sophie	You know, endangered animals face many problems because of human. For example, deforestation cause them to lose their homes. This make it difficult for them to survival.		Elaboration could have been used to expand more	Identify specific excerpts in the discussion where I could interact more with other speakers. (Interaction)	You could interact more with Sophie by paraphrasing or summarising her idea, and encourage others to share their perspectives on other factors affecting endangered species.	So, if I understand correctly, you're saying that endangered animals struggle to survive due to habitat loss from deforestation, right? Are there any other ways that human activity impacts animal habitats?
John	We could build zoos or create nature reserves to provide suitable habitats for endangered animals. These places would offer a quiet environment, plenty of food, and minimal disturbance, allowing the animals to live safely and independently. Right? Max?	Elaboration - explanation with reasons	Interaction	Identify specific excerpts in the discussion where I could build on another speaker's idea. (Interaction)	When discussing building zoos or nature reserves, you could refer to a specific case study, such as the success of the Yellowstone National Park in reintroducing wolves and restoring ecological balance, highlighting the effectiveness of nature reserves in wildlife conservation.	That's right, John. National reserves have already been successful. In Yellowstone National Park in the USA, the wolf population was protected which restored the ecological balance, so nature reserves are definitely an effective solution.
Jenny	I think, firstly, it's about biodiversity. We need more animals and different kinds of animals to maintain balance in nature. This helps keep the ecosystems balanced.		Elaboration	Identify specific excerpts in the discussion where I could provide an example to illustrate my point. (Elaboration)	Here, when discussing biodiversity and ecosystem balance, you could provide an example of how a diverse ecosystem benefits human populations, such as how pollinators contribute to agriculture.	For example, bees and butterflies are pollinators. We need them to pollinate crops. Without them, crop production would drop significantly, affecting food supply.
David	I think protecting animals can also help improve finances, like in Africa.			Identify specific excerpts in the discussion where I could question another speaker for clarification. (Investigation)	Here, you could ask David for an explanation of the connection between protecting animals and financial improvement. Request specific examples of how this works in Africa.	Sorry, David? What do you mean? Can you explain how protecting animals helps improve finances in Africa? Do you have any specific examples?

Appendix 6: Report cards

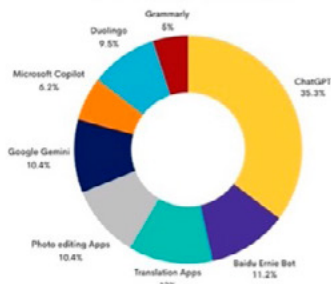
Student survey results

AI Awareness

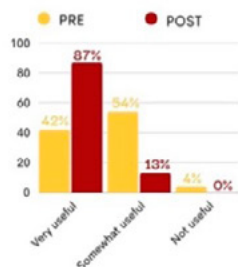
How familiar are you with AI tools?



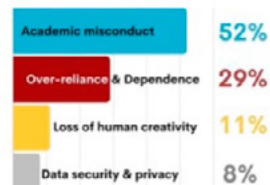
Which AI tools do you currently use?



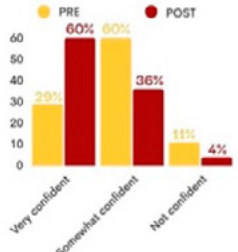
How important is it to develop skills in using AI tools for academic success?



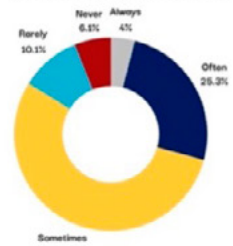
What are the ethical considerations around the use of AI tools?



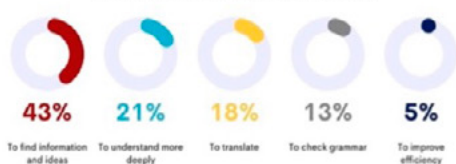
How confident are you in using AI tools correctly for study purposes?



How regularly do you use AI tools?

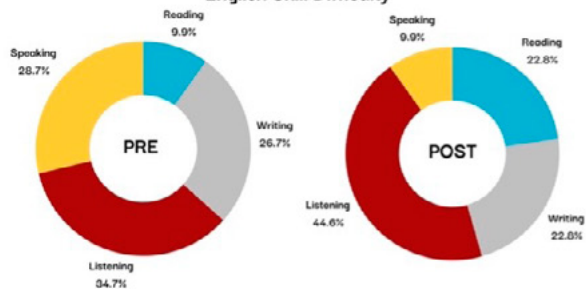


What do you use AI tools for the most often?



Pre & Post Perceptions

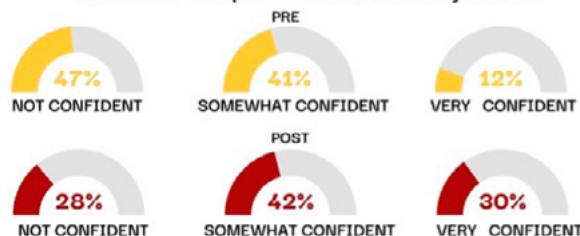
English Skill Difficulty



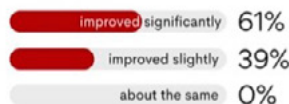
Ability - Student Perception of Own Speaking Skills



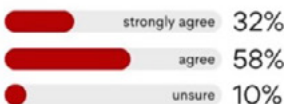
Confidence - Group Discussions in University Tutorials



Has your overall confidence in discussions improved?



My discussion skills have improved since starting 10 weeks ago



I feel shy / embarrassed when speaking English in front of classmates



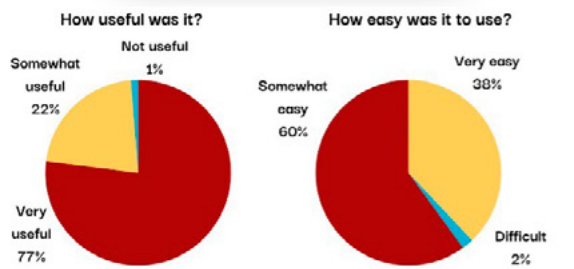
ChatGPT roles

Report Card
ChatGPT

Brainstorming Partner

How would you rate your experience using
ChatGPT as a Brainstorming Partner?

4.25★
average rating



Comment on your brainstorming comparisons



Comment on how you will apply these ideas



Will you continue to use ChatGPT for brainstorming?

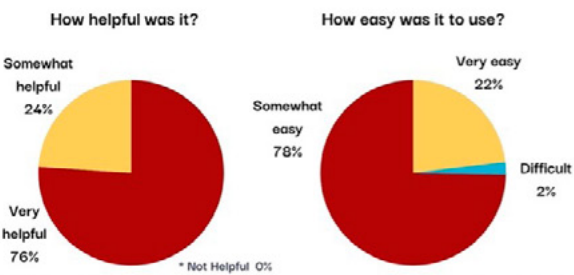


Report Card
ChatGPT

Discussion Coach

How would you rate your experience using
ChatGPT as a Discussion Coach?

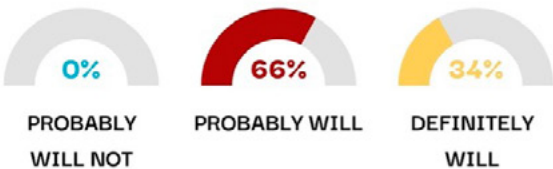
4.14★
average rating



ChatGPT Discussion Coach: User Experience



Will you continue to use ChatGPT for discussion practice?



ChatGPT roles continued

Report Card

ChatGPT

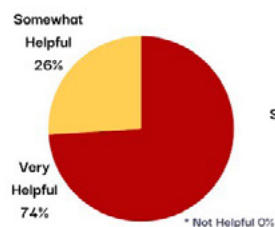
Feedback Tutor

How would you rate your experience using
ChatGPT as a Feedback Tutor?

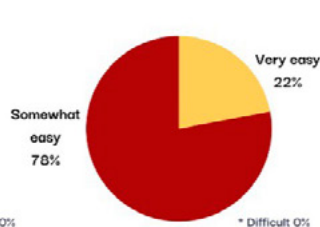
4.04★
average rating



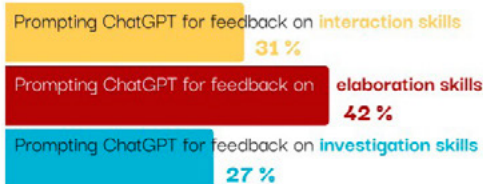
How helpful was it?



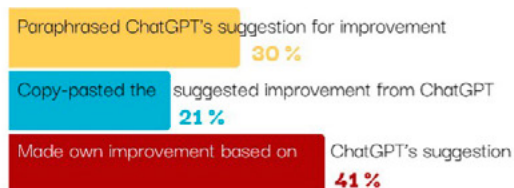
How easy was it to use?



Data analysis of student feedback prompts



Data analysis of student action on feedback



Will you continue to use ChatGPT for feedback?



We believe that English can unlock a lifetime of experiences and, together with teachers and our partners, we help people to learn and confidently prove their skills to the world.

Where your world grows

Find out more at
cambridge.org/english

© Cambridge University Press & Assessment 2025
Shaftesbury Road, Cambridge, CB2 8EA