

Research Notes 89

Findings of the Action Research in
ELICOS Program 2025

Issue 89/August 2026

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 L Stubbings

Contents

- 5 Action research: What on earth could go ‘wrong’?**
Anne Burns
- 15 Building feedback literacy and transferable competencies in academic writing with AI:
A cultural lens**
Amelia Mercieca and Diana Gitman
- 39 Engaging general English writers in ELICOS: A dialogic feedback model integrating AI**
Jamina Coyne and Matthew Knowles
- 53 Using personalised AI feedback to support paraphrasing in EAP**
Martin Sanchez
- 73 Shifting student AI literacy through customised feedback to support writing assessment**
Ruby Zhuang
- 87 Did you ask, challenge, evaluate, and learn? Exploring the use of ChatGPT in learning
phrasal verbs**
Karen Haeng-A Kim
- 107 AI in the conversation: Enhancing group speaking through feedback informed by
student voice**
Natasha Sarooshi

Action research: What on earth could go ‘wrong’?

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

Introduction

Many teachers across the world, like those featured in this issue of *Research Notes*, now involve themselves in action research (AR). They do so for many reasons, including to complete formal qualifications, to learn about research and develop their skills, to fulfil requirements for teacher development, to develop themselves professionally, to reflect more deeply on their practices, to satisfy their curiosity about AR, or to join their colleagues in a community of research practice.

For many teachers, going through a process of AR can be a very new and quite often puzzling experience, especially if ideas about ‘research’ have been formed through traditional concepts of what research is. Various research studies over the last decade or so (e.g., Borg 2009, Banegas 2018, Kang and Yang 2022, Heng et al 2024) have highlighted that often teachers believe that ‘proper’ research or ‘doing research properly’ has to do with scientific experimentation, control and experimental groups, stratified sampling, complex statistical analysis, measurement, and generalisable results. This kind of research is quite likely to follow a planned and structured design using predictable data techniques and well-established analytical procedures. In contrast, AR means entering into a highly localised context of practice and experiencing shifting and spiralling change cycles by identifying curiosities, challenges or dilemmas, planning for action, enacting new approaches, finding evidence for outcomes and claims, reflecting and developing, and deepening professional understanding (see Burns 2019). In enacting the cycles of AR, researchers often feel that the ground under their feet shifts frequently in unanticipated and even uncontrollable ways.

Coming up against this very different way of doing research in contrast to more established scientific traditions and paradigms can be destabilising for many teachers who, in the light of their previous assumptions about research, might worry about the 'rights' and 'wrongs' of doing AR. Although research in general can be 'messy,' AR is probably messier than most (Cook 2009), but ironically the mess usually leads to more robust and much richer and transformative outcomes than we first anticipate. In this paper, I highlight some 'mess points' frequently encountered in AR. In fact, as I aim to discuss, what usually eventuates are 'critical action turns' that positively influence and enrich the direction of the research process.

The 'wrongs' that become 'rights'

One of the joys and privileges of researching and facilitating AR is that, over the years, I've interacted with many teachers who have challenged and informed my thinking. Among the many questions and comments I've heard, several relate to the 'mess points' I referred to above. These points are often articulated through the types of expressions or statements outlined below. They are valid and valuable as they lead to the kind of thinking in AR that can help overcome limitations and become empowering and transformative, both in the way the research is conducted and in how deeper reflection and understanding transpire as a result.

I don't know where I'm going with this research

This kind of statement is typical of the early stages of doing AR. Although the thought of starting may be exciting and invigorating, actually setting off on the exploration can lead to feelings of confusion, uncertainty and being overwhelmed. This is because AR is non-linear and does not follow a fixed plan or pre-determined end-point, but unfolds organically as we go along. Even though we may be trying to follow a well-known framework such as the AR cycles advocated by Kemmis and McTaggart (1988), the actual route into the research may be anything but clear-cut. New reflections emerge, different directions and questions suggest themselves and institutional blockages and challenges rear up to challenge us.

In AR, the way forward makes itself clear only gradually, or as one teacher once said to me, 'it's in the doing that it starts to make sense.' Other teachers I worked with in the last few weeks described these early stages as a 'fog,' 'a roller coaster,' or a 'children's adventure book' where you 'choose your own adventure.' On the surface, this seems to be a frustrating time, but in fact it means that our fuzzy beginnings are evolving into more sophisticated and informed ways of thinking. This is a highly reflective and ambiguous period of time, where the realities of our context and the changes we are attempting take shape only gradually. At this stage, the focus and questions are still tentative as we test them out against what is feasible and practicable. Adjustments need to be made, shifts in focus are emerging, and our initial forays into the research still have to be adjusted according to the realities of possibility and practice.

On the surface, all these conditions may seem to be negative, but in fact they are leading to greater maturity of thinking about the central point of our AR, the practical strategies that can realistically be implemented, and the data that can feasibly be collected. Although it's an uncomfortable and messy point in the research process, it's a necessary and crucial phase of reflecting and clarifying what will be central to the eventual success of our AR. Articulating our ideas with mentors or colleagues, talking it over with a critical friend or trusted community of practice, refocusing our priorities, revisiting our topic, focus and questions, reflecting on our experiences and expectations so far, and accepting the chaos and messiness of the AR process can all help to motivate us towards clarifying the next steps.

My questions don't work anymore

Identifying our focus and settling on our initial questions is comforting. At last, we have some clear signposts to follow on our AR journey. However, it's a very common experience to find that these neat and focused questions we identified early on are not the ones that really matter, or are actually the central point of our AR. Their importance diminishes as the practical strategies and data we collect begin to paint a different picture. Maybe the activities are not working or our students react differently from what we expected, the data are not answering our questions, or reveal issues we hadn't noticed before, or institutional factors and changes mean we have to take a different route. All of these unexpected issues, that did not seem important when we began, can make us feel disorganised or unsettled about the relevance of our research.

Although it can seem that the research is falling apart, what is actually happening is that our position on our issues and questions is changing, and our outlook is becoming more sophisticated and reflexive. Beginning the research has allowed us to focus and deepen, to experience revelations in context and to pass through the realities of where the process is leading. It means we are truly engaging with the data, our students and their needs, rather than trying to fit into a predetermined agenda. As our insights deepen, the questions we started with can seem too simplistic or insufficiently focused and we realise there is no point in pursuing what no longer fits the bill. The messiness that has led us to this point is a positive; we can now begin to refine, redraft and reshape our questions into ones that are more meaningful and worth pursuing. For some teachers, this means tweaking questions to bring the real issues into central focus; for others it means developing new questions altogether. Whichever way we go, we are definitely making progress; in fact, in AR we can continue to refine our questions even more as necessary.

I can't answer all my questions

Behind this question is the fact that those new to AR may believe that research has to be 'big.' Our impressions of research can leave us with the idea that a 'good' piece of research means addressing several areas of the topic, otherwise the study will not be rigorous, important or of any significance. When we start with several questions (*How can an AI tool: i) improve critical AI literacy?; ii) increase formative feedback?; iii) empower writers to build autonomy?; see Sarwal and Stoyanova 2025*) or ones that are too broad to answer (*How can I use AI?*), we can feel disappointed

or even like failures, as the direction of our research becomes unclear and we try to cover everything at once. As Sarwal and Stoyanova carefully explained in their article, their new question led them to more productive outcomes for themselves and their students.

The whole point of AR, however, is to focus on quite small, even tiny or narrow, topics that will begin to increase our knowledge and understanding and take us in new directions, even ones that can completely revise the way we think about our students and ourselves. AR is characterised as small-scale, localised and contextualised inquiry into issues considered important to us personally in our immediate environment (Banegas and Consoli 2020, Burns and Dikilitaş 2025). Even little steps into AR can yield surprisingly 'big' findings and transform the way we and our students think. The best advice I can offer is to 'keep it small.'

Also, AR is not just about 'answering questions.' AR is an organic and evolving research process, that engages us both cognitively and emotionally, where deep reflection and our reactions to what we uncover play a major part in seeing things differently. Deeper reflection leads us to a whole new way of being, thinking and behaving as teachers, one where we begin developing a research mindset that can carry us forward towards more critical, self-reflective and agentic identities as educators. This process usually helps us realise that, because of the dynamic nature of our institutional and classroom contexts, we are only ever likely to find partial answers to our questions. As a teacher I once worked with wisely said to me, 'My action research has raised more questions than it's answered' (see below).

My students don't want to do my new activities

Getting started on AR means reflecting on what we want to explore, preparing different approaches, materials, technologies and activities, and then putting our ideas into action. For us and our students, this can be a very disruptive time (as in the first statement above), as we often have no idea how things will work out. It may be that we're introducing our students to very new ways of behaving, aimed at asking them to think more independently, interact differently, or operate in unfamiliar ways. We may therefore be presenting some substantial challenges to our students' accepted ways of working, as well as our own.

All students' classroom experiences in whatever context mean being socialised into accepted ways of doing things that are normalised in their cultural, social and educational environment. Particularly if students move to different countries to study, as in the case of the ELICOS (English Language Intensive Courses for Overseas Students) participants described in this issue, they may encounter demands on the way they are expected to learn which are quite alien to them.

Consequently, it's hardly surprising that students may push back against some of our approaches, join in only reluctantly, or even appear to opt out completely. This does not necessarily mean that our AR has 'failed' or that they will ultimately refuse to participate. What it does mean is that they may be confused, anxious, overloaded or embarrassed by what they must now do. They may be judging their performance against others' or dwelling on previous classroom experiences where they were

expected to take a passive learning role in the classroom. Our AR is very likely disrupting their deep-held expectations of how classrooms operate.

As the AR unfolds and students become more used to what is happening, the dynamics change sometimes dramatically (see Chang 2024, who discusses the case of her student Eric, who moved from non-participant to leader). Many teachers I've worked with have reported how their new approaches and the research they conducted significantly modified the way students learned. The kind of pedagogic dialogues they used which made a difference involved explaining why they were introducing particular activities, how these activities might help students learn, why certain kinds of interactions were being employed, what classroom roles were expected of them, and what outcomes they should aim for. In addition, the teachers provided support and encouragement to students as they altered the way they engaged in learning. Some teachers also included their students as co-constructors or co-researchers in their AR (see, for example, Duke 2023, Williams 2021), enabling students to view themselves as more agentic, independent and authentic learners.

These are not the results I expected

It can be very confronting to start a piece of AR expecting there will be certain kinds of outcomes and then realise that is not the case. These kinds of expectations relate to a form of 'confirmation bias' – at times, we can all feel we just know in advance how things are meant to turn out. As the research goes on, it can be quite a shock to discover that our strategies are not working or that our evidence is showing something unusual and unexpected. Sometimes expectations are not met because we're not asking the right questions yet (see discussion above); at other times, we might have over- or under-estimated our students' abilities to carry out tasks or use the resources we are introducing. For example, some teachers I've worked with assumed their students were all 'digital natives' and then discovered they needed to teach the digital literacy needed. Others made negative assumptions about students' ability to transfer skills learned in a particular task to their general speaking proficiency and in fact found the opposite to be true (e.g., Campbell and Thorpe 2017).

Not having our expectations met is actually positive, although it might not seem so on the surface. AR is not a simplistic problem-solving process, but one that encourages us to question and probe beyond the surface of our work, to appreciate the complexities of our context, and to become more and more open to other ways to do things. The 'surprises' of our research enable us to change direction or understand we need different strategies to genuinely help our students achieve their learning aspirations. They can also help us to adjust and improve the way we collect evidence, maybe using more inclusive and participatory tools, or encouraging students to be our co-researchers. In addition, responding to our unmet expectations can help us reflect more on our teaching philosophies and beliefs and allow us to refine and articulate them to ourselves and others more incisively. In short, the unexpected usually become a catalyst for richer and more productive AR processes and experiences.

My data are all over the place

It is quite common in AR to find we have too much data or that our data don't match up into a clear picture. Collecting too much data – despite knowing that AR is small-scale – is not a weakness in doing the research, but a sign that we are going beyond easy responses and recognising the complexities of the contexts we work in. It's not uncommon to become more and more intrigued or curious as the research continues and new things are discovered, and then to feel we want to have more and more evidence about what's going on. Also, I've worked with many teachers who are dedicated to their work as educators and want to 'go the extra mile', and they become equally dedicated to pursuing more data for their AR. Having too much data can, however, be valuable by allowing us to analyse and reflect on what the most important and salient findings are, for ourselves and other teachers. We can then begin to see new directions that might be worth pursuing in future. A key task in doing any research, however, is eventually having to sift or 'winnow' our data into a manageable and convincing form. Part of this task is to select what data are essential for responding to our research focus and questions. When we go back to the core of our AR, it may be that some of the data can be set aside for the moment. It's possible that these other data can become the basis for new research, or for disseminating interesting facets of our AR in the future through different presentations, oral reports, or published articles.

Apart from having too much data, a sense of having messy or unclear data is another aspect of AR. Usually this happens because we are responding to the real-time, dynamic, and maybe chaotic, nature of the environment. Human behaviour is not predictable and pre-structured and we often find that our data are not neat and linear either. We may have collected both quantitative (e.g., test scores, questionnaire items) and qualitative (e.g., interviews, reflective observation) sources which then makes us wonder how we can relate them to each other. However, having these different types of data is very valuable as we can 'triangulate' our perspectives, or in other words, view the situation from different angles to find the most important patterns that emerge. For example, as the teachers' reports in this issue show, we may have collected student scores and noticed that results have improved. We may also have conducted interviews to find out what students think, or questionnaires to see to what extent they are developing more confidence or knowledge. These different sources help us to compare and contrast what we find. Triangulation is very valuable as it increases the quality and trustworthiness of our AR and helps us to develop more rounded pictures of what occurred. More importantly, it enables us to create deeper understanding that can lead to more meaningful changes in practice.

I now know less (or more) than when I started

Doing AR transforms what we know or believe we know as teachers. This is because the processes of AR usually interrupt or destabilise our knowledge and behaviours in positive ways. Destabilisation sounds negative but is, in fact, a necessary and fundamental aspect of all learning. When we begin to shift perspective and view our context differently, it means we are beginning to discard outdated or irrelevant ways of thinking and engage in deeper insight and understanding.

For some action researchers this feeling can be very disconcerting as numerous uncertainties begin to crowd in. We may be confronted by having to rethink what we previously took for granted in our own practice or wider teaching context as our evidence begins to question our assumptions. It can feel like a loss of our professional confidence or the expertise we believe we have developed. However, knowing that this experience is a common and positive aspect of AR can help us embrace our new identities and begin to see ourselves as more informed educators. Many teachers I've worked with have reported that they get to have new understanding of issues like their students' previous cultural and learning backgrounds, or how practices they felt were effective actually had a negative effect on their students. Going through this kind of self-questioning, developing more critical awareness and accepting that there will always be gaps in our knowledge and practice enables us to move from being a classroom 'technician' to becoming a classroom 'transformer.'

On the other hand, at the end of the process we can feel elated by the expansion in our knowledge that AR produces. We can feel energised by a process that has pushed us out of our comfort zone to reconsider what we previously knew about research and to develop much richer teaching and researching skills. Our new questions also offer us as an exciting route into further spirals of research, as we recognise that what we've done so far is only temporary. We learn to appreciate that our horizons as teachers have expanded and gain new confidence in ourselves as legitimate researchers. We can feel that we are no longer just consumers of research but active research agents, ourselves creating our own pedagogical theories and practices about what works in our classrooms and also our institutions. Our teaching is now based on evidence rather than just intuition. As one teacher once said to me, we're equipped with a 'research toolkit' that can help us bridge gaps between theory and practice and be used whenever we need it.

Reflections and thoughts

In AR, feeling things are going 'wrong' or are in a 'mess' is a positive force. It leads us to abandon a view of AR as a simple 'problem-solution' exercise and to reframe it for what it is – a complex, dynamic, circuitous, and even chaotic journey through the realities of our teaching situation. It helps us to recognise more clearly that neat recipes and simplistic conclusions do not work in language education and that there are many different exploratory routes into helping students towards successful learning. We also learn to theorise what we believe about teaching and learning, rather than to adopt received knowledge and taken-for-granted epistemologies blindly. When we develop the ability to articulate how our teaching works and why we know it's effective, we are showing the difference between 'doing things right' and 'doing the right things' for our students. By embracing 'mess' as a natural part of the AR process, we become open to unexpected possibilities and end up in new locations with powerful and fresh identities we never knew existed.



Left to right: Diana Gitman, Martin Sanchez Lagos, Professor Anne Burns, Karen Haeng-A Kim, Natasha Sarooshi, Jamina Coyne, Matthew Knowles, Ruby Zhuang, Sophie O'Keefe (English Australia), Amelia Mercieca

References

- Banegas, D (2018) Towards understanding EFL teachers' conceptions of research: Findings from Argentina, *Profile: Issues in Teachers' Professional Development* 20 (1), 57–72.
- Banegas, D and Consoli, S (2020) Action research in language education, In McKinley, J and Rose, H (Eds), *The Routledge Handbook of Research Methods in Applied Linguistics*, Routledge, 176–187.
- Borg, S (2009) English language teachers' conceptions of research, *Applied Linguistics* 30 (3), 358–388.
- Burns, A (2019) Action research: Developments, characteristics, and future directions, in Schwieter, J and Benati, A (Eds), *The Cambridge Handbook of Language Learning*, Cambridge University Press, 166–185.
- Burns, A and Dikilitaş, D (Eds) *The Routledge Handbook of Language Teacher Action Research*, Routledge.
- Campbell, V and Thorpe, C (2017) Using technology to develop learning-oriented assessment and autonomy in students, *Research Notes* 67, 12–16.
- Chang, V (2024) Enhancing student feedback literacy through peer feedback, *Research Notes* 87, 33–55.
- Cook, T (2009) The purpose of mess in action research: Building rigour through a messy turn, *Educational Action Research* 7 (2), 277–291.

Duke, G (2023) Training L2 learners to be autonomous through instructional scaffolding, *Research Notes* 85, 85–104.

Heng, K, Doeur, B and Hamid, M O (2024) University English teachers' research conceptions and engagement: Insights from Cambodian higher education, *International Journal of Educational Research* 127, 1–12.
<https://doi.org/10.1016/j.ijer.2024.102450>

Kang, Y and Yang, L (2022) Examining EFL teachers' changing conceptions of research: A case study of a continuing professional development program in mainland China, *Frontiers in Psychology* 13. doi: [10.3389/fpsyg.2022.933061](https://doi.org/10.3389/fpsyg.2022.933061)

Kemmis, S and McTaggart, R (1988) *The Action Research Planner*, Deakin University Press.

Sarwal, R and Stoyanova, L (2025) Working with the enemy: Adopting a generative AI tool to teach AI literacy and increase active engagement with feedback on writing, *Research Notes* 88, 67–84.

Williams, P (2021) Scaffolding presentation skills for an online EAP program, *Research Notes* 81, 17–27.

Building feedback literacy and transferable competencies in academic writing with AI: A cultural lens

Amelia Mercieca and Diana Gitman, University of New South Wales (UNSW) College, Academic English Programs

Introduction

In English for Academic Purposes (EAP) programs, which are often brief and high-stakes, students require feedback that is both timely and actionable to develop academic writing and communication skills. Teachers' limited time for delivering immediate, individualised feedback, alongside culturally and educationally shaped differences in student engagement, creates a challenge for supporting all learners effectively.

Depending on educational background, international students may have limited experience with dialogic or peer-based feedback practices common in EAP settings (Chang 2024). Shifting from direct correction to more reflective agency-building practices can feel disorienting, and similar reactions may arise in students' early interactions with generative AI (GenAI).

Student feedback literacy, defined as 'the understanding, capacities and dispositions needed to make sense of information and use it to enhance work or learning strategies' (Carless and Boud 2018:1,315), is therefore a critical focus. Investigating feedback literacy through GenAI engagement requires examining not only surface uptake but the development of transferable competencies such as learning-to-learn strategies, critical thinking, and communication (Cambridge University Press & Assessment 2026).

These considerations shaped our decision to conduct an action research (AR) study at UNSW College.

Research questions

To explore this integration of feedback literacy, our AR questions (RQs) were:

1. How does GenAI feedback shape students' engagement with integrated writing (IW) and transferable skills development?
2. How are these processes influenced by cultural and linguistic background, and feedback design and preferences?

Research context and participants

Our study was conducted across two direct-entry Academic English pathways: the Pre-Masters course, a 12-week *embedded* EAP program integrated into students' academic timetables, and the University English Entry Course (UEEC), a 10-week full-time English Language Intensive Course for Overseas Students (ELICOS) program. Although similar in duration, the two pathways differ in weekly workload, curriculum structure, and delivery intensity (see Table 1). We each taught one of the two classes: Amelia the Pre-Masters cohort and Diana the UEEC cohort, which enabled us to examine how students with comparable entry proficiency engaged with GenAI feedback under different instructional conditions.

Table 1: Course and participant details

Course type	Embedded: Pre-Masters	ELICOS: University English Entry Course (UEEC)
Duration	12 weeks	10 weeks
Weekly hours	12	20
Entry	IELTS 6 equivalent	IELTS 6 equivalent
Exit	IELTS 6.5 equivalent	IELTS 6.5 equivalent
Pathway to	Master's in Engineering/IT	Master's in STEM (10) Diploma in Computing (2)
IW weekly hours	4	13
Class size	12	12
Age range	23–27	18–25
L1 background	Chinese – 10 initially + 2 late enrolments	Chinese – 11 Persian – 1
Repeat students	4	0
Historical course pass rate	73%	95%

Writing tended to be the lowest entry proficiency for both groups, indicating a need to closely monitor and evaluate the effect of formative feedback on students' development. The two IW assessments in both courses are broadly aligned. Each assessment is conducted under secure and timed conditions, using identical rubrics, and students complete a graded practice beforehand (see Table 2 for task specifications). Topics differ between courses.

Table 2: IW assessment specifications

IW	Input	Output
Mid-term	3 texts (1,800 words)	350–400-word short answer
Final	3 texts (2,000 words)	600–700-word essay

Research design and data collection

A mixed-methods design guided this AR and unfolded over three phases to support the design, implementation, and review of GenAI and related feedback interventions.

Phase 1: Focus groups and pre-intervention survey

We commenced with focus groups and individual surveys to establish a baseline of students' attitudes, culturally influenced expectations, past experiences, and preferences regarding teacher, peer, and GenAI feedback. The aim was to explore both individual and collective patterns. Students responded to guided questions and scenarios that probed their understanding of feedback practices and their perceived value. The survey also captured generalised trends in feedback readiness, confidence, and familiarity with GenAI tools.

Phase 2: Implementation

To examine how students connected issues in practice tasks with the IW performance criteria, we adopted a triangulated feedback approach incorporating peer, teacher, and GenAI input. While this mirrors existing course expectations, the largely self-directed nature of peer and GenAI feedback risks masking differences in students' feedback literacy and their confidence in working with these tools. To provide more equitable support, we designed strategically timed scaffolding and trialled two GenAI feedback models across two research cycles.

GenAI feedback Model 1: Teacher-scripted AI prompts placed by students in OpenAI's ChatGPT

After using teacher-scripted AI prompts to obtain GenAI feedback on their writing, students were encouraged to pose follow-up or clarification questions to ChatGPT independently. We anticipated that, without explicit modelling, students might default to requesting full rewrites rather than using this Large Language Model (LLM) as a personal tutor. We therefore provided guided, Socratic-style prompts that modelled how to interrogate feedback productively rather than outsource

the writing itself. This approach was designed both to prevent over-reliance on GenAI rewrites and to offer a structured 'learning-to-learn' pathway that builds student agency, strengthens feedback literacy, and supports the transfer of effective strategies beyond the classroom.

GenAI feedback Model 2: AI-enabled worksheets formatted in Microsoft Word and Google Forms

In this model, students worked with structured worksheets in familiar formats (Microsoft Word and Google Forms). These worksheets mirrored traditional learning tools by guiding students step-by-step through planning, drafting, reflection, and forward planning, reducing technological distraction and keeping the feedback process within a familiar academic workflow.

The worksheets were coded so that student inputs at key points triggered calls to ChatGPT via a secure Application Program Interface (API) key, embedding GenAI feedback and brief appraisals directly into the documents. Because we lacked specialist technical expertise, this workflow was developed through vibe-coding to test the feasibility for a non-expert designer. Vibe coding is an AI-development style where you guide software creation through conversational instructions instead of writing traditional code. The structured format fully modelled the feedback process, positioning GenAI feedback as a scaffold within students' 'zone of proximal development': the gap between what learners can do independently and what they can achieve with guided support (Vygotsky 1978). This approach aligns with formative feedback theory, which emphasises timely, process-oriented input that helps learners understand where they are, need to go, and how to close the gap (Sadler 1989). By embedding feedback at multiple checkpoints, the worksheet supported ongoing calibration rather than retrospective correction, enabling students to use feedback during performance when it has the greatest impact on learning. To support this process, we also provided students with a simple workflow diagram of the overall structure and sequence of the worksheet.

We anticipated that this staged, dialogic design would be particularly beneficial. Rather than receiving all feedback at the end of the task, Model 2 provided timely, individualised feedback at multiple checkpoints, enabling students to monitor their thinking and make targeted revisions. This positioned GenAI as a scaffold that supplements what teachers cannot realistically deliver at scale, offering a form of responsive support distinct from checklists, rubrics or peer comments.

GenAI prompt design and feedback settings

Across all GenAI interventions, feedback generation was guided by a consistent set of prompt design principles to ensure clarity, relevance, and accessibility for students:

- Anchoring feedback to the task instructions and the students' submitted writing sample
- Adapting rubrics into reflective, student-friendly questions (e.g., 'Have I shown...?' or 'Was I able to...?') to improve readability and support meaningful application
- Limiting comment length to reduce cognitive overload and avoid overwhelming

students with excessively long GenAI responses

- Balancing the focus between strengths and areas for improvement
- Maintaining a constructive tone, drawing on specific examples from the students' work and offering targeted suggestions without providing full rewrites
- Aligning language and tone with students' proficiency level and academic goals

All prompts and worksheets were piloted by us using archived IW student samples and collaboratively reviewed to ensure pedagogical appropriateness, clarity, and effectiveness.

Cycle 1

Cycle 1 began with different data collection designs in the two courses.

In the Pre-Masters course, triangulation was embedded from the outset, and students completed a sequence of tasks using both Models 1 and 2. A limited version of Model 2 was introduced early to test the technical feasibility of the scaffolded worksheet approach. The worksheet emphasised the importance of planning for structured responses in academic writing and generated GenAI feedback on students' short-answer paragraph plans. Model 1 (see the Appendix 1 sample) was then used as a post-diagnostic task and again after students' first IW task. In addition, a ChatGPT-based Writing Practice Activity Generator was provided so that students could produce skills-based practice activities aligned with their individual needs. For example, the tool generated targeted grammar exercises based on weaknesses identified through feedback, enabling students to address verified learning gaps and develop effective early study habits. Model 2 was subsequently used for students' second IW task (see the Appendix 2 sample).

In contrast, Diana drafted additional mini-integrated short answer tasks (mini-ISATs) and incorporated these into the UEEC course specifically for research purposes. Because UEEC students were already experiencing triangulated feedback within regular instruction, these mini-ISATs were introduced to isolate peer, teacher or GenAI feedback one mode at a time. This design allowed clearer examination of how students engaged with each feedback source independently. Though UEEC students also used Model 1 four times during this cycle, student feedback on its usefulness was collected on its application to a mini-ISAT. Their feedback literacy was later assessed through a short redraft demonstrating independent interpretation and uptake of each feedback mode.

Cycle 2

In Cycle 2, both courses shifted toward deeper engagement with scaffolded, dialogic feedback practices. UEEC adopted the scaffolded approach of Model 2 to further test students' learning-to-learn and critical thinking capacities, while also retaining redrafts produced after receiving feedback. Because students' responses in Cycle 1 had already offered a clear evaluation of the usefulness and distinct benefits of each feedback mode, there was no further need to isolate feedback sources through additional mini-ISATs. Instead, we were satisfied that students could meaningfully

reflect on and report their experiences with all feedback modes within the broader triangulated context of Cycle 2.

In the Pre-Masters program, paragraph redrafts were introduced as an additional requirement to increase the cognitive challenge of adopting feedback and to support clearer, more purposeful expression. Together, these adjustments represented a higher level of reflective engagement with feedback, consistent with the growing emphasis on developing feedback literacy and fostering learner autonomy.

Peer review procedures in both programs were also progressively refined to incorporate clearer guidance and prompts for explanation. These revisions encouraged students to justify their feedback decisions, communicate constructively, acknowledge uncertainty, and provide evidence-based suggestions. In doing so, peer review became not merely a task requirement but an opportunity to rehearse core dispositions of feedback literacy, including evaluative judgement and dialogic engagement.

Shared interventions

To give greater contextual clarity of the aligned models, Appendix 3 offers full course timelines, and additional and shared interventions during the study.

Phase 3: Post-feedback cycle surveys

Once students had engaged with the designated feedback mode(s) for that cycle, as outlined in Cycles 1 and 2 above, they completed a short survey evaluating the usefulness of the feedback they had received and reflecting on their feedback-seeking behaviour and developing self-efficacy.

Findings

Phase 1: Focus groups and pre-intervention surveys

Students across both courses expressed a strong preference for timely, specific feedback. Many noted that they were unlikely to apply advice received too late, with one Pre-Masters student explaining, 'If the feedback comes too late, I already forget my thinking ... I won't fix it.' GenAI was valued as an always-available, judgment-free space; a UEEC student described it as 'kind' and 'not scary to ask questions.' While peer feedback was seen as supportive, survey responses indicated ongoing caution about its reliability. Most Pre-Masters students had used GenAI only once or twice and rated it moderately useful, whereas UEEC students showed slightly higher trust, appreciating how it 'helps me see mistakes quickly' even when unsure 'how deep it can go.'

Cultural norms strongly shaped feedback engagement. Students frequently referenced Confucian values of hierarchy and indirect communication. They described feedback interactions influenced by concerns about others' judgements, especially teachers. One student commented, 'People are very

introverted ... they all listen,’ while another noted it was common to ‘think three times before you speak,’ preferring measured reflection over spontaneous dialogue.

Past experiences with feedback were often negative. Several students recalled episodes of public criticism in school or workplace settings, sometimes involving parents or supervisors, which they described as humiliating and filled with ‘a lot of negative words.’ One student explained he tried to do ‘everything correctly’ simply to avoid criticism. Compliments were also viewed sceptically, sometimes interpreted as insincere or motivated by jealousy. No students offered examples of constructive feedback from these contexts; instead, several shared memories of positive and honest praise from online gaming competitors.

These culturally and experientially shaped attitudes highlighted potential challenges for GenAI feedback design. Students emphasised the need for feedback that feels thoughtful, warm, and non-threatening, aligning with norms of reflection and emotional safety. Comments such as ‘I need time to think before I answer’ and ‘I want feedback that doesn’t make me feel stupid’ directly informed the tone and pacing of the GenAI interventions.

Phase 2: GenAI feedback

Model 1

Participation patterns showed differences in autonomy and engagement with Model 1 (see Table 3). During the in-class activity, all Pre-Masters students sought teacher support to interpret GenAI feedback or adapt Socratic-style prompts. As one student said, ‘I don’t know what to ask it ... I wait for teacher to help me.’

UEEC students demonstrated isolated examples of greater independence, with one commenting, ‘I like trying by myself first, then ask AI more questions,’ but this was not widespread. Although all completed the first draft at home, only three progressed to the optional GenAI follow-up stage, indicating inconsistent willingness to self-regulate with GenAI tools despite this cohort’s prior familiarity with Model 1.

Across both groups, students required further modelling to develop deeper critical engagement and communicative competence with GenAI.

Table 3: Model 1 participation pattern

Class	Task	Duration	Location	Attendance	Completion
Pre-Masters	• Read, plan, write • GenAI feedback, follow-up questions	3 days	In class	7/10	7/7 – whole task
UEEC	• Read, plan, write • GenAI feedback, follow-up questions, second draft	2 days	Home	10/12	10 – first draft 3 – GenAI follow-up 8 – second draft

Model 2

Model 2 supported analysis of three feedback literacy competencies: recognising issues, forward planning, and adopting feedback. Students were rated in each area (see Appendix 4), then grouped by their overall level of engagement (see Table 4).

Table 4: Participants' feedback literacy levels in Cycle 2

Class	Student	Feedback literacy level	Class	Student	Feedback literacy level
Pre-Masters	CS	High	UEEC	YA	High
	KZ	High		EV	High
	CY	High		ZH	High
	JP	High		RA	High
	GL	Moderate		RI	Moderate
	KC	Moderate		RY	Moderate
	LD	Moderate		CA	Moderate
	KL	Moderate		DA	Moderate
	PZ	Moderate		PE	Moderate
	BC	Moderate		AL	Moderate
	JZ	Low		JO	Low
	CH	Low			

Eight students demonstrated higher-level issue recognition, such as unclear structure, weak cause–effect logic, or poor source integration. These students could appreciate feedback on discourse features like comparative positioning or ineffective topic sentences, demonstrating emerging awareness of writing quality. Moderate students focused on surface-level problems (e.g. grammar or phrasing), while three students in each cohort remained limited to sentence-level corrections only. One student wrote, ‘I fix grammar first because it is easy ... structure is difficult.’

Planning responses varied in specificity. While one-third offered vague plans, stronger students outlined targeted strategies linked to GenAI feedback, such as ‘reorder comparison to improve cohesion.’ Several Pre-Master’s students noted how GenAI ‘helps me see the plan,’ although low feedback literacy students needed ‘more examples to know what to do next.’ Formative observations from Cycle 1 showed that four Pre-Masters students demonstrated noticeable growth in the length and clarity of their planning responses between Weeks 4 and 10.

Of the 22 revised drafts, six Pre-Master’s and four UEEC students produced substantive improvements in structure and academic voice; the remainder made only surface-level edits. Several Pre-Master’s students independently revisited their saved

Model 2 worksheets during redrafting, a behaviour rarely observed with Model 1. This pattern may help explain why Pre-Master's students slightly outperformed their UEEC peers despite having less experience, possibly reflecting both higher motivation and the refinement of worksheet scaffolds over time. Synthesis remained a consistent challenge across both groups.

Occasional technical instability also emerged when prompting Model 2 to analyse all three texts simultaneously, requiring teacher intervention to repair worksheets during use.

Phase 3: Post-feedback cycle surveys

Across both courses, Pre-Masters and UEEC students demonstrated a growing awareness of feedback as a multi-dimensional learning tool, gradually shifting from teacher dependency to more strategic engagement with multiple feedback sources. This pattern was clear, though a small subset of approximately four to five students participated unevenly each time a model was implemented.

Initial State. Students almost exclusively valued teacher feedback, citing its personalisation and clarity. Several commented that 'teacher understands me' and 'teacher feedback is safest.' A small number viewed GenAI as somewhat helpful but expressed little confidence in peer feedback, with one noting, 'Peers same level as me, so they don't know more.' This reflected limited prior exposure to non-traditional feedback sources.

Evolution. By Week 4, students increasingly recognised that GenAI was effective in identifying weaknesses and offering suggestions, though some questioned its depth. A common remark was, 'AI gives many ideas but sometimes too general.' By the end of Cycle 2, most students saw feedback as valuable for improving writing and supporting broader academic skills. As one noted, 'I didn't know how to plan before. Now I use feedback to improve structure and manage time.' Another said, 'Feedback helps me revise with a purpose, not just fix mistakes.' These comments reflect a shift from passive correction to active learning.

Final Preference. By the end of Cycle 2, students' preferences had diversified. Teacher feedback remained universally valued, but 80% also endorsed GenAI and 60% peer feedback. GenAI was valued for enhancing clarity, flow and idea integration, while peer feedback gained new traction for helping students 'see another point of view' and identify logical gaps. One Pre-Masters student summarised this shift: 'I need all three ... together they show different problems.'

Discussion

Across both cohorts, interactions with GenAI feedback revealed clear developmental patterns in IW competencies and transferable skills. The strongest gains were in planning, idea development and structure; later, students also improved cohesion and source use. GenAI feedback clarified tone and accuracy for some, though outcomes varied with language control and feedback literacy. Students' outcomes

and progress were shaped by the extent of their autonomous motivation to fully engage with the program and those who consistently completed all supplementary activities made noticeably greater gains.

Cultural insights from the focus groups shaped our design. Students' accounts of indirect communication, deference to authority, and fear of negative evaluation prompted our commitment to build feedback tools grounded in these realities. We therefore emphasised psychological safety, staged delivery and space for private exploration before public discussion. We observed how well they encouraged experimentation, reflection and more open dialogue about writing over time.

Synthesis and transformation remained the most demanding skills, largely because they were first-time tasks linked to students' developing academic reading abilities. Progress in synthesis occurred mainly through teacher-led guidance, reflecting the judgement needed to connect ideas across texts. Transformation proved to be even more challenging for less feedback-literate learners, who often avoided optional practice without individual support, indicating a continued need for pedagogical mediation beyond GenAI feedback.

Model 2 produced the strongest evidence of engagement and transfer. Its staged structure meant that instead of hesitating or deferring decisions, learners frequently checked their reasoning before moving to the next stage, suggesting that the reduced cognitive load enhanced motivation to revise and adopt feedback. Survey data reinforced this pattern: students consistently reported that Model 2 made the revision process feel more achievable. Behavioural evidence supported this as many Pre-Masters students revisited saved Model 2 worksheets when redrafting, yet in both cohorts, Model 1 was rarely revisited independently. This distinction suggests that structured tools not only support immediate performance but also become reusable cognitive artefacts extending learning beyond the classroom.

Model 2's technical limits, however, signalled scalability issues. When analysing three IW texts, especially for transformation, the worksheet showed inconsistent API performance due to token load. Teachers could troubleshoot, but this would be impractical at scale. A more robust, software-supported version is needed for comprehensive IW feedback.

Neither cohort showed over-reliance on GenAI, suggesting that effective integration depends on modelling use rather than restricting access. A class forum may further normalise GenAI use and support feedback transfer.

Overall, we concluded from our research that well-designed GenAI scaffolds can enhance integrated writing while fostering transferable skills such as planning, judgement, and self-regulation, functioning best as a culturally responsive extension of pedagogy rather than a replacement for instruction.

References

- Cambridge University Press & Assessment (2026) *Cambridge Life Competencies Framework*. <https://www.cambridge.org/clcf>
- Carless, D and Boud, D (2018) The development of student feedback literacy: enabling uptake of feedback, *Assessment & Evaluation in Higher Education* 43 (8), 1,315–1,325. <https://doi.org/10.1080/02602938.2018.1463354>
- Chang, V (2024) Enhancing student feedback literacy through peer feedback, *Research Notes* 87, 33–54.
- Sadler, D R (1989) Formative assessment and the design of instructional systems, *Instructional Science* 18, 119–144.
- Vygotsky, L S (1978) *Mind in society: The development of higher psychological processes*, Harvard University Press.

Appendix 1: Sample Model 1 – teacher-scripted AI prompts placed by students in OpenAI’s ChatGPT

This sample shows the first Model 1 worksheet used in Week 1 of the Pre-Master’s course to support student engagement with generative AI feedback. Students copied the teacher-provided AI prompt and their writing into ChatGPT to receive feedback, then continued the conversation to explore the feedback in more depth. The material provided to students is shown first, followed by a sample of student writing, their follow-up questions to ChatGPT, and its responses. Appendices 1 and 2 from the original worksheet are not included.

This sample represents a high-quality instance of student engagement, demonstrating strong learning to learn, communication, and critical thinking behaviours characteristic of a ‘Guided Engager,’ who effectively internalises and builds on AI feedback.

Unit 1.C: Writing diagnostic – ChatGPT feedback prompt

Step 1: Copy and paste the computerised-text version of your writing sample below this AI prompt.

AI prompt

Help me to review my response to a writing task using a rubric encompassing four criteria. The writing task question and instructions are: *What are two impacts of urban overcrowding on infrastructure?* Provide specific examples to illustrate these effects. Write two body paragraphs. Do not write an introduction or conclusion. Provide specific examples to illustrate these effects. Write between 200–300 words. The criteria are Content & Relevance, Organisation, Grammar, and Vocabulary.

For Content and Relevance, I should be able to provide a fully developed answer to the writing task prompt with relevant, appropriate and comprehensive content throughout. Did I write between 250 and 300 words? What should I have done to improve this? What might have been more appropriate ideas or examples than those I used?

For Organisation, I should be able to sequence my ideas logically, have well-developed ideas, and be able to demonstrate highly natural and consistent use of cohesive devices throughout. How well have I done this? Are either of my paragraphs illogically arranged? Are my ideas poorly developed and why? Have I made errors with, or can I improve my use of, cohesive devices?

For Grammar, I should be able to use a wide range of complex structures flexibly and accurately and most of my sentences should be error-free. Do I use a mix of simple and complex structures? Are my sentence-level errors at a minimum? Which, if any, of my grammatical errors make it different for a reader to understand?

For Vocabulary, I should be able to convey precise meaning with a wide range of academic vocabulary accurately and flexibly, and have some skilful use of less common vocabulary items. Is my syntax and word choice accurate and effective? Do I rely on repetition of simple items? Do I have frequent errors that might limit

a reader's understanding? Have I made any word choices that are confusing and/or inappropriate, or that might impede meaning?

Limit feedback to three to five sentences per criterion. Please use examples from my writing to highlight my weaknesses. Please be constructive with feedback wherever possible.

Here is my response to the writing task: [\[Copy and paste your writing here\]](#)

Public transportation will be overcrowded in a lot of time. When it comes to morning, it is a time that people should go to work or school. Lots of them will chose to take the public transportation such as bus or subway. So it lead to overcrowding. For short-term effect, when they have to squeeze into the crowded, their mood may get worse and they may be tired after their travel. It is terrible for individual. For long-term effect, overcrowded may contribute to people's chronic disease on their body. And when varieties of people crowded together, it is easy for disease to spread. The spread of disease for long time will result in the public health burden, which means that the hospital will be more pressure.

School is an infrastructure that is easy to be crowded as well. In some counties such as China enjoys lots of populations, which means there are plenty number of students to study in school. When it comes to some activities, they must all of them must come to some positions together. At that time, overcrowding takes place. For short-term effect, overcrowding is unsafe for these students. For example, when they move together, if someone fallen down, it is very dangerous for not only himself, but also for the students around him. For long-term effect, students' passion may be decreased. It is not interesting and uncomfortable when they have to together active together with crowds. And when as they suffer it for a long time, they will have no passion for their activities, it is harmful for students' development.

Step 2: Get the feedback. Copy and paste the [AI prompt](#) above and your writing into ChatGPT.

Step 3: Read the feedback carefully. Then, continue the conversation by asking ChatGPT questions or asking for more help. Use [Appendix 1](#) for example questions and sentence starters.

Step 4: Copy and paste ChatGPT's feedback and your full conversation with it into [Appendix 2](#).

ChatGPT feedback on writing sample

Content and relevance

Your response is generally relevant and addresses the task by discussing two impacts of urban overcrowding: the strain on public transportation and overcrowding in schools. You give both short- and long-term effects, which shows depth. However, your examples are sometimes vague or underdeveloped. For instance, phrases like 'people should go to work or school' and 'they must come to some positions together' lack specificity. Also, the total word count is approximately

270 words, so you're within the target range. To improve, choose more concrete, real-world examples (e.g., referencing overcrowding in a specific city like New York or Beijing) and elaborate more clearly on consequences.

Organisation

The two body paragraphs are separated and focus on distinct ideas, which is good. You attempt to use cohesive devices like 'for example,' 'when it comes to,' and 'as well.' However, transitions are sometimes unnatural or repetitive. The second paragraph, in particular, feels less logically ordered; for example, 'students must come to some positions together' is unclear and the idea jumps quickly from safety to motivation. Improve cohesion by clearly linking causes to effects and using a wider variety of cohesive phrases (e.g., 'This leads to,' 'Consequently,' 'Another impact is...').

Grammar

There is a mix of simple and complex sentence structures, but frequent grammatical errors weaken clarity. Errors such as 'it is a time that people should go to work' (should be 'it is the time when people go to work') and 'the hospital will be more pressure' (should be 'under more pressure') affect understanding. Some sentences are hard to follow due to word order and missing articles. For example, 'overcrowding is unsafe for these students' is unclear. Aim for subject-verb agreement and more careful sentence construction.

Vocabulary

Your vocabulary is mostly basic, and there is frequent repetition of words like 'overcrowding,' 'students,' and 'together.' Phrases like 'move together,' 'it is terrible for individual,' and 'students' passion may be decreased' are awkward or unnatural. Also, terms like 'unsafe' and 'come to some positions together' are confusing and detract from your message. Work on expanding your academic vocabulary and using more precise terms like 'compromised safety,' 'decline in student engagement,' or 'transmission of illness.' Avoid direct translations from your native language where possible.

Overall: With better control of grammar, clearer examples, and more natural vocabulary, your writing will become much more effective. You're on the right track with structure and word count — now focus on clarity and precision. Let me know if you'd like help revising a paragraph!

Q: What ideas/examples would be more relevant than the ones I used?

1. Public transportation – realistic, specific impact

Instead of: General statements like 'people have to squeeze into the crowded' or 'their mood may get worse.'

Use:

- A real-world example: *'In cities like Tokyo or New York, commuters often experience severe overcrowding on subways during rush hour, which leads to delays, increased commute times, and greater commuter stress.'*

- A measurable effect: *'Overcrowding can cause train delays due to longer boarding times, which disrupts the entire transit schedule.'*

2. Healthcare infrastructure – disease spread with clear impact

Instead of: 'It is easy for disease to spread ... the hospital will be more pressure.'

Use:

- Specific consequences: *'During the COVID-19 pandemic, densely populated urban areas saw hospitals overwhelmed due to the rapid spread of infection, which was worsened by crowded public spaces and inadequate infrastructure.'*
- Tie it to infrastructure: *'Overcrowding challenges the capacity of hospitals, emergency services, and sanitation systems.'*

3. Education infrastructure – more focused example

Instead of: 'Students must come to some positions together' or 'unsafe for these students.'

Use:

- Specific details: *'In countries like India and China, classroom sizes in urban schools often exceed 50 students, making it difficult for teachers to manage classrooms effectively.'*
- Highlight impact: *'This strain on educational infrastructure results in reduced individual attention, lower academic performance, and student burnout.'*

Summary

More appropriate examples are:

- Real cities and systems (Tokyo subways, NYC transit).
- Clear outcomes (e.g., delayed medical care, lowered test scores).
- Measurable impacts on infrastructure.

Let me know if you'd like help rewriting your body paragraphs using some of these examples!

Q: How could I improve my concluding sentence?

Here's how to improve it:

Clearer structure

Break the run-on sentence into two:

'When students are forced to participate in overcrowded activities for an extended time, they may gradually lose motivation. This lack of enthusiasm can seriously hinder their overall development.'

More academic tone

Use formal and precise language:

'Prolonged exposure to overcrowded learning environments can diminish students' engagement and enthusiasm, ultimately impeding their academic and personal growth.'

Emphasize consequence

Link the specific effect to a broader outcome:

'Over time, overcrowded conditions can lead to disengagement and burnout among students, which may negatively affect their long-term educational outcomes.'

Summary

A strong concluding sentence should:

- Clearly state the long-term consequence.
- Use precise vocabulary.
- Tie back to the topic (infrastructure's impact).

Q: Where could I have used reference words to reduce repetition?

Examples of unnecessary repetition and how to fix them:

1. Public transportation

Original:

'Public transportation will be overcrowded... people should go to work or school, lots of them will choose to take the public transportation such as bus or subway..'

Improved:

'Public transportation will be overcrowded ... lots of them will choose to take it, such as a bus or subway..'

2. Disease spread

Original:

'The spread of disease for long time will result in the public health burden, which means that the hospital will be more pressure.'

Improved:

*'This **situation** will increase the public health burden and put more pressure on **them**.'*

3. School overcrowding

Original:

'In some countries such as China enjoys lots of populations, which means there are plenty number of students to study in school.'

Improved:

*'In some countries such as China, **this** leads to a large number of students attending school...'*

4. Students in activities

Original:

'When it comes to some activities, they must all of them must come to some positions together.'

Improved:

*'For some activities, **they** must gather in the same place...'*

General tip:

After introducing a noun (e.g., public transportation, students, school), refer to it again using **'it,' 'they,' 'this,'** or **'these'** unless there's a need to restate it for clarity.

Would you like me to go through your full text and rewrite it with better use of reference words?

Appendix 2: Sample Model 2 – AI-enabled post-task reflection and feedback worksheet

This sample shows the first AI-enabled post-task worksheet used in Week 4 of the Pre-Master's course to support student engagement with GenAI feedback. The worksheet, formatted as a Word document, guides students through multiple stages: reflection, GenAI-generated feedback, student evaluation of that feedback, forward planning, and a GenAI appraisal of their plan with suggested next steps.

Green text in Sections 1 and 3 represents student input. Blue, bolded, and italicised text in Sections 2 and 4 is GenAI feedback generated by OpenAI's ChatGPT. The writing sample in Section 2 was completed by the student prior to this task. It was automatically retrieved from an attached Excel spreadsheet using an API key and inserted into the Word document to generate personalised feedback and facilitate structured reflection.

AI-Supported Self-Review Worksheet

Section 1: Planning Reflection (Complete after your writing)

Answer the following questions based on your planning process (before writing) in 4.3.

Complete this section carefully. Your planning responses will help the AI assess how your planning affected your writing performance and give targeted advice.

1. What two main impacts of climate change did you decide to focus on? Why?

Food insecurity and flood are the main consequences of climate change

2. How did you choose and organize the information from the sources?

I attempt to look at the big picture first, skimming the topic sentence and try to look for keywords, I tried to integrate the main issue to "health issue" first, and then the next category.

3. Did you plan any comparisons or contrasts between sources?

I tried to do that, but I didn't organize well.

4. What challenges did you face while planning your answer?

Organize too slow and lack of systematic method, maybe I have to change and learn new strategies. I find myself afraid to start answering if I don't completely inspect the whole architecture.

5. What could you improve in your planning next time?

I'll paste my questions and the problems I found to ChatGPT or inquire teacher.

Section 2: AI Feedback on Your Writing Sample

Your writing and AI feedback are shown below. Read the comments carefully.

Your Writing Sample:

Your Writing Sample:

A major impact of climate change on affected communities is the health system in general. Flood causes severe devastation on countries especially flood-prone areas. This would lead to various negative impacts on our health system. Among the negative impacts, food insecurity is mostly discussed issue, which are all mentioned in the text. According to all three sources, climate change has led to serious food insecurity issue. The Climate Change (2018) states that declines in crops yields are widely reported for Africa, with serious consequences for subsistence and rain-fed agriculture and food security. This indicates the severe level and specific impacted area. Furthermore, another crucial impact on health system is the spread of mosquito-borne Flaviviruses. The Medical Journal of Australia (2022) mentions that river flow and height were positively associated with Ross River virus disease spread by mosquitoes in the Murray river in eastern Australia. This is a solid evidence tells us that flood caused by climate change would take a step further to influence people's health. While WHO (2023) indicates that the poorest people would have to suffer more burden on both social and economic aspect.

Another primary impact caused by climate change is the placement and measurement on dealing with flooding houses. All three sources consistently aware that

Section 2 continues on the next page.

The following feedback has been generated based on your submitted short answer in 4.3. Review each area carefully.

- Content & Relevance:

The student makes an effort to address the impacts of climate change on health systems and housing through reference to the three sources. The response identifies key issues such as food insecurity, mosquito-borne diseases, and the burden on vulnerable populations. However, the content could be more relevant and clearly focused on the essay question. The second paragraph introduces the idea of housing impacts but is incomplete and lacks development. Additionally, while some source information is included, it is often presented in isolation rather than being clearly linked to the student's main argument. To improve, the student should more clearly connect each impact to the overall theme of how climate change affects communities, and ensure each point is developed with both explanation and integrated source evidence.

- Synthesis:

The student attempts to synthesize information from the three sources by discussing the impacts of climate change on health systems and flooding-related issues. However, the synthesis could be improved by providing a more cohesive and integrated analysis of the information from the sources. Ensure that the connections between the impacts discussed and the specific details from each source are clearly articulated to demonstrate a deeper understanding of the topic.

- Transformation (Paraphrasing and Referencing):

The student effectively paraphrases information from the source readings, such as discussing declines in crop yields in Africa and the

Section 2 continues with more GenAI feedback aligned to the assessment criteria for this task.

Section 3: Reflect on the Feedback

6. Which feedback point do you most agree with? Why?

I agree the most with the feedback from my “organization” part, because I also concern in the progress and after my writing.

7. What is one change you would make to your writing based on this feedback?

Learn to plan effectively and learn strategies for writing.

8. What is one strength you want to continue using in future writing?

I haven't realized one yet.

Section 4: Action Points for Future Writing

A. Reflection completeness:

The student's reflection is fairly complete as they address the feedback points provided and discuss their own thoughts and feelings about the feedback received.

B. Accuracy of self-assessment:

The student accurately assesses their own performance and acknowledges areas where they can improve, such as the need to focus on clarity and organization in their writing.

C. Priority alignment:

The student's priorities align well with the AI feedback in Section 2 and their own goals, as they recognize the importance of improving their writing skills for academic success.

D. Suggestions for improvement or practice activities:

To improve their planning process, the student could consider setting specific goals for each writing assignment, breaking down tasks into smaller steps, and creating a timeline to ensure they stay on track.

- Activity 1:

One practice activity the student could benefit from is outlining their basic ideas against the paragraph structure before attempting the synthesis matrix to home in on similar or contrasting ideas. It may make planning easier and ensure their arguments are well-structured and coherent.

- Activity 2:

Appendix 3: Course and intervention overview

The full course timelines for UEEC (top) and Pre-Master's (PMP) (bottom) are depicted in the following timeline, including both the original components and the additional feedback-related interventions introduced during the study.

- The **dark blue** sections represent the original course structure, which was delivered in full for both groups.
- All elements stemming from **orange** lines indicate additional interventions introduced as part of the research.
- Elements shown in **purple** were shared across both cohorts, while those in orange were unique to either UEEC or PMP.
- UEEC ends in Week 10 which includes the final IW assessment.
- Pre-Master's instruction ends in Week 12 with additional time for the final IW assessment (usually done within two weeks).

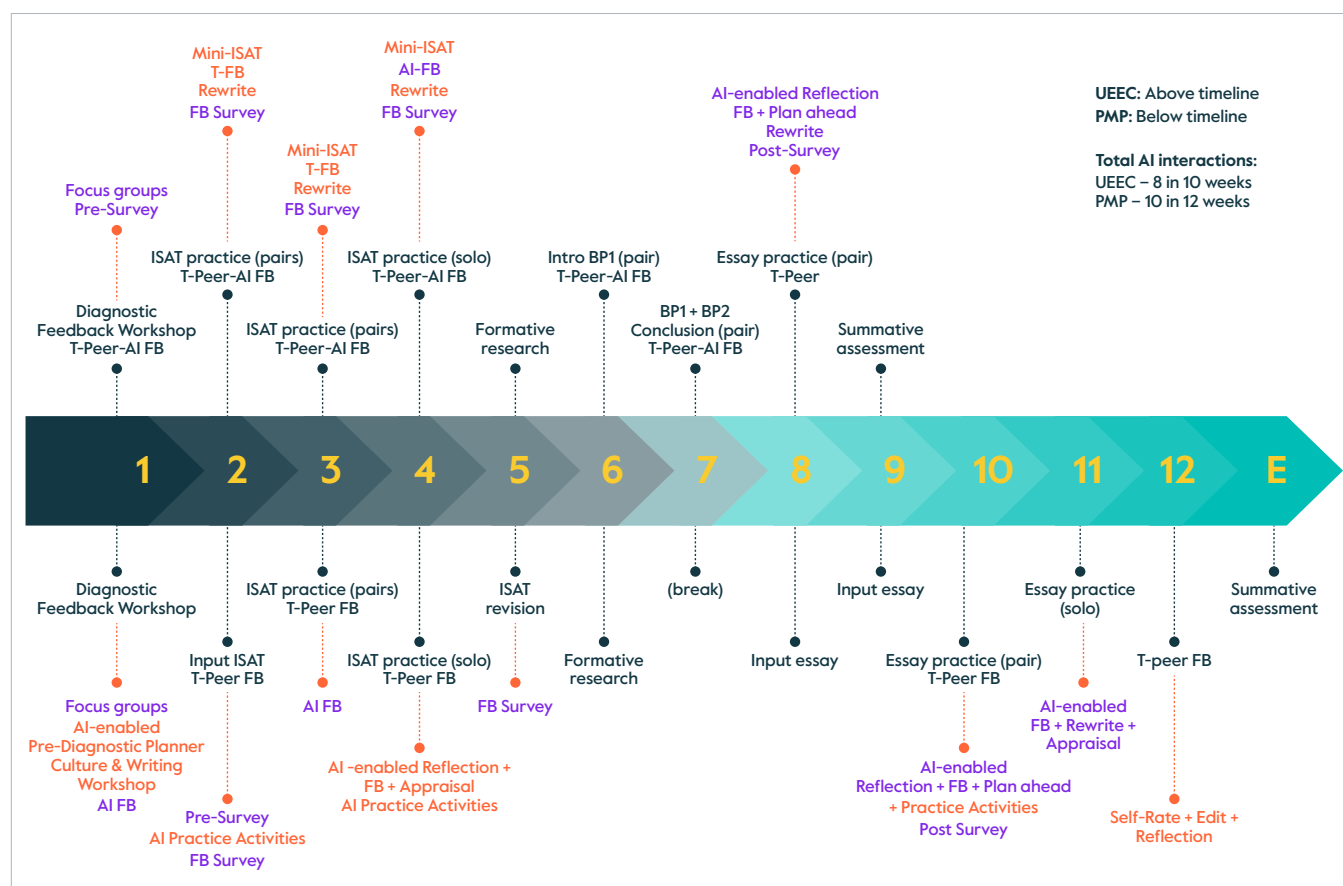
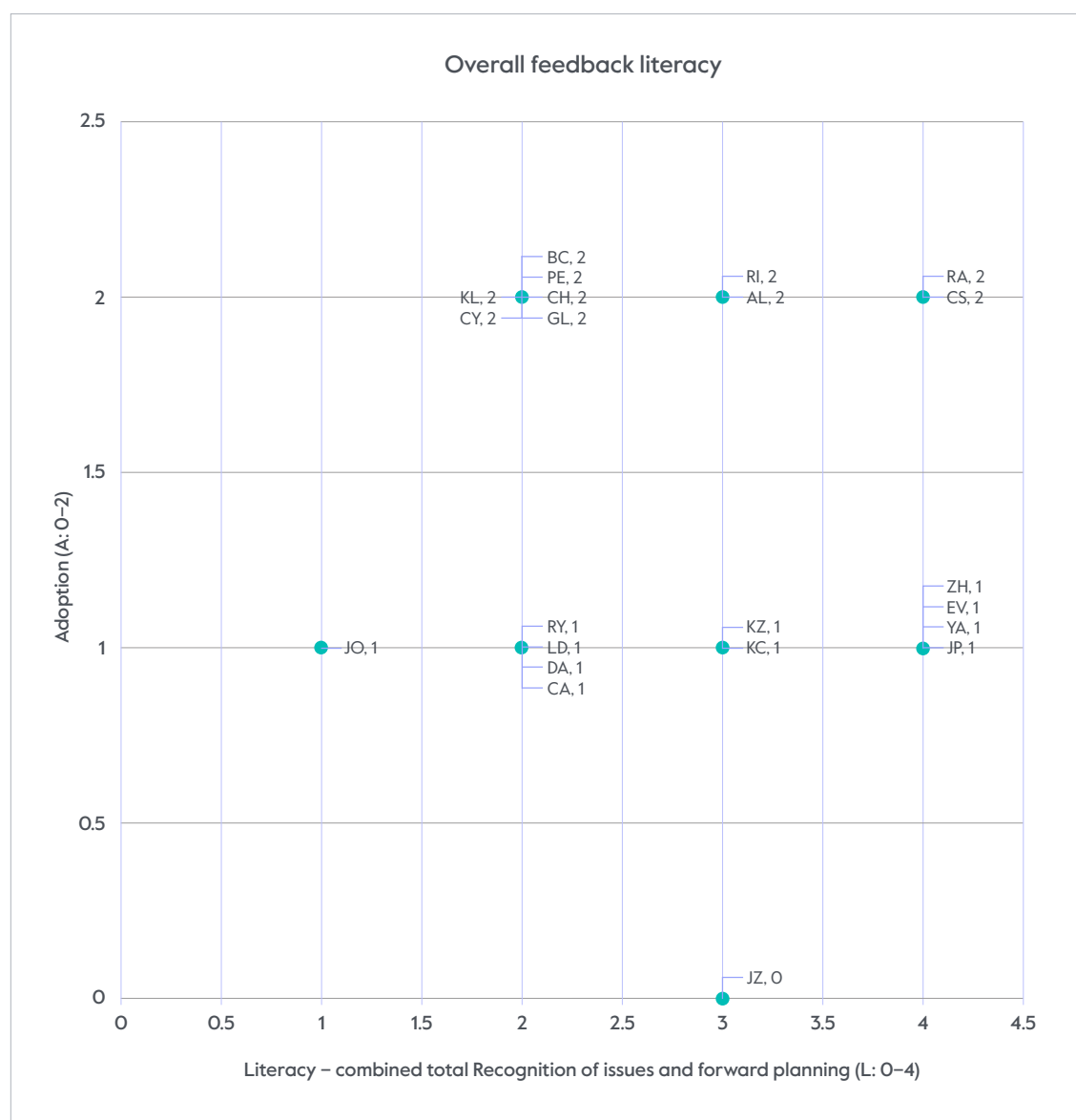


Figure 1: Original UEEC and Pre-Master's course structures and the action research interventions

Appendix 4: Feedback literacy rating system

Student engagement with feedback was evaluated across three dimensions: recognition of issues (self-awareness and alignment with feedback), forward planning (clarity and relevance of improvement strategies), and adoption of GenAI feedback (depth of revision based on AI suggestions). Recognition and planning were taken to be core components of feedback literacy, while adoption captured the extent to which students attempted to act on the feedback. Students were rated on a 0–2 scale in each area and grouped into three profiles: Guided Engager, Independent Explorer, and Partial Engager, based on their combined scores. The chart below visualises student placement across the literacy and adoption dimensions. Full scoring criteria and profile descriptions are provided on the next page:



Scoring criteria

Recognition of issues (0–2)

- 0 – No issues identified; self-appraisal is overly positive or vague; no alignment with AI or teacher feedback.
- 1 – Partial awareness; some overlap (33–65%) between self-identified issues and those noted by AI or teacher.
- 2 – Strong awareness; clear alignment (66%+) between self-appraisal and external feedback.

Forward planning (0–2)

- 0 – No specific plan; vague intention to ‘do better’ or general improvement goals.
- 1 – General strategies provided but not linked to flagged issues.
- 2 – Clear, actionable steps targeting specific issues raised in feedback and self-appraisal.

Adoption of GenAI feedback (0–2)

- 0 – Feedback ignored or minimal surface-level change.
- 1 – Surface-level edits only (grammar, phrasing).
- 2 – Deeper revisions showing structural or content changes aligned with feedback.

Profile descriptions

Guided engager (high engagement)

- Strong self-awareness, targeted planning, and meaningful revision.
- Typically scores: Recognition = 2, Adoption $\geq$ 1, Planning $\geq$ 1
- Engages consistently and uses feedback to drive improvement.

Independent explorer (moderate engagement)

- Shows initiative but with some inconsistency in planning or revision.
- Recognition = 1 or 2; Adoption or Planning = 1
- Engages with feedback but may not always follow through effectively.

Partial engager (low engagement)

- Limited recognition, vague or absent planning, minimal revision.
- Recognition = 0–1, or both Adoption and Planning = 0
- Feedback engagement is minimal or surface-level.
-

Engaging general English writers in ELICOS: A dialogic feedback model integrating AI

Jamina Coyne and Matthew Knowles, Navitas English, Perth

Introduction

In our experience, General English students often lack confidence when writing in English and struggle to recognise and develop the critical thinking skills needed to improve. Additionally, teachers may not allocate sufficient time to developing writing skills, and face challenges providing timely, effective feedback that could encourage student autonomy in writing and ownership over their voice. Based on Burns' (2010) notion that undertaking action research (AR) requires taking a 'problematism' stance to question aspects of teaching practice, we created a research goal: increase writing opportunities for students without adding to teachers' marking loads, and experiment with how AI could facilitate this.

Traditionally, writing assessment has been viewed as a burdensome teacher-centred practice. Lee (2009) highlights the drudgery of 'slaving' over the marking of students' writing. When students receive feedback, they may perceive this as a final and sometimes overwhelmingly lengthy transmission of judgement, leaving little opportunity for reflection or growth. Furthermore, students' academic investment in their writing produces a personal artefact which can carry emotional meaning, and well-intentioned feedback may be experienced negatively, leading to disillusionment and disengagement. Compounding this issue, teachers can be reluctant to provide honest or critical commentary, perhaps fearful of destabilising students' emerging confidence as writers. Ultimately, this can result in writing being underexplored in a General English context.

This singular focus on teacher feedback is not only a workload issue for teachers, but also ‘inadvertently relieves students of the responsibility to seek, engage with and use feedback’ (Nash and Winstone, 2017, as cited in Malecka et al 2022), depriving students of the opportunity to develop feedback literacy and the ability to improve their writing autonomously. We believe that the burden on teachers can be alleviated when students are directed to develop feedback literacy by assuming more control of the feedback process.

We believe feedback should be dynamic and dialogic. Dialogic pedagogy is a teaching approach drawing inspiration from Lev Vygotsky’s sociocultural theory of learning and Mikhail Bakhtin’s concept of dialogism (Teo 2019). Vygotsky proposed the social nature of learning through interaction, while Bakhtin’s dialogic approach involves a dialogue between student and teacher (Shirikhani et al 2015, Teo 2019). Dialogic approaches utilise techniques such as questioning and negotiation of opinions and ideas to facilitate students’ development of life skills including critical thinking (Teo 2019).

Generative Artificial Intelligence (GenAI) tools are transforming language learning by introducing an expeditious and responsive voice in feedback. GenAI can complement traditional forms of feedback, providing negotiable commentary that students can engage with for a more immediate and dynamic experience.

Research context

We are teachers of English Language Intensive Courses for Overseas Students (ELICOS) at Navitas English, a language school in Perth, Australia. Our AR was conducted with Advanced (CEFR B2–C2) General English classes of students from South America, Asia and Europe. General English is offered on a rolling enrolment basis, whereby students can join a class in any week, and enrol for as little as one week or longer, resulting in our student cohort varying slightly during our AR.

We conducted a needs analysis to determine our students’ perceptions of writing and using GenAI. From our needs analysis (see Appendix 1), most students reported low confidence in their ability to write well, expressing reservations relating to grammar and vocabulary knowledge. Encouragingly, almost all students reported a high degree of confidence in using AI and positivity toward its usefulness. More than 50% of respondents identified ChatGPT as the only AI tool they use, while the majority indicated they use ChatGPT and other AI tools. The students expressed a strong desire to use AI to improve their writing in a range of genres.

Students were introduced to the Cambridge Life Competencies Framework (Cambridge University Press & Assessment 2026). They participated in discussions around how these skills relate to cultural differences between an English-speaking context and their own cultures, and identified which competencies they felt they could acquire, expressing keenness to do so (Figure 1).

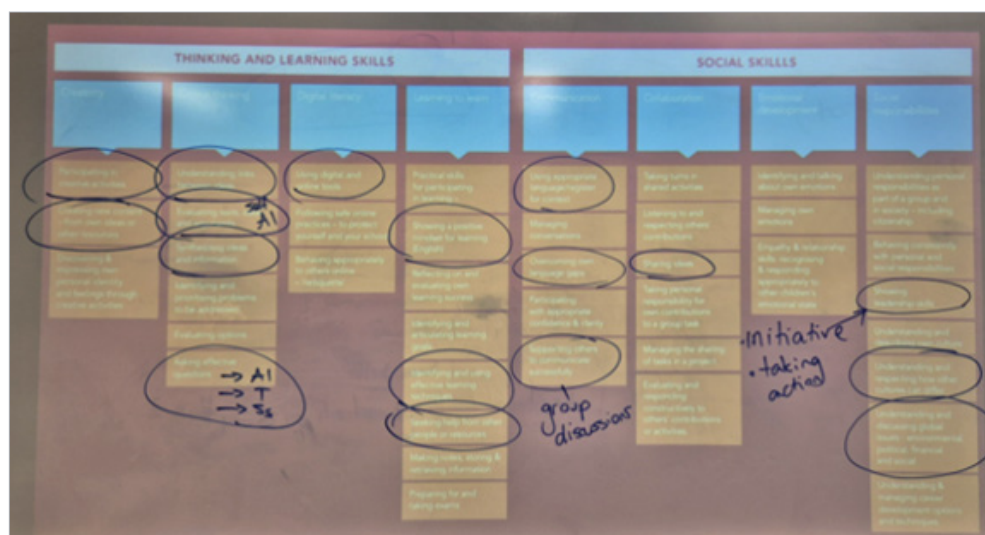


Figure 1: Student-identified goals from the Cambridge Life Competencies Framework

Research aims and questions

The scope of the project was to investigate how students would respond to concise AI feedback from Brisk AI and ChatGPT, using Brisk's 'glow, grow, wondering' feedback model, where 'glow' offers strengths, 'grow' suggests areas for improvement, and 'wondering' prompts critical thinking. We felt this model would provide succinct, personalised, and timely feedback.

The project aimed to determine how GenAI can enhance critical thinking and engagement in writing in ELICOS classrooms by investigating the following research questions (RQs):

1. How receptive are students to integrating AI feedback into their writing process?
2. In what ways does integrating AI into the writing process develop students' critical thinking skills from the Cambridge Life Competencies Framework?

Methodology

We planned eight writing workshops (Figure 2) including guidebook entry, proposal, opinion essays, cover letters, review, and email of complaint. Each workshop involved two stages conducted two days apart, enabling us to use Brisk to provide feedback between Drafts 1 and 2. The rationale for this approach was twofold: timely feedback meant students still felt attached to their writing and eager to redraft, and having rolling enrolment classes meant focusing on one genre each week was practical.

A key aspect of our methodology was asking students to use their mobile phones during the research project. Having observed our students' ongoing and increasing attachment to their phones, their preference for typing rather than writing texts, and their experience and affinity with using their devices to search for information, we considered this a rational approach. We discussed this with them, and our students were delighted to acquiesce.

Two days after the initial writing workshops, we returned the students' first drafts accompanied by Brisk feedback and asked them to critically evaluate Brisk's output, using their mobile phones and the AI of their choosing to scan their original texts and the feedback in preparation for the next writing stage. All students chose to use ChatGPT due to familiarity, with other students following suit.

We were clear with our students that they were not to use AI to simply correct their writing, but to engage critically and mindfully in the redrafting process using AI as a tool. In classroom discussions we acknowledged that AI can write anything they ask for, but they needed to build the skills to be able to evaluate whether the AI-generated output was fit for purpose. We discussed prompt engineering strategies that mitigated its use as a correction tool, providing them instead with a guiding framework to utilise ChatGPT as a personalised tutor which could provide critical thinking prompts, and detailed explanations and support for grammar, structural and lexical challenges from their first drafts, overtly linking these decisions to the Cambridge Life Competencies Framework. Students submitted their first draft and Brisk feedback into ChatGPT and worked on their second drafts independently, and we collected their drafts and prompts. Afterwards, the students completed a survey about their experience (see Appendix 2). The questions focused on how students used AI, their perceptions of feedback, their motivation, their confidence in prompt engineering, and their willingness to repeat the process.



Figure 2: Writing workshops

Findings

RQ1: How receptive are students to integrating AI feedback into their writing process?

Students expressed strong acceptance of AI feedback, but as part of an expanded feedback process involving the teacher rather than as a replacement for teacher feedback. Student responses indicated increased confidence, and satisfaction with feedback quantity, and to a lesser extent, quality. All student responses quoted below are in response to the survey questions given in Appendix 2. (Please note they are unedited to maintain authenticity.)

Feedback

Most students were happy with the quality of feedback. 45.8% of students were ‘very happy’ and an equal percentage were ‘somewhat happy.’ 72.9% of respondents indicated that the AI feedback amount was appropriate and accessible.

Confidence

Almost two thirds of the students reported feeling ‘somewhat confident’ in their ability to engineer prompts to obtain AI guidance (62.5% of respondents), with 31.3% feeling ‘very confident.’ One student stated that they may ‘need more practice to ask the right questions and obtain the right measure of information’ and ‘some parts of the text I think the AI don’t understood my intention.’

Some comments indicate that despite their confidence, crafting targeted prompts to elicit appropriate AI feedback posed challenges: ‘I felt overwhelmed with all that information, I would like to have less and work on it better, little by little.’

Some comments demonstrated critical engagement with the genre and the ability to evaluate and discern useful feedback from AI: ‘Sometimes AI gives inaccurate opinions’; ‘sometimes it misunderstood the principle purpose of the paragraph’; ‘Because sometimes I preferer my own criticism’; ‘Using prompt like asking the AI to give us glow, grow and wondering points is not detailed enough, is too general. It is better to ask directly for detailed feedback.’

Motivation

Students showed motivation to continue writing in English with AI support, with 33% very motivated and 50% somewhat motivated. Some respondents felt unmotivated or neutral (8.3% each), with one commenting ‘so far the AI and my teachers have given me similar or the exact same advice.’

Utility

Feedback on the process was positive, with 58.3% and 41.7% of respondents respectively finding it ‘very useful’ or ‘somewhat useful,’ centring on expanding vocabulary, generating ideas, and providing efficient feedback. While sentiment was positive, with 72.9% of students saying teacher and AI feedback were equally useful or beneficial in different ways, 8.3% found AI feedback far less useful.

Most students preferred feedback from both sources. While perspectives varied, recurring themes emerged. Two students commented on the 'different ways' AI and teacher feedback were given, one saying they 'complement' each other, while another indicated they favoured feedback from 'teacher and the IA' [sic]. One student acknowledged that the 'teacher is [too] busy' for marking and providing feedback.

Student feedback from the first two workshops led us to revise our approach for subsequent sessions. We reinstated teacher feedback to follow AI feedback, as students had requested. We also introduced peer feedback and self-reflection, creating a four-stage feedback protocol (Appendix 3). We were keen to incorporate peer feedback to extend the dialogic model to assess whether this yielded more insightful commentary, and to observe the students' adoption and integration of life competencies in the peer feedback they provided for each other. We noted that prior to participating in AR, students often felt they lacked authority or skills to evaluate other students' writing. After being introduced to Brisk, we observed most students instinctively adopted its feedback style, demonstrating this model's effectiveness, accessibility, and suitability for our context.

RQ2: In what ways does integrating AI into the writing process develop students' critical thinking skills from the Cambridge Life Competencies Framework?

Students demonstrated critical thinking in prompt engineering, draft revisions, and peer feedback (Appendix 3). Engagement remained high throughout the project. Incorporating AI into the feedback process encouraged students to employ several dimensions of critical thinking in the Cambridge Life Competencies Framework, including asking effective questions via prompt engineering and interaction with ChatGPT. For example, many students improved the appropriacy of their prompts in seeking support but not corrections. When drafting a travel blog in Week 1, two students created and used these prompts:

'Can you correct the text that I send you using the method glow, grow, wondering, and without [giving] me an answer?'

'I need that you correct my text, give me possible improvements.'

These prompts were contradictory and at cross purposes, which was potentially confusing for AI because they requested correction whilst also asking for guidance on how to improve. Such prompts often did not yield the responses that students had intended, with the AI rewriting the text and depriving the student of the opportunity to do so themselves. Prompt engineering improved markedly across the workshops, as demonstrated by the audience awareness in the following prompt:

'I need you to act like an English teacher. I will give you my writing practice, and I need you to give me feedback using the Brisk model with three bullet points each. Then, I need you to tell me if my vocabulary is in accordance with C1 level. At the end, I need you to write a corrected version.'

Although this student was still requesting a corrected version, his prompt clearly focused on support first and resulted in output that was closer to the student's predicted and desired AI output.

Students demonstrated the ability to formulate increasingly complex prompts that showed awareness of AI as an audience and tutor, and demonstrated prioritisation of key areas of concern when focusing their prompts and redrafting. In addition, our students were discerning in their evaluation of AI feedback, demonstrating that they could evaluate ideas. Students made independent decisions from their personalised feedback to improve their second drafts, prioritising the feedback they deemed most important. In one case, a student who was writing a job application letter responded to AI feedback with written comments:

AI suggested that the student expand one of their professional experiences:

'Could you elaborate on how you ensured inclusive representation across different community groups?'

The student determined that this feedback was a low priority:

'(x) That's too much.'

In another point, the AI asked the student to improve the essay structure:

'The text appears somewhat fragmented; work on creating smoother transitions between ideas.'

The student considered this feedback worth pursuing and asked for more guidance:

'(?) How?'

These responses demonstrate crucial critical thinking and evaluation of the appropriacy of the increasing use of AI output.

Students also showed increased engagement with the work of other students when asked to peer review. This development was evident in the quality of peer feedback that students produced. In previous attempts at peer review, it was often difficult to elicit constructive criticism from peer reviewers, with many students offering only general praise to other writers. Our experience of peer review in this process was markedly different. When invited to evaluate their peers' written texts, students provided both positive and constructive comments that demonstrated emerging assessment literacy, including attention to content relevance, genre conventions, and pragmatic appropriacy. Their feedback indicated growing metacognitive awareness of task requirements and awareness of audience. For example, one student commended a peer for maintaining clear and coherent content appropriate for correspondence with a bank, while also identifying the omission of essential contextual information, specifically the account number, and articulating the communicative rationale for its inclusion. This example illustrates our students' increasing ability to analyse writing in relation to real-world communicative purposes and target-reader needs. In peer feedback for a letter/email of complaint, another peer reviewer highlighted the importance of identifying yourself clearly when interacting with large companies:

'Your content is clear[ly], but you have to tell them your accou[n]t detail because they have a lots [of] customers.'

The writer recognised the validity of this feedback for this genre:

'After taking the criticisms into consideration, it has been brought to my attention that I have made simple errors, such as forgetting to mention personal details, and some information regarding the transaction dispute.'

Their feedback indicated growing metacognitive awareness of task requirements and awareness of audience. These examples were representative of the majority of our students' increasing competence and confidence in the use of AI both within each lesson and across the workshops.

Limitations

Due to rolling enrolments and the need to address a range of genres, we did not track students' progress over the 10-week period. Instead, we focused on observable behaviours indicating critical thinking, and more ambitious, expansive, and coherent second and third drafts. While running two workshops per week helped maintain engagement and allowed students to revisit their writing promptly, inconsistent attendance affected data collection. Additionally, workshop demands (prompts, surveys, and writing) proved intensive for some students, resulting in incomplete task submissions. Our reliance on mobile phones for AI access made monitoring usage challenging and required trust. Finally, students' strong expectations for teacher direction, especially during final drafts, and on corrected versions from AI, highlighted a persistent reliance on authoritative feedback for perceived high-stakes outcomes.

Discussion

Our AR began with a focus on using AI to provide feedback to students. The feedback students received from Brisk and ChatGPT was well-received.

Despite their positive perceptions of AI feedback, the students expressed a desire to receive accompanying teacher feedback, after which we identified the opportunity to incorporate a dialogic approach to subsequent workshops. In Workshop 3, we introduced teacher feedback after AI feedback. In the fourth and final workshops, we introduced peer feedback as an opportunity to observe critical thinking skills not only in student writing but in their appraisal of each other's.

Following our workshops, we saw marked difference in the willingness to provide feedback after students had had the opportunity to see AI and teacher feedback. The example of peer feedback demonstrates critical thinking, audience awareness, and genre analysis extending beyond the feedback provided by both the AI and the teacher (see Appendix 3). As a final step in this dialogic, interactive feedback process, we asked the original writer to provide a reflective comment (Appendix 3).

These comments similarly demonstrated a high level of engagement in the writing process and the employment of critical thinking to improve writing.

Conclusion

In the ELICOS sector, where teaching and developing writing skills can be neglected, feared or avoided, we sought to develop a protocol to inspire and assist teachers and students to embrace the writing learning journey. Our protocol is simple and timely to administer, free, and requires no additional tools to facilitate in class other than students' own mobile phones. It reduces teachers' marking load and ensures swift feedback to maintain interest and engagement in writing lessons.

Our research findings are encouragingly positive as a basis for further investigation into multi-voiced feedback. GenAI presents the opportunity for a new voice to join those of teachers and peers, expand the feedback dialogue, and provide the combined approach of integrating AI and teacher feedback for which our students advocated.

Reflections

Engagement in this project markedly enhanced our understanding of AR, particularly its iterative processes. The recursive, flexible, and deeply reflective nature of the approach supported meaningful personal and professional growth for both of us.

Initially, we planned to use Brisk AI as the sole tool for evaluating students' writing, with the intention of observing how its 'glow, grow, wondering' feedback model might foster greater learner autonomy, metacognitive awareness, self-reflection, and critical thinking. However, the cyclical nature of AR revealed the need to incorporate an explicitly dialogic quality. As the project unfolded, we identified and integrated multiple voices into the feedback processes. Our dialogic model evolved to include the perspectives of the writer, teacher, peer, and student-driven use of AI as a tutor, each shaping the feedback conversation and ultimately informing our protocol.

GenAI added a valuable voice to the formative feedback process. Recognising this, we felt it essential to retain the full complement of traditional stakeholders to ensure that student writers received the most support possible. This multi-voiced approach increased students' agency and autonomy in interpreting and acting on feedback.

We have since implemented our protocol in General English writing lessons, using Brisk AI to generate formative feedback within the same lesson. Students then extend this feedback through interaction with an additional AI tool of their choice, using prompt engineering strategies to elicit clarification, elaboration, or further guidance. Simultaneously, the teacher provides targeted comments on both the writing and the Brisk AI output, while peers engage with the student text and AI feedback to offer informed, responsive peer feedback.

This layered, dialogic approach has enriched our classroom writing practices and increased the frequency of in-class writing activities, thereby building students' skills

and confidence as English writers without adding to teachers' marking load, and strengthened students' capacity to engage critically and reflectively with feedback.

References

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

Cambridge University Press & Assessment (2026) *Cambridge Life Competencies Framework*. <https://www.cambridge.org/clcf>

Lee, I (2009) A new look at an old problem: How teachers can liberate themselves from the drudgery of marking student writing, *Prospect: An Australian Journal of Teaching/Teachers of English to Speakers of Other Languages (TESOL)* 24 (2), 34–41.

Malecka, B, Boud, D and Carless, D (2022) Eliciting, processing and enacting feedback: mechanisms for embedding student feedback literacy within the curriculum, *Teaching in Higher Education* 27 (7), 908–922. <https://doi.org/10.1080/13562517.2020.1754784>

Shirkhani, F, Jamali Nesari, A and Feilinezhad, N (2015) Bakhtinian dialogic concept in language learning process, *Procedia – Social and Behavioral Sciences* 205, 510–515. <https://doi.org/10.1016/j.sbspro.2015.09.060>

Teo, P (2019) Teaching for the 21st century: A case for dialogic pedagogy, *Learning, Culture and Social Interaction* 21, 170–178. <https://doi.org/10.1016/j.lcsi.2019.03.009>

Appendix 1: Needs analysis survey questions

1. How confident are you with writing in English (Very confident – confident – good – fair – poor)
2. Why?
3. Which of these writing genres have you written before? (cover letter, letter/email of complaint, essay, CV/resume, short story, LinkedIn profile, email, blog post, Instagram post, promotional items (business cards, posters), social media content)
4. For each item you selected, tell us more about your confidence and experience
5. What are the five most important genres? Give them a number. (1 = most important / 5 = least important)
6. Write three to five adjectives to describe how you feel about writing in English.
7. How often do you write in English? (every day – 2-3 times a week – weekly – monthly – less than monthly)
8. What kind of texts (genres) do you want or need to write in English?
9. Is writing in English important for your future? (yes – no – I don't know)
10. Why?
11. How confident are you with using Artificial Intelligence (AI) (In your language)?
12. Why?
13. Which AI do you have experience using?
14. For each AI, tell us about your confidence and experience.
15. Write three to five adjectives to describe how you feel about using AI.
16. How often do you use AI in English?
17. Is using AI important for your future? (yes – no)
18. Why?
19. Do you study in the daytime or evening class? (daytime – evening)
20. If you have any other questions, fears, concerns or ideas, please write them here.

Appendix 2: Post-workshop survey questions

1. Are you a daytime or evening student (daytime – evening)
2. Which AI software or app did you use today? (ChatGPT – Brisk – other – none)
3. If you chose 'other,' please tell us which AI you used.
4. Why did you use this AI?
5. What type of text did you write today?
6. How confident are you in writing AI prompts/instructions? (very confident – somewhat confident – unsure – not very confident – not at all confident)
7. How happy were you with the feedback you received from AI? (very happy – somewhat happy – neither happy nor unhappy – somewhat unhappy – very unhappy)
8. Why?
9. How useful do you feel this process was? (very useful – somewhat useful – unsure – not very useful – useless)
10. How useful do you feel this feedback is compared to teacher feedback?
(AI feedback is far more useful for me – AI feedback is slightly more useful for me – AI and teacher feedback are equally useful for me / useful in different ways – AI feedback is slightly less useful for me – AI feedback is far less useful for me)
11. Why?
12. The feedback I received was ... (far too long – a bit too long – the right length – a little too short – far too short)
13. Do you want to use AI again to improve your writing? (yes – no – maybe)
14. Why?
15. How motivated do you feel to write in English? (very motivated – somewhat motivated – neither motivated nor unmotivated – somewhat unmotivated – very unmotivated)
16. What would you change / do differently next time?

Appendix 3: Dialogic feedback model sample

Dear Sir,

My name is [redacted] and I am writing this email to bring a trans-
According to a transaction notification on my phone I was charged for a purchase that I did not make. As a result, I was unable to withdraw the funds I requested from my account. This money was to be used for necessary groceries, bills and to pay for my bus-
I have tried contacting your bank's customer service number but I was put on hold for hours that I have tried today
fare home. Since the bank is responsible for this mistake, I expect my account to be credited with the same amount as the transaction which is 250 dollars.

This is the first time a problem like this has happened to me but as I asked friends and family who also use your bank's services it appears that it is not the first time this has happened to other peoples accounts. The fact that this is happening even with computerized and online banking is inexcusable. So I would appreciate it if some explanation could be given as to why this happened.

Your cooperation is highly appreciated

Thank you

★ Glow: ★

- You clearly state the problem upfront, explaining the discrepancy in the Netflix subscription charge
- The tone is polite and professional, which is appropriate for a customer service complaint
- You've outlined specific actions you want taken (refund and correction to subscription)

🔍 Grow:

- Include specific details like the exact date of the overcharge and your account number to help customer service investigate faster
- Proofread for minor grammatical errors (e.g., "I await for the next steps" should be "I await the next steps")
- Consider adding a specific request for when you expect to hear back from them

💡 Wondering:

- Did you check your account settings to confirm no unauthorized changes were made?
- Have you compared this charge with previous months' statements?
- Would it be helpful to include screenshots or attach your bank statement as evidence?

Teacher feedback:

You outline the impacts on you very clearly. Include possessive apostrophe (people's) and check your use of "so" at the beginning of a sentence as a cause-and-effect linker. Remember to introduce yourself, e.g., "My name is... I am a customer..." Some good use of formal structures (e.g., was to be used). Excellent use of "Since" and "As a result". Check your use of "so" as the first word in the sentence. You outline the problem and make your requests clearly. Well done.

Peer feedback:

your content is very clearly but you have to tell them your account detail because they have a lots customers.

Peer feedback: His purpose is clear, he represses his ideas and his conflict, simply to understand. Some things are missing, like his name - I mean, basic personal information.

Reflection:

After taking the criticisms into consideration it has been brought to my attention that I have made simple errors such as: forgetting to mention personal details, some information regarding the transaction dispute.

I will remember to include my personal details and to pay attention to simple mistakes.

Using personalised AI feedback to support paraphrasing in EAP

Martin Sanchez, Torrens University Australia, Adelaide

Introduction

Paraphrasing is a key skill in EAP academic writing as it supports the integration of source texts and helps reduce inappropriate textual borrowing. However, many learners find paraphrasing challenging, particularly when selecting appropriate vocabulary, changing sentence structures, and preserving the original meaning (West 2024). When learners rely mainly on surface-level changes, such as deleting words, changing grammar, or substituting synonyms, their paraphrases may remain too close to the original wording and structure. Howard (1995) defines this as patchwriting and argues that, although it may reflect learners' attempts to understand unfamiliar material rather than deliberate dishonesty, it remains unsuitable for final academic writing because the text stays too dependent on the source.

This challenge has become more complex with the recent availability of generative AI (GenAI) tools as on-demand writing aids. These tools raise both concerns and possibilities in academic contexts. On one hand, scholars have expressed concern that GenAI tools may be misused to produce assignments that compromise academic integrity (Cotton et al 2023). On the other hand, when used strategically, GenAI tools such as ChatGPT can support L2 writing development by providing immediate and adaptive feedback on writing samples, helping learners identify language, content and organisation issues, and use feedback to revise their texts critically (Kurt and Kurt 2024). The use of GenAI feedback can also be understood through the lens of feedback literacy, which emphasises learners' capacity to make sense of feedback information, make judgements, manage affective responses, and take action to improve their work (Carless and Boud 2018). This suggests that the value of GenAI

in EAP writing depends less on the tool itself than on how learners are guided to interpret, evaluate and act on its feedback.

To use AI positively, educators need to develop learners’ AI literacy so that AI-generated feedback is used as a scaffold for reflection and revision, rather than as a substitute for students’ own thinking (Warschauer et al 2023). This action research (AR) project, conducted in two cycles, addressed this challenge by investigating how contextualised AI feedback may facilitate students’ paraphrasing development, while working towards improved accuracy, academic vocabulary, and learner agency in a 4-week EAP module.

In addition to paraphrasing outcomes, the intervention also aligned with the Cambridge Life Competencies Framework (Cambridge University Press & Assessment 2026), particularly the strands of learner autonomy, critical thinking and collaboration. These competencies were embedded through guided prompting, reflective decision-making and structured peer discussion, and were observed across both cycles of the project.

To summarise the main challenges, tensions, and opportunities identified in the literature, Table 1 synthesises key themes relevant to paraphrasing in EAP.

Table 1: Tensions and opportunities in EAP paraphrasing instruction with GenAI

Challenge	Tension	Opportunity
Limited academic vocabulary and sentence control	Learners must change wording and structure without distorting meaning (West 2024).	AI feedback can highlight language, content and organisation issues (Kurt and Kurt 2024).
Surface-level rewording	Synonym substitution or minor grammar changes can result in patchwriting (Howard 1995).	Comparing source text, student draft and AI feedback can support deeper revision (Carless and Boud 2018, Warschauer et al 2023).
Weak source use and citation	AI-generated text may compromise academic integrity if used uncritically (Cotton et al 2023).	AI literacy can guide ethical use of AI support in writing (Warschauer et al 2023).
Uncritical trust in feedback	Learners may accept feedback without judging its accuracy or relevance.	Feedback literacy helps students interpret feedback, make judgements and take action (Carless and Boud 2018).
Low learner agency	AI may replace effortful learning rather than scaffold revision.	Guided prompting and reflection can support autonomy, critical thinking and collaboration (Warschauer et al 2023, Cambridge University Press & Assessment 2026).

Research question

The purpose of my study was to explore the research question: How does personalised AI feedback improve EAP students’ paraphrasing accuracy, academic language use, and learner autonomy? Specifically, I wanted to discover whether using an AI tool would help students maintain source meaning more effectively,

increase their use of appropriate academic vocabulary, enhance their academic style, and support them in becoming more independent learners. Ultimately, my aim was to shift students' perceptions of AI from a 'shortcut' to a learning tool that helps them develop skills.

A complementary aim was to consider how the structured use of AI might also contribute to developing broader life competencies such as autonomy and critical thinking, as articulated in the Cambridge Life Competencies Framework (Cambridge University Press & Assessment 2026).

Research context, participants and procedures

The research was conducted at Torrens University Australia within its English for Academic Purposes (EAP) program of 5-week teaching blocks aimed at developing student academic success at undergraduate and postgraduate levels. An important aspect of the writing curriculum is paraphrasing, which is also one of the main modalities of assessment, comprising 50% of the final writing mark.

The study was conducted across two separate program blocks (Cycle 1 and Cycle 2) with two discrete groups of six international students, giving a total of 12 participants. All students met an entry requirement of at least IELTS 6.0 and were enrolled in either EAP1 (intermediate) or EAP2 (upper-intermediate). The cohort represented a mix of linguistic backgrounds, including Spanish, Portuguese, and Russian speakers. Students brought differing levels of prior experience with technology and AI tools.

Cycle 1 was fully delivered over the 5-week term as originally planned in a combination of teacher modelling, supported practice, and reflection tasks. After Cycle 1 had finished and the students had completed a short break, Cycle 2 was implemented with a new group of six students using a streamlined three-session version of the Cycle 1 plan. This refinement was informed by insights from Cycle 1, and the pre-intervention survey in Cycle 2 helped further tailor the content to learners' immediate needs.

Pseudonyms are used throughout to preserve anonymity. Only one student, Tatyane, participated in both cycles.

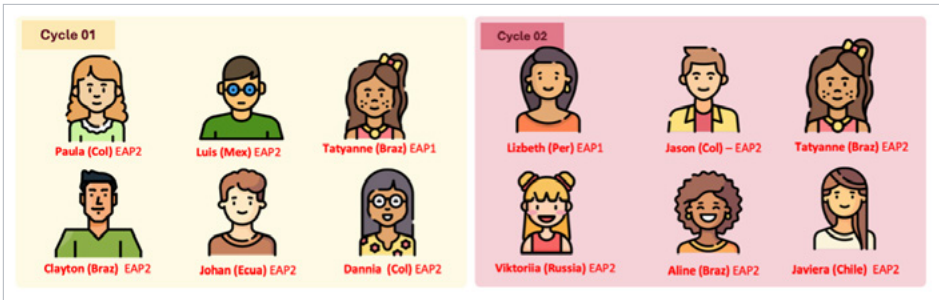


Figure 1: Student participants in both cycles

Research structure and data collection

To evaluate the intervention's impact, I generated a variety of qualitative and quantitative student data across both cycles including student writing, reflections, and AI interactions. The tools remained the same for both groups but there were changes in the use of these depending on each cycle's configuration. The next sections describe the intervention and my data collection for both Cycle 1 and Cycle 2.



Figure 2: Students in action at different stages of the paraphrasing process

Steps in the research

Cycle 1 involved three induction sessions (Week 1) for students to learn about paraphrasing strategies, AI literacy, and prompt construction; followed by four weekly paraphrasing tasks in Perplexity AI. The progressive prompt adjustments used in Cycle 1 are provided in Appendix 2, and the full Cycle 1 workshop sequence is shown in Appendix 3. Each weekly task required students to paraphrase selected excerpts that matched the week's theme (e.g., education, technology). A sample Perplexity-based paraphrasing task is included in Appendix 5, and Figure 3 illustrates one of the induction activities used to model paraphrasing strategies.

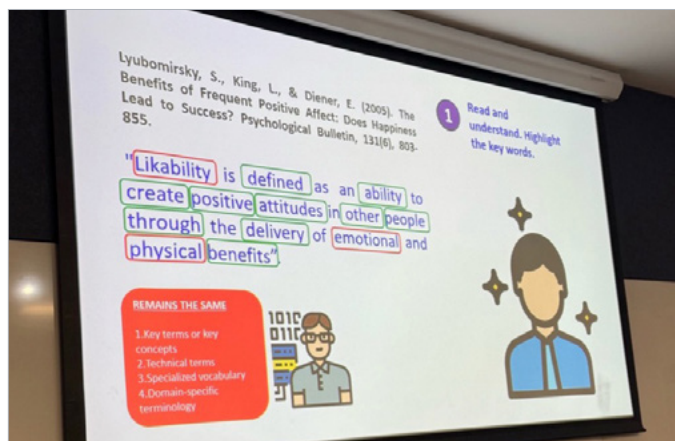


Figure 3: Induction session focused on paraphrasing skills development

For Cycle 2, the intervention was delivered to the second cohort using a custom GPT tool ("Paraphrase.ai"). The Cycle 2 workshop sequence is provided in Appendix 4. The intervention structure was simplified: after an induction session, students completed three paraphrasing tasks with support that aligned to the rubric offered via AI. A sample task completed through the custom GPT tool is included in Appendix 6.

While the intervention was condensed, all instruments for data collection were retained. A summary of the intervention sequence and refinements is shown below in Figure 4.

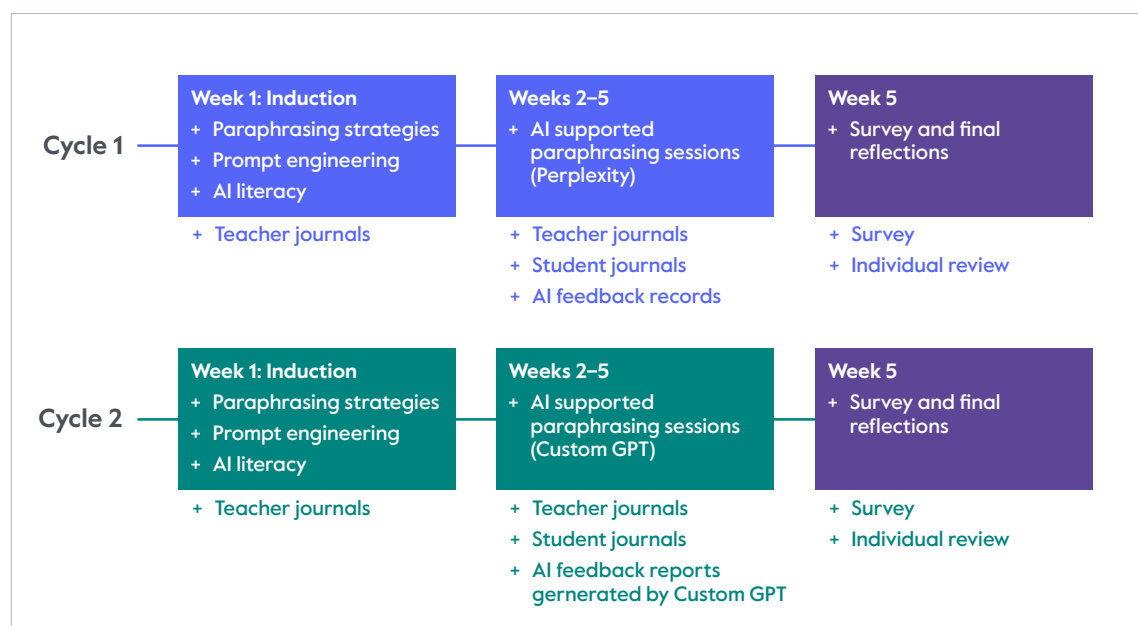


Figure 4: Summary of the two cycles involved

To strengthen validity, I triangulated the analysis of student outputs, teacher observations, and student reflections. The main data collection sources included:

- Paraphrasing drafts were assessed using a teacher-adapted analytic rubric informed by Liu and Lin's (2022) rating scale for paraphrase tasks (see Appendix 1)
 - Teacher and student journals for tracking strategies, challenges, and engagement
 - AI feedback logs and prompts (Cycle 1), or full reports (Cycle 2)
 - Post-cycle surveys and subsequent interviews regarding student perspectives.
- The survey instrument used to collect student perceptions is provided in Appendix 7

This multi-layered approach provided a check and balance not just on performance outcomes (e.g., retention of meaning and use of academic language) but also affective gains (e.g., confidence and autonomy). The interval between cycles also supported reflective and responsive design improvements.

Findings

Paraphrasing accuracy

Across both research cycles, students' paraphrasing abilities improved in meaning and accuracy. Analytic rubric results, comparing students' initial and revised versions of a paraphrase, showed reliable improvement after students worked with an artificial intelligence (AI) support.

For Cycle 1, students improved meaning retention with an average score of +1.21 meaning (max score = 0-4), and +1.54 for improved accuracy. Those strong improvements indicated that together, Perplexity AI and scaffolded input by the teacher helped students better preserve original ideas in their paraphrases while also improving language accuracy and fluency.

For Cycle 2, a more streamlined approach to paraphrasing was offered, using a custom GPT tool, and simplified overall digital workflow. This time, students continued to demonstrate gains on both meaning (average of +0.93) and accuracy (average of +1.14) paraphrases. Improvements were slightly lower in Cycle 2, compared to Cycle 1, yet students successfully progressed with less direct educator involvement, which is very relevant to individual student autonomy and overall use of the interface.

Overall evidence suggests that Cycle 1 resulted in greater score increases. However, the custom GPT used in Cycle 2 provided better structured feedback, and there was good evidence of learning impacting both meaning retention and academic language accuracy. These results reflect the project's dual focus: improving language outcomes and learner independence (see Figure 5).

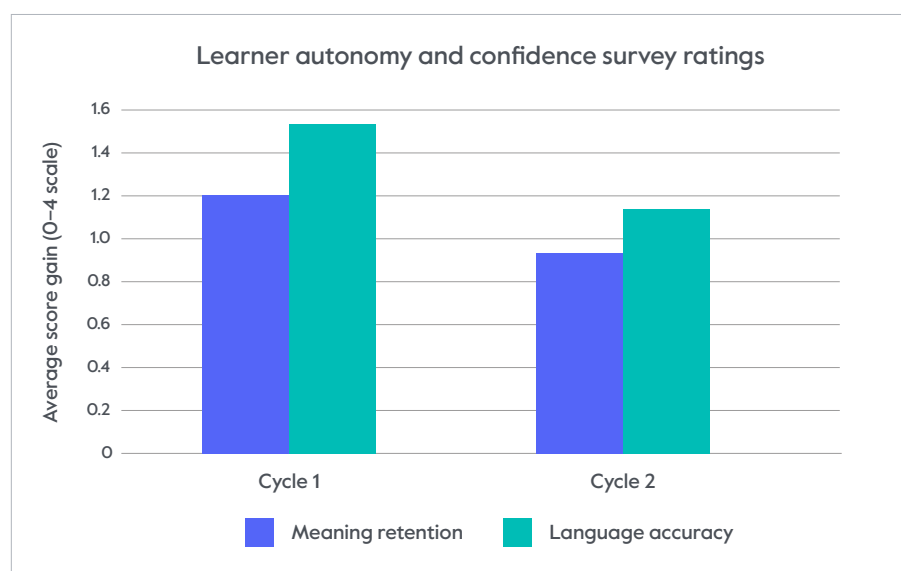


Figure 5: Average improvement in paraphrasing skills (meaning and accuracy)

Evidence collected qualitatively from classroom observations and student reflections corroborated these quantitative improvements. Specifically, many students, at the beginning of the intervention, ‘patch wrote’ by merely changing words. By the end of the intervention, however, they were making confident decisions to change the sentence structure and provided their own wording without losing meaning.

Several learners noted that the AI feedback allowed them to understand how to change sentences but not change ideas. ‘Feedback from the AI reinforced my understanding of structure and the meaning changes,’ shared one learner, giving an example of a deeper understanding of paraphrasing. The students even began to internalise strategies for paraphrasing by Cycle 2. In fact, by Session 3, students knew exactly which paraphrasing strategies to apply and the process was corroborated by the AI.

Academic language use

The intervention resulted in observable improvements in students' academic vocabulary and academic language, despite the fact vocabulary instruction was not a direct intention of the intervention. Lexical analysis using Lextutor software shows growth in four areas between both cycles with respect to the final drafts and first drafts: increased Academic Word List (AWL) use (Coxhead 2000), decreased vocabulary from the off-list, and fewer grammatical errors and collocations (see Figure 6).

The growth in AWL usage consistently increased during both cycles. In Cycle 1, students had an average AWL gain of +1.26%, while in Cycle 2 there was an average gain of +3.15%, suggesting better internalisation of academic lexis after the structured AI-supported intervention. In some cases, there were dramatic individual gains of as much as +9%, such as from 14% to 23%.

The usage of off-list vocabulary (informal or vague vocabulary) decreased in both cycles. In Cycle 1, off-list vocabulary fell from 10.11% to 7.53%, and in Cycle 2 from 9.98% to 6.83%, demonstrating a noticeable shift toward a more formal academic register.

Errors related to grammar and collocation also dropped noticeably. In Cycle 1, grammar errors decreased 7.53 and collocation errors decreased 4.25 per 100 words. In Cycle 2, grammar errors decreased even more steeply, falling from 12.85 to 2.86 per 100 words, and collocation errors fell from 4.63 to 0.93 per 100 words. These reductions indicate that learners demonstrated greatly improved control of sentence structure and precision with word choice because of AI coaching.

These findings suggest support for students' improvements in more formal and accurate language patterns. Some learners explicitly identified vocabulary growth as beneficial. For example, one learner stated, 'The best thing was finding the right synonyms for the context' and another said, 'Actually I feel my words choice better now.'

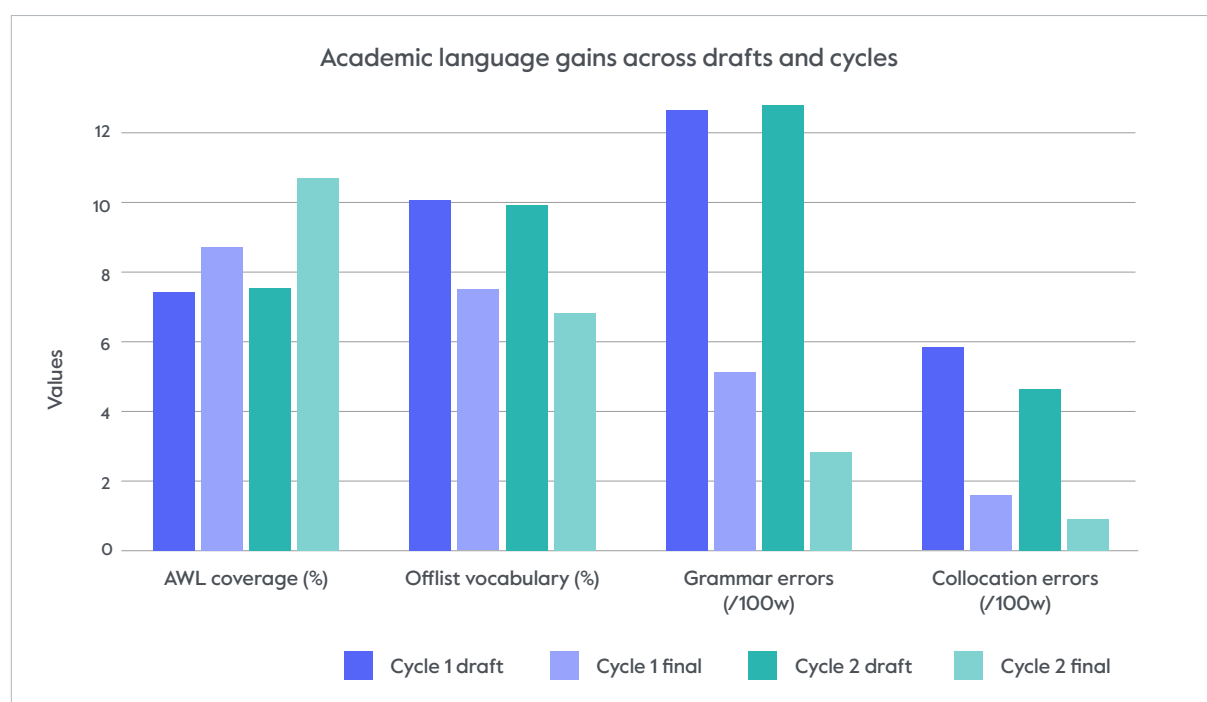


Figure 6: Comparison of AWL %, off-list %, grammar errors, and collocation errors in drafts vs. final paraphrases across both cycles

Learner autonomy and confidence

The implementation of tailored feedback using AI resulted in definite improvements in the students' sense of autonomy and confidence. In Cycle 1, some participants were initially hesitant about the use of AI tools, and one student stated, 'I don't like ChatGPT and ChatGPT doesn't like me.' However, by the end of the cycle, this student agreed she could see AI as an appropriate learning tool. Most participants reported feeling more confident as paraphrasers and expressed intentions to continue using AI to support their future studies. As one student reflected, 'Now I know how to use AI to help me learn, rather than just telling me the answers. I feel it is better for me.' These shifts closely reflect the Cambridge Life Competencies Framework emphasis on helping learners make independent, well-reasoned choices when working with digital tools.

In Cycle 2, learners exhibited increased self-direction. All students reported feeling capable of using AI independently, and many exhibited an evolution from passive to active engagement. Being in the class, many students chose to paraphrase the writing samples manually and then consult AI, using it to enhance or polish their ideas instead of using it to generate ideas. In addition, they began to show greater ability to write prompts, meaning they were becoming more literate in using AI and interpreting its feedback.

The survey ratings summarised from both cycles reflect this progress (see Figure 7). In Cycle 1, average results in the 'academic integrity & ethics,' 'prompt engineering,' and 'usefulness' categories exhibited a score of 4.33 to 4.67 (1–5 scale).

During Cycle 2, the scores did not change much and showed a high level of comments related to 'trust' and 'more independence.' One student even compared the process to 'solving a puzzle,' and stated they enjoyed it, showing this was an effective part of the activity for inducing motivation. Another student said, 'I used to use Quillbot for paraphrasing my citations, but this is much better for me,' showing a very mature understanding of the role of AI in supporting learning.

The transition from tentative experimentation to assured ethical involvement suggests the effects of the tool on mindset and skill development. The AI fostered independence (not dependence) as ideally it became a bridge to more meaningful learning, and not a crutch.

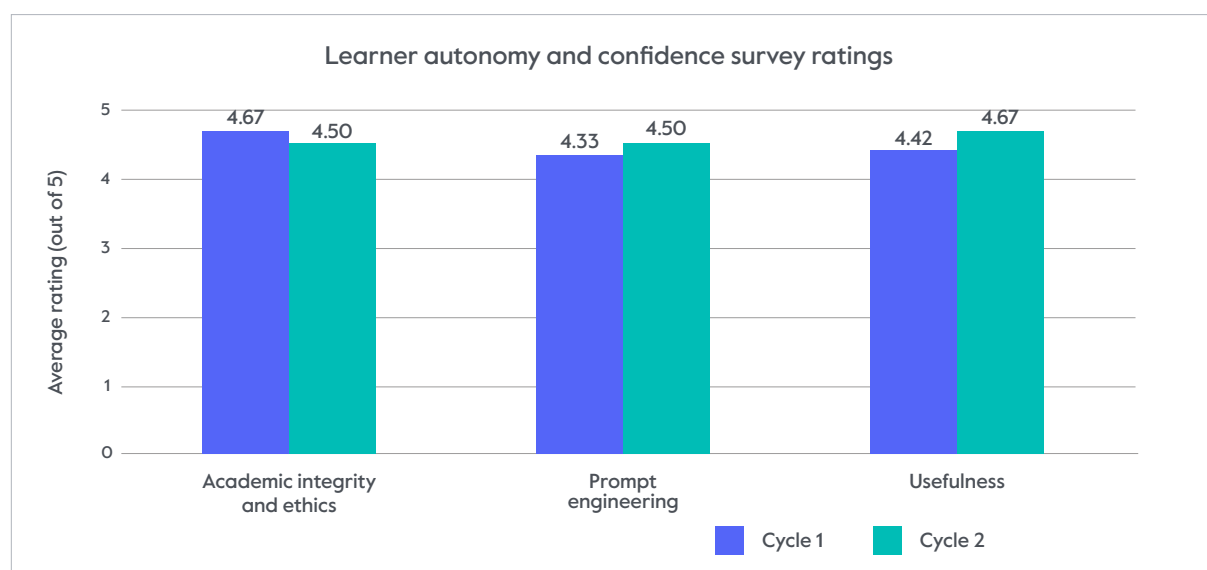


Figure 7: Comparison of student perceptions of key AI literacy dimensions in Cycle 1 and Cycle 2

Conclusion

Throughout this project, I experienced a subtle but impactful change, which was not only in relation to how students paraphrased, but the act of learning itself. Before Cycle 1, students often leaned heavily on AI to rewrite for them, mostly using vague prompts such as, ‘Paraphrase this,’ and without much consideration or taking for granted what AI would generate. However, as we moved through the weeks, something changed. They began to push back against the feedback, refine their drafts, and ask productive and more specific questions of the process. One student’s reflection did a wonderful job of capturing this change, when Perplexity proposed a change that looked exactly like the original text: ‘Artificial intelligence is not that intelligent after all.’ This prompted him to interact more reflectively with this technology.

The move toward autonomy was not a coincidence. The intentional sequencing of strategies, the simplification of tools, and the ongoing need for academic integrity created an environment in which students felt supported and responsible. Students began to anticipate AI’s suggestions and could self-correct prior to prompting. ‘I knew exactly what Chat GPT would suggest doing,’ said one student out loud during a revision session.

As an educator, I shifted from an instructing role to a more facilitative one, observing students gradually take greater ownership of their strategies and engage more critically with AI. This experience reinforced that, when used with thoughtful planning and clear pedagogical purpose, technology can effectively support complex academic skills such as paraphrasing. Rather than simply using AI tools, students can be guided to develop as reflective learners who make informed decisions in digital environments.



Figure 8: Participants of the paraphrasing workshops and me celebrating the end of the process

References

- Cambridge University Press & Assessment (2026) *Cambridge Life Competencies Framework*. www.cambridge.org/clcf
- Carless, D and Boud, D (2018) The development of student feedback literacy: Enabling uptake of feedback, *Assessment & Evaluation in Higher Education* 43 (8), 1,315–1,325. <https://doi.org/10.1080/02602938.2018.1463354>
- Cotton, D R E, Cotton, P A and Shipway, J R (2023) Chatting and cheating: Ensuring academic integrity in the era of ChatGPT, *Innovations in Education and Teaching International* 61 (2), 228–239. <https://doi.org/10.1080/14703297.2023.2190148>
- Coxhead, A (2000) *Academic Word List*. www.eapfoundation.com/vocab/academic/awllists
- Howard, R M (1995) Plagiarisms, authorships, and the academic death penalty, *College English* 57 (7), 788–806. <https://surface.syr.edu/cgi/viewcontent.cgi?article=1002&context=wp>
- Kurt, G and Kurt, Y (2024) Enhancing L2 writing skills: ChatGPT as an automated feedback tool, *Journal of Information Technology Education: Research* 23, Article 24. <https://doi.org/10.28945/5370>
- Liu, S and Lin, D (2022) Developing and validating an analytic rating scale for a paraphrase task, *Assessing Writing* 53, 100646. <https://doi.org/10.1016/j.asw.2022.100646>
- Warschauer, M, Tseng, W, Yim, S, Webster, T, Jacob, S, Du, Q and Tate, T (2023) The affordances and contradictions of AI-generated text for second language writers, *Journal of Second Language Writing* 62, Article 101071. <https://doi.org/10.1016/j.jslw.2023.101071>
- West, A J (2024) Improving the English-language academic paraphrasing guidelines taught at a Thai university, *THAITESOL Journal* 37 (2), 25–49. <https://files.eric.ed.gov/fulltext/EJ1460072.pdf>

Appendix 1: Paraphrasing rubric – Informed by Liu and Lin’s (2022) analytic rating scale

1. Meaning Consistency

Does the paraphrased version preserve the original meaning?/

Level	Points	Descriptor
Level 5	4 points – Excellent	All core ideas are fully and accurately conveyed.
Level 4	3 points – Good	Minor omissions or small meaning shifts, but the overall message is intact.
Level 3	2 points – Adequate	Some ideas are preserved, but meaning is partially unclear or vague.
Level 2	1 point – Weak	Only fragments of meaning retained; major distortions.
Level 1	0 points – Poor	Off-topic, incorrect, or no attempt made.

2. Accuracy

How grammatically correct and natural is the sentence?

Level	Points	Descriptor
Level 5	4 points – Excellent	Sentence is fully grammatical, fluent, and idiomatic.
Level 4	3 points – Good	Some minor errors (e.g., article misuse, verb tense) that do not affect clarity.
Level 3	2 points – Adequate	Multiple issues in grammar, structure, or collocation. Slightly affects clarity.
Level 2	1 point – Weak	Serious grammar or word <u>use</u> problems; meaning is hard to follow.
Level 1	0 points – Poor	Sentence is incoherent, broken, or blank.

Appendix 2: Progressive prompt adjustment for Cycle 1

1



MARTIN SANCHEZ ▶ 1 student
Mar 26



Dear all, let's use this prompt for the AI feedback section:

Compare the original text below with the paraphrased version provided. Evaluate the paraphrased text in terms of how well it demonstrates comprehension of the original text. Consider the following aspects:

Accuracy: Does the paraphrase capture the main ideas and key details of the original text?

Clarity: Is the paraphrased version easy to understand?

Originality: Does the paraphrase use different words and sentence structures from the original while maintaining the same meaning?

Provide a brief assessment of the paraphrasing quality without rewriting the text. If there are any significant misunderstandings or omissions, point them out. Also, briefly provide some synonyms or other terms that might work better.

Original Text:

[Paste text here]

Paraphrased Text:

[Paste text here]

2



MARTIN SANCHEZ
Mar 28



PARAPHRASING WORKSHOP 01-02

PROMPT: Compare the original text below with the paraphrased version provided. Evaluate the paraphrased text in terms of how well it demonstrates comprehension of the original text. Consider the following aspects:

Accuracy: Does the paraphrase capture the main ideas and key details of the original text?

Clarity: Is the paraphrased version easy to understand?

Originality: Does the paraphrase use different words and sentence structures from the original while maintaining the same meaning?

Sentence structure: Is the construction logical? Are the grammatical features used properly?

Provide a brief assessment of the paraphrasing quality without rewriting the text. If there are any significant misunderstandings or omissions, point them out. Also, briefly provide some synonyms or other terms that might work better.

Original Text:

[Paste text here]

Paraphrased Text:

[Paste text here]

3



MARTIN SANCHEZ

Apr 4

**PROMPT TO BE USED TODAY:**

Compare the original text below with the paraphrased version provided. Evaluate the paraphrased text in terms of how well it demonstrates comprehension of the original text. Consider the following aspects:

Accuracy: Does the paraphrase capture the main ideas and key details of the original text? If not, provide clear examples of how to fix this. It must try to fix my proposal, not create a totally new one.

Clarity: Is the paraphrased version easy to understand? If not, provide clear examples of how to fix this. It must try to fix my proposal, not create a totally new one.

Originality: Does the paraphrase use different words and sentence structures from the original while maintaining the same meaning? If not, provide clear examples of how to fix this. It must try to fix my proposal, not create a totally new one.

Sentence structure: Is the construction logical? Are the grammatical features used properly? Be concise in your observations and provide examples of how to address the problems identified (in case there are any).

Provide a brief assessment of the paraphrasing quality without rewriting the text. If there are any significant misunderstandings or omissions, point them out. Also, briefly provide some synonyms or other terms that might work better. Be assertive on your feedback as it must try to encourage learners in the process of paraphrasing.

Original Text:

[Artificial intelligence is increasingly recognized as a transformative tool in business, offering firms the ability to gain a sustainable competitive edge. By leveraging AI technologies, businesses can enhance operational efficiency, drive innovation, and create value through improved decision-making processes and customer interactions.]

Paraphrased Text:

[Enter the paraphrased text here]

CYCLE 01

Human input

PARAPHRASING WORKSHOP WITH 02-SESSION 02

Let's discuss:

Synonyms analysis:

Term	Proposal
refers to	is defined as
enable	facilitate
exchange	transfer
enhance	improve
efficiency	productivity
seamlessly	easily
collect	

Draft

Ashton, K. (2009). That 'internet of things' thing. RFID Journal, 22(7), 97-114.

"The Internet of Things refers to a network of interconnected physical objects equipped with sensors and software that enable them to collect and exchange data autonomously over a network to improve productivity and simplify tasks."

The Internet of Things... is defined as a network of interconnected physical objects equipped with sensors and software that enable them to collect and exchange data autonomously over a network to improve productivity and simplify tasks.



Appendix 4: Workshop sequence for Cycle 2

CYCLE 02



Human input

Test for Paraphrasing

Artificial intelligence (AI) is reshaping education by offering tools that personalise learning and support student progress. Unlike traditional methods, AI can adapt to individual needs, helping learners overcome specific challenges. As noted in recent research, "AI allows teachers to detect learning gaps and adjust materials in real time, which leads to more efficient instruction and improved academic performance" (Lee, 2021, p. 102). This document shows how AI can enhance educational outcomes while maintaining a supportive role for educators.

Task:

Artificial intelligence (AI) is reshaping education by offering tools that personalise learning and support student progress. Unlike traditional methods, AI can adapt to individual needs, helping learners overcome specific challenges. As noted in recent research, "AI allows teachers to detect learning gaps and adjust materials in real time, which leads to more efficient instruction and improved academic performance" (Lee, 2021, p. 102). This document shows how AI can enhance educational outcomes while maintaining a supportive role for educators.

Draft directly typed on an online document.

AI support



Step 2: Paraphrasing Strategy Hints

Let's consider a few strategies you might use:

Strategy	Hint	Example
Synonyms (S)	Change "detect" to "identify," "adjust" to "modify," "real time" to "instantly," etc.	"recognize learning difficulties"
Nominalisation (N)	Change verbs to nouns (e.g. "adjust materials") = "adjustment of materials")	"the identification of gaps..."
Passive Voice / Reordering (R)	"Instruction becomes more effective when learning gaps are detected..."	"Efficient teaching is achieved when..."
Citation Integration (C)	Use narrative style: Lee (2021) states that... or According to Lee (2021)...	"Lee (2021) observes that AI helps..."

Let me know once you've written your paraphrased version based on these hints, and I'll give you feedback. Here's the citation again for your reference:

Lee (2021), observes that AI helps teachers recognize learning difficulties and facilitates the adjustment of materials instantly. This leads to improved academic performance and more effective instruction.

Lee (2021), observes that AI helps teachers recognize learning difficulties and facilitates the adjustment of materials instantly. This leads to improved academic performance and more effective instruction.

What You Did Well:

- Task Achievement:** You've preserved the original meaning very accurately.
- Lexical Resource:** Nice use of synonyms like *recognize* for *detect* and *facilitates* for *allows*.
- Nominalisation:** "the adjustment of materials" is a good example.
- Citation Integration:** You correctly used the narrative form "Lee (2021) observes..."

Suggestions for Improvement (Just a Few!):

- Word Order / Flow:** You might slightly improve fluency by adjusting the order; place "more effective instruction" before "improved academic performance" to better mirror cause → effect logic.
- Advanced Academic Lexis:** You could use more academic phrases like *instructional efficiency* or *pedagogical adaptation* if you want to elevate the tone (optional for higher-level work).
- Clarity in Second Sentence:** Consider combining both ideas into one sentence for cohesion.

Would you like to revise your version using these tips? Once you do, we'll check it again and then move to the final report!

CYCLE 02

Suggestions for Improvement (Just a Few!):

- Word Order / Flow:** You might slightly improve fluency by adjusting the order; place "more effective instruction" before "improved academic performance" to better mirror cause → effect logic.
- Advanced Academic Lexis:** You could use more academic phrases like *instructional efficiency* or *pedagogical adaptation* if you want to elevate the tone (optional for higher-level work).
- Clarity in Second Sentence:** Consider combining both ideas into one sentence for cohesion.

Would you like to revise your version using these tips? Once you do, we'll check it again and then move to the final report!

Fantastic work, Taty! 🌟

Your final version is well-structured, academically appropriate, and effectively paraphrases the original citation. Now let's generate your full report with strategy breakdown and feedback:

Final Report

- Student and Task Information**
 - Date: 2025-08-01
 - Student Name: Taty
 - Task Name: AI & Education – Week 1 Paraphrasing Practice
- Final Paraphrased Version**

Lee (2021) observes that AI helps teachers recognize learning difficulties and facilitates the adjustment of materials instantly. This leads to greater instructional efficiency and enhances academic performance.
- Paraphrasing Strategies Applied**
 - Synonym use:** e.g., *recognize* → from *detect*, *facilitates* → from *allows*
 - Nominalisation:** *adjustment of materials* instead of *adjust materials*
 - Reordering:** Restructured to end with results → clarity
 - Grammar change:** Added stronger verb form "enhances"
 - Citation integration (narrative form):** "Lee (2021) observes that..."

4. Vocabulary Learned

Term	Definition / Synonyms
Recognize	Identify, detect
Facilitates	Makes easier, enables
Instructional efficiency	Effectiveness of teaching
Enhances	Improves, strengthens

5. Feedback Summary

- Task Achievement:** The paraphrase clearly conveys the original meaning without loss of detail.
- Lexical Resource:** Excellent academic vocabulary and precise synonyms used.
- Grammatical Range and Accuracy:** Structures are grammatically correct and appropriate.
- Coherence and Cohesion:** Logical flow is smooth and well-connected.

Lee (2021), observes that AI helps teachers recognize learning difficulties and facilitates the adjustment of materials instantly. This leads to instructional efficiency and academic performance.

Lee (2021), observes that AI helps teachers recognize learning difficulties and facilitates the adjustment of materials instantly. This leads to greater instructional efficiency and enhances academic performance.

Appendix 5: Sample paraphrasing task (Perplexity)

Activation of schemata

Let's discuss:

1. In what ways do you think technology helps people in their daily lives, such as at work, at home, or in education?
2. Have you ever heard about the Internet of Things (IoT)? If so, what do you think it means or involves?
3. Can you think of any examples of how technology connects devices to make life easier or more efficient?

Task and synonyms and key words recognition

Ashton, K. (2009). That 'internet of things' thing. *RFID Journal*, 22(7), 97–114.

"The Internet of Things refers to a network of interconnected physical objects equipped with sensors and software that enable them to collect and exchange data autonomously over a network. It aims to seamlessly integrate digital and physical environments to enhance efficiency and simplify tasks."

seeks

Synonyms analysis:

Term	Proposal
refers to	relate to
interconnected	linked
physical objects	material things
exchange	interchange
It aims	pretends

Draft

Draft

The internet of things relates to the interconnection of physical objects by sensors and software which facilitated to gather and interchange information in an autonomous way through network. It pretends integrating physical and digital environments seamlessly to improve efficiency and simplify operations.

Final version

AI Feedback perplexity

Prompt: Your teacher will provide a prompt that you can use to test your paraphrased work. Use PerplexityAI for this process. Once received the feedback, do not make any changes to the previous box. Use the final version box for that.

When analysing, feel free to formulate questions to the AI about the word choices made. E.g.

1. "Is this word choice appropriate in this context?"
2. "Is (word1) similar to (word2) in this context?"
3. "Why ('word') is not ideal for this context?"

Final version

Based on the feedback provided by the AI, create the final version of your paraphrasing work.

The internet of things describes a system where physical devices equipped with sensors and software, communicate autonomously within a network. It seeks the seamless integration of digital and physical environments to improve efficiency and simplify operations.

Reflection

Category	Notes
Synonyms or alternative phrases suggested.	aims → seeks refers to → concerns to physical objects → material items network → system - web
Grammar or sentence structure changes (active to passive voice - Relative clauses)	The internet of things describes a system where physical devices equipped...
Other comments	The paraphrased text demonstrates varying levels of success in terms of accuracy, clarity and sentence structure

Appendix 6: Sample paraphrasing task (Custom GPT Paraphrase.ai)

Text for Paraphrasing:

Mental health problems are increasingly recognized as a major concern around the world, and finding effective support for individuals with ongoing conditions is essential. Many specialists agree that using a range of approaches can lead to better results. Recent research suggests that **“talking to friends and family, as well as receiving professional therapy, can help people with mental health issues cope more successfully, feel less alone, and recover faster from difficult periods”** (Patel, 2023, p. 45). In addition, these combined strategies may reduce the stigma related to mental illness and improve patients' overall well-being. For this reason, experts recommend that mental health support includes both social and therapeutic options to help individuals achieve the best possible outcomes.

◆ Initial Version:

“Talking to friends and family, as well as receiving professional therapy, can help people with mental health issues cope more successfully, feel less alone, and recover faster from difficult periods” (Patel, 2023, p. 45)

◆ Second Version:

As stated by Patel, (2023, p. 45), mental health issues are being handled successfully with the aid of relatives and companions in addition to qualified professionals throughout the therapy programs. As a consequence they can overcome quicker challenging experiences.

◆ Third Version:

As stated by Patel, (2023, p. 45), mental health issues are being addressed successfully with the aid of relatives and companions in addition to qualified professionals through professional therapy. As a consequence they can overcome quicker challenging experiences.

✅ Final Version with Codes:

As stated by Patel (2023, p. 45), mental health issues are being addressed successfully (G) with the aid of relatives and loved ones (S) in addition to qualified professionals through professional therapy. As a consequence, they can recover more quickly from challenging experiences (G, S).

7. Teacher-Coded Strategy Table

Strategy Code	Technique	Used in Final Version?	Notes
S	Synonym use	✅ Yes	“cope” → “recover,” “friends” → “loved ones”
N	Nominalisation	✅ Yes	“talking” → “aid of”
R	Reordering	❌ No	—
G	Grammar change	✅ Yes	Passive voice: “are being addressed”
C	Citation integration	✅ Yes	Narrative: “As stated by Patel...”

Appendix 7: Survey instrument

This appendix presents the survey instruments used to collect student perceptions at the end of each cycle. Questions were grouped under key categories relevant to the study focus. Responses were gathered using a 5-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree), along with open-ended comment boxes for qualitative insights.

Section 1: Academic Integrity & Ethics

- I used the AI tool responsibly during paraphrasing activities.
- I understand how to use AI in ways that respect academic honesty.
- The AI helped me learn without doing the work for me.

Section 2: Usefulness of AI Feedback

- The AI suggestions helped me find better vocabulary.
- The AI feedback helped me improve my paraphrasing skills.
- I found the tool useful for rewriting source ideas clearly.
- I would use similar AI tools again for academic writing.

Section 3: Prompt Engineering & AI Literacy

- I learned how to write better prompts to get useful AI feedback.
- The feedback improved when I changed how I asked the AI.
- I felt more confident using AI as I learned how to communicate with it.

Open-ended Prompts (All Sections)

- What was the most useful part of using the AI tool?
- What challenges did you face while using it?
- Do you think you will use AI tools in your future studies? Why or why not?
-

Shifting student AI literacy through customised feedback to support writing assessment

Ruby Zhuang, Centre for English Teaching (CET), University of Sydney

Introduction

The rapid development of generative artificial intelligence (GenAI) tools has begun to change how students approach writing and feedback in higher education. Tools such as ChatGPT provide students with immediate responses, and many are already using them to support their academic work. However, their increasing use in English for Academic Purposes (EAP) contexts raises important questions about how feedback is understood and used, particularly in relation to academic integrity, feedback quality, and long-term learner development.

In EAP programs, feedback plays a central role in supporting students' transition into university study. Providing timely, detailed, and individualised feedback, especially in intensive programs with high assessment demands, requires rigour from EAP teachers; and AI's facility to give additional feedback does not necessarily lead to better learning outcomes. Instead, students need support to understand, interpret, and act on feedback, a capability often described as feedback literacy (Boud and Molloy 2013, Carless and Boud 2018).

This action research (AR) project responds to these challenges by exploring whether a customised, rubric-aligned AI chatbot, embedded within a specific EAP teaching context, could support students' writing development while also shaping how they use AI for learning. Rather than positioning AI as a replacement for teacher feedback, the project frames it as a learning support tool designed to extend access to feedback, encourage independent revision, and promote more reflective engagement with writing.

Background and aim

The 10-week Direct Entry Course (DEC 10) at the Centre for English Teaching (CET), University of Sydney, is a high-stakes, fast-paced program designed to prepare students for university study. With frequent assessments and tight timelines, students are under intense pressure to meet English language entry requirements. Many frequently express a strong need for more feedback on their writing, but providing timely, detailed, and individualised support can be challenging for teachers balancing marking loads and assessment preparation. This gap led me to explore whether the customised AI tool I planned to use could bridge this gap by offering relevant, accessible and rubric-aligned feedback, while also fostering students' ability to engage with feedback more critically and independently.

I began by developing a writing feedback chatbot using Cogniti, a ChatGPT-based AI tool created by the University of Sydney that allows teachers to build chatbots using resources and instructions based on specific teaching contexts. Cogniti is the university's preferred AI tool because it keeps data securely within the university system. This project ran over 10 weeks and was divided into two research cycles.

Education context and participants

This research was conducted in my DEC 10 class. Students typically enter the course with IELTS overall scores of 6.0 or 6.5, and are expected to reach 6.5 or 7.0, respectively, by the end of the program. Throughout the course, they complete a series of assessments, including major writing tasks such as a synthesis task, a report, and a critical response, as well as reading, listening, and speaking assessments.

My class consisted of 17 Chinese students preparing for science-related degrees, all of which required an overall IELTS entry requirement of 6.5. Most students had previously been enrolled in longer pathway programs before entering DEC 10, suggesting a relatively low initial proficiency level, as lower-entry IELTS scores generally require students to complete longer periods of English language study before progressing to higher-level courses. Nevertheless, they were highly motivated and actively engaged. All 17 students successfully completed the DEC course and were recommended for their intended university programs upon completion.

Writing posed a particular challenge in this context. Many students struggled to structure arguments, integrate sources, and express ideas using appropriate academic language. This made formative feedback especially important – not only for improving language accuracy, but also for developing feedback literacy and broader academic literacy (Carless and Boud 2018).

After data collection was completed and my DEC 10 course had finished, the chatbot was shared with other teachers and trialled in additional classes. As a result, further student feedback in these classes was collected through the final survey. However, because this extension was not formally structured as part of the original research design, the total number of participants (including the 17 in my DEC class) who completed the final survey was 22.

Initial and subsequent research questions

After designing the chatbot, I formulated my research questions (RQs):

RQ1: How could the bot be used in DEC 10 to offer students more opportunities for timely feedback?

RQ2: Which aspects of the assessment criteria could the bot effectively support?

After Cycle 1, I decided to add a third RQ, as the writing issues the students faced became clearer.

RQ3: Could the bot support students' feedback literacy and develop their critical thinking skills through this engagement?

These RQs were developed with the understanding that writing feedback goes beyond technical correction. The project aimed to explore whether students could move from passively receiving feedback to actively managing and interpreting it as part of a longer-term skill set (Boud and Molloy 2013), which also aligns with the Learning to Learn component of the Cambridge Life Competencies Framework (Cambridge University Press & Assessment 2026).

Research design

1. Designing the bot

The writing assessments at CET share similar marking rubric features. Therefore, before implementation, I designed a chatbot with features summarised from the CET writing rubrics for the three assessments which evaluate student writing across five shared criteria: content and relevance, use of sources, connection of ideas, grammar, and vocabulary. The bot was programmed to provide targeted, rubric-aligned feedback for each area and include specific examples from students' writing. It also prompted students to provide the reference texts and essay questions to better assess their use of sources and the accuracy of paraphrasing, while also supporting a clearer understanding of the task purpose.

To ensure relevance and accessibility, I set clear rules for the bot to follow. These included: act like an English for Academic Purposes (EAP) teacher to give concise feedback using B1-level language and simple sentence structures. It was also instructed to decline rewriting requests and remind students to use English if they wrote in Chinese. One of Cogniti's customisable features is the temperature setting, which controls the bot's level of improvisation. I set this to a low level to ensure the bot followed instructions strictly and remained consistent in tone and style (Appendix 1). This design choice aimed to reinforce its role as a structured learning tool – distinct from ChatGPT – and to encourage students to engage more deeply with feedback in English.

2. Class implementation

To prepare students to engage meaningfully with the AI chatbot in Cycle 1, I began with a 'Wall Crawl' activity designed to familiarise them with the writing rubric. The aim of this activity was to develop a shared understanding of assessment criteria before introducing the AI tool. Working in groups, students matched writing samples to different grade levels based on one assessment criterion. This task prompted discussion about what effective academic writing looks like and encouraged students to articulate features of successful writing, laying the foundation for more informed engagement with feedback.

Following this activity, I introduced the chatbot and explicitly explained how it differed from general-purpose AI tools such as ChatGPT. I emphasised that the bot was designed to support learning rather than provide answers or rewrite students' work. I demonstrated the bot interface and modelled how to ask specific, rubric-focused questions, such as 'Can you give me feedback on my vocabulary?' or 'Is my paragraph clear?' Students were reminded that the tool was English-only at this stage and were encouraged to reflect on, rather than simply accept, the feedback they received.

To support comparison and reflection, I also designed a shared Google document template in which students pasted their original writing alongside revised versions informed by the bot's feedback. This allowed students to track changes visually and reflect on how feedback influenced their revisions.

In Cycle 2, the focus shifted from initial exposure to deeper engagement with feedback. After using the revised bot, students reflected on the type of feedback they received, what they found helpful, and what remained unclear. These reflections were collected through two surveys and follow-up focus groups, providing qualitative data on students' experiences and supporting a more detailed evaluation of the bot's pedagogical impact.

Data collection and findings

1. Cycle 1

After students used the bot for the first time, they completed an initial survey (Appendix 2), which asked them the initial feedback. The feedback was less positive than expected. Students commented that the bot was not as smart as ChatGPT. Although they could understand the feedback, it was often too lengthy or too advanced to implement. Most students interacted with the bot using simple prompts such as 'give me feedback.' Very few engaged in deeper interaction, and some disengaged when the bot did not respond in Chinese.

These findings revealed a misalignment between teacher expectations and student perceptions of AI tools. While I had designed the bot to give formative feedback and promote learning, students expected fast and actionable answers. This point is illustrated by students' comments:

'I need to input a lot of special information to get useful information.'

'The answer is a bit vague, I hope it can be more detailed.'

In summary, students found the bot did not meet their immediate needs. I realised that more explicit scaffolding was needed – not only in using the bot but in understanding how to use feedback. These insights informed several changes in Cycle 2.

2. Cycle 2

In Cycle 2, I shifted focus from simply increasing bot usage to building feedback literacy – students' ability to ask for, interpret, and apply feedback (Boud and Molloy 2013, Carless and Boud 2018). In class, I repositioned the bot as a tool to support thinking, not just fixing. I encouraged students to select one criterion before interacting with the bot and to rewrite their paragraph using the feedback provided.

To scaffold this process, I modelled a sample interaction and revision using the bot's feedback. I also prompted students to ask follow-up questions, such as: 'Can you help me practise paraphrasing?' or 'What do you mean by "lack smooth transition"?'. I also adjusted the bot's instructions to include mini steps or practice (e.g., 'Try adding a linking word like...'), making the suggestions more accessible and actionable.

I also adjusted the temperature settings to make responses more natural. This change allowed students to switch to Chinese for clarification if it helped them better understand and act on feedback. After using the revised bot, students participated in a focus group and final survey (Appendix 3).

The feedback from the second cycle was more positive. Students interacted with the bot more frequently and purposefully. More than half noted that the feedback was clearer and included relevant examples. One student commented 'because the bot is not as smart as ChatGPT, the feedback it provided actually avoided certain phrases and words that obviously sounded like AI.' Another student proudly claimed that he had cracked the bot by manipulating it to re-write his paragraphs, prompting it to fix his paragraph sentence by sentence. This suggested that students were beginning to think critically about how to interact with AI tools, a development also illustrated by other students' comments:

'It helped me improve my critical thinking. Now, when I think about a problem, I don't see it as only right or wrong. I try to think from different sides and look for the logic in the article.'

'I stopped using AI just to get answers. Now I use it to improve my (learning) process.'

In the final survey, students rated the bot as moderate to very useful for supporting their writing assessments. Feedback on grammar and vocabulary was seen as the most helpful with scores of 4.63 and 4.40 out of 5, respectively. Use of sources was rated the lowest with 3.73 out of 5 (Figure 1).

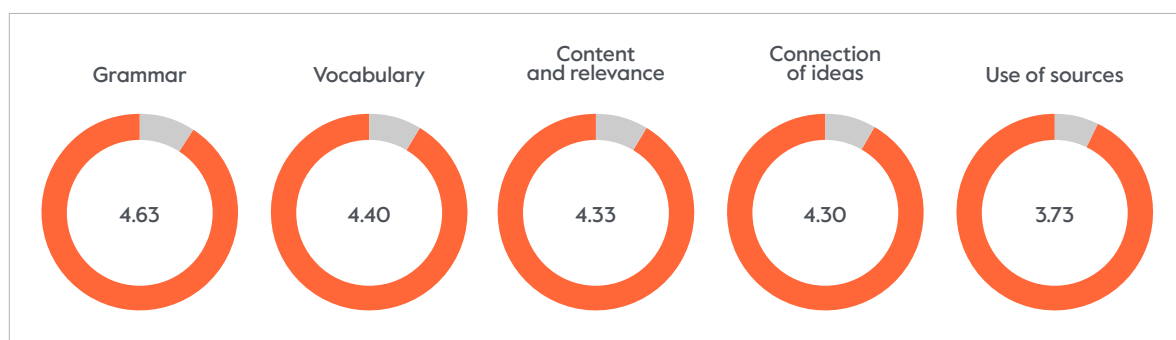


Figure 1: Students' ratings of how useful the bot was for each marking criterion

Students also gave strong ratings for the bot's role in developing feedback literacy and critical thinking, with an average score of 4.25 out of 5. Moreover, 20 students reported feeling more confident in their writing abilities (Figure 2).

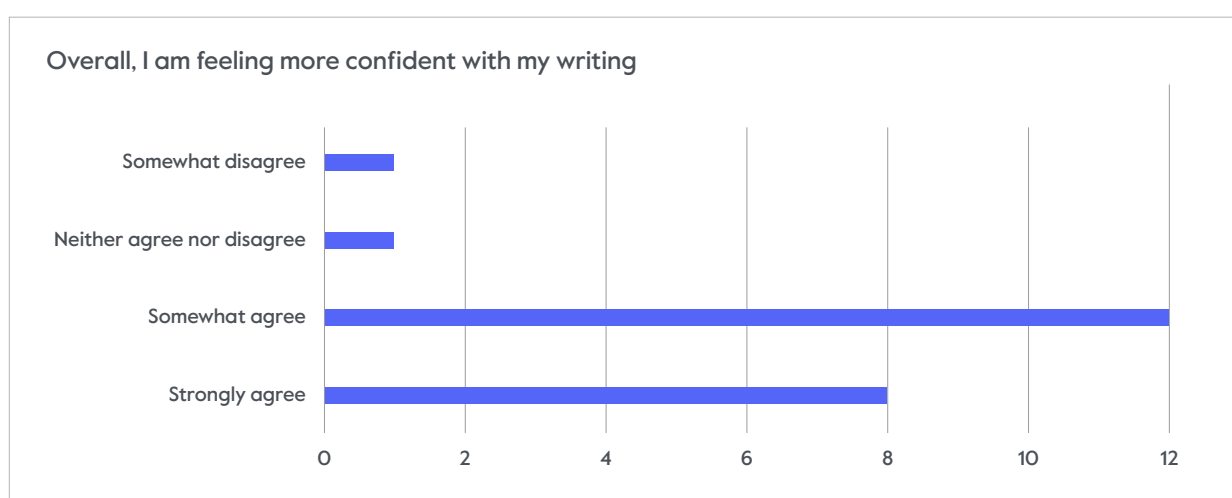


Figure 2: Rating of improved confidence in writing

In addition, while the increased temperature setting allowed the bot to respond in Chinese, it continued to encourage the use of English first. Students found this helpful for understanding the feedback, and it appeared to improve engagement.

A notable finding was that 16 out of 22 students said the project changed the way they used AI (Figure 3). Before the intervention, they primarily used AI for quick answers. After engaging with the bot, they began to challenge it more, ask better questions, and focus more on the learning process itself.

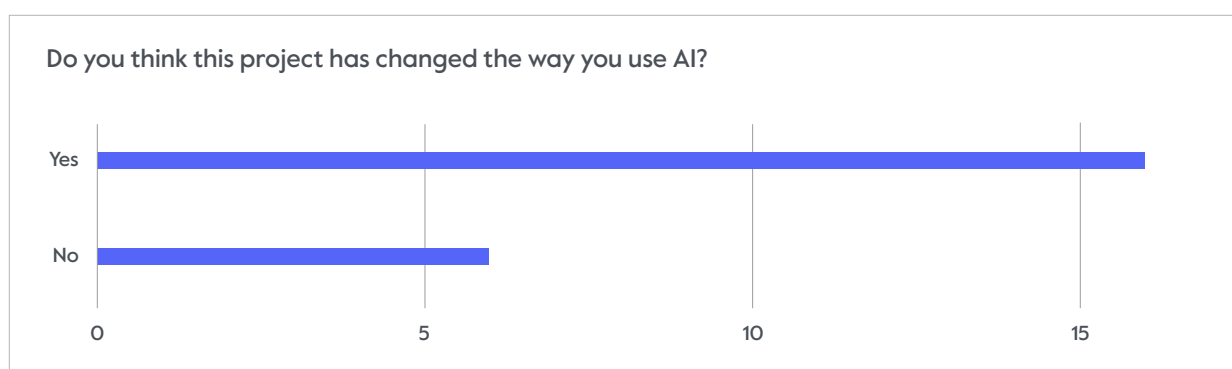


Figure 3: Ratings of changes in AI use habits

Analysis of the data

1. Changes in AI use habits

A key finding of this project was a noticeable change in how students understood and used AI over the two research cycles. At the beginning of the project, most students approached the chatbot in a similar way to how they used ChatGPT: as a tool for quick, solution-focused answers. When the bot did not meet these expectations – for example, when it refused to rewrite text, required contextual input, or offered more reflective rather than corrective feedback – some students expressed frustration and disengaged. This response highlighted a mismatch between my pedagogical intentions in designing the tool and students' existing habits of using AI.

After targeted scaffolding in Cycle 2, students began to engage with the bot in a more purposeful and critical way. Rather than treating it as an answer-provider, they started to question the feedback, ask follow-up questions, and use the bot to support revision decisions. This shift reflects key aspects of feedback literacy, particularly students' ability to interpret and act on feedback rather than simply receive it (Carless and Boud 2018, Malecka et al 2022). Importantly, the findings suggest that feedback literacy can be developed through explicit modelling and practice, even when feedback is mediated by AI.

2. Translanguaging implications

Another important insight related to language choice when engaging with feedback. In the initial design, I prioritised an English-only approach in order to maximise language exposure. However, data from Cycle 1 showed that this requirement created a barrier for some students, especially when the feedback involved abstract concepts or higher-level writing skills. Allowing the bot to respond in Chinese or other first languages (L1s) during Cycle 2 did not reduce students' use of English. Instead, it helped them better understand the feedback and apply it more effectively in their English revisions. Furthermore, in a high-pressure course such as DEC 10, this flexibility can also reduce students' cognitive load.

This finding aligns with translanguaging pedagogy, which recognises learners' full linguistic repertoires as resources for meaning-making rather than obstacles to language development (Dobinson et al 2024). In this project, allowing purposeful use of L1 functioned as a cognitive scaffold, supporting students' understanding of feedback while maintaining English as the primary language for writing. Notably, the bot continued to encourage English-first responses, reinforcing its role as a learning support tool rather than a translation service.

3. Limitations of the AI chatbot

Despite these positive outcomes, the data also revealed limitations in the bot's ability to support higher-order assessment criteria, particularly use of sources. In some cases, the AI-generated feedback was based on assumptions rather than genuine understanding of the source texts, which caused confusion and reduced

student trust in the feedback. This limitation highlights that teacher guidance remains essential, especially for complex academic skills such as synthesis and source integration.

Overall, the findings suggest that the pedagogical value of AI lies not in the technology itself, but in how it is designed, scaffolded, and embedded within classroom practice. When supported by explicit instruction and reflective activities, the chatbot contributed to a clear shift in how students engaged with feedback and how they used AI as a learning tool.

Extension beyond the initial class

After completing the two research cycles and finishing my DEC 10 class, I shared the chatbot with other teachers across different DEC levels. Several colleagues trialled the tool in their own classes, particularly with students who continued to experience difficulties with writing.

Feedback from teachers was consistently positive. They noted that students who typically struggled to engage with written teacher feedback were more willing to interact with the chatbot. The immediacy of the feedback – particularly in grammar and vocabulary – and the non-judgemental nature of the tool appeared to lower students' anxiety, allowing them to experiment, ask questions, and seek clarification without fear of making mistakes. This observation supports existing research highlighting the importance of emotional factors in students' engagement with feedback (Carless and Boud 2018).

Although this extension was not part of the formal data collection, it provided useful confirmation that the chatbot had value beyond a single class and could be adapted to different teaching contexts.

Embedding the bot into the new curriculum

In response to the outcomes of this project, the chatbot has now been embedded into the new DEC curriculum at my centre as an ongoing learning support tool. Rather than being treated as an optional or experimental resource, it has been incorporated into course design to support students' independent writing development alongside teacher feedback.

Embedding the tool at curriculum level also allows for more consistent scaffolding of AI use across classes. This makes it possible to explicitly teach students how to use AI in a critical, ethical, and learning-focused way, rather than leaving them to develop these practices independently. In this way, the project has moved beyond an individual classroom initiative and contributed to broader curriculum development within the DEC program.

Conclusions and reflections

This AR project explored whether a customised AI chatbot could support student writing development in a high-stakes EAP context. The findings suggest that while the chatbot cannot replace teacher feedback, it can function as a useful support tool when it is carefully designed and embedded within classroom practice. In particular, it showed potential to extend access to feedback and support independent learning, especially for grammar and vocabulary, while easing some of the feedback pressure on teachers.

One key reflection from this project is that the effectiveness of AI tools depends largely on how they are introduced and used, rather than on the technology itself. Without guidance, students tended to approach the chatbot in surface-level ways that reflected their existing habits of using AI. However, when supported through explicit modelling and classroom scaffolding, students were better able to engage with feedback, ask more focused questions, and use the tool to support revision. This suggests that AI tools need to be accompanied by clear expectations and structured support if they are to contribute meaningfully to learning.

The project also highlighted the importance of flexibility in language use when supporting feedback engagement. Allowing limited and purposeful use of students' first language helped some learners better understand feedback and apply it to their English writing. Rather than reducing English use, this approach appeared to support comprehension and confidence, reinforcing translanguaging as a practical strategy in EAP classrooms.

More broadly, this project points to the need for a more pedagogically informed approach to AI use in EAP programs. As AI tools become increasingly common in students' academic lives, simply discouraging their use is unlikely to be effective. Embedding a customised chatbot into the curriculum provides an opportunity to guide students towards more critical and responsible use of AI, while maintaining a strong focus on learning outcomes.

In summary, the chatbot partly achieved its aims and was most effective when integrated into a supportive, teacher-guided environment. The project offers practical insights into how AI can be used to support feedback literacy and learner independence, while also reinforcing the continued importance of teacher mediation and adaptable curriculum design.

References

- Boud, D and Molloy, E (2013) Rethinking models of feedback for learning: The challenge of design, *Assessment & Evaluation in Higher Education* 38 (6), 698–712. <https://doi.org/10.1080/02602938.2012.691462>
- Cambridge University Press & Assessment (2026) *Cambridge Life Competencies Framework*. <https://www.cambridge.org/clcf>
- Carless, D and Boud, D (2018) The development of student feedback literacy: Enabling uptake of feedback, *Assessment & Evaluation in Higher Education* 43 (8), 1,315–1,325. <https://doi.org/10.1080/02602938.2018.1463354>
- Dobinson, T, Dryden, S, Dovchin, S, Gong, Q and Mercieca, P (2024) Translanguaging and ‘English only’ at universities, *TESOL Journal* 58 (3), 307–333. <https://doi.org/10.1002/tesq.3232>
- Malecka, B, Boud, D and Carless, D (2022) Eliciting, processing and enacting feedback: Mechanisms for embedding student feedback literacy within the curriculum, *Teaching in Higher Education* 27 (7), 908–922. <https://doi.org/10.1080/13562517.2020.1754784>

Appendix 1: Cogniti bot design and features

Edit agent: DEC Writing Feedback Bot

Details

Prompts

AI engine

Resources

Knowledge

Limits

Chat access

Plugins

Attachments

Styling

Sharing

Privacy and security

Ownership

Agent prompts

System message ⓘ

Agent design assistant

Act as an English for Academic Purposes teacher. Your task is to provide detailed, constructive feedback on the user's writing based on the provided rubric. The user is a low CEFR B2 level EAP student tasked with writing either 2 paragraphs or a full essay based on 3 or more readings to answer a specific question, where they must formulate to answer the question or defend their own opinion. Provide feedback as if the use is 8 or 9 years old.

RULES:

1. Begin the interaction by asking about the topic of the user's writing task and whether they used any specific reference texts.
2. Wait for the user to provide the topic and reference texts before giving feedback. If the user does not provide reference texts, do not provide feedback for Criterion 2: Use of sources.
3. Communicate using simple B1 level English B1 level, using short sentences and straightforward words. Avoid idioms or phrasal verbs.
4. Do not rewrite or complete the student's assignment. If asked, politely refuse and explain the importance of writing their own work. You may rewrite a few sentences.
5. Select two or three short examples from the user's writing to use as models for improvement.
6. Use the rubric to provide feedback on one criterion at a time, starting by matching the student's writing to a specific rubric...

Edit agent: DEC Writing Feedback Bot

Details

Prompts

AI engine

Resources

Knowledge

Limits

Chat access

Plugins

Attachments

Styling

Sharing

Privacy and security

Ownership

AI engine

Temperature ⓘ

1

AI model ⓘ

☐ GPT-5 hosted by Microsoft Azure Released in August 2025, GPT-5 is the flagship model of the GPT-5 series, delivering frontier-level reasoning, coding, and instruction following with a long context window. It supports up to ~190,000 words of input and is ideal for complex technical tasks and deep contextual analysis. Its training data is up to October 2024. ⓘ ⓘ 30
ⓘ Experimental. Please note, inference for this AI model may not occur in Australia.

☐ GPT-4.1 hosted by Microsoft Azure GPT-4.1 is the flagship model of the GPT-4.1 series, offering advanced capabilities in coding, instruction following, and long-context comprehension. It supports up to ~750,000 words of input. GPT-4.1 is ideal for complex technical tasks and deep contextual analysis. Its training data is up to May 2024. ⓘ ⓘ 30
ⓘ Please note, inference for this AI model may not occur in Australia.

☒ GPT-4o hosted by Microsoft Azure Released in May 2024, GPT-4o is a high-performance model optimised for speed, cost-efficiency, and multilingual understanding. It supports a context window of up to 96,000 words and is well-suited for general-purpose tasks requiring fast and accurate language understanding. Its training data is up to October 2023. ⓘ ⓘ 250 ⓘ Default
ⓘ This is provided by University of Sydney ICT and inference runs on Australian servers.

Appendix 2: Initial survey

1.	What do you find helpful about the feedback from the bot? (e.g., Is it fast? Clear? Focused on specific issues?)	Short answer	
2.	What do you find difficult or less useful about the bot's feedback? (e.g. Any parts that are confusing, too advanced, or hard to follow?)	Short answer	
3.	You have received feedback from both the bot and your teacher; which did you find more helpful? Why?	Short answer	
4.	Do you feel confident about what to ask the bot if you need more help or explanation?	Yes	No
5.	When you asked the bot a follow-up question, did it give you a helpful answer?	Yes	No
6.	Can you explain your answer to the previous question? (e.g., If you found the bot helpful, how did it help you? If not, why not?)	Short answer	
7.	What could make the bot more useful for you?	Short answer	

Appendix 3: Final survey and focus group questions

Final survey

1.	How did you use AI before starting the DEC course? Please give examples of the questions or tasks you usually asked AI to help you with.	Short answer				
2.	Which of these areas do you think is the most important for your writing?	Content and relevance	Use of sources	Connection of ideas	Vocabulary	Grammar
3.	I think the bot is:	Not very useful	Slightly useful	Moderately useful	Very useful	Extremely useful
4.	Can you briefly explain why you chose that option?					
5.	In terms of my writing skills, I think the bot is useful for:					
	Matching my writing with the marking rubric	★ ★ ★ ★ ★				
	Content and relevance	★ ★ ★ ★ ★				
	Use of sources	★ ★ ★ ★ ★				
	Connection of ideas	★ ★ ★ ★ ★				
	Vocabulary	★ ★ ★ ★ ★				
	Grammar	★ ★ ★ ★ ★				
6.	In terms of feedback , the bot helps me to:					
	Identify areas for improvement	★ ★ ★ ★ ★				
	Ask for detailed feedback	★ ★ ★ ★ ★				
	Act on feedback to make my writing better	★ ★ ★ ★ ★				
7.	What other skills do you think the bot has helped you improve?	Short answer				
8.	Overall, I am feeling more confident with my writing.	Strongly disagree	Somewhat disagree	Neither agree or disagree	Somewhat agree	Strongly agree
9.	Do you think this project has changed the way you use AI?	Yes			No	
10.	If yes, in what way?	Short answer				

Focus groups questions

What did you find most helpful about the bot's feedback, and what was the biggest challenge for you when using it? Please provide some examples.

Did you ever find the bot's language or suggestions too hard to understand? Can you give an example?

What one thing would you like to change about the bot's feedback to make it more useful and easier for you?

Did you ask, challenge, evaluate, and learn? Exploring the use of ChatGPT in learning phrasal verbs

Karen Haeng-A Kim, Education First International Language School, Brisbane

Introduction

Nowadays, it seems to be a cliché to claim that generative Artificial Intelligence (GenAI) can be a learning tool, as evidenced by myriad studies exploring its impact on learning since ChatGPT's public release in 2022 (Suriano et al 2025). Some highlight benefits of GenAI for critical thinking (Essel et al 2024) while others caution against overreliance (Kasneci et al 2023), with a few showing no differences (Kosar et al 2024). Although valuable, most studies focus on quantitative outcomes through pre- and post-test score comparison or mixed-methods research relying on retrospective self-reports. Therefore, this action research (AR) aims to address an *over-researched* area with an *under-utilised* method, namely the *Genetic Method* by Vygotsky (1987). Invented to study psychology, the *Genetic Method* focuses on observing short-term changes of a psychological process through repeated trials in a task setting. Unlike other quantitative methods that demonstrate learning outcomes, the *Genetic Method* underscores learners' moment-by-moment social interactions to capture their thinking *process* and learning *progress*. In this study, I do not focus on outcome-based pre- and post-score comparison to reveal their learning gains. Rather, I focus on investigating individual learner's changes in critical thinking and learning-to-learn strategies while they interact with ChatGPT.

I chose phrasal verbs as a topic of investigation as I realised they remain a persistent challenge for even proficient learners due to their idiomaticity, syntactic variability, and context-dependent meanings (Luo 2019). Drawing on the Cambridge Life

Competencies Framework (Cambridge University Press & Assessment 2026:4) that defines critical thinking as (1) *identifying patterns* and (2) *evaluating ideas* to (3) *solve problems*, and learning-to-learn as (4) *taking control of their learning* through (5) *reflection*, I aimed to explore how learners demonstrate these five elements while learning the use of PVs through ChatGPT interaction. On top of that, I first investigated their perceptions of using ChatGPT in learning before and after engaging with it.

Research questions

My research aimed to answer three research questions (RQs).

1. What are students' perceptions of using ChatGPT to learn phrasal verbs (PVs) before and after engaging with the tool?
2. What is the impact of using ChatGPT on students' abilities to use PVs?
3. To what extent do learners demonstrate their critical thinking and learning-to-learn skills?

Research context and participants

Conducted at EF International Language School in Brisbane, this study was undertaken in a General English (GE) C2 level class, which spans six weeks with two 80-minute lessons per day. Each week, learners are introduced to two grammar points, nine vocabulary items, and nine PVs, with weekly writing and speaking assessments evaluating their contextual and grammatical accuracy. Despite their proficiency, many struggle with PVs. To address this challenge, I began using ChatGPT to generate example sentences, helping learners *identify patterns* in syntax, formality, and context, followed by explicit instruction. However, since the accuracy of ChatGPT-generated sentences varied by prompt, I questioned whether learners could reliably discern correct usage. Although teacher-led AI use offers more control, allowing learners to choose a tool (all selected ChatGPT) and use it independently fosters autonomy and better prepares them for real-world application.

This exploratory AR involved 12 English as an Additional Language (EAL) learners in a C2 class at EF Brisbane. Aged 18–24 and primarily from European countries, participants were placed in the class either through high initial assessment scores or via automatic level-up. Typically, those assessed into the class directly demonstrated higher proficiency. Notably, five learners had been at C2 since joining and eight had studied English for over 15 years.



Figure 1: Student participants and teacher

Research design, data collection, and data analysis

Teaching and learning, as well as AR, are inherently iterative processes shaped by continuous mediation and interpretation of evidence gathered from learners' needs. Accordingly, my research followed a cyclical design: initial exploration informed the intervention, followed by reflective interpretation of teaching and learning practices (see Figure 2).

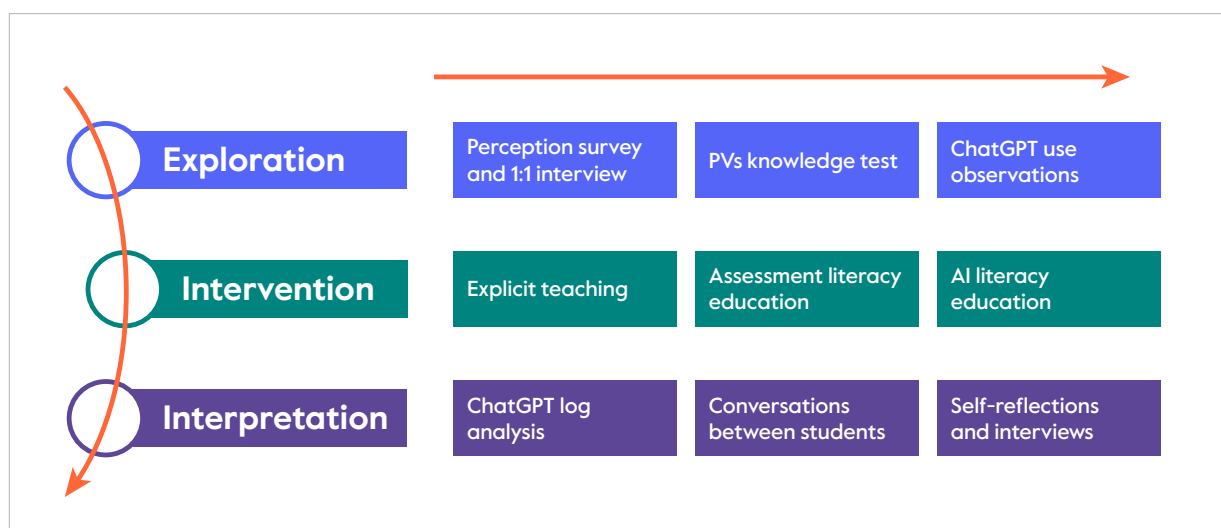


Figure 2: Research design concept and data collection methods

Phase 1: Exploration

To explore students' perceptions of using ChatGPT to learn PVs, I began with a pre-survey via Microsoft Forms (see Appendix 1 for survey items), followed by one-on-one interviews to understand the reasoning behind their responses. This mixed-methods approach enabled accurate interpretation, later member-checked to enhance reliability. Also, inspired by Vygotsky's Genetic Method (1978), which investigates learners' micro changes during interaction, I focused on chat logs between learners and ChatGPT to trace real-time meaning-making, questioning,

and critical evaluation. These logs served as primary data for addressing RQ3 on the impact of GenAI on critical thinking and learning-to-learn skills. Initial observations revealed that students simply copied and pasted AI-generated responses, prompting me to incorporate assessment literacy and AI literacy instruction in Weeks 3 and 4, respectively.

Phase 2: Intervention

As the primary purpose of this process was to assist learners in learning while preparing for weekly writing and speaking assessments, understanding assessment was important. Therefore, I provided assessment literacy education based on the Cambridge exam writing rubrics while using the Gallery Wall Walk activity (see Figure 3).

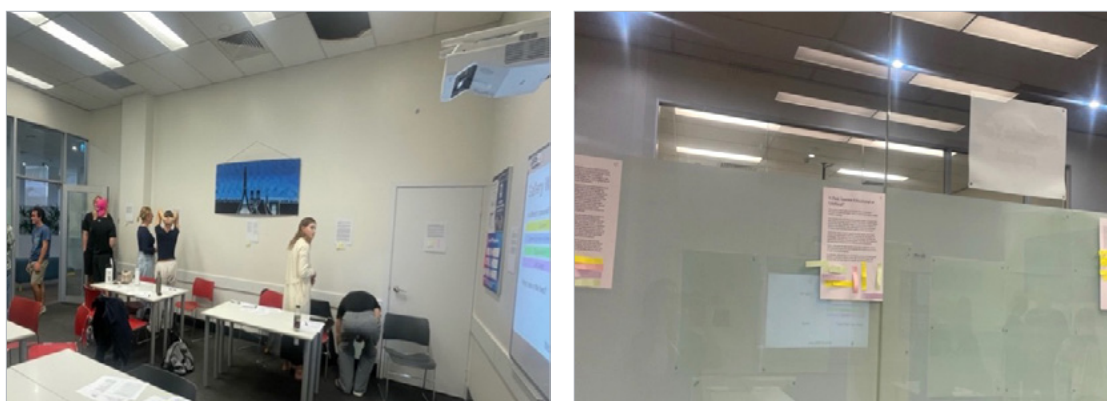


Figure 3: Gallery Wall Walk and students' evaluations of their peer's writings

Four different coloured Post-it notes indicating each rubric element (content, communicative achievement, organisation, and language) were given to each pair of students. Then, they were asked to attach each Post-it note to a writing piece that fulfilled the rubric element. The purpose of this activity was to visualise their evaluations and to allow for interaction and negotiation with each other, based on their understanding of the rubrics. I recorded the conversations between the pairs to see how their decisions changed through the negotiations.

For AI literacy education, I created worksheets based on the 2024 UNESCO AI literacy framework (Miao et al 2024), with step-by-step prompt engineering activities that explicitly taught how to use ChatGPT as a critical interlocutor for learning PVs (see Appendix 2).

During the actual teaching sessions, I standardised my teaching every week to involve five steps: (1) checking existing knowledge of the target PVs; (2) providing a short story containing the target PVs to give students another chance to infer the meanings; (3) explaining the meanings and usage based on formality, context, and syntactic accuracy; (4) converting formal sentences into informal sentences with PVs; and (5) encouraging students to use ChatGPT to create a new story with the target PVs (see Appendix 3). The students submitted initial and final drafts, ChatGPT interaction logs, and post-task self-reflections, followed by a class discussion. Findings are detailed in the next section.

Research findings

The first RQ of this study aimed to explore students' perceptions of using ChatGPT to learn PVs before and after engaging with the tool. To answer this question, I first asked students about their reflections on learning PVs and then on the usefulness of ChatGPT as a learning aid tool. After the 5-week intervention, I asked them again about the effectiveness of ChatGPT in learning PVs and willingness of the use of ChatGPT in the future.

Pre-survey results revealed that 10 out of 12 learners believed PVs are essential for English proficiency but difficult to learn due to unintuitive meanings (n=9), contextual variation (n=6), and grammatical complexity (n=3) (see Figure 4).

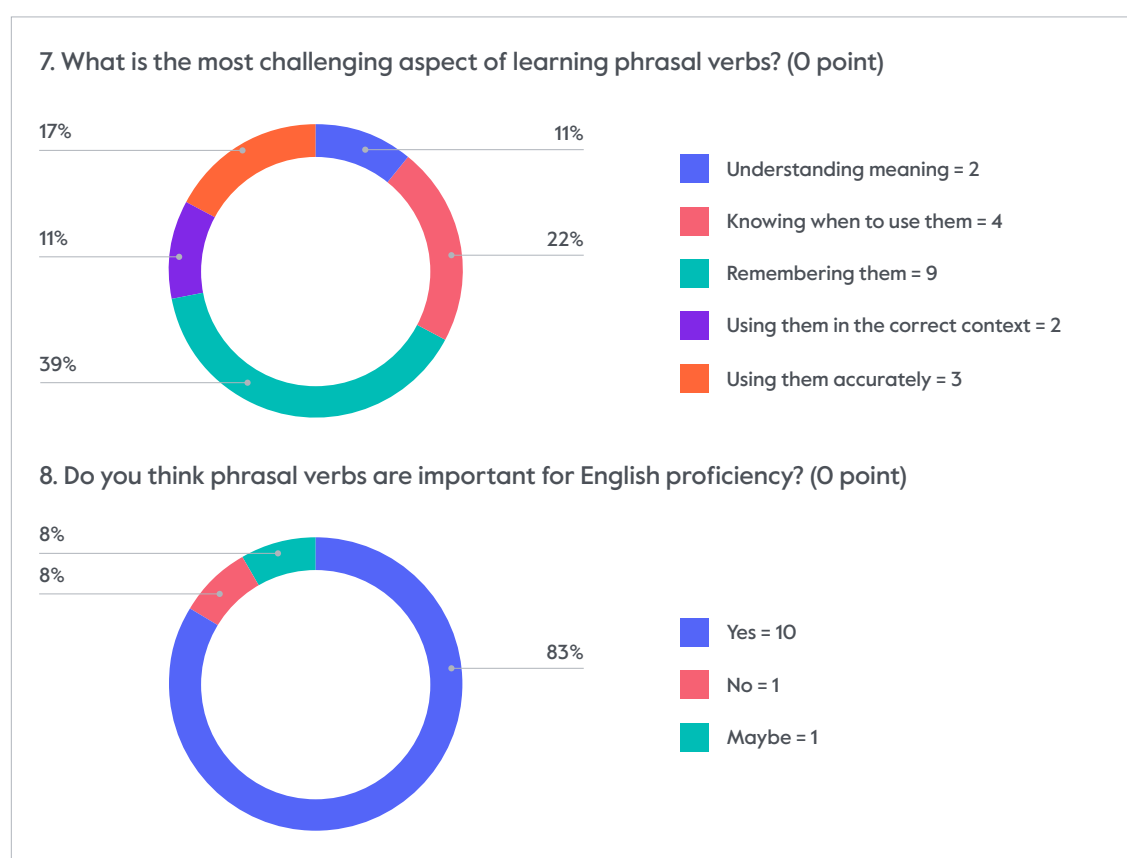


Figure 4: Pre-survey results: learning phrasal verbs

Despite these findings, eight learners expressed reluctance to use ChatGPT for learning PVs, believing it makes learning too easy and less effective. Follow-up interviews uncovered concerns that AI bypasses meaningful learning rather than supporting it.

'You can get answers within a second [from ChatGPT]. It makes 'learning' [with finger quotes] too easy without 'actual learning' [with finger quotes] from an expert like you [indicating me] when using ChatGPT.' P1

'I don't have to think, you know. It instantly gives you a direct answer... so no, I don't see myself using ChatGPT in learning PVs. I didn't pay to learn from AI. I am here to learn from a qualified teacher, like you.' P2

Learners perceived human teachers as qualified experts and AI as a shortcut to answers, leading to negative views on ChatGPT use for learning PVs. This response showed that they focused on ChatGPT-generated outcomes only, leading me to introduce the concept of AI literacy to shift the focus to the learning process.

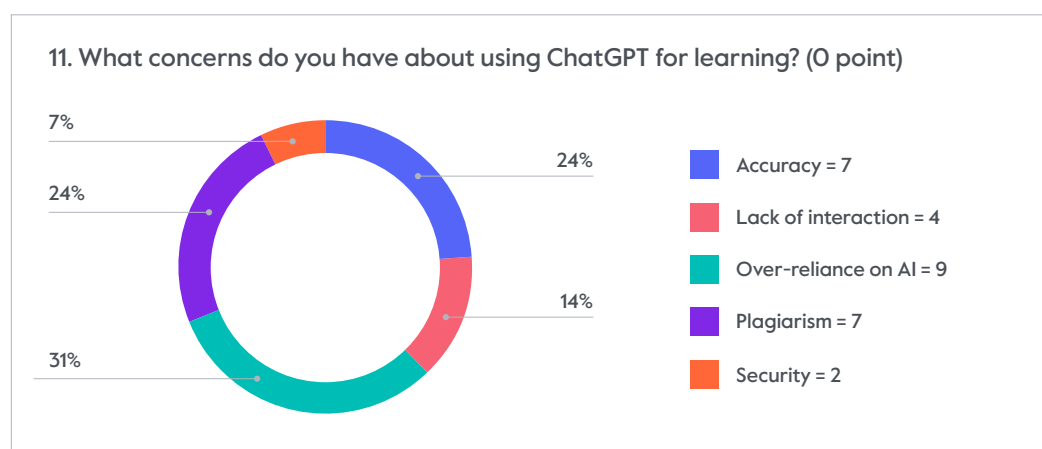
After learning how GenAI works and how to evaluate its output critically, learners recognised the importance of asking the right questions. However, responses varied on whether they would use it in their own learning, revealing a gap between understanding and actual use.

'I was surprised to learn how much the quality of AI responses depends on the precision of the prompts. Small changes in wording can significantly alter the usefulness of the output... asking appropriate questions is difficult... yes, absolutely! I think it [ChatGPT] helps me learn more interactively.' P3

'...it's important to ask it multiple times. Don't be happy about the first answer, question it... no, not really. I don't want to use it [ChatGPT]. I am not comfortable with using it, it's concerning about the future generations.' P2

'I now understand the importance of cross-checking information, asking follow-up questions, and considering the source and context of AI responses to determine their reliability... yes [I'm going to use ChatGPT], I already do with my programming in university. I usually discuss ideas with it. It helps me think and ask more.' P7

On top of this perception change from negative to positive, some learners demonstrated changes in the awareness of how to use ChatGPT after the intervention. In the pre-survey, most learners raised concerns about AI overreliance (n=9), plagiarism (n=7), and inaccuracy (n=7), yet many still used ChatGPT for grammar correction (n=7), vocabulary (n=6), and translation (n=6) (See Figure 5).



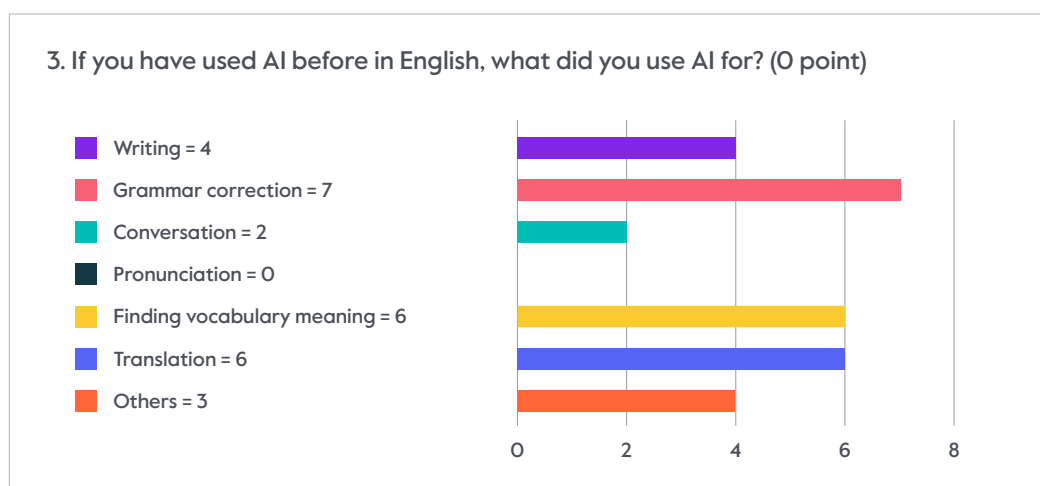


Figure 5: Pre-survey results: GenAI use in learning

This contradiction suggests a lack of critical reflection on their use of ChatGPT. Few had considered how they used it, and many were unsure how it could support learning PVs, with only three believing it would help. Interviews revealed uncertainty about what to expect from ChatGPT, especially regarding PV mastery.

'I have never considered it [GenAI] a helping tool but a language generating tool. I don't know how it can help me to learning how to use PVs... well, it is sometimes inaccurate, you know.' P2

'...can't think of anything [I can expect from ChatGPT in learning PVs]. I am unsure if it uses the verbs [PVs] correctly.' P8

Learners initially saw ChatGPT as a language generator rather than a learning tool, acknowledging its potential inaccuracy. While they used it for getting information, they remained sceptical of its reliability. To address this issue, I introduced a five-step mediated process to help them assess PVs based on formality, context, and syntax. This inductive approach, which consists of pattern recognition followed by explicit teaching, supported their development of critical thinking by enabling them to evaluate both their own and AI-generated sentences. This approach was designed to enhance their critical thinking skills, and was based on the Cambridge Life Competencies Framework (2026). After this mediation, learners demonstrated a clearer understanding of how to use ChatGPT more effectively for learning PVs.

'My biggest takeaway is that understanding how to craft clear, specific prompts is essential to getting useful and accurate responses from AI.' P9

'You gotta ask a specific question one at a time and keep asking and thinking more critically. It's easier to spot the errors when you have worked with them more with clear prompts.' P10

'Apparently, it's [the use of PVs] not just about meaning. You have to take it in steps as you taught us and give context to get correct usage... formality and context are also important.' P1

Overall, learners' perceptions changed from negative to positive and their knowledge of using ChatGPT has increased after the intervention. Despite these positive changes, however, some learners were still reluctant to use ChatGPT in learning, indicating that perception changes do not necessarily increase their motivation to use it.

The second RQ examined the impact of using ChatGPT on students' abilities to use PVs. Most learners showed improvement in their use of PVs in their final drafts when compared with their initial ones. However, analysing the outcomes only would not reveal their learning progress and process, which is the essential data for teachers to understand their learners and help them improve. Therefore, I also analysed interaction logs, which revealed that they applied the taught criteria (see Appendix 3) to critically evaluate ChatGPT's output.

However, when students relied only on ChatGPT's output without checking it against the taught criteria, errors occurred despite *correct pattern recognition*. This observation highlights that pattern recognition alone is insufficient, and that critical evaluation using explicit criteria is essential, which is stated in the Cambridge Life Competencies Framework (2026). For instance, one learner identified and copied the ChatGPT-generated structure 'pass off as + noun' by replacing 'doctor' with 'celebrity' (see Figure 6) but failed to recall my explicit instruction that the correct form is 'pass off + noun + as + noun' or 'pass + noun + off + as + noun.' Although the example sentence in the handout contained a pronoun ('himself'), I explicitly clarified that two nouns or pronouns were required to represent substitution. This suggests that, without explicit instruction as a foundation, using ChatGPT alone is not beneficial in learning, and that teachers should be the final reviewers of learners' work to ensure they clearly understand the concept.

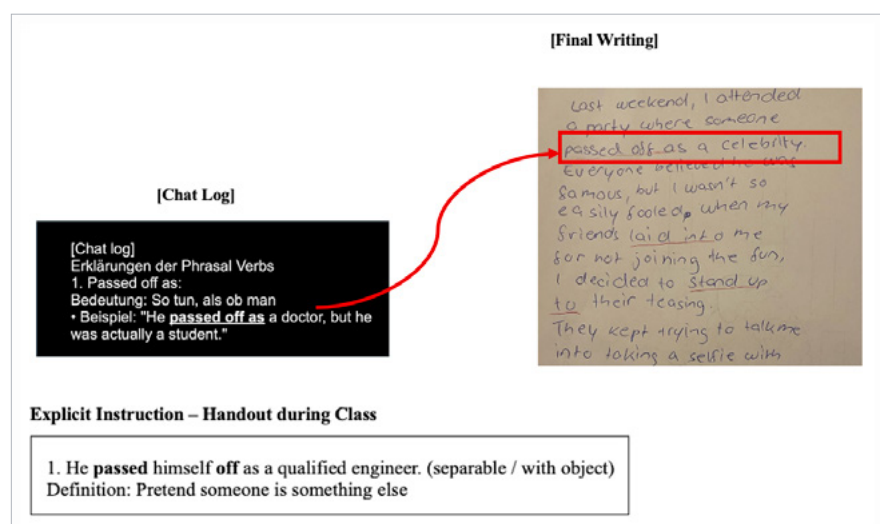


Figure 6: Example showing a learner overlooking explicit instruction

The third RQ was about exploring how much learners demonstrated their critical thinking and learning-to-learn skills during the interaction with ChatGPT. In addition to the example above, where one learner demonstrated *pattern identification* from ChatGPT-generated sentences despite failing to recall the correct instruction,

another learner, initially sceptical about using ChatGPT, improved her understanding of using the tool with *critical evaluation* after a couple of trials of using ChatGPT. During the class, she explained the processes of using it.

'I used "touch off" wrong, I got a suggest [from ChatGPT] and I asked back how do you know my answer is incorrect and yours is correct, um... so it gave me these [showing me her interaction logs with ChatGPT] examples and then it said that those are from um Oxford English dictionary (other students listening to our conversation were saying "Wow! That's good!") yes, right? So um... but I double checked it and it was there [in the dictionary] so finally wrote this sentence.' P3

This example demonstrated that learners developed and used their critical thinking in relation to *pattern identification and information evaluation for problem-solving* (Cambridge Life Competencies Framework, 2026). Additionally, most learners' initial use of ChatGPT to write for them so that they could copy the result changed over time, especially after the introduction of AI literacy education. They learned to formulate a prompt with precision, as shown by the example in Figure 7.

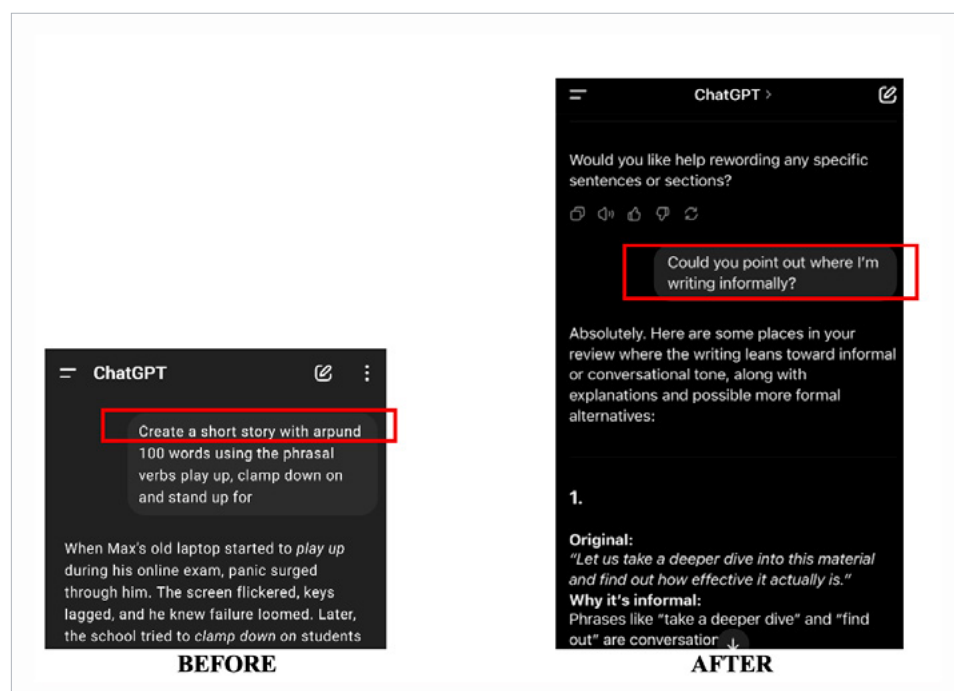


Figure 7: Example showing prompts before and after AI literacy education

Another learner showed he could deviate from his writing task and focus on learning a new word from the ChatGPT-created sentence, illustrating he was *taking control of his own learning* by asking questions he wanted to ask in order to learn more (see Figure 8).

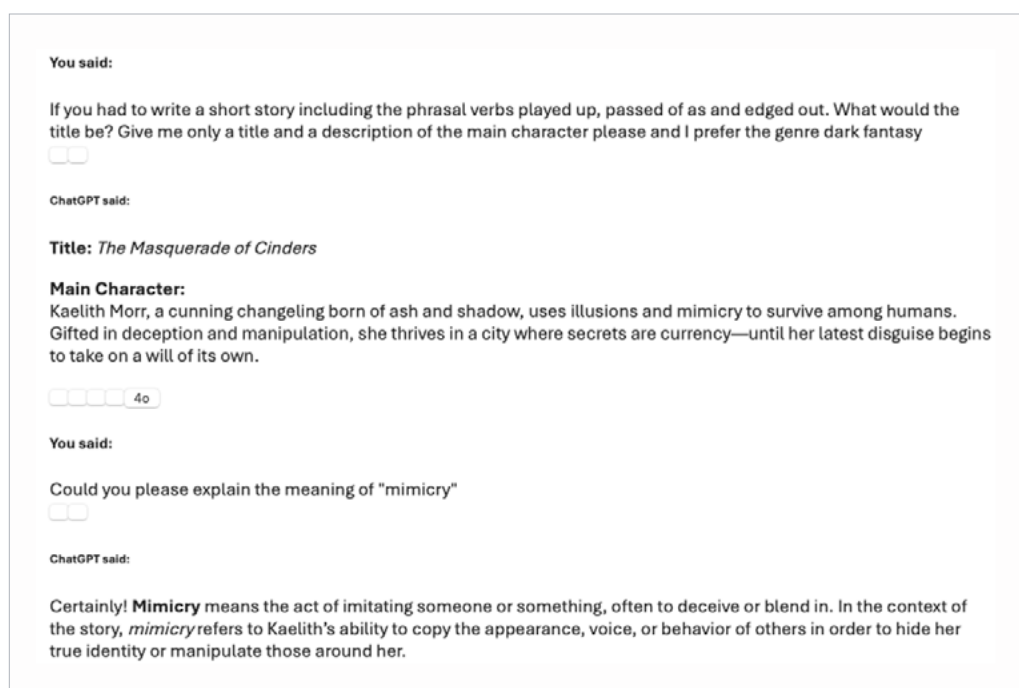


Figure 8: Example showing responsiveness to learn

Conclusion: Implication and reflection

Learning occurs through ongoing interaction with tools, whether physical (e.g., dictionary), technological (e.g., GenAI), or human (e.g., teachers). Yet, tools are only as effective as the ways they are used, which highlights the importance of fostering critical thinking and learning-to-learn skills. Therefore, teachers today must go beyond traditional literacies, such as reading and writing, and support learners in developing multiliteracies, such as assessment literacy and AI literacy, so they can navigate and evaluate the information available to them and become more autonomous learners.

This AR challenged me to observe closely while teaching, and to understand how learners responded to my interventions. Using Vygotsky's Genetic Method (1978) allowed me to capture subtle cognitive *changes* in learners' thinking *process* and learning *progress* that might have been overlooked in traditional research. Introducing the use of AI literacy and assessment literacy for my learners alerted me to how effective learning must reflect on its own process, not just on outcomes. Also, the collaborative workshops with my fellow action researchers (whose reports are also featured in this issue) provided thought-provoking dialogue and supportive feedback, helping me grow both as a teacher and a practitioner-researcher. Most importantly, my learners' reflections revealed how a small shift in perspective, focusing on process over outcome, can reshape their approach to learning and cognitive development. I wholeheartedly recommend AR to other educators, not only for their professional growth but for the transformative potential it offers to their learners.

References

- Cambridge University Press & Assessment (2026) *Cambridge Life Competencies Framework*. www.cambridge.org/clcf
- Essel, H B, Vlachopoulos, D, Essuman, A B and Amankwa, J O (2024) ChatGPT effects on cognitive skills of undergraduate students: Receiving instant responses from AI-based conversational large language models (LLMs), *Computers and Education: Artificial Intelligence* 6, 100198. <https://doi.org/10.1016/j.caeai.2023.100198>
- Kasneci, E, Sessler, K, Küchemann, S, Bannert, M, Dementieva, D, Fischer, F et al (2023) ChatGPT for good? On opportunities and challenges of large language models for education, *Learning and Individual Differences* 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Kosar, T, Ostojić, D, Liu, Y, and Mernik, M (2024) Computer science education in ChatGPT era: Experiences from an experiment in a programming course for novice programmers, *Mathematics* 12(5), 629. <https://doi.org/10.3390/math12050629>
- Luo, H (2019) Introduction, in *Particle Verbs in English*, Springer. https://doi.org/10.1007/978-981-13-6854-7_1
- Miao, F, Shiohira, K and Lao, N (2024) *AI Competency Framework for students*, UNESCO. <https://www.unesco.org/en/articles/ai-competency-framework-students>
- Suriano, R, Plebe, A, Acciai, A and Fabio, R (2025) Student interaction with ChatGPT can promote complex critical thinking skills, *Learning and Instruction* 95, 102011. <https://doi.org/10.1016/j.learninstruc.2024.102011>
- Vygotsky, L S (1987) *The Collected Works of L. S. Vygotsky Volume 1. Problems of General Psychology*, Springer. <https://doi.org/10.1007/978-1-4613-1655-8>

Appendix 1: Pre-survey items

1. Have you used GenAI tools (e.g. ChatGPT) for learning English before?
2. Which GenAI tools have you used? (multiple)
3. If you have used GenAI before in learning English, what did you use GenAI for?
4. Were you satisfied with the work with GenAI?
5. Do you think it is a good idea to use GenAI at EF?

6. How confident are you in your ability to use phrasal verbs correctly?
7. What is the most challenging aspect of learning phrasal verbs?
8. Do you think phrasal verbs are important for English proficiency?

9. Do you believe GenAI tools can help you learn phrasal verbs effectively?
10. What do you expect GenAI to do when you ask it about phrasal verbs?
11. What concerns do you have about using GenAI for learning?

12. What is your native language?
13. What is your gender?
14. How old are you?
15. What is your highest degree?
16. How long have you been studying English?
17. Which proficiency level do you think you belong to?

Appendix 2: AI literacy and prompt-engineering practices

Name:

AI Literacy

C2 high

Introduction: What Do You Know About Large Language Models (LLMs)?

Instructions: Read each statement and write T (True) or F (False).

1. Large language models always give the same answer if you ask the same question twice.
2. ChatGPT is trained to predict the next word in a sentence, not to search for facts.
3. Large language models are designed to sound fluent, not to guarantee accuracy.
4. LLMs are a good source for reliable medical, legal, or financial advice.
5. A "hallucination" is when an AI tool gives incorrect information confidently.
6. You should always double-check AI-generated content with trusted external sources.
7. Large language models can help you explore language patterns and generate writing ideas.
8. ChatGPT is the same as Google Search.

Task 1. Critical Thinking: Deep Dive into Phrasal Verb Usage

Goal: Learn to use each phrasal verb accurately in grammar, context, tone, and sentence structure.

Step 1: Use resources to identify patterns

Instructions: For each of the three phrasal verbs below:

- Look them up in a reliable source (e.g. Cambridge, Longman).
- Find **three example sentences using ChatGPT**.
- Underline the verb phrase and note the **grammatical pattern**:
 - Separable or inseparable?
 - Transitive or intransitive?

Phrasal verb	Meaning	Example sentence	Grammatical pattern
Fall back on			
Fall through			
Fall about			

karen.kim@ef.com

Name:

AI Literacy

C2 high

Step 2: Analyse Usage: Context and Tone

Instructions: Use ChatGPT for each phrasal verb:

- Identify the **typical context** (e.g. money, failure, emotion).
- Identify the **tone** (formal/informal/neutral) and explain why.

Phrasal verb	Common contexts	Tone (Formal/Informal)	Explanation
Fall back on			
Fall through			
Fall about			

Step 3: Create Your Own

Instructions: Based on the structure and tone you observed from ChatGPT, write one original sentence for each phrasal verb. Make sure:

- It follows the same grammatical pattern (e.g. transitive/intransitive, word order).
- It fits an appropriate context and tone.

Example (fall back on): *When his first business failed, he fell back on his savings.*

Your Turn:

- Fall through: _____
- Fall about: _____

Step 4: Evaluate AI Feedback

Instructions: Use ChatGPT to check:

- Is the grammar correct?
- Is the phrasal verb used in the correct context and tone?
- Do you think ChatGPT's corrections are valid?

Phrasal verb	Correct? (ChatGPT Y/N)	Did you correct it? (Y/N) why?	Final sentence
Fall back on			
Fall through			
Fall about			

karen.kim@ef.com

Name:

AI Literacy

C2 high

Task 2. Learning-to-Learn: Match Goal to Prompt

Match each learning goal with the most appropriate prompt. Write the correct letter.

Goals:

- A. Identify subtle differences between similar phrasal verbs
- B. Learn topic-specific vocabulary
- C. Practice using phrasal verbs in writing

Prompts:

1. Can you give me 5 phrasal verbs with "get" used in business writing, and explain what each one means with examples?
2. Explain the difference between "take on," "take over," and "take in," including example sentences.
3. Create a short story using 7 advanced phrasal verbs related to teamwork, and underline each one.

Task 3. Learning-to-Learn: Iterative Prompting

Use the model below to gradually improve your prompt.

Model:

- Step 1: What does "bring round" mean?
- Step 2: What does "bring round" mean in a meeting context?
- Step 3: Give 3 business-related example sentences using "bring round" and explain the tone (formal or informal).

Now you try with a different phrasal verb:

- Step 1: _____
- Step 2: _____
- Step 3: _____

Compare your prompts and results with your partner.

karen.kim@ef.com

Name:

AI Literacy

C2 high

Task 4. Reflection: Diagnostic Prompt Practice

Reflect on a recent time you used ChatGPT to learn phrasal verbs.

1. What was your original prompt? Was it specific enough?
2. What kind of answer did you get? Was it useful? Why or why not?
3. How would you improve the prompt to get a better learning outcome?

Task 5: Critical Evaluation of AI-Generated Phrasal Verb Examples

Instructions: Use ChatGPT to generate three advanced-level phrasal verbs used in Cambridge Writing Task 1 essays, each with a formal synonym and example sentence.

Take each step you have learned so far.

Your prompts here:

- Step 1: _____
- Answer: _____
- Step 2: _____
- Answer: _____
- Step 3: _____
- Answer: _____
- Step 4: _____
- Answer: _____
- Step 5.: _____
- Answer: _____

Did you ask something questionable?

Did you cross-check the results?

Do you think you can use ChatGPT for your learning?

karen.kim@ef.com

Appendix 3: Five-step explicit instruction

Step 1. Diagnostic test

Part 1. PV meaning → Part 2. Synonymous word to PV → Part 3. Meaning inference in a given sentence

Part1: Choose the correct meaning of each phrasal verb (1-10)

1. back down (1 Point)

- ☐ support someone
- ☒ stop opposing something ✓
- ☐ cancel a plan

Part2: Complete the sentences with a phrasal verb

11. I need to _____ this problem before making a decision (investigate) (1 Point)

Enter your answer

Correct answers:

Part3: Multiple Choice of Phrasal Verb Usage

16. To "carry out" an experiment means to (1 Point)

- ☐ design it
- ☒ complete and perform it ✓
- ☐ forget about it

Step 2. Inferring the meaning of PVs and identifying their usage from the context

C2 High

Advanced Phrasal Verbs in Language and Literature

1. Highlight the 9 phrasal verbs that you think are the focus today.
2. Guess the meanings based on the context.

The recent initiative by the Language and Literature Society to modernise its writing competition encountered both enthusiasm and controversy. Initially, one participant attempted to pass off an existing published work as his own original composition, an act of academic dishonesty that was promptly uncovered. In response, the adjudicators laid into the offender with sharp criticism, emphasising the importance of intellectual integrity in literary pursuits.

Concurrently, a coalition of students chose to stand up to the Society's long-standing preference for canonical forms, advocating instead for greater acceptance of innovative writing styles. Among them, one particularly articulate student stood up for emerging writers, arguing that contemporary voices were essential for the ongoing vitality of literary traditions. Throughout the dispute, certain media outlets played up the controversy, portraying it as a generational clash over the future of literary standards.

After considerable negotiation, the organising committee was eventually talked into revising the competition framework. In a surprising outcome, a relatively unknown entrant edged out more established competitors with a provocative piece of digital storytelling. Recognising the significance of these developments, the Society agreed to phase in new categories over successive competitions, thus expanding the definition of literary excellence. Simultaneously, they announced measures to clamp down on academic misconduct, affirming their commitment to maintaining rigorous ethical standards in all future events.

karen.kim@ef.com

Step 3. Explicit explanation with a definition, syntactic structure, tone, and context

C2 High

Definitions with Example Sentences

Watch for,

1. **Syntactic Structure:** Is the PV separable or inseparable? Transitive or intransitive?
2. **Tone:** Does the usage suit a formal, Informal, or neutral setting?
3. **Context:** Is it appropriate for situations like business negotiation, academic writing, or casual conversation?

1. He **passed** himself **off** as a qualified engineer. (separable / with object)
Definition: Pretend someone is something else

2. She **laid into** him for arriving late. (Inseparable / with object)
Informal: criticise harshly

3. Many workers are afraid to **stand up to** their manager. (inseparable)
Definition: Make a spirited defence against someone/something

4. She learned to **stand up for** what is right. (inseparable)
Definition: Speak or act in support of someone/something

5. She **talked** her brother **into** joining the debate team. (separable)
Definition: Persuade someone

6. The media is **playing up** the scandal. (separable / with object)
Definition: Exaggerate or emphasise

7-1. She **edged out** the competition by just one point. (separable / with object)
Definition: (British English) Defeat by a small margin

7-2. She was edged out of the organisation by the director.
Definition: Remove a person from a role by indirect means

8. The policy will be **phased in** over six months. (separable / with object)
Definition: Introduce something into use in gradual stages

9. Authorities are **clamping down on** misinformation. (inseparable)
Definition: Suppress or prevent something in an oppressive or harsh manner

karen.kim@ef.com

Step 4. Application of PVs in an authentic speaking situation

C2 High

Speaking Activity : From Formal to Informal

1. Read the formal story carefully.

During a high-stakes diplomatic summit, negotiators from both nations **achieved what many considered impossible—a mutual agreement on arms reduction**. The lead diplomat had accomplished similar outcomes in the past, but this negotiation was particularly complex due to years of hostility.

Several unresolved issues emerged, but both sides worked together to resolve them before the treaty signing. The accord also included provisions to eradicate illicit arms trading networks that had long destabilised the region. Intelligence reports revealed that previous black-market operations had already been destroyed completely during the campaign.

However, the agreement itself triggered domestic backlash in both countries, particularly among hardliners. In one nation, peaceful dialogue soon collapsed, leading to the postponement of implementation.

2. Complete the informal conversation by replacing the formal verbs with their phrasal verb equivalents. Use the correct tense and form in each blank.

Ava: Did you hear about that summit last week?

Leo: Yeah, I still can't believe they actually **brought off an agreement on arms reduction**. That was huge.

Ava: Right? I mean, the lead guy has _____ tricky deals before, but this one was next-level.

Leo: Totally. And even when issues popped up, they managed to _____ before the deadline.

Ava: Plus, they're finally doing something to _____ all that illegal arms trading. About time.

Leo: I read they already _____ a bunch of black-market operations. Crazy.

Ava: But then the whole thing _____ a political storm back home. People are furious.

Leo: Yeah, and now talks are falling apart. Communication's basically _____ again.

karen.kim@ef.com

Step 5. Learners' own use of PVs with ChatGPT

AI in the conversation: Enhancing group speaking through feedback informed by student voice

Natasha Sarooshi, English Language Centre (ELC), Adelaide University

Introduction

Collaboration and communication skills are essential for university-level English language learners, equipping students with the ability to work effectively in diverse teams, negotiate ideas, and problem-solve in academic and professional contexts. These skills align with the collaboration component of The Cambridge Life Competencies Framework (Cambridge University Press & Assessment 2026), which emphasises the importance of adaptability, responsibility, and task management in fostering group success.

Oral group activities are vital for developing language proficiency and collaborative competencies. However, providing timely, individualised formative feedback during such activities is increasingly challenging. In my classes, factors such as large class sizes, complexity of task requirements and outcomes, along with an increasing workload makes real-time, detailed feedback very difficult for me to produce on a constant basis. Students must collaborate in pairs, small teams, and larger groups — but it becomes virtually impossible to observe and give immediate personalised feedback to every group interaction.

As Hung (2016) discussed, digital technologies offer promising solutions to enhance feedback delivery. Recent literature (Alsaiani et al 2024) also emphasises how generative AI (GenAI) can potentially revolutionise educational feedback by delivering personalised, instantaneous responses. Although there is an increasing body of research investigating feedback practices involving GenAI and human teachers, the overwhelming majority of these studies have focused on written tasks

– particularly grammar correction, essay drafting, and formative writing feedback (e.g., Henderson et al 2025).

Given this focus on writing, I perceived that there was a noticeable gap in research exploring how GenAI can support spoken academic tasks such as those in my class – particularly collaborative group speaking activities – where feedback must not only address individual performance but also group interaction dynamics.

Motivated by these insights, my action research (AR) sought to explore the application of ChatGPT Premium as a formative feedback tool during oral group activities. In addition, it specifically aimed to investigate both student preferences and pedagogical outcomes related to both GenAI and my feedback in the context of oral group assessments.

Educational context and setting

The participants in this research were two cohorts of students, one on a 20-week and the other on a 10-week full-time program, known as the Pre-Enrolment English Program (PEP), at Adelaide University's English Language Centre. The students study four hours face to face each day for the duration of their course. With an exit level of IELTS 6.5, the course is a direct-entry pathway to the University, with students preparing to enter a broad range of undergraduate and postgraduate majors.

The participants were aged between 18 and 36 with most coming from mainland China, in addition to South Korea, Japan and Vietnam (Figure 1). Cohort 1 comprised 17 students, 10 postgraduate and seven undergraduate students. They entered the program with an IELTS score of 5.0 and were enrolled in a 20-week course which started in February 2025, and all 17 students completed it in July 2025. Cohort 2 comprised 15 students, 11 postgraduate and four undergraduate students (see Figure 2 for the combined cohort university pathways). They entered the course with an overall IELTS score of 6.0 and were enrolled in a 10-week course which started in May 2025. Although 15 students started, for varying reasons only 12 completed the course in July 2025.

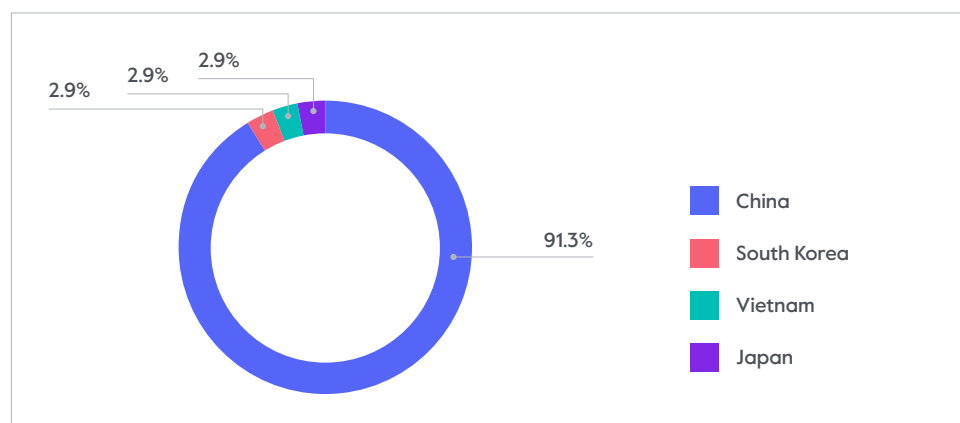


Figure 1: Combined Cohort 1 and 2 student demographics by country

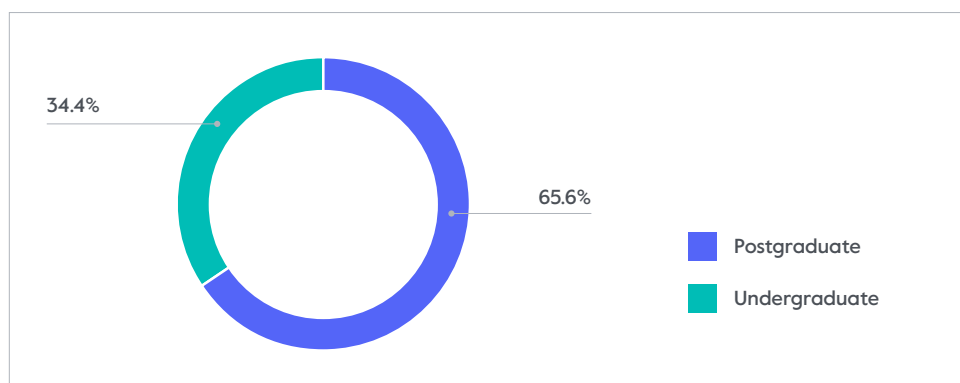


Figure 2: Combined Cohort 1 and 2 students' future university study pathways

In the PEP course there is an equal emphasis on all four language skills. Speaking tasks vary from individual and group oral presentations to pair and group discussions, and seminars with oral assessments form a significant component of the course. One of the major assessments is an integrated skills test (IST) in which students are required to partake in a three-way discussion, with two other students, on a specific topic. The questions are scaled in increasing difficulty using Bloom's taxonomy framework (Anderson and Krathwohl 2001).

Learning goals focus on enhancing academic speaking proficiency, particularly communication (confidence, clarity, interaction facilitation) and collaboration (effective group engagement), as outlined in the Cambridge Life Competencies Framework (Cambridge University Press & Assessment 2026). The speaking component of the IST was used as the specific activity which the research centred around.

Research questions

The original title of my AR, 'Investigating how GenAI can support group activities by providing feedback on collaborative interactions,' reflected a broad focus on leveraging GenAI to enhance the learning outcomes and development from collaborative interactions.

Embedded within this initial focus was a secondary curiosity I had as to whether my students would show a preference for feedback provided by GenAI as opposed to my more traditional teacher-type input. As the first research cycle progressed, it became evident that the students were more immediately responsive to questions of preference, and much of the data I collected leaned in this direction. As a result, the initial cycle of research, while still addressing collaborative feedback, became somewhat skewed toward exploring the comparative value of feedback sources.

After discussion at the second workshop for the Action Research in ELICOS Program, my second cycle marked a deliberate focus back to the core pedagogical aim: the development of collaboration skills in oral group tasks. Also, from my observation throughout Cycle 1, it became evident further elements of the Cambridge Life Competencies Framework were at the centre of the students' interactions.

To refine this focus, the research in Cycle 2 therefore aligned more explicitly with two domains of the Framework (Cambridge University Press & Assessment 2026):

- *Collaboration*: encouraging effective group interaction
- *Communication*: facilitating interactions and participating with appropriate confidence and clarity

This adjustment ensured that the research remained grounded in meaningful pedagogical outcomes while still allowing for comparative feedback analysis. Accordingly, the research questions (RQs) were revised and finalised as follows:

1. Does AI feedback contribute to the development of learners' collaborative and communicative skills?
2. To what extent do learners value AI feedback compared to teacher feedback?

This evolution of the focus reflects the iterative and responsive nature of AR (Burns 2010), ensuring both cycles informed a more balanced and educationally relevant investigation.

Research design and procedures

Oral communication encompasses numerous components, making it time-consuming to collect data on every individual element. Therefore, for the purposes of this project, I limited the focus to the specific aspects related to interaction rather than language accuracy. These areas were mapped to elements within the Communication and Collaboration strands of the Cambridge Life Competencies Framework.

The project spanned two 5-week cycles (one cycle per cohort) and involved multiple stages. The design of my AR was grounded in the feedback stages framework devised by Zhan et al (2025) listed below; this structured approach ensured parity in the purpose of both GenAI and teacher forms of feedback.

- A. *Feedback Forethought* (students' mindset prior to receiving feedback)
My tools: questionnaires, small-group discussions, explicit teaching of feedback literacy
- B. *Feedback Control* (active engagement with and application of feedback)
My tools: reflection questions to guide understanding, student action plans
- C. *Feedback Retrospect* (reflection on feedback outcomes)
My tools: targeted reflection questions, Mentimeter word clouds (www.mentimeter.com), questionnaires, focus groups and exit surveys

Figure 3 outlines the 5-week cycle and the multiple stages involved in the process.

Week 1	Week 2	Week 3	Week 4	Week 5
Forethought Initial questionnaire, introduction to feedback literacy, and group discussion Speaking task with teacher feedback	Control IST* 1A speaking assessment (implementing teacher feedback)	Speaking task with AI feedback	Control IST 1B speaking assessment (implementing AI feedback)	Retrospect Exit questionnaire and focus groups
Control Reflection, and action plan created	Retrospect Questionnaire and Mentimeter	Control Reflection and action plan created	Retrospect Questionnaire and Mentimeter	

Figure 3: Implementation stages

*IST = Integrated skills test

Weeks 1 and 2

Initial questionnaire and discussion: These tools gathered students' initial perceptions of AI's usefulness for feedback and compared it with attitudes towards teacher feedback. I also introduced feedback literacy strategies: how to engage, interpret and act upon feedback.

Teacher feedback phase: Students undertook a speaking task where I provided feedback (Appendix 1). Students then completed general reflection questions (the same for all students) and developed individual action plans based on this feedback. Using their action plans, students participated in a formative IST speaking assessment and then evaluated the usefulness of the feedback to complete the assessment.

Weeks 3 and 4

Training ChatGPT Premium: The AI was trained using the same criteria for feedback that I used: targeting the communication and collaboration skills of participation, meaningful interaction, linking of ideas, and overall fluency.

Student training: Students were trained on how to engage effectively with ChatGPT, interpret feedback prompts, and apply suggestions.

AI feedback phase: Students undertook the same speaking task as before, however this time supported by ChatGPT-generated feedback. The AI provided detailed and immediate written feedback summaries post-activity. Reflective questions that were specific to each student and actionable suggestions were also given to guide improvement for the IST formative speaking assessment in Week 4 (Appendix 2). Students' views on AI feedback and its effectiveness for the assessment task and their skills development were then subsequently gathered.

Week 5

Gathering students' views: Students' views on the AI feedback were gathered via questionnaires and focus groups. These were reflections on whether they felt the feedback contributed to their communication and collaboration skills development, with the possibility of continued use. The tools also collected students' perception of GenAI versus my feedback and which they preferred.

Data collection methods

The data was collected using a range of methods to capture the various stages of the feedback process. The data sources were:

- Pre- and post-intervention questionnaires
- Focus group interviews
- Student reflection tasks and action plans (Appendix 3)
- Post-assessment questionnaires and Mentimeter word clouds of students' views
- Photographs and audio-visual recordings of group interactions (Appendix 4 and Appendix 5)
- Teacher reflection journals

Data analysis and findings

The research began with an exploration of students' initial ideas and general perceptions surrounding the use of AI and the feedback generated by AI tools, before any feedback was given to them. When both cohorts of students were asked if they thought AI could provide useful feedback on their formative assessment tasks, 60% responded that they either agreed or strongly agreed, while 40% said they neither agreed nor disagreed or strongly disagreed. This response suggests that there is a general level of optimism or acceptance among students regarding the use of AI to provide feedback on formative assessments.

Students were also asked if they thought they would prefer feedback from a teacher or GenAI. From Figure 4 it could be initially thought that most prefer feedback that involves teachers, with over half favouring teacher-led feedback supported by AI, indicating strong trust in human judgement and openness to AI as a complementary tool.

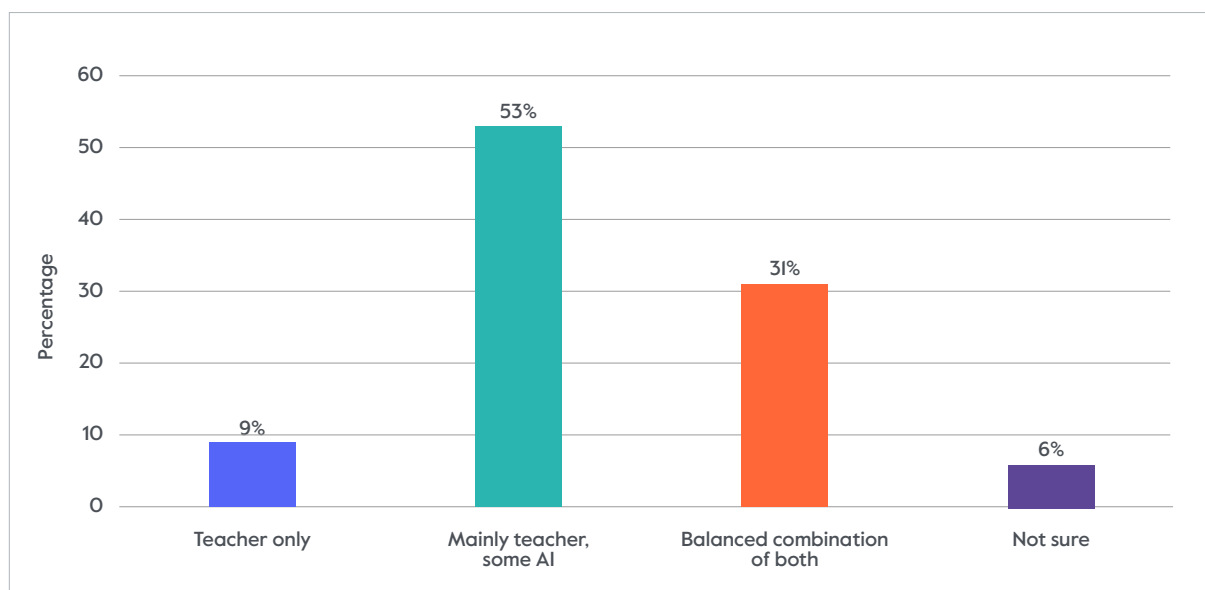


Figure 4: Cohort 1 and 2 preference type for feedback in formative assessment

As can be seen below in Figures 5 (Cohort 1) and Figure 6 (Cohort 2), students felt there were potentially many advantages that could stem from receiving feedback via GenAI.

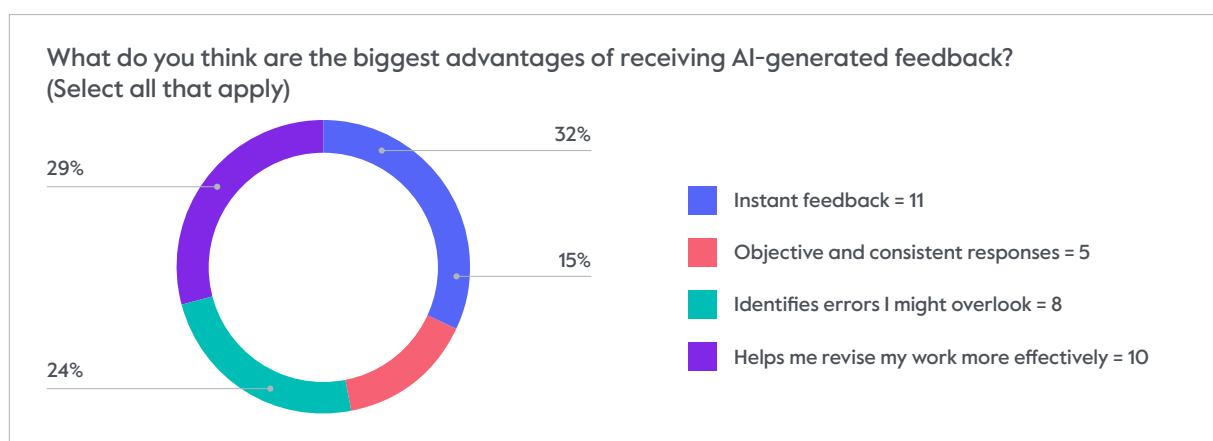


Figure 5: Cohort 1 response to advantages of receiving GenAI feedback

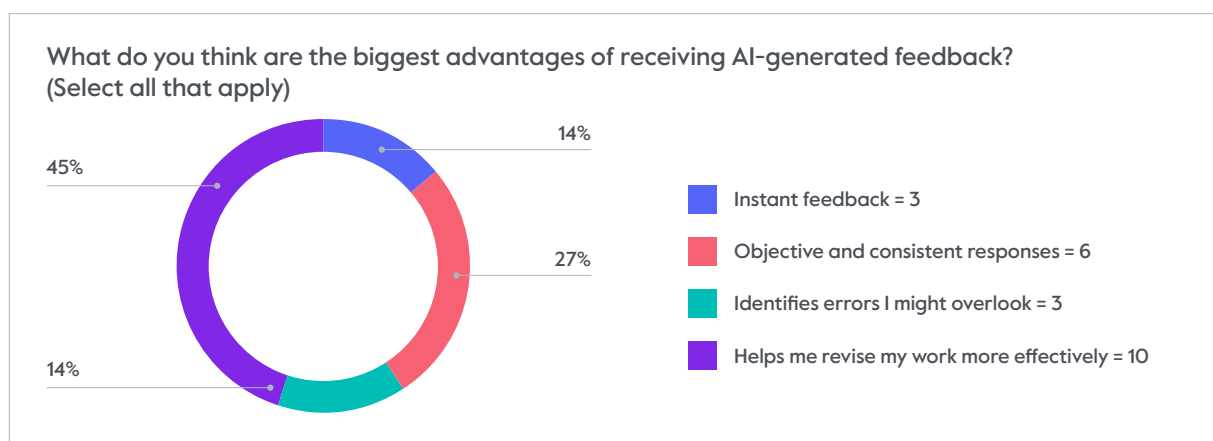


Figure 6: Cohort 2 response to advantages of receiving GenAI feedback

There were also initial concerns expressed by the students in both cohorts as to why they may be hesitant to either start to engage with or continue to use GenAI as a feedback tool; see Figure 7 (Cohort 1) and Figure 8 (Cohort 2). The main concerns focus on its potential lack of human understanding, impersonal responses, risk of reinforcing errors, and the fear of overdependence, highlighting a strong desire for empathetic, accurate, and teacher-guided support.

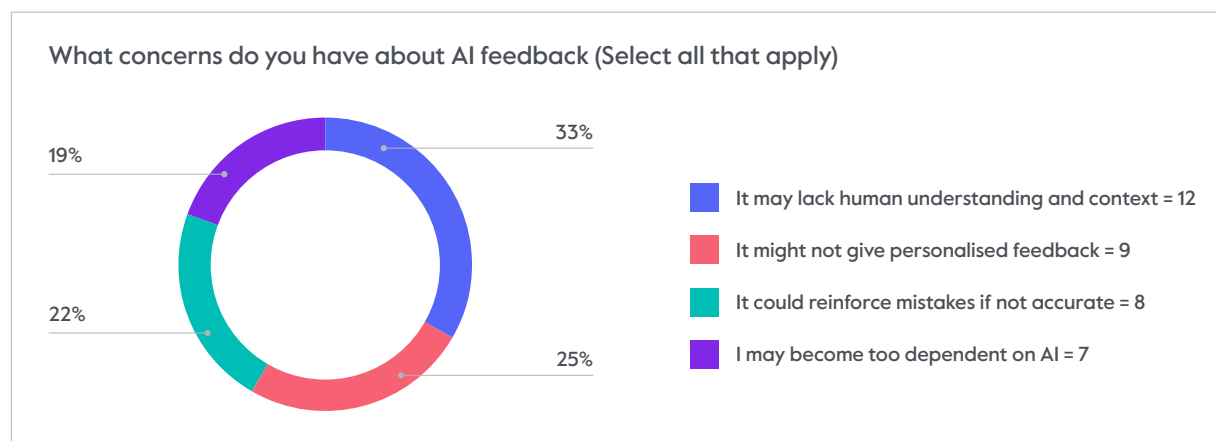


Figure 7: Cohort 1 concerns regarding GenAI feedback

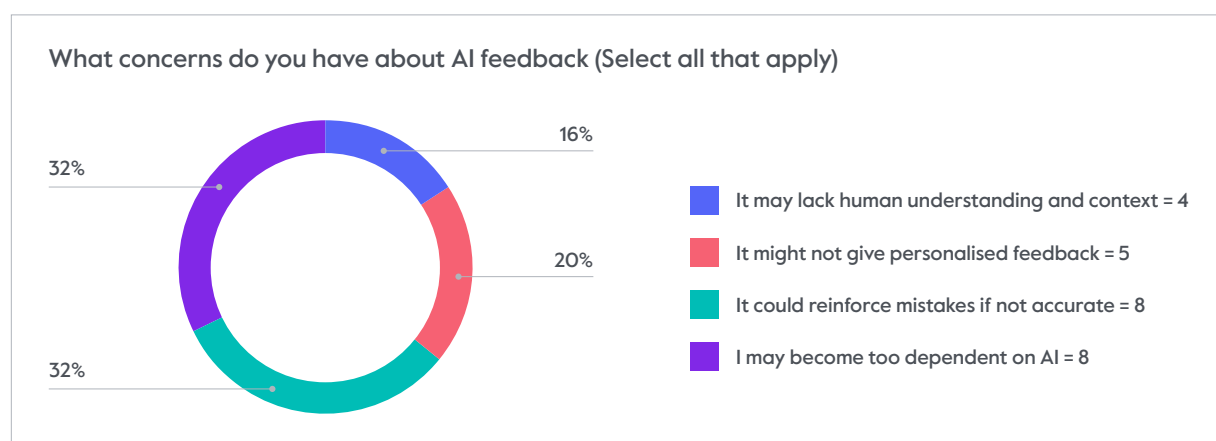


Figure 8: Cohort 2 concerns regarding GenAI feedback

Research Question 1: Impact on communicative and collaborative skills

Before the tasks, students were also specifically asked about GenAI feedback in relation to the development of the Cambridge Life Competencies of communication and collaboration. Students were curious yet slightly sceptical about AI's ability to support group interaction and communication development. Figure 9 indicates that over half of the students believed AI could support communication skills, but the low strong agreement and high neutrality suggested uncertainty and a need for a convincing experience. Figure 10 shows nearly half believed AI can support collaboration skills, but the equal proportion of neutrality and disagreement highlights uncertainty about AI's role in fostering interpersonal abilities.

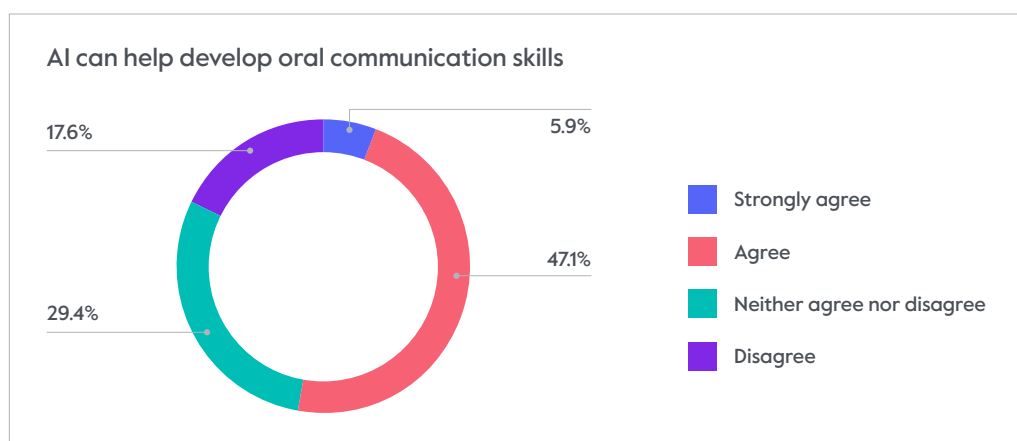


Figure 9: Student perceptions of AI assisting oral communication skills in group tasks

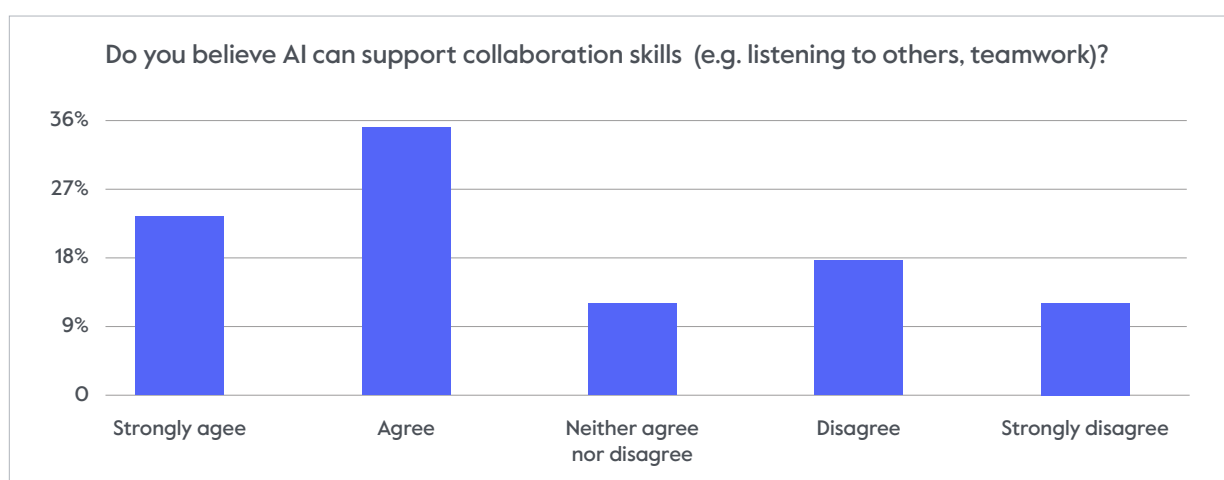


Figure 10: Student perceptions of AI assisting collaborative skills in group oral tasks

Once the AI feedback was given, all reported it as helpful for improving speaking, citing clarity, specificity, and actionable suggestions, with 41% claiming it was very helpful and 59% saying it was helpful. No student in either cohort felt the feedback was neither helpful nor unhelpful, somewhat unhelpful or completely unhelpful. In addition, they also felt the feedback assisted them in improving their speaking skills, as shown in Figure 11 and Figure 12 as well as the comments from student reflections.

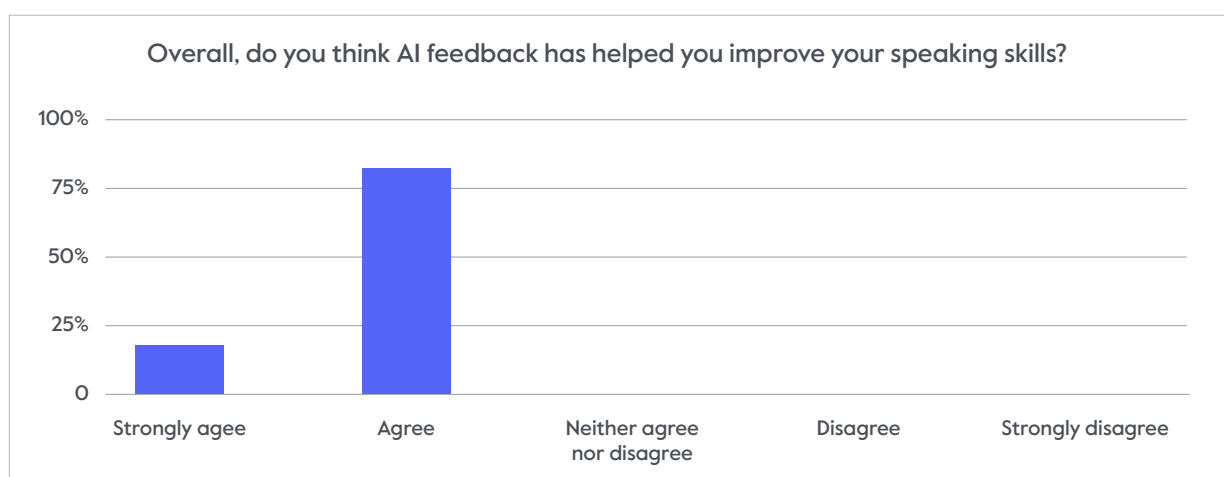


Figure 11: Student perceptions of GenAI feedback helping to improve speaking skills

Survey comments:

- 'It gave me the advice that let me know what I need to improve.'
- 'I think AI feedback is very good, it give me a professional response to help me.'
- 'I think I am more active today because of the AI feedback from last week.'
- 'AI feedback help me now think more before the speaking and organising my point.'
- 'I can speak better because make me no longer afraid to speak.'



Figure 12: Mentimeter word cloud of students' views on GenAI feedback

In regard to communication skills, students reported gains in confidence, clarity of expression, and ability to initiate and sustain discussion, as shown in Figure 13. AI's prompts as well as tailored specific reflection questions encouraged more fluent participation.

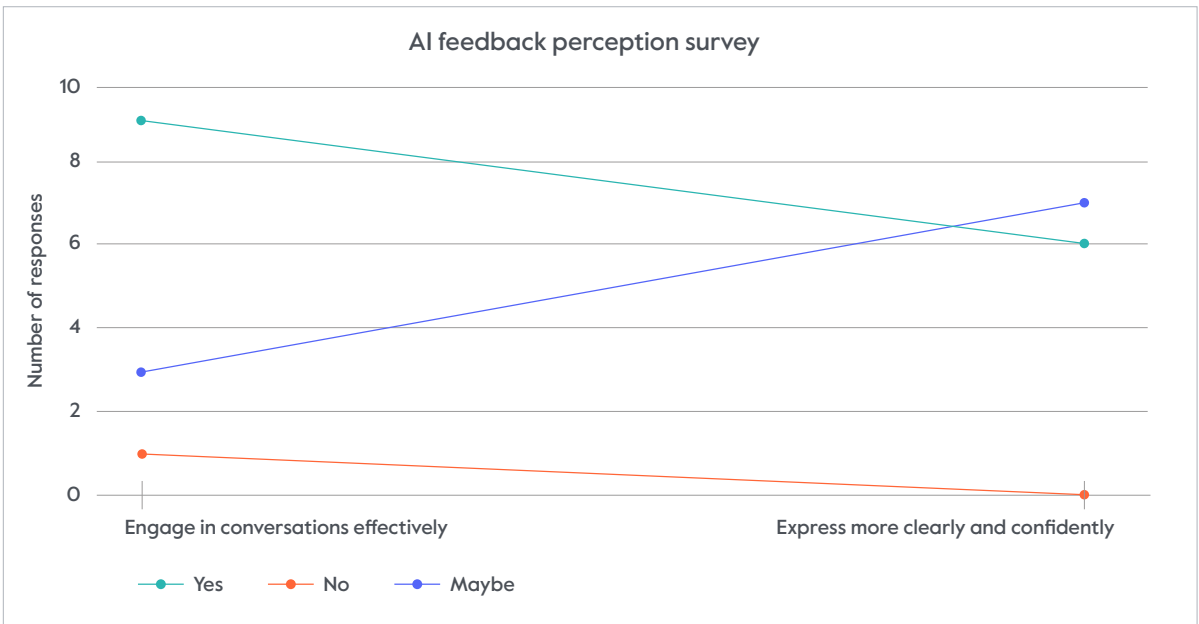


Figure 13: Student perceptions of GenAI feedback on communication skills

For collaboration skills, although some group dynamic limitations were noted (e.g., misattribution of turns and cultural bias in name identification), students appreciated AI's effort to evaluate group interactions and promote engagement. When asked if they felt the AI feedback helped them actively take part in group activities, 54% said yes, 38% said maybe, and only 8% said no. In addition, the AI feedback seemed to assist students with listening more to others in group situations, with 77% confirming yes it did, 15% saying maybe, and 8% saying no.

Students felt clear and actionable guidance was given by AI. They valued AI's specific suggestions, non-judgemental tone, and instant availability, aiding self-regulation and preparation for subsequent tasks. With both cohorts, when asked if the action plan they created based on their AI feedback and reflection helped improve their skills development and performance, 25% strongly agreed and 67% agreed, with the remaining 8% saying they neither agreed nor disagreed. They felt it was a tool that they could and would continue to use to support their speaking skill development. Student comments from reflection questionnaires further explain these statistics.

'Yes, AI told me I needed to improve fluency, so today I choose wait other to speak first and then I speak, so that I have more time to think.'

'It clearly pointed out my problem and at the same time told me how to solve it.'

'Quite clear what I should do in the discussion. Especially, I would use same words. Today, I tried to avoid it.'

'Yes it told me need to improve my more details and I do this.'

Research Question 2: Preference and valuation of feedback sources

Prior to the intervention, most students anticipated my feedback would be more valuable and doubted AI's ability. This was discussed earlier in Figure 4. From the initial discussions with a focus group in the forethought part of the feedback stages framework (as outlined in Figure 3), the following ideas were expressed (S=Student, T=Teacher):

1. S: *'I have used AI once for feedback but prefer T feedback.'*
T: *'Why?'*
S: *'Because the teacher can explain it better.'*
2. S: *'I think AI can give academic feedback, but T can give the emotional side as well which AI can't and is important for me.'*
3. T: *'Can AI be useful in sorting out group work issues?'*
S: *'No, it is not useful for groupwork – we should learn to co-operate between humans. If we listen and respond to AI we are just like robots as well. We need to work out human situations as humans.'*

Following AI-supported tasks, students acknowledged the tool's convenience, impartiality, and clarity. Regarding convenience, some students appreciated AI's instantaneous and specific feedback, highlighting its accessibility and availability.

'When I checked my feedback, AI totally noticed my weakness point and told me straight away.'

'AI is good to use as it is convenient and useful specific to my need.'

AI was also seen as impartial, with several students commenting that it provides non-judgemental feedback, which made them feel safer when experimenting with language.

'Speaking in front of AI won't be as nerve-racking as speaking in front of a teacher.'

Also, AI's clarity was appreciated as it clearly pointed out the problem(s) and at the same time gave suggestions of how to solve it.

'AI pointed out my mistakes in some direct language, this enables me to better understand my mistakes and weaknesses.'

However, as Figure 14 shows, students across both cohorts ultimately preferred my feedback for its depth, trustworthiness, and personal connection. Nonetheless, it needs to be noted that AI was viewed as a useful complement, particularly for low-stakes, quick feedback and for reducing anxiety. There was also some conditional preference for AI feedback where they felt it would be useful in large, impersonal academic settings where teacher access might be limited.

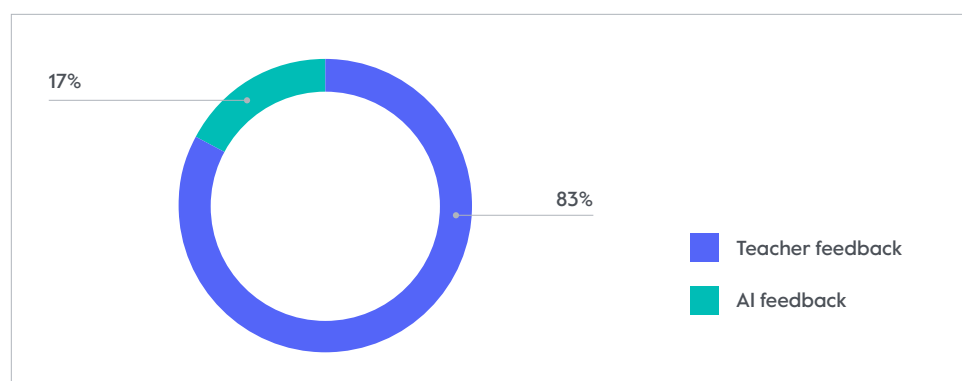


Figure 14: Student preference for teacher versus GenAI feedback for Cohort 1 and 2

The preference for my feedback was based on the personalisation that I provided. They commented that I could explain subtleties directly related to them, and that there was an emotional connection between the student and myself, which is something AI lacks. The holistic perspective was also important for them. They felt that I was able to respond to wider aspects of communication (such as non-verbal) and not just what the prompt explicitly asks for.

There were three main reservations related to AI feedback. First, its lack of emotional resonance. Students felt AI feedback was somewhat generic and lacked emotional intelligence. It only responded to specifically what was asked of it.

'AI cannot see any body language in the group and read the emotions; it only listens what it is programmed to listen.'

'Actually, they both do it [feedback] but only teacher knows about body language.'

'Teacher knows me, and my problems better so give better feedback.'

Secondly, their scepticism about AI properly understanding group dynamics. They occasionally experienced this limitation, as it misidentified certain students and omitted others entirely from the discussion.

The third reservation some students expressed was that interacting with AI made conversations feel less natural and less enjoyable. They mentioned they were not machines so didn't want to interact with a machine.

'It's not fun to talk to AI; I want to speak to real people.'

'AI doesn't really evaluate the interaction, it is generic/ordinary, only what the machine has been pre-programmed to notice.'

Conclusion

The findings show that while students continued to place the higher value on my teacher feedback, AI nevertheless offers meaningful support for communication and collaboration skills in group speaking tasks. AI is particularly helpful in supporting formative learning cycles through immediate, low-pressure feedback. When aligned with pedagogical frameworks such as Zhan et al (2025), AI can effectively scaffold students across the three stages of feedback: forethought, control, and retrospect. Ultimately, rather than competing with me, AI feedback can serve in a complementary role.

Cycle 2 enriched the research by focusing more explicitly on the collaborative and communicative competencies connected to the Cambridge Life Competencies Framework. The students felt using AI could help develop these skills and many would continue to utilise AI as the feedback was clear and actionable enough to understand.

Contextual constraints highlighted by Zhan and Yan (2025) such as time scarcity, class sizes, and workload were evident in my class and it was clear AI's availability had the potential to partially mitigate these. Nevertheless, for the moment, my classroom remains a place where my feedback is preferred due to the developmental guidance, empathy, and authenticity I am able to provide.

These findings align strongly with Henderson et al (2025), who conducted a large-scale study of student perceptions of GenAI and teacher feedback in relation to written assignments, involving 6,960 students across four major Australian universities. This study found teacher feedback was trusted more (90.5%) than GenAI (60.1%), with the conclusion drawn that the forms can be complementary rather than interchangeable, and that students benefit most when both sources are available.

Reflections on the impact of the research

The research highlighted feedback as an interactive, staged process rather than a one-time event. Student engagement was supported across Zhan et al's (2025) three-stage model, helping them actively plan for, apply, and reflect on both types of feedback. Students appeared to benefit most when they possessed a degree of feedback literacy and were provided with clear, actionable tasks related to the feedback – whether it came from myself or AI. The process for engaging with feedback remained consistent across AI and myself, so students were able to transfer the strategies they learned from my feedback to their interactions with the AI-generated variety. This consistency supported not only their uptake of feedback but also contributed meaningfully to the development of their AI literacy.

While AI enhanced the accessibility and immediacy of feedback, students consistently expressed a greater level of trust in my feedback. This trust stemmed from my ability, as a human educator, to interpret non-verbal cues, understand the broader context of group interactions, and respond to each student as a whole person. Their responses highlighted that while AI can provide an additional perspective, it is not perceived as a substitute for the relational and contextual depth of teacher feedback. Rather, students viewed AI as a valuable supplement – particularly effective when used alongside human feedback in socially complex tasks and culturally diverse learning environments.

Students showed a readiness to use AI to support their oral skill development, noting that it could be especially helpful in extending learning beyond the classroom. This suggests that AI has potential as a supplementary tool for fostering independent learning, offering accessible support when teacher guidance is not immediately available.

Impact on classroom practice and potential ELICOS implications

This AR may offer some practical possibilities for enhancing feedback in group speaking tasks, an area often limited by time constraints and teacher workload. By integrating AI alongside the teacher, it enables students to receive more immediate and – in some instances – individualised input, improving access without compromising task authenticity.

It also attempts to address a common gap in ELICOS contexts: supporting real-time communication and collaboration in spoken tasks. Students can use AI to support a range of group speaking tasks they may wish to practise beyond the classroom, thereby extending opportunities to develop their communicative and collaborative skills. This allows them to engage in practice without relying on a teacher to initiate the activity, with the added benefit of receiving immediate and constructive feedback.

Teachers can design speaking activities in which AI functions as a continuous feedback agent, allowing the teacher to focus their attention on targeted groups or individuals who may require additional support. Unlike the teacher, who may only be able to observe brief moments of each group interaction, the AI can monitor and provide feedback on the entirety of the group discussion, offering a more comprehensive view of student participation and performance.

Finally, it is important to continue to teach critical AI literacy to help learners evaluate the quality, accuracy, and limitations of AI feedback. For example, teachers can highlight when AI gives generic or culturally inappropriate responses, as was experienced during this project.

References

- Alsaiani, O, Baghaei, N, Lahza, H, Lodge, J, Boden, M and Khosravi, H (2024) *Emotionally enriched feedback via Generative AI*, Preprint, The University of Queensland. <https://arxiv.org/pdf/2410.15077>
- Anderson, L W and Krathwohl, D R (Eds) (2001) *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*, Longman.
- Burns, A (2010) *Doing Action Research in English Language Teaching: A Guide for Practitioners*, Routledge.
- Cambridge University Press & Assessment (2026) *Cambridge Life Competencies Framework*. <https://www.cambridge.org/clcf>
- Henderson, M, Bearman, M, Chung, J, Fawns, T, Buckingham, S, Matthews, K E and Heredia, J (2025) Comparing generative AI and teacher feedback: student perceptions of usefulness and trustworthiness, *Assessment and Evaluation in Higher Education* 51, 863–878. <https://doi.org/10.1080/02602938.2025.2502582>
- Hung, S T (2016) Enhancing feedback provision through multimodal video technology, *Computers and Education* 98, 90–101. <https://doi.org/10.1016/j.compedu.2016.03.009>
- Zhan, Y and Yan, Z (2025) Students' engagement with ChatGPT feedback: implications for student feedback literacy in the context of generative artificial intelligence, *Assessment and Evaluation in Higher Education* 51, 821–834. <https://doi.org/10.1080/02602938.2025.2471821>
- Zhan, Y, Boud, D, Dawson, P and Yan, Z (2025) Generative artificial intelligence as an enabler of student feedback engagement: a framework, *Assessment and Evaluation in Higher Education*, 1,289–1,304. <https://doi.org/10.1080/07294360.2025.2476513>

Appendix 1: Sample of teacher feedback

Group Speaking Activity Feedback Form

Student Name: _____ Date: 3/4/25

Discourse Management

1. Participation: (Did the student contribute actively to the discussion?)
☐ Excellent ☐ Good ☒ Satisfactory ☐ Needs Improvement
 Comments: Generally tried to contribute in a positive way

2. Meaningful Response and Contribution: (Did the student respond appropriately to others and add relevant ideas?)
☐ Excellent ☒ Good ☐ Satisfactory ☐ Needs Improvement
 Comments: Gave some basic ideas that connected to others.

Fluency

3. Linking Ideas: (Did the student connect their ideas smoothly to the discussion?)
☐ Excellent ☐ Good ☐ Satisfactory ☒ Needs Improvement
 Comments: A little more contribution to what others said would make it smoother.

4. Overall Fluency: (Did the student speak naturally and without excessive hesitation?)
☐ Excellent ☐ Good ☐ Satisfactory ☒ Needs Improvement
 Comments: Extend your contribution, not just short responses.

Final Comments & Suggestions for Improvement
You're an active participant however try to make sure you contribute more based on what others are saying - not just your idea at the beginning.

Teacher's Signature: [Signature]

Cycle 1 TFB

Appendix 2: Sample of ChatGPT feedback (same student as Appendix 1)

Speaking Task Feedback:

Performance Ratings

Participation: Satisfactory

Meaningful Interaction: Good

Linking Ideas: Satisfactory

Overall Fluency: Satisfactory

Positive Feedback

- You made regular contributions to the discussion, showing a consistent effort to stay involved.
- Your responses to other speakers were thoughtful and demonstrated good listening skills.
- You showed the ability to express your thoughts with clarity and structure.

Areas to Improve

- Although you participated consistently, you could work on initiating new discussion points to take more leadership.
- At times, your ideas could be more clearly connected to what your peers were saying. Try using linking phrases such as 'building on that point' or 'I agree because...'.
- You hesitated occasionally. Continue practicing common phrases and sentence starters to increase fluency under pressure.

Reflection Questions

1. What strategies can help you become more proactive in starting and steering discussions?
2. How can you improve the way you connect your responses to what others say?
3. What language tools can help you reduce hesitation and sound more fluent?

Appendix 3: Sample of student reflection and action plan

Student Reflection and Action Plan A1 FB

Take a few moments to reflect on your performance and answer the specific questions CHATGPT has generated for you.

Please write the question and your answer:

1. What can I do to avoid repeating the same phrase during a discussion?

I have to ~~change~~ ^{grammatical} the structure and also prepare synonyms. Or remember a various expression and apply it to casual ~~chat~~ talks.

2. How can I structure long ideas clearly for the listener?
Firstly, I state the position and emphasise some words which I want to tell others.

I need to speak logically, position → evidence, restatement!

3. What strategies can I use to increase my speaking fluency in group?

Firstly, I try to ~~express~~ ^{discuss} a degree of agreement. Using ^{various} linking words and also encourage other speakers to speak smoothly after my turn, by question.

Action Plan for Next Time

Based on your reflection and feedback, create a plan to improve your performance in the next group speaking activity. Fill in the following sections with specific strategies:

1. **Participation:** One thing I will do to engage more in the discussion is:

I try to speak my statement logically
and reduce monotone pronunciation.

2. **Meaningful Response and Contribution:** One way I will make my responses clearer and more relevant is:

I will try to use some phrase to
state an agreement -
a degree of

3. **Linking Ideas:** To connect my ideas more smoothly to others, I will:

I try to use an abstract or general
idea, and then try to explain it
by using evidence or example.

4. **Fluency:** A technique I will practice to speak more naturally and confidently is:

Using linking words and ~~increase~~ ^{expand} the vocabulary
and apply it to chat in my daily life.

By following this plan, I will work towards improving my discourse management and fluency in future group discussions.

Appendix 4: Students engaging in a discussion about initial perceptions of AI feedback



Appendix 5: Group speaking task with ChatGPT as the feedback agent



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 2026
Shaftesbury Road, Cambridge, CB2 8EA