

6 Essential Practices for Developing Teaching Resources with Generative AI

'Teaching Students to Think
With AI, Not Through It:
A Practical Gen-AI-Supported Writing
Feedback Design for EAP Classrooms

Centering the Social:
What we learnt from our
international language exchange.

IN THIS ISSUE

06 | **FEATURE**
Six Essential Practices for Developing Teaching Resources with Generative AI

22 | **REFLECTION**
How English Language Teachers Can Use Critical Thinking to Unpack Everyday Ideology?

28 | **REFLECTION**
Centering the Social: What we learnt from our international language exchange.

38 | **REFLECTION**
Taking ownership of the language classroom: its design

46 | **REFLECTION**
Implementing Deep Processing Theory and Differentiated Instruction with an AI-aided Vocabulary Learning Material

60 | **RESEARCH REPORT**
Boosting Feedback Effectiveness: A Study on EAP Writing

74 | **RESEARCH REPORT**
Evaluating the Accuracy and Reliability of AI Detector Results: A Comparative Analysis

92 | **IN PRACTICE**
Teaching Students to Think With AI, Not Through It: A Practical Gen-AI-Supported Writing Feedback Design for EAP Classrooms

104 | **REVIEW**
Gamification in Review Classes: Enhancing Student Engagement with Wayground

WELCOME TO THE latest edition of LTiC Magazine!

It is clear that language education in China is being shaped by educators who refuse to settle for a one-size-fits-all approach. Across these pages, our contributors bridge the familiar and the new by grounding their practice in established pedagogy while experimenting with digital tools and reflective strategies. What unites their work is a shared commitment to learner empowerment. The articles in this issue also address some fundamental questions: How do we integrate technology responsibly without losing human oversight and connection? How can we create classroom environments and physical spaces where both teachers and students thrive? How can we meaningfully engage students in extra-curricular activities? And how can we provide feedback that truly makes a difference? This issue offers thoughtful, practical answers to these questions. I hope you find as much inspiration in these articles as I have!

As always, LTiC accepts articles in Mandarin Chinese, Japanese and Spanish, as well as English. If you are interested in contributing in any of these languages, please email: Helen. Beech@xjtlu.edu.cn

— Helen

FEATURE

6 Essential Practices for Developing Teaching Resources with Generative AI



Yuri Yarin, MA in TESOL, is a Language Lecturer in the English Language Center at XJTLU's Global Cultures and Languages Hub. He is currently the principal investigator on a funded project called "Teacher Evaluative Judgment in Material Creation with GenAI".

Yuri.Yarin@xjtlu.edu.cn

Introduction

57 % OF HIGHER education institutions now consider AI a strategic priority, according to the 2025 EDUCAUSE report on the AI landscape in education, an increase from 49% the previous year (Robert, et al., 2025). A specific component of these strategies (indicated by 59% of respondents) is the active encouragement of staff and faculty to use AI tools. In other words, educators across the board are being encouraged to incorporate AI into learning. However, as professionals who have attempted using AI tools, we know that there are pitfalls and limitations in these tools and that they must be used carefully to ensure that the teaching content aligns with educational aims.

This article outlines the findings of a research study that specifically looks to answer the question: How do teachers use their professional judgment when using AI to create teaching resources (e.g., lesson materials, texts, online content, etc.)?

While this question sounds theoretical, it has significant practical implications. First and foremost, it asks how teachers

learn from their peers. Secondly, it has the potential for improving teachers' confidence as professionals. Thirdly, and most critically to institutions, this question is key to further adoption from skeptical or reticent teachers. Fourthly, it calls attention to both the effort of teachers developing these skills, and the value that it produces, by ensuring that institutions are able to meet their goals. Thus, this study was developed to answer the critical questions that emerge from the need to ensure pedagogical quality of the teaching resources we provide to our students.

The study described in this article used informed grounded theory as described by Thornberg (2012), which means that there was an iterative process of reviewing the data and the literature. To make this clearer, this study will begin with outlining this methodology. This will then be followed by a brief review of the literature, and the findings. A visual representation of the six practices is then provided as a practical guide with additional explanation for readers to consult. Finally, the six practices are discussed considering the implications.

Methodology

This study, employing informed grounded theory methodology (Thornberg, 2012), initiated with a literature review noting that studies (Van den Berg & du Plessis, 2023; Almuhanha, 2025; Koraishi, 2023) consistently make calls for evaluation but do not provide practical steps of how this happens. Responding to this gap, focus groups were held asking teachers how they conduct evaluations. 12 English for Academic Purposes teachers from Xi'an Jiaotong-Liverpool University participated in May 2025. Participants represented diverse teaching experience (1 to 20+ years), materials development experience (less than 1 year to 10+ years), and AI familiarity (no experience to 100+ hours). Focus groups were audio-recorded, transcribed verbatim, and de-identified.

The initial inductive coding identified that teachers described actions of evaluative professional judgment, which was organized into six practices: interrogating, verifying, revising, selecting, analyzing, and delimiting.

Literature on professional practice and analytical frameworks was then reviewed to validate these practices. Additionally, these six practices were developed with Jabareen's (2009) methodology for frameworks by focusing on action

verbs. The validation helped refine the operational definitions of the codes for comprehensive recoding, resulting in 157 codes across 129 excerpts.

The Literature

As a first look at scholarship, this study looked at articles focusing on implementation attempts from customized chatbots to printed worksheets. Eight studies were identified that looked specifically at implementing AI for these purposes and all eight studies, from one degree to another, made a very specific call to teachers to carefully evaluate and assess all AI tools and outputs before implementation (Almuhanha, 2025; Denny, Kumar & Giacaman, 2023; Dhamija & Dhamija, 2025; Karpouzis, Pantazatos, Taouki & Meli, 2024; Koraishi, 2023; Lee & Song, 2024; Peachey & Crichton, 2024; Van den Berg & du Plessis, 2023). However, none of these studies examine the concrete steps needed for evaluation thereby creating the need for empirical data.

With the data collected, inductive coding revealed the actions: Interrogating, Verifying, Revising, Selecting, Analyzing and Delimiting. These were found to be supported in professional practice literature in the following ways. Both Interrogating and Revising are found as actions of professional judgment in Schön's (1983)

work on reflective professional practice, where they are discussed as concepts of knowing-in-action and reflection-in-action. Bearman, Dawson, Boud & Ajjawi (2024) conceptualize evaluative judgment as making judgments about work quality, which corresponds directly to the practices of Analyzing and Selecting. Eraut's (2000) tacit professional knowledge explains how teachers set boundaries for appropriate tool use (Delimiting).

There are also several analytical framework studies that help support the practices teachers discussed. Mishra, Warr & Islam (2023) identify five AI characteristics—protean, opaque, unstable, generative, and social—that require teachers to resist anthropomorphizing. This requires engaging analytical thinking, grounding the practices of Interrogating AI's reasoning and Analyzing its systematic patterns. Sun, Sheng, Zhou & Wu (2024) provide an eight-category taxonomy of errors frequently found in AI-generated content. This taxonomy establishes support for the importance of Analyzing to recognize common error patterns and Verifying to check for them. Dickey and Bejarano (2024) demonstrate that evaluation operates throughout content generation: Verifying occurs during review stages and Revising during refinement stages. Chen, Zou, Xie, Cheng & Liu (2024) show that AI capabilities and risks require simultaneous attention across multiple levels, which maps onto Selecting (balancing affordances and

risks), Analyzing (operating across multiple levels), and Delimiting (determining when and where AI is appropriate at the teaching model level).

The above literature, combined, makes it clear that evaluation is essential. AI tools are powerful and capable of producing effective content that can lead learners to achieve their goals in a personalized manner. However, this cannot be accomplished without careful and meticulous oversight of professional, highly trained teachers who are prepared for this process.

Research Findings

The final coding of the transcripts yielded 157 codes derived from 129 excerpts corresponding to six identified professional judgment practices that teachers employ when working with AI-generated materials: Interrogating, Verifying, Revising, Selecting, Analyzing, and Delimiting. All six practices appeared in both focus groups, with Verifying appearing in all 12 participants' discourse. The most frequently discussed practices were Delimiting (44 instances, 28% of codes) and Analyzing (39 instances, 24.8%). Cross-speaker consistency and appearance across both focus groups suggest these represent shared professional behaviors rather than individual approaches. Differences between speakers suggest that pro-

professional experience and experience with AI can make a difference in these practices. Teachers described deploying multiple practices simultaneously rather than sequentially, indicating that professional judgment operates as an integrated repertoire adapted to specific contexts. Table 1 summarizes the findings.

Practice	Empirical Support	Theoretical Support	Teacher examples
Interrogating	14 (8.9%)	Reflection-in-action (Schön, 1983); Resisting anthropomorphization (Mishra et al., 2023)	"I usually ask: 'where did you get the sources?'" "I ask 'why did you put this in?'" "I ask AI to adjust CEFR levels, then I ask what it changed."
Verifying	32 (20.4%)	Evaluative judgment (Bearman et al., 2024); Implementation studies (Van den Berg & du Plessis, 2023; Almuhan, 2025)	"You do have to be quite careful and definitely check what it outputs. It's not just the complexity, but things like ... a specific number of words." "I will do the questions myself ... and see if there's any mistakes, or if it's unsuitable for the students."
Revising	15 (9.6%)	Knowing-in-action (Schön, 1983); Pedagogical knowledge (Mishra et al., 2023)	"It uses words like 'foster,' you can hear its' voice,' so I rewrite to remove that." "I spend almost as long checking it, fixing it."
Selecting	13 (8.3%)	Making judgments about work quality (Bearman et al., 2024); Balancing affordances and risks (Chen et al., 2024)	"Brainstorming works if you evaluate and select, using your own judgment." "AI might give 15 activities, with eight being terrible ideas that wouldn't work."
Analyzing	39 (24.8%)	Internal understanding of quality (Bearman et al., 2024); System 2 thinking (Mishra et al., 2023)	"It's based on human thoughts because it's read previous human thoughts on the Internet." "AI doesn't hold an opinion. If you're talking to someone, you have different beliefs. You can feel the passion of their belief system, whereas with AI it's [simplistic pros and cons]"

Practice	Empirical Support	Theoretical Support	Teacher examples
Delimiting	44 (28.0%)	Professional capability (Bearman et al., 2024); Determining when/where AI is appropriate (Chen et al., 2024)	"I think we have to be a role model [to our students] in how to use AI in an appropriate way." "I feel that we've kind of fallen into... how we should use it ... rather than questioning why we're using it."

Table 1 – Summary of findings

Note: Percentages represent proportion of 157 total codes. All six practices appeared in both focus groups.

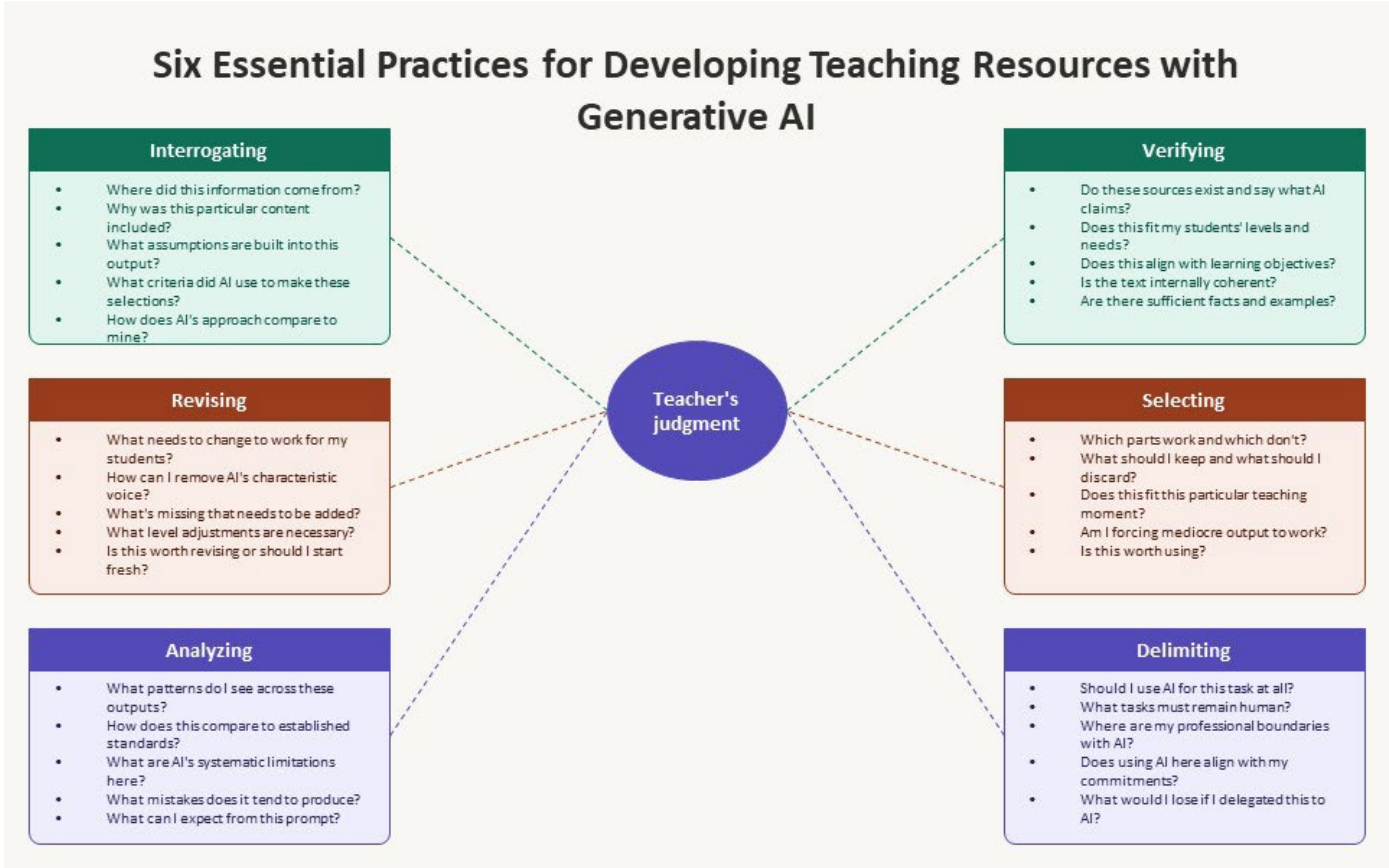


Figure 1 – Framework of Six Essential Practices for Developing Teaching Resources with Generative AI

The Six Practices of Professional Judgment

This section is support for Figure 1, which is a visual and practical presentation of the six practices that teachers employ when working with AI-generated materials. Below, an explanation is provided of what teachers do, how it appears in practice and potential challenges.

Practice 1: Interrogating

Interrogating means asking questions of or about AI to understand its reasoning, sources, and decision-making processes. Teachers engage in this practice by questioning how AI approached a task or made particular analytical choices, examining where AI obtained its information and whether those sources are credible, and probing why AI included specific elements, examples, or explanations. They also assess whether AI content is suitable for particular pedagogical aims and examine what standards or metrics AI applied when making selections or assessments.

In practice, teachers ask AI questions either directly in the chat interface or pose questions to themselves as they review outputs. This questioning helps them understand the

source of information, the reasoning behind decisions, and the criteria AI appears to have applied. These insights enable teachers to modify their prompts more effectively and make better judgments about the validity and usefulness of outputs. Interrogating works particularly well as an initial step before verifying and analyzing, as it surfaces assumptions and decision points that require closer scrutiny.

The main challenge is treating AI’s explanations as actual reasoning rather than recognizing them as pattern-based responses. AI generates explanations based on patterns in its training data, not from genuine understanding of its own processes. Teachers must maintain critical distance when evaluating AI’s self-explanations.

The Forms This Practice Takes:

- Analytical decisions:** How AI approached a task or analytical choices.
- Information provenance:** Where AI got its information and whether those sources are credible.
- Content choices:** Why AI included specific elements, examples, or explanations.
- Appropriateness of use:** Whether AI content is suitable for a pedagogical aim.
- Evaluation criteria:** What standards or metrics AI applied.

Practice 2: Verifying

Verifying means checking AI outputs against external standards for accuracy, appropriateness, and specifications. This practice takes multiple forms: fact-checking content and validating sources, checking fit for learning objectives and student needs, checking coherence, comparing AI outputs against established materials such as textbooks or other trusted sources, verifying against professional pedagogical criteria, and checking whether outputs match the original request in terms of complexity, format, or length.

Experienced teachers develop systematic verification routines based on standards and expectations they have defined. The critical insight from teachers’ experience is dependent on the teacher’s expertise

as a professional. It cannot be reduced to a simple checklist because it requires nuanced judgment about what constitutes appropriate quality for specific teaching contexts.

The main challenge is verification fatigue, the potential to reduce checking effort over time. Teachers can gain a sense of how much real effort may be needed with AI assistance by a verification routine that meets their standards. By referring to their routine, they can better obtain a sense of how long a task might take. Anticipating errors by recognizing patterns, can help speed up this process.

The Forms This Practice Takes:

- Accuracy:** Fact-checking content and validating sources.
- Appropriateness:** Checking fit for learning objectives and student needs.
- Comparative verification:** Comparing AI outputs against established materials (e.g. textbooks or other sources).
- Quality standards:** Verifying against professional pedagogical criteria.
- Specification verification:** Checking match to request (e.g. level, format, or length).

Practice 3: Revising

Revising means modifying AI outputs to better meet the expected standard. Teachers revise AI materials by improving pedagogical design to enhance teaching effectiveness, editing to remove AI’s characteristic voice and make materials sound authen-

tically human, adjusting difficulty level by modifying language or task complexity to match student capabilities, correcting factual mistakes, grammatical problems, or logical inconsistencies and adapting content for specific student populations.

The reality of working with AI is that outputs will require modification, and that it may require multiple rounds of iterative refinement. Teachers who have a strong sense of AI's capabilities, can anticipate issues and prepare themselves to address them.

The challenge with revising is similar to veri-

fication – what may look like minor revisions can consume large periods of time unnecessarily. Teachers should combine this practice with delimiting to decide whether revising is worth the time. Understanding in which stages of production (e.g. brainstorming) AI can be most helpful, can maximize AI's affordances and avoid wasting time.

The Forms This Practice Takes:

- Pedagogical Design:** Adapting activities to enhance teaching effectiveness.
- Authenticity:** Removing AI's vague and impassionate voice.
- Task or Text Difficulty:** Modifying tasks and texts for complexity.
- Errors:** Correcting factual mistakes, phrasing problems, or logical inconsistencies.
- Learner context:** Customizing content for specific student populations.

Practice 4: Selecting

Selecting means making deliberate choices about which AI outputs to use and which to discard. This practice operates in two primary ways: choosing specific portions or elements from larger AI-generated outputs while discarding others, and determining whether and which AI outputs fit particular teaching situations, student populations, or pedagogical purposes.

Teachers exercise selection when they scan through multiple AI-generated suggestions and identify the one or two

worth developing, when they extract useful components from otherwise flawed outputs, and when they make binary decisions about whether generated materials meet minimum thresholds for use.

The main challenge is applying professional expertise to recognize when outputs do not meet standards. Familiarity with professional standards (e.g. learning outcomes, knowledge targets, and understanding learners' context), can help teachers be productive at selecting.

The Forms This Practice Takes:

- Brainstorming and discarding:** Using AI to provide options from which to choose.
- Context appropriateness:** Making choices when revising AI outputs to fit situations, students and pedagogical purpose.

Practice 5: Analyzing

Analyzing means understanding and recognizing how AI works in order to accomplish teachers' task aims. Teachers take note of recurring features or characteristics across AI outputs, understanding AI's systematic strengths and weaknesses. This helps to identify issues with quality, pedagogical approach, as well as examining features like logic and coherence.

Teachers discussed this practice in significant detail during focus groups, demonstrating how they managed their thinking to prepare themselves for what AI would produce. They described identifying patterns in AI's tendencies, recognizing characteristic weaknesses, and developing the ability to spot errors quickly. They also emphasized that evaluating AI

outputs requires fundamentally different approaches than evaluating materials from trusted published sources. Teachers demonstrating more sophisticated analysis combined this practice with others such as interrogating to understand AI's reasoning, verifying to check specific claims, or delimiting to determine when analysis reveals that AI is inappropriate for certain tasks.

The pitfall is overestimating one's own judgment capabilities. Even experienced teachers can miss problems in seemingly polished AI outputs. Teachers should be prepared to raise the level of rigour when checking AI outputs, and make note to learn from their experience.

The Forms This Practice Takes:

- Pattern identification:** Recognizing recurring features or characteristics across AI outputs.
- Quality assessment:** Evaluating appropriateness and effectiveness of materials.
- Limitation analysis:** Understanding the strengths and weaknesses of generative AI.
- Comparative analysis:** Comparing AI outputs against human-created materials.
- Structural analysis:** Examine organizational features, coherence, and logical flow.

Practice 6: Delimiting

This practice involves reaching conclusions about which task types are suitable or unsuitable for AI assistance, judging when human input supersedes any value AI could contribute (such as tasks requiring understanding of context), setting limits based on pedagogical principles, educational purposes, and professional identity (such as modeling certain practices for students), and defining acceptable ways to deploy AI within teaching practice.

This practice becomes sophisticated when teachers establish particular conditions and parameters, allowing AI usage under specific circumstances for tasks they would not otherwise be able to perform or when they recognize AI as making meaningful

enhancements such as handling certain mechanical work.

The main challenge is institutional pressure to use AI, which can become overwhelming. Teachers should aim to understand the wider perspectives of best practices in education as well as in their specific field and aim to ensure their applications fit those. Where specific use cases are proposed, those use cases should be tested systematically and evidence should be documented to demonstrate either positive or negative outcomes. This grounds delimiting decisions in professional evidence rather than popular enthusiasm.

The Forms This Practice Takes:

Suitability: Making conclusions about appropriate and inappropriate task types for AI assistance.

Affordances and limitations: This means judging when human interaction supersedes any value AI could contribute (e.g. social interaction).

Professional boundaries: Setting limits based on pedagogical principles, educational purposes, and professional identity (e.g. setting examples for students).

Defining parameters: Acceptable ways to deploy AI within teaching practice.

Discussion and Implications

This framework demonstrates that while AI tools can enhance teaching practice when preparing lesson resources, realizing those benefits requires professional skill. The six practices operationalize professional judgment, making visible what teachers do when working with AI-generated materials and providing a foundation for professional development and institutional policy.

The framework supports professional development, firstly, by enabling deliberate practice combined with reflective action (Schön, 1983). Teachers can apply the framework when developing teaching resources with AI or creating AI agents for student use. In both cases, the practices guide critical questions: what capabilities to select, how to adjust outputs to meet requirements, how to verify materials produce intended results, and how to explain tools' appropriate use to students.

Second, the framework builds confidence for teachers uncertain about implementing AI. By specifying concrete practices rather than abstract principles, it provides actionable steps for navigating the complexity that generative AI inherently holds (Mishra et al., 2023). Where AI's opacity blocks understanding, the framework shows what to interrogate and what to verify.

Third, the framework acknowledges the professional work required. While implementation studies and technology proponents claim AI makes teachers' work easier and faster, this framework demonstrates that highly trained teaching professionals are needed to operationalize AI and realize its promised affordances.

Finally, this framework places attention on what teachers do to make AI applications work. While implementation studies and proponents claim that AI can make the teachers' work easier and faster, the reality is not so clear. This framework demonstrates that highly trained teaching professionals are needed to operationalize AI and bring about the affordances it promises, a claim made not only in this study but widely supported by relevant literature.

This article addresses a central contradiction in AI and education: AI-generated content offers multiple affordances but requires careful professional judgment to achieve intended effects. With competing claims from AI proponents and institutions eager to stay at the forefront of educational innovation, nuance is easily lost. This framework restores that nuance by placing teachers and their professional judgment at the center. Based on the evidence presented here, teachers and institutions should consider this framework for professional development and policies that draw out AI's benefits while mitigating its shortcomings through skilled professional practice.

References

Almuhanna, M. A. (2025). Teachers' perspectives of integrating AI-powered technologies in K-12 education for creating customized learning materials and resources. *Education and Information Technologies*, 30, 10343–10371. <https://doi.org/10.1007/s10639-024-13257-y>

Bearman, M., Tai, J., Dawson, P., Boud, D., & Ajjawi, R. (2024). Developing evaluative judgement for a time of generative artificial intelligence. *Assessment & Evaluation in Higher Education*. <https://doi.org/10.1080/02602938.2024.2335321>

Chen, X., Zou, D., Xie, H., Cheng, G., & Liu, C. (2024). Two decades of artificial intelligence in education. *Educational Technology & Society*, 27(1), 1–32. [https://doi.org/10.30191/ETS.202401_27\(1\).0001](https://doi.org/10.30191/ETS.202401_27(1).0001)

Denny, P., Kumar, V., & Giacaman, N. (2023). Conversing with Copilot: Exploring prompt engineering for solving CS1 problems using natural language. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education* (pp. 1136–1142). ACM. <https://doi.org/10.1145/3545945.3569823>

Dhamija, D., & Dhamija, A. (2025). Understanding teachers' perspectives on ChatGPT-generated assignments in higher education. *Journal of Interdisciplinary Studies in Education*, 14(1), 38–62.

Dickey, M. D., & Bejarano, A. (2024). GAIDE: A framework for using generative AI to develop educational materials. *TechTrends*, 68, 748–757. <https://doi.org/10.1007/s11528-024-00966-1>

Eraut, M. (2000). Non-formal learning and tacit knowledge in professional work. *British Journal of Educational Psychology*, 70(1), 113–136. <https://doi.org/10.1348/000709900158001>

Jabareen, Y. (2009). Building a conceptual framework: philosophy, definitions, and procedure. *International journal of qualitative methods*, 8(4), 49–62.

Karpouzis, K., Pantazatos, D., Taouki, J., & Meli, K. (2024). Tailoring education with GenAI: A new horizon in lesson planning. In *Proceedings of the IEEE EDUCON 2024*. IEEE. <https://doi.org/10.31219/osf.io/9b4su>

Koraishi, O. (2023). Teaching English in the Age of AI: Embracing ChatGPT to optimize EFL materials and assessment. *LET Journal*, 3(1), 55–73.

Lee, Y., & Song, H. (2024). AI implementation in educational settings: A systematic review. *Educational Technology Research and Development*, 72(2), 445–467.

Mishra, P., Warr, M., & Islam, R. (2023). TPACK in the age of ChatGPT and generative AI. *Journal of Digital Learning in Teacher Education*, 39(4), 235–251. <https://doi.org/10.1080/21532974.2023.2247480>

Peachey, N., & Crichton, H. (2024). AI and the future of education: Teaching in the age of artificial intelligence. *British Council*. <https://doi.org/10.57884/B8CD-WF42>

Robert, J., Muscanell, N., McCormack, M., Pelletier, K., Arnold, K., Arbino, N., Young, K., & Reeves, J. (2025). 2025 EDUCAUSE Horizon Report: Teaching and learning edition. *EDUCAUSE*. <https://library.educause.edu/-/media/files/library/2025/5/2025hrteaching-learning.pdf>

Schön, D. A. (1983). *The reflective practitioner. How professionals think in action*. Basic Books.

Sun, Y., Sheng, D., Zhou, Z. *et al.* (2024). AI hallucination: towards a comprehensive classification of distorted information in artificial intelligence-generated content. *Humanit Soc Sci Commun*, 11, 1278. <https://doi.org/10.1057/s41599-024-03811-x>

Thornberg, R. (2012). Informed grounded theory. *Scandinavian Journal of Educational Research*, 56(3), 243–259. <https://doi.org/10.1080/00313831.2011.581686>

Van den Berg, G., & du Plessis, E. (2023). ChatGPT and generative AI: Possibilities for its contribution to lesson planning, critical thinking and openness in teacher education. *Education Sciences*, 13(10), 998. <https://doi.org/10.3390/educsci13100998>

REFLECTION

How English Language Teachers Can Use **CRITICAL THINKING** to Unpack Everyday Ideology?



Gareth Owen, is a Senior English language lecturer at Xi'an Jiaotong-Liverpool University in Suzhou, China, specializing in English for Academic Purposes for undergraduates.

Gareth.Owen@xjtlu.edu.cn



ENGLISH LANGUAGE TEACHING often focuses on fluency, accuracy and communicative competence but rarely questions what learners are actually being taught to say and why. By applying critical thinking to everyday expressions, idioms, and conversational norms, English teachers can empower students to question the hidden values, biases, and assumptions embedded in English to become more conscious and ethical users of the language.

At A Glance

Treat idioms and common expressions as discussion points, not just vocabulary to memorise.

Briefly explore their imagery, origins, and possible implications.

Ask guiding questions such as:

- What picture does this phrase create?
- Where might it come from?
- Could it be interpreted differently?

Incorporate simple activities such as analysing the origins of idioms, debating phrases such as “boys will be boys”, or suggesting alternative expressions.

Introduction: Rethinking the Role of the English Teacher

I have often told my students to use idioms or metaphors in their spoken English, especially when preparing for a speaking test. When talking about solutions, say “Kill two birds with one stone”. If you are talking about something that went against you, tell them you were “sold down the river”. Expressions like these can push up scores for lexical resource and vocabulary use, but what values are we passing on, whether intentionally or not? By enabling our students to become critical thinkers, we can help them question the language they want to use, not just try to use it fluently.

The words, idioms and expressions we choose reflect our particular ways of seeing the world. Some of these ways are so ingrained that they have become normalised, even when they reinforce outdated stereotypes or power imbalances. As English teachers, we are not just teaching grammar and vocabulary, we are also, often unconsciously, transmitting cultural narratives and value systems. If we can unpack how everyday language carries these hidden messages, we can help students become more critical and self-aware communicators.

Language as a Carrier of Ideology

Language has never just been a collection of neutral words and grammar rules. It is the way culture, values and even power are highlighted. One of the main ways to spread how a society views and understands the world is through its language. When we speak, we are not only exchanging information, we are also reinforcing certain perspectives, although quite often, we do not realise we are doing so!

When referring to someone as an “illegal alien”, it frames them as a criminal. However, if they are referred to as an “undocumented migrant”, it shifts the focus to a lack of legal paperwork. The choice of words guides how the listener interprets reality, subtly supporting one worldview over another.

In the English language classroom, these choices matter. When we teach idioms, we may be passing on more than just expressions, but we may also be transmitting the histories and values attached to them. Some phrases carry legacies of racism, sexism or violence, even if their original meanings have faded from everyday awareness.

Recognising language as a carrier of ideology means acknowledging that what we teach can

never be culturally empty. This gives us the opportunity, and I believe the responsibility, to help students see beyond surface meaning and to question why certain expressions exist. They can then decide whether they wish to adopt them and thus transform language learning into an act of critical engagement.



Figure 1. How do words shape our thinking?

Note. Image generated using ChatGPT. The image was generated in response to this prompt: 'Create an image that represents this article', and was one panel of the AI-generated four-panel composition.

Everyday Sayings That Reinforce Power Structures

Some ideologically loaded language hides in plain sight in the sayings we might use every day. These expressions are often passed down uncritically, presented as harmless, yet they can subtly reinforce existing power structures

by framing them as social norms.

The phrase, “boys will be boys”, is often used to excuse male misbehaviour, normalising gender stereotypes and making it harder to challenge toxic masculinity. Expressions like “if you work hard, you’ll succeed” reinforce the idea of meritocracy, downplaying the role of privilege or systemic bias in shaping opportunities. It ignores structural barriers such as poverty or lack of access to education and could be used as an argument against inequality.

Idiomatic expressions such as “sold down the river” originate from the U.S. slave trade, where enslaved people were literally sold further south down the Mississippi River. Many speakers today use these without any awareness of their origins. When such idioms are taught or encouraged in the classroom without context, we risk perpetuating outdated or discriminatory worldviews.

As teachers, the challenge is not to ban idiomatic expressions but to approach them critically. If we can consider their origins and implications, we can model how to critically question language and allow learners to make informed decisions about whether to use it. This transforms such language into opportunities for cultural reflection and ethical choice.

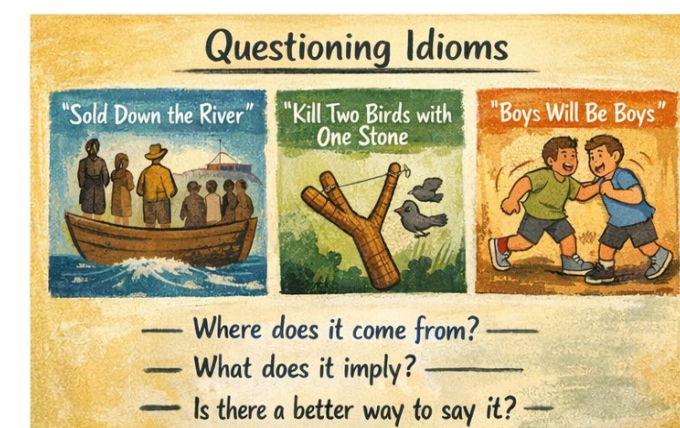


Figure 2. Questioning idioms

Note. Image generated using ChatGPT. The image was generated in response to this prompt: 'Create an image that represents this article', and was one panel of the AI-generated four-panel composition.

Language That Naturalises Inequality

Not all problematic expressions directly reference social groups or power hierarchies. Many are more subtle and embed ideas of violence into everyday speech. Over time, these metaphors become so familiar that their imagery is invisible and unchallenged.

A common example is “kill two birds with one stone”. In most contexts, speakers are not consciously imagining animal cruelty, yet the image of killing is normalised through casual repetition. Similarly, “shoot yourself in the foot” and “pull the trigger” carry violent

connotations to firearms, but are used freely in business and education. The more these expressions circulate, the more they desensitise us to their original images of



Figure 3. Rethinking Language

Note. Image generated using ChatGPT. The image was generated in response to this prompt: 'Create an image that represents this article', and was one panel of the AI-generated four-panel composition.

harm.

When such language is taught without critical discussion, learners may adopt it as part of "authentic" English without questioning what it normalises. For English teachers, this is an opportunity to highlight how metaphor shapes perception. By exploring or asking learners to think of alternatives, we can show that communication can be vivid without relying on imagery rooted in harm. The aim is not to sanitise language into blandness, but to make its impact visible so students can choose their words with intention.

Challenging the Language We Use

Recognising that language carries hidden ideologies is the first step, and the real challenge lies in how we respond. As English language teachers, we can play a pivotal role in helping students develop the critical thinking skills needed to consider this.

One key strategy is raising awareness. Encourage students to question familiar idioms, expressions, and phrases: Where did this come from? What assumptions does it carry? This can be done alongside language proficiency.

Critical thinking exercises, such as comparing idiomatic expressions with their origins, exploring alternatives or discussing the impact of certain phrases, can turn passive learners into active language users. For example, students might debate the meaning and implications of "boys will be boys" or create a list of idioms they find problematic and suggest more inclusive replacements. This can encourage engagement and empower our students, which in turn can improve motivation and allow students to express themselves more authentically. This approach not only improves linguistic competence but also nurtures global citizenship and empathy.

Practical Steps to Critically Examine Language

Read the following scenario.

During a university trip to an art gallery, a group of first year, male students are shouting, pushing each other, and making rude comments about some of the paintings. Some other students look concerned and look towards the lecturer who is on the trip. The lecturer responds by saying "Boys will be boys."

Consider and discuss these questions with your partner.

- Do you think the male students are behaving well or badly?
- Does the lecturer try to stop the male students?
- Does what the lecturer says explain or excuse the behaviour? Why?
- Do you think the lecturer would say or do something different if it was female students behaving in a similar way?
- Could the lecturer say "girls will be girls"? Why?

Conclusion

Everyday language is far from neutral and can carry the weight of history, ideology, and power. As English language teachers, we have a unique opportunity and I believe a responsibility to help students see beyond surface meaning, uncover the hidden assumptions embedded in common expressions, and make conscious choices about the words they adopt.

By applying critical thinking to idioms, sayings, and conversational norms, we can move language learning to more ethical engagement. This not only enriches students' communication skills but also empowers them to participate thoughtfully in a diverse and complex world.



Centering the Social:

What we learnt from our international language exchange



Jonathan Tillotson is a Senior Language Lecturer at XJTLU. He has presented at numerous international conferences on a range of topics. His current research interests include the benefits for language acquisition of student-centred, peer-to-peer international language exchange.

Jonathan.Tillotson@xjtlu.edu.cn



James Stubbs is a Language Lecturer in the Academic Literacies Centre at XJTLU. He is currently working alongside the school of Humanities and Social Sciences. His interests include multilingualism and the role of social presence in language acquisition and learning.

James.Stubbs@xjtlu.edu.cn



Aaron Christopher Kelly is a language lecturer at XJTLU and has worked in Cambodia, Australia, the UK, and China. His professional interests include fostering autonomous language learning and language teaching methodologies.

Aaron.Kelly@xjtlu.edu.cn



Part 1: The Experience

THE XJTLU INTERNATIONAL Language Exchange has been running at XJTLU since September 2023. It gives international and domestic Chinese students the opportunity to socially interact in a relaxed, informal environment, helping each other to practice and improve their English and Chinese, while also making new friends and connections with students around the world. It was created with the idea that learning is best seen as a social endeavor which requires people to communicate and co-construct knowledge in a social setting. Whereas the recent trend has been towards phone APPs which learners use in isolation from each other, the language exchange offers a counter narrative in which learning is heavily contextualized, wholeness and process are centred rather than focusing on specific skill-building, and learners develop the real-world ability of co-constructing meaning. For all these reasons, we established the language exchange based on the principals of the Community of Inquiry

(Garrison, Anderson & Archer, 2001), a pedagogical framework designed to emphasise collaboration in constructivist learning through teacher, social, and cognitive presence.

The model shows that for learning to be meaningful and learners to be engaged, there needs to be cognitive presence which for us meant challenge at the right level; a teacher presence which in our case really meant the presence of other learners who were expert in the language (remember students in the language exchange move between English and Chinese), and social presence.

XJTLU is an English as a Medium of Instruction (EMI) university, and most classes are delivered in English. For non-native English speakers they must take credit bearing English language classes as part of their Year 1



curriculum. For non-Chinese speakers, Chinese classes are part of theirs. The main goal of the language exchange was to group these learners together and to create a space and opportunity for them to work together to provide 'real life' social use of their target language and to provide Comprehensible Input (Krashen, 1982), too, or in other words, language that the students will mostly understand, but not necessarily all. The language exchange takes place every semester, one evening a week, over between four and ten weeks. The student numbers vary but have so far peaked at over 100 students.

For us, the student experience of being in the language exchange is paramount. As aforementioned, all the students already attend formal language education, and we wanted to provide a different model of learning from this in which students could aid each other's development and to emphasise the social aspects of learning as per Vygotsky's (1978) Socio-cultural theory, which argues that learning is inherently social, driven by discussions with more knowledgeable others (MKO). The reflections from the participants came from semi-structured interviews and a survey, and the key findings are as follows.

1.1 Social Learning is Enjoyable

First of all, the social aspect of the language exchange was something which was

continually mentioned by the participants. One international participant noted that they had extensively used APPs, but it is 'way more enjoyable when you have someone else to talk to.' They all discussed how they had become friends with people on the exchange and these friendships had permeated other aspects of their university life. One group of students, for example, met regularly to socialise and eat together while using the target languages. Another interesting breakaway group was a mathematics group who had taken the core ideology of the language exchange and applied it to their mathematics, and they would meet and use both English and Chinese to solve maths problems from their classes together. The social aspect and the chance to learn from each other were something that students commented on frequently as a positive with another international student stating that 'the main goal of this activity is to help people be friends outside' and stated that they had had challenges forming meaningful friendships with Chinese students until the language exchange.

1.2 Reduced Anxiety

Many students saw the language exchange as a safe, low-stakes environment which contrasted to the often-hypercompetitive nature of modern education. Some spoke in detail about how the language exchange was effective in reducing their language anxiety and boosting their confidence. One

participant stated that 'this programme has really shaped a lot of people, I noticed from the first meeting that some people weren't as social as I expected them to be but then by the last session, they were just speaking out loud'. This shows that the space created was successful in reducing anxiety in a cross-cultural environment and this happened due to a number of factors. They felt that the scaling back of teacher presence and the lack of any evaluation from others meant that they were free to learn what they wanted and needed from their peers. It was mentioned that although staff were present in the language exchange, they were seen as facilitators rather than teachers and their role was to set up tasks. There was no error correction, feedback, and very little input from the teachers present, which allowed the participants to relax and enjoy the process of learning from and with peers. One participant stressed how enjoyable this experience was as 'you feel so free to say whatever you want because you're talking with people your own age and you're basically choosing the topic'.

1.3 Multilingualism Aids Learning

The sessions were all conducted in a multi-lingual environment where students could move between languages drawing from translanguaging pedagogy which allows learners to 'make full use of their repertoires, at times, to understand, assimilate and

interpret' (Anderson, 2024:75) content. While the stages of the tasks were set up to focus on a particular language, the students were welcome to use whichever language to further the communication. This was noted as positive as the students welcomed being experts, or as Vygotsky (1978) describes, the 'more knowing other' (MKO), in their own language and being able to support their peers. One Chinese student mentioned that 'there are not many chances to speak in Chinese and English, and I need to express my thoughts better [sic]'. This meant that we observed conversations where utterances were conducted in different languages to help convey and co-construct meaning as another student mentioned that 'it is still hard for me to speak in English' so they enjoyed the opportunity to use their full repertoire of languages. For the final stage of the task, cross-talk was encouraged where students would use their strongest language to communicate and provide input for their partner. This meant that they had to interpret what was being said in their target language and respond in their strongest language. One student stated that in this stage they often used English to 'help my partner understand the Chinese' and felt that their group didn't 'follow the rules'. While the students initially had difficulty with this stage and it was common for them to code-switch between languages, but after some practice, they became more accustomed to it and naturally began translanguaging to co-construct meaning.

Part 2: An Analysis of the Experience Using Theoretical Underpinnings.

2.1 Learning but not linear

There has been a growing sense across fields as diverse as neuroscience (Damasio, 1994; 1999) cognitive psychology (Bruner, 1966), epistemology (Latour, 2005), education (Davis and Sumara, 2014) and even economics (Kahneman, 2011) that knowledge and learning are not lineally organised. Instead, they evolve in a complex web of interconnections at both an explicit and implicit level. Language acquisition got there a long time ago in the form of the Input Hypothesis (Krashen, 1982). This states that when learners are given a rich input of language, which they can understand through both their knowledge of the language and the context, language acquisition occurs. In terms of the language exchange, this context could mean the scaffolding of conversation partners, images or gestures, prior schemata of how the world works and what people say in such situations or the patient, collaborative reframing of ideas.

The implicit web of knowledge students acquire grows incidentally to understanding and incrementally so that there is often no moment at which students realise a word or concept is known but instead a

slow evolution of their understanding. The language exchange puts forward the idea that the richest source of this input isn't YouTube videos, Netflix or APPs but social interaction with others.

The challenge we face in the 21st century is that technology and particularly screens now play the central role that social interaction once had (Haidt, 2024). While most learning throughout human history has been mediated by social experience (Vygotsky, 1978), our students' mental models (Van Dijk, 2017) of what learning looks like are more likely to include APPs than dialogue and conversation with others. The premise behind these APPs is a simple one, operant conditioning (Skinner, 1953) techniques, like small rewards, are used to keep the learner engaged with monotonous gamified tasks that split language into tiny linear and supposedly learnable chunks. The language exchange embraces a more complex view of learning, following on from the Community of Inquiry framework (Garrison, Anderson & Archer, 2001) which emphasizes the need for cognitive challenge not rote learning, the presence of more- knowledgeable others and a social presence.

2.2 Should learning be social?

APPs generally only involve other people, if they involve them at all, so that learners have somebody to compete with. However, the Community of Enquiry Framework (Garrison, et al. 2001) suggests that real cognitive engagement requires a social presence, and so isolating learners from each other will lead to a lack of focus on tasks. It may be the case that some learners have gone on year-long Duolingo streaks of placing words in spaces, but what language acquisition and learning require is real investment in meaning making (Krashen 1982). So, rather than being isolated, we should be actively immersed in collaboration with others. Additionally, as Deci and Ryan (2004) pointed out, learning which is operantly conditioned, like Duolingo, lacks real autonomy because the source of the engagement is not the self but the rewards. To be autonomous they claim, activities need to be done for intrinsic reasons like the joy we find in them or for more internalized extrinsic motivations. This means that we do something because it speaks to who we feel we are or ought to be. Following on from this, the language exchange takes a very different approach to APPs and leverages the joy people feel when interacting with students from other cultures and the importance students place on having friends from other cultures. Perhaps the term here

is that they are invested (Norton, 2016) in the activity. This social investment leads to increased focus which creates a self-driven engagement rather than an externally motivated habit.

In addition, while we live in the modern world of screens and gamified rewards, our brains evolved to be social (Cacioppo, 2022). Removing other people while we perfect knowledge alone can lead to what Curran (2023) calls an epidemic of perfectionism and anxiety. The students who participated frequently observed that the social participation and acceptance others had for their personal linguistic practices reduced their language anxiety.

2.3 A multilingual environment

Looking at what Krashen (1982) said about language acquisition being the result of comprehending input might lead us to question why the Language Exchange is a multilingual event. Surely having only one language would maximize the input of that language and therefore increase the amount of language acquisition. This, however, was not the position of Krashen (1997) who said that students in multilingual classes acquired language more quickly than those in monolingual classes. To explain this, we can look to the concept of Translanguaging (Garcia, 2011) which suggests we are wrong

to assume that simply exposing ourselves to larger amounts of a single language by prohibiting other languages within a space will lead to greater acquisition of that language. It is rather the richness of communication that increases language acquisition. The communication within the language exchange is rich because multilingual students are using their full sets of linguistic resources to create meaning socially.

Part 3: What we have learnt and some tips

After conducting the language exchange over four semesters, we have learnt that while students need and expect structure and direction, guidance should not be too prescriptive. Single topics for each session, spread over a handful of simple slides, attractively illustrated, with open-ended questions work best in motivating students to engage with one another. Ultimately, if language learning aims to equip students for communication, it follows that their motivation is highest when their interest in the content is maximized. Therefore, unlike in regular classes, where topics through which language is taught can be dry, impersonal and academic (in accordance with syllabus

requirements) we sought to increase motivation by only selecting subjects that would directly connect with their everyday lives; topics they were familiar with, had opinions about, and wanted to hear others discuss. So we chose topics like food, intercultural differences, hobbies, travel, memes, sports, festivals, video games, hometown highlights, embarrassing moments (and how they've handled them), and desired campus changes. This activated language such as to turn it into a bridge across which their desire to connect with others about interesting matters could be realized. In this way, language was no longer primarily a deposit of knowledge, learnt for its own sake, as on an APP, or as part of an assessment regime, but a dynamic tool for social interaction.

Typically, despite decades of theoretical emphases on task-based student-centred learning, the imperative imposed upon learning by assessment related outcomes has meant students still tend to take a back seat to the teachers in their learning process. And yet, if we survey the number of student-led, extra-curricular societies at XJTLU, many students are keen to take the lead in relation to activities that interest them. If we add to this the importance placed at university upon teaching entrepreneurship, critical thinking and leadership, we can conclude that encouraging students to take a leading role in their learning process is desirable for their academic and personal development,

as well as something we can expect many to be enthusiastic about – if the context is right.

The language exchange taps into and encourages this capacity for leadership. It enables students themselves to become teachers. After all, all Chinese students are fluent in Chinese and at XJTLU most of our international students are fluent or near-fluent in English. Students, therefore, possess an expertise that, though this should be obvious in a transnational university, might not have occurred to them. Knowing that you have much to give and a service to provide to your peers, we believe, can enhance self-esteem and confidence, which again contributes positively to motivation in learning environments.

One key finding of our research, on the other hand, was that students can still feel intimidated by the presence of experts in the discussion group. But what is meant here was something different. In an experiment we conducted one week, teachers were asked to join groups as equals so they could reflect on their participation in deliberative discussions to gain insight into their teaching methodology. While this experience was enlightening for the teachers, the students tended to withdraw and become silent, implicitly expecting teachers to dominate as they were used to experiencing in their timetabled classes. Even though the teacher

gave no instructions, it was as if the standard classroom hierarchy had been re-imposed. What became apparent was that the problem wasn't expertise per se. Rather it was that the language exchange's culture of equality had been disrupted. When the teacher was absent, yes, each student knew they were merely a novice or learner in their L2, be it English or in Chinese. But they also knew that it was the same for the others, for whom their role was to act as teacher. No student had any official authority, and this easy back-and-forth of expertise amongst equals avoided embarrassment and loss of face. Everyone knew it was ok to make mistakes, without which learning is inevitably constrained.

In addition, we found that moving students between tables helped increase the social contact that students sought to make during our sessions. Forging intercultural connections had always been a guiding principle behind the language exchange, in addition to language learning.

Finally, in keeping with the original intention that these sessions be fun and enjoyable, we found it was particularly effective if students worked together to complete a particular task. For example, we asked them to compile playlists of their favourite music, make short videos or, in what was perhaps the most engaging sessions of all, build towers and bridges made out of spaghetti.

The conclusion we drew from this was that students love to be actively involved but need to feel that the right kind of permission has been given to them to behave in this autonomous way.

If we were to sum up what the language exchange can give students, in contrast both to the over-governed classroom and solitary tech-centred learning via APPs we could say it provides autonomy without isolation. APPs, it's true, unlike the classroom, can give a student control, even total control, over their learning process. They can tweak it and turn it on or off at will. But this, in fact, involves too much autonomy. It also denies, as we have seen, the essentially social nature of language. The student's only companion is an artificial super-expert who neither knows nor cares about them. In the language exchange, however, autonomy operates at the group-level and is dynamically social and egalitarian. While teachers oversee the meta-structure, students are in charge and create the conditions for the co-construction of meaning.

The main challenge we have in the language exchange has to do with the demographics of the student body. While XJTLU is a transnational university, domestic Chinese students vastly outnumber those from overseas, which means we face a deficit of students whose stronger language is English. Hopefully, as international student numbers continue to grow post-Covid, this problem will be alleviated. That said, a problem may also concern promotion on campus.

We have experienced no difficulties recruiting enough Chinese students. Even from one email requesting interest we have always been inundated with student interest but we are always looking for new ways to reach out to the international student body. It would be unfortunate if, given that they have travelled all this way to experience China and are required to learn Chinese, that international students should miss out on the opportunity we provide only because they have not been informed about it.



References

- Anderson, J. (2024). Translanguaging: a paradigm shift for ELT theory and practice, *ELT Journal*, 78 (1), 72-81, <https://doi.org/10.1093/elt/ccad057>
- Bruner, J. S. (1966). *Toward a Theory of Instruction*. Harvard University Press.
- Cacioppo, S. (2022). *Wired for love: A Neuroscientist's journey through romance, loss and the essence of human connection*. Hachette UK.
- Curran, T. (2023). *The perfection trap: Embracing the power of good enough*. Simon and Schuster.
- Damasio, A. R. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. Gosset/Putnam.
- Damasio, A. R. (1999). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. Harcourt.
- Davis, B., & Sumara, D. (2014). *Complexity and education: Inquiries into learning, teaching, and research*. Routledge.
- Deci, E. L., & Ryan, R. M. (Eds.). (2004). *Handbook of self-determination research*. University of Rochester Press.
- Garrison, D. R., Anderson, T., & Archer, W. (2001). Critical Thinking, Cognitive Presence, and Computer Conferencing in Distance Education. *American Journal of Distance Education*, 15 (1), 7-23, <https://doi.org/10.1080/08923640109527071>
- García, O. (2011). *Bilingual education in the 21st century: A global perspective (2nd ed.)* Wily Blackwell.
- Haidt, J. (2024). *The anxious generation: How the great rewiring of childhood is causing an epidemic of mental illness*. Penguin.
- Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux.
- Krashen, S. (1982). *Principles and practice in second language acquisition*. Pergamon Press.
- Krashen, S. D. (1997). Why bilingual education? *ERIC Digest*. <https://files.eric.ed.gov/fulltext/ED403101.pdf>
- Latour, B. (2005). *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford University Press.
- Norton, B. (2016). Learner investment and language teacher identity. In G. Barkhuizen (Ed.) *Reflections on language teacher identity research* (pp. 88-94). Routledge.
- Skinner, B. F. (1953). *Science and human behavior*. Macmillan.
- Van Dijk, T. A. (2017). *Discourse and power*. Bloomsbury Publishing.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Taking Ownership of the Language Classroom: Its Design



Glenn Howlett is a Senior Language Lecturer and Deputy Head of the Academic Literacies Centre for the School of Mathematics and Physics at Xi'an Jiaotong-Liverpool University. His research interests include factors influencing the motivation of language learners in the classroom.

Glenn.Howlett@xjtlu.edu.cn



AS LANGUAGE LECTURERS, we commonly focus on providing the best learning environment for our learners whilst paying attention to various extrinsic and intrinsic factors in the classroom. Outside of rearranging seating in our classroom, we often have limited options to cater for a range of teaching activities from workshops to individual consultations and tutorials whilst being mindful of sharing the same teaching space with others. This article will explore possible options when presented with the opportunity to redesign the language classroom.

Background

To provide greater context, the Global Cultures and Languages Hub at XJTLU was established in 2025 and is home to three dynamic centres: the Academic Literacies Centre (ALC), the English Language Centre (ELC), and the Modern Languages Centre (MLC). The aim of the ALC is to support all academic departments across its campuses with

English language and academic skills, falling outside of the regular ELC support offered to all students in their initial year of undergraduate study in the form of credit-bearing English for Academic Purposes (EAP) modules. For this to be possible, ALC lecturers were embedded in departments to provide topic- and skills-specific support to students. As a result, each ALC team was allocated a single classroom for non-timetabled language teaching activities in the academic department's own teaching building. This article is based on the experience of the ALC team allocated to support the School of Maths and Physics (SMP).

Considerations

It quickly became evident that the ALC team supporting SMP would need to offer a range of non-timetabled English language teaching activities including workshops, tutorials, group and individual support sessions. Additionally, this ALC team also decided to keep the teaching space open to



students to fully use and access as a self-study space with some hours being manned by an ALC lecturer to offer informal support. Therefore, this space not only needed to be flexible, but also visible to all SMP students as a place to seek English language support.

When deciding how best to arrange the SMP ALC teaching space, research conducted by Byers, Imms, and Hartnell-Young (2018), which explored how an Innovative Learning Environment (ILE) may impact learners and their academic performance, proved to be a useful starting point. Although this study acknowledged that numerous other school factors must be considered outside of the ILE, it nevertheless highlighted the potential for ILEs to influence a teacher's approach and offer a greater variety of layout options as opposed to the layout of a traditional classroom. This research also highlighted that learners perceived ILEs to offer a more active, engaging, and personalised classroom environment compared to other teaching spaces. This was of great importance since students will spend the majority of time in lectures and seminars meaning an ILE may appear to be more welcoming and attractive to learners. These are also important considerations for our ALC classroom.

Design

This article will focus on the factors which could be altered such as the layout and

furniture arrangements which do not require any permanent changes to the classroom or formal renovation. If given a chance to make permanent changes or support the specific design of the ILE during its initial construction, the SMP ALC team would have drawn on research which highlights a classrooms' impact on learners as a result of its sound quality and its acoustic environment (Karim and Khairuddin, 2025); colour schemes utilised (Duyan and Isik, 2024); and comfortable seating (Costa, 2025) to make more meaningful and fundamental changes to the space.

Number of areas or stations

The first consideration was to decide the number of different areas in the ALC classroom. Due to limited space in a classroom with a maximum capacity of thirty, it was decided that 4 zones would be created. These were:

- **Zone 1: Teaching Area.** The regular area for the teacher to be stationed at the front of the traditional classroom including the whiteboard, computer desk, projector screen, and teaching podium.
- **Zone 2: Flexible Teaching Area.** An area taking up the bulk of the central

classroom space with five tables for students to sit around to allow for collaboration and natural communication.

- **Zone 3: Informal Study Area.** The back section of the classroom with sofas and tables creating booths offering an alternative area when it is not necessary to focus on the teacher and projector screen.
- **Zone 4: Self-study Area.** The introduction of a high desk with bar stools to the right side of the classroom.

These areas are somewhat similar to those proposed by Byers et al. (2018) as shown in Figure 1 .

Maximizing space in the ALC classroom

The ALC classroom possesses an unusual and unique L-shape with a complete working area of approximately 69m². As the room was set up as a traditional classroom with rows of tables, the odd angles created by the walls and two pillars meant that much of the space was not utilised. As highlighted in Figure 2, replacing rows of tables with smaller round tables, and including Zones 3 and 4 along the back wall and side windows created much more space to move freely around the ALC ILE. Unlike rows of bulky rectangular tables, as shown in Figure 3, the

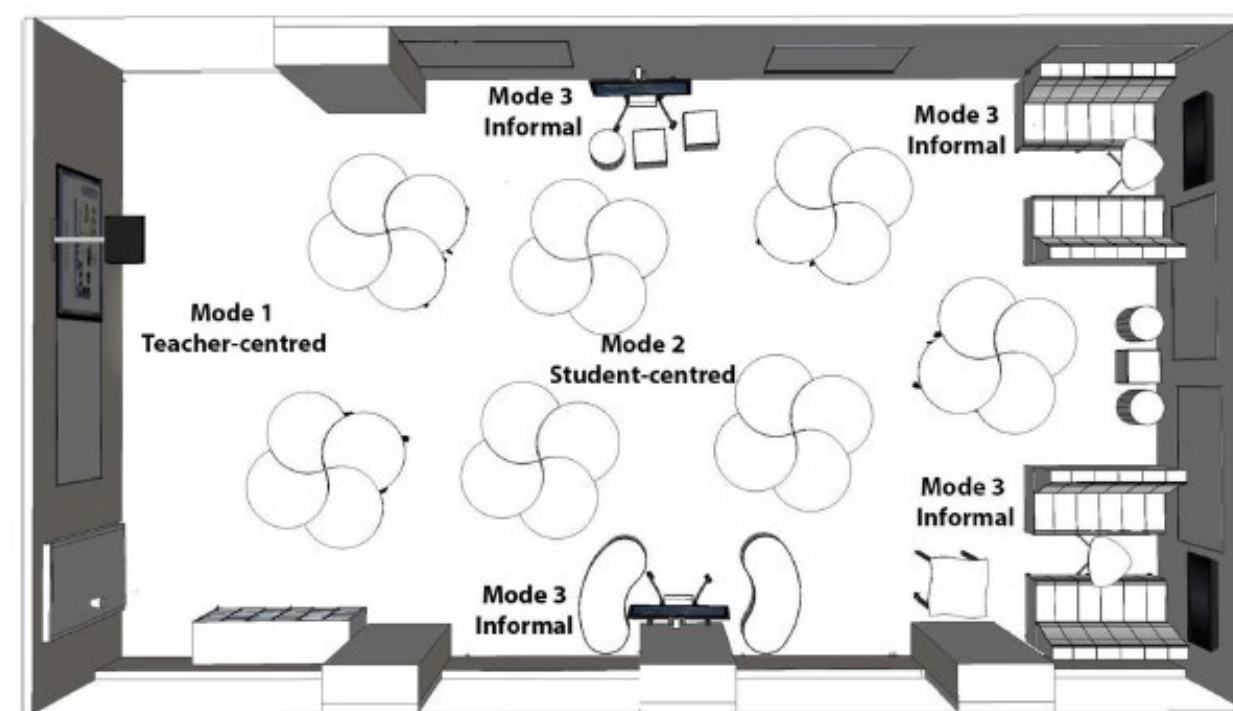


Figure 1. Retrofitted classroom in the design of an ILE (Byers, Imms, and Hartnell-Young, 2018)

round tables can easily be moved to the side to create open space for activities. Such flexibility can attract learners as they can take greater ownership of the study space (Costa, 2025), meaning this room may be utilised more for personal study.

Due to the limitation of classroom size, few ideas were considered practical from research by Detyna, Ogunbase, and Briffa (2025) such as their Project and Gather Spaces. Both require a larger area of seating, which is often circular and without the restriction of tables to ensure more natural communication. However, their examples of Collaboration and Discussion Spaces did inform the design of our teaching space. Firstly, with regards to the Collaboration and Discussion Spaces, comfortable seating with tables can ensure students can work together in relatively smaller spaces. Secondly, their Discussion Space highlighted how the spatial design of the room could create effective zones and how to more effectively utilise space around the perimeter of the room as shown in Figure 4. All areas of our ALC classroom are in direct eyesight of the sole projector screen which is also of great importance for our room compared to larger ILEs which often contain numerous screens.



Figure 2. Photograph of the ALC classroom in the design of an ILE. Photo by author.



Figure 3. The desks commonly used in this School at XJTLU. Photo by author.



Figure 4. Discussion Space (Detyna et al., 2025)

Activities held in the ALC classroom to benefit staff and students so far...

As a result of the flexibility offered by our redesigned ILE containing four zones as opposed to the traditional seating highlighted in Figure 3, a greater variety of activities can be held more comfortably to the benefit of students in the SMP as the design facilitates movement between different zones of our classroom:

- Workshops and small group seminars focusing on Academic skills related to Writing, Speaking, Listening, and Reading
- Review sessions to focus on citation and referencing conventions
- Informal speaking sessions aimed to increase confidence when using English
- Small group assessment preparation support (e.g., for the delivery of a presentation in English)
- Ad hoc informal drop-in support sessions for students to seek advice related to an English language task, activity, or assessment
- Professional Development sessions

Recommendations

As we proceed through our second semester following the inception of the ALC, a review will be conducted into the non-timetabled, English language activities held in this ALC classroom. In our context, each ALC team and classroom connected to a specific Academic School or Academy may find clear differences between what activities are requested by faculty and students, and what restrictions are placed on potential support given the team size and classroom dimensions. If presented with the opportunity to redesign a classroom into a teaching space to support students, it is highly recommended that the classroom is converted into a more flexible space such as an ILE. For our own context at XJTLU, this allows for a clear and visible base for language teachers to be stationed inside a specific School or Academy and a more attractive and comfortable space will increase the likelihood of students opting to find and take advantage of such English language-related support. If following similar design principles for the creation of an ILE, the following points should be considered:

1. Is it likely that the intended non-timetabled teaching activities can be realistically held or delivered in this room when considering its flexible layout and capacity limitations?
2. If different teaching and learning stations are created in this room, could they negatively impact any activities being held?
3. Will such a flexible teaching space provide additional opportunities to better support your students which are currently not offered or feasible?
4. Is the teaching space on offer being better utilised as a flexible ILE space compared to having a more traditional layout such as rows of fixed desks and seats?
5. Will it be more convenient for students to work together and collaborate in group activities if you alter the design of the classroom?

It must be noted that although classrooms with a unique shape may benefit from being redesigned as an ILE, the space has to be large enough to accommodate different working stations for students.

References

- Byers, T., Imms, W., & Hartnell-Young, E. (2018). Comparative analysis of the impact of traditional versus innovative learning. *Studies in Educational Evaluation*, 58, 167-177. doi: <https://doi.org/10.1016/j.stueduc.2018.07.003>
- Costa, A.R. (2025). Homing into school to create the ideal classroom: young people want to combine the best of home and school learning environments. *CoDesign*, 21(1), 74-94. doi: <https://doi.org/10.1080/15710882.2024.2363925>
- Detyna, M., Ogunbase, A., & Briffa, H. (2025). A socio-material approach to investigating classrooms: student engagement in an innovative learning environment. *Learning Environments Research*, 28, 473-502. doi: <https://doi.org/10.1007/s10984-025-09529-0>
- Duyan, F., Isik, G. (2024). The evaluation of the impact of computer classroom wall colors on students' perception in the context of color components. *Megaron*, 19(1), 61-74. doi: <https://doi.org/10.14744/megaron.2024.96562>
- Karim, A.A. & Khairuddin, K.F. (2025). The Impact of Classroom Acoustic Environments on The Behaviour of Special Needs Students: A Systematic Literature Review. *e-Bangi Journal of Social Sciences & Humanities*, 22(2), 525-534. doi: <https://doi.org/10.17576/ebangi.2025.2202.42>

Implementing Deep Processing Theory & Differentiated Instruction with an AI-aided Vocabulary Learning Material



Yu Hu is an EAP Associate Language Lecturer at Xi'an Jiaotong-Liverpool University, Suzhou, China. His interests include narrative writing, vocabulary acquisition, AI-aided material design, and integrating cultural elements into pedagogy to enhance student engagement.

Yu.Hu@xjtlu.edu.cn

At Xi'an Jiaotong-Liverpool University (XJTLU), students study in an English-medium environment: classroom activities, homework, and assessments are all conducted in English. Although most Chinese students have learned English from primary school through senior high school, they have mainly acquired the most frequent words used in everyday English and still lack academic English. As a result, their written texts are often not formal enough, and their spoken utterances are not concise.

Under these circumstances, the Academic Word List (AWL) becomes helpful. Compiled in 2000 by Averil Coxhead at Victoria University of Wellington, New Zealand, the AWL contains 570 word families frequently used in academic texts, divided into 10 sublists by frequency: sublist 1 has the most frequent words, sublist 10 the least frequent, with 60 words per sublist except sublist 10, which has 30. By learning the AWL, students acquire essential vocabulary to read, write, speak, and listen effectively in academic contexts, bridging the gap between general English and discipline-specific language.

The current learning mode is as follows: (1) teachers upload the sublists in PDF format onto the Learning Mall—XJTLU's online learning platform; (2) students download and familiarize themselves with sublists one, two, and three in Semester One and with sublists four, five, and six in Semester Two. Over the course of a 13-week semester, teachers design a range of 10-minute learning and assessment activities to help students review these words irregularly. In Week 10, one mock test is administered, and two weeks later students sit the final vocabulary test. For proactive, goal-oriented, self-regulated students who employ effective learning strategies, the current AWL learning material and learning mode are sufficient. For others, however, the learning experience still requires scaffolding.

This article reviews the successes and limitations of current approaches to using the AWL sublists, proposes a guided self-learning approach, and explores supplementary activities to support it. The key takeaway is that the set of materials I have created can be used to transform AWL word learning from mechanical memorization into an interactive, scaffolded, and differentiated experience, both for self-learning and in-class instruction.

1 Reflecting on the Current Approach to AWL Learning

1.1 Effective Practices

One positive effect of the current arrangement is that learner autonomy has been effectively cultivated. With a clear timeline for how many words to learn each month, week, or day, students can pace their learning according to their individual styles and take responsibility for time management, avoiding last-minute cramming.

Another benefit is that students are encouraged to develop a personalized vocabulary approach. Unlike traditional high-school lists that provide words with fixed definitions and parts of speech, the AWL sublists present only target words, requiring students to consult dictionaries independently. This exposes them not only to definitions but also to phonetic symbols, synonyms, antonyms, sample sentences, idioms, CEFR levels, and register information. The process of searching, exploring, and selecting is likely to stimulate deeper

cognitive engagement and may help learners move beyond textbook learning to develop personalized strategies, which may benefit them when transitioning to major-specific and low-frequency academic vocabulary.

Finally, mastery of the high-frequency words across all 10 AWL sublists contributes to improved academic speaking and writing. Even students who primarily memorized Chinese definitions demonstrated greater conciseness and accuracy in essays and spoken coursework. This is because AWL items themselves are concise expressions and often have concise Chinese equivalents. When translating from Chinese to English, learners may retrieve a concise Chinese expression, map it to the corresponding AWL item, and convey complex, precise meanings using fewer, more accurate English words. For example, 'Several factors contribute to this situation' is more concise than 'There are many different reasons that cause this situation'. Similarly, 'The data are analyzed and interpreted' demonstrates more precise academic expression than 'We look at the data and say what it means'.

1.2 Identified Limitations

Despite the positive outcomes of the current approach, several notable weaknesses warrant attention. First, some students become trapped in mindless repetition.

Even highly industrious learners who review AWL words daily or every two days may still struggle to use them fluently. The issue is not repetition itself—research shows that even well-structured programs require seven or more encounters for a word to become usable (Schmitt & McCarthy, 2002, p. 74)—but what is being repeated. Many students review only Chinese definitions, or at best definitions with parts of speech, keeping them in a decontextualized learning environment and unable to produce sentences with appropriate collocations. In more concerning cases, students fail to memorize meanings even after ten rounds of repetition, let alone use the words in sentences, reflecting Craik and Lockhart's (1972) observation that increased study time does not necessarily improve memory.

Another limitation concerns the tendency to overlook pronunciation. After learning words from the AWL sublists, students repeatedly encounter them in academic reading materials and use them in essay-writing tasks. However, some students make little effort to improve their pronunciation accuracy, either because they lack strategies for independent pronunciation practice or because they require more engaging ways to practise it. This neglect may lead to unnecessary difficulties in daily communication as well as in speaking assessments.

2 A Possible Framework for Guided Self-Learning

Guided by the above reflections and informed by deep processing theory (Craik & Lockhart, 1972; Craik & Tulving, 1975), I leveraged AI tools such as ChatGPT, Suno, and Canva to develop a 720-page AWL learning resource. Figure 1 is a screenshot of the AWL Sublist-6.

Each word from the AWL sublists occupies one page, and each page is divided into eight panels. Each panel provides students with an opportunity to deepen their understanding of the target word. However, some words are not associated with any idioms, and some do not belong to commonly used word families. In these cases, the Word Family panel was replaced with a cognitively beneficial word game, and the Idiom(s) panel was replaced with a Root & Affix(es) panel, see Figure 2.

To encourage students to further practice newly learned words, I added a supplementary page after each vocabulary page. The screenshot (Figure 3) illustrates this.

CEFR Level

capable

Definition(s)

adj.: able to do things effectively and skillfully, and to achieve results
Translation:
Look up the translation(s) if needed.

Pronunciation(s)

/ˈkeɪ.pə.bəl/

Check your pronunciation using an electronic dictionary.

1. Tongue Twister: Cleverly capable cats can carefully carry countless colorful cups, creating chaotic commotion calmly continuously.

2. Transcribe the IPA back into a sentence: /hi: ɪz ˈkeɪ.pə.bəl əv ˈhʌndrɪd ˈdɪf.ɪ.kəlt tɪks ˌɪn dɪ ˈpɛn.dənt.li/

Practice your pronunciation with these two tasks.

Word Family

n. capability /ˌkeɪ.pəˈbɪl.ə.ti/

- With the new machines we finally have the capability to do the job properly.

adv. capably /ˌkeɪ.pəˈbəl.i/

- All of the main supporting actors play their roles capably.
- His office is run quietly and capably by Mrs. Lau, who has worked with him for 12 years.

Use these words to enhance your syntactic variety.

Collocations

adj.

- highly capable students
- a capable research assistant
- capable of independent work
- capable of critical analysis
- a technically capable team
- capable problem-solving skills
- capable of handling pressure
- capable academic performance

Make your language sound more fluent and natural.

Synonyms and/or Antonyms

Synonyms

- suitable /ˈsuː.ə.bəl/
- qualified /ˈkwɒl.ɪ.faɪd/
- competent /ˈkɒm.pɪ.tənt/
- able /eɪ.bəl/

Antonyms

- incompetent /ˌɪn.kəmˈpi.tənt/
- unfit /ʌnˈfɪt/
- poor /pɔː/
- inept /iˈnept/

Make your language more expressive and nuanced.

Idiom(s)

in someone's capable hands [humorous]

Meaning: being dealt with by the person mentioned

- I'm going away next week, so I'll be leaving everything in your capable hands.
- Jade said not to worry about the presentation—it's in Ryan's capable hands now.

Help you connect more deeply with native speakers.

Usage on Campus

[Informal Use - Conversation]

Isa: I think Lily is totally capable of leading the presentation.

Jack: Yeah, she's super capable and always prepared.

[Formal Use - Academic Writing]

- The results suggest that the model is capable of predicting learner outcomes.
- Students capable of autonomous study tend to perform better in advanced courses.

Communicate effectively on campus, not just in exams.

Usage in a Song

capable

Let the song help you review and enjoy the rhythm of learning.

AWL Sublist - 6

Figure 1. A screenshot of a sample from the 720-page AWL learning material.)

Argue

Some people believe the Dazu Rock Carvings are primarily religious art, while others think they are also important historical records. Which view do you agree with, and why?

Dazu Rock Carvings

Diaoyu Fortress

Dazu Rock Carvings

(Chongqing)

Figure 3. A screenshot including topic-specific pictures.

CEFR Level

whereas

Definition(s)

conj.: compared with the fact that; but
Translation:
Look up the translation(s) if needed.

Pronunciation(s)

/ˈweɪ.əz/

Check your pronunciation using an electronic dictionary.

1. Tongue Twister: Wendy wondered wildly, whereas Walter wisely weighed worrisome warnings, weaving witty wordwork wonderfully.

2. Transcribe the IPA back into a sentence: /ɪt ɪz ɪn ˈfɜːr ˈkɒfɪ ˈweɪ.əz ɦɪ ˈpɪ ˈfɜːr ɦɪ ɪn ɒ ˈmɔːr.nɪŋ/

Practice your pronunciation with these two tasks.

Letter Lab Challenge

Choose two to seven letters from 'w, h, e, r, e, a, s' to create twelve words.

Enhance your word-building skills. Foster creative thinking.

Collocations

conj.

- whereas others prefer online
- whereas some choose face-to-face
- whereas students attend regularly
- whereas professors assign homework
- whereas some focus on theory
- whereas others gain experience
- whereas outcomes vary significantly
- whereas participation is optional

Make your language sound more fluent and natural.

Synonyms and/or Antonyms

Synonyms

- because /biˈkɔːz/
- since /sɪns/
- now /naʊ/
- seeing /siːɪŋ/
- considering /kənˈsɪd.ərɪŋ/
- inasmuch as /ɪnˈæz mʌtʃ ɔːz/
- as long as /æz lɒŋ ɔːz/
- 'cause /kɔːz/

Make your language more expressive and nuanced.

Root & Affix(es)

Root: Old English hwær meaning 'where' + as meaning 'as'

Meaning: The root elements indicate a contrast or comparison between two clauses or situations.

Improve your ability to decode and generate words.

Usage on Campus

[Informal Use - Conversation]

Eva: I love going to lectures, whereas Emma skips most of them.

Arlene: Yeah, whereas you actually enjoy sitting through three hours of notes!

[Formal Use - Academic Writing]

- The majority of students prefer hybrid learning, whereas a minority insist on fully in-person classes.
- Whereas prior studies focused on exam performance, this research emphasizes overall student engagement.

Communicate effectively on campus, not just in exams.

Usage in a Song

whereas

Let the song help you review and enjoy the rhythm of learning.

AWL Sublist - 5

Figure 2. A screenshot including Letter Lab Challenge and Root & Affix(es).

The three screenshots illustrate the basic layout of the AWL material. Although learners still need to encounter a word, at least, seven to fourteen times to use it effectively, the eight panels provide multiple, varied exposures. Since simply covering many aspects of a word does not guarantee mastery and students need cognitively meaningful processing, I shall explain the logic behind the eight-panel design here. This design aligns with Bloom’s Taxonomy (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956), which outlines a hierarchy from lower- to higher-order thinking skills. For example, the Definition(s) panel corresponds to Remember; Root & Affix(es) to Understand; Collocations and

Synonyms/Antonyms to Apply and Create; and picture pages (e.g., Figure 3) to Analyze, Evaluate, and Create. Theoretically, learning a word across these panels promotes deep processing, increasing retention and addressing the limitation in Section 1.2 where some students fail to memorize words. To tackle decontextualized learning, I compiled sample sentences for each AWL word and added a Usage on Campus panel, dividing examples into informal and formal use. This contextualized practice enhances the likelihood that students can retrieve and apply words in campus life.

To address the second limitation mentioned in Section 1.2, two panels were created:

Pronunciation(s) and Usage in a Song. The Pronunciation(s) panel begins with the target word's pronunciation, progresses to a tongue-twister task to encourage sentence use, and concludes with an exercise analyzing phonetic symbols. This panel emphasizes the importance of phonetic symbols, which facilitate accurate pronunciation of words like sword, helluva, and petit four, especially when consulting a native speaker is impractical, even in an English-medium university. The Usage in a Song panel aims to improve comprehension through repeated exposure to English in context, a method supported by Temur (2021). Unlike typical pop songs, the songs in this material are AI-generated using Suno, with prompts designed to repeat each target word coherently. Where possible, song content aligns with our EAP (English for Academic Purposes) seminar learning outcomes; for example, a song for the word 'abstract' focuses on essay abstracts, covering steps, structure, and common errors. Listening to these songs allows students to practice pronunciation and absorb language learning tips. Meanwhile, by applying the Pomodoro Technique—a time-management method where learners work in focused 25-minute intervals, followed by a short 5-minute break and a longer break after four sessions—students can make meaningful use of their breaks, for example, by listening to AWL-focused songs.

3 Beyond Self-learning

In an era in which students are supported by a wide range of AI tools, some may argue that effective learning can be achieved without regular class attendance. However, it would be premature to conclude that classroom instruction has therefore become unnecessary.

Empirically, students at different proficiency levels face distinct challenges when learning independently. Foundation-level learners may struggle with phonetic symbols; intermediate learners may fail to distinguish between formal and informal lexical items; and advanced learners may underestimate seemingly simple words while still being unable to produce idiomatic expressions. As a result, students may gradually disengage from the deep processing of vocabulary. In such cases, teachers play a critical role in guiding learning, designing tasks, organizing activities, scaffolding understanding, and providing timely feedback.

From a theoretical perspective, research suggests that learners make faster progress and sustain motivation for longer when guided by teachers (Long, 1983; Willis, 2009). Moreover, vocabulary—and the



Figure 4. A screenshot of sample cards

English language more broadly—serves not only to convey meaning but also to facilitate interaction, communication, and discourse. Activities involving authentic communication function as vehicles through which these language functions are realized (Richards & Rodgers, 2014). Human instructors, rather than AI tools, are better positioned to design and adapt activities to the needs of specific learner groups, because they can avoid the pitfalls of one-size-fits-all models by applying principles of differentiated instruction. As Tomlinson (1999) observes, uniform in-

structional approaches often contribute to learner failure and demotivation; by the same logic, a one-size-fits-all educational game is unlikely to succeed.

Drawing on these theoretical insights, I designed a card game to promote interpersonal communication while providing students with appropriately differentiated challenges. Sample cards from the game are shown in Figure 4.

The following steps explain how the game works.

Step 1: Grouping

The teacher organizes students into groups of three or four. Each group is provided with six decks of cards, each with a different focus:

- Deck One: English definitions of AWL words
- Deck Two: Pronunciation tasks
- Deck Three: Collocation(s)
- Deck Four: Synonyms and/or antonyms
- Deck Five: Roots & Affix(es) or Idiom(s)
- Deck Six: Topic-specific pictures

Also, each group is provided with five sets of coupons in denominations of \$100, \$150, \$200, \$250, and \$300 (20 of each). Sample coupons from the game are shown as Figure 5.

Step 2: Preparation

Each group member selects one deck based on their language proficiency and/or interests. Allowing students to make this choice themselves is critical, because both Krashen's Monitor Theory and the Input Hypothesis (Larsen-Freeman & Long, 2000) suggest that learners are more likely to complete tasks that are appropriately challenging rather than those that go

far beyond their current level of language proficiency. Members then shuffle their chosen decks and place them face down on the desk before the game begins. Because groups contain only three or four members, not all six decks may be used in a single round. Regardless of the deck combination, students take turns in either clockwise or counterclockwise order.

Step 3: Playing

Group members decide who starts the game. Players take turns drawing one card from their own deck and completing the task on the card. Other group members evaluate the response, and the teacher acts as arbiter when necessary. For example, (1) a student holding Deck One reads the definition(s) aloud and identifies the target word. If the answer is correct, the student receives a \$100 coupon. If the student fails and another member provides the correct answer, the coupon is awarded to that member. If there is disagreement, the teacher confirms the answer. (2) A student holding Deck Two reads a tongue twister aloud as quickly and accurately as possible. Other members time the performance. Challengers may attempt the task, and the fastest speaker receives a \$100 coupon. (3) For Deck Three, the student must produce a grammatically correct sentence using one of the given collocations within 60 seconds.

Success earns a \$100 coupon; if challengers succeed after an initial failure, the winner receives a \$200 coupon. (4) For Deck Four, the student produces sentences using either one synonym and one antonym, or two different synonyms, within two minutes. Minor grammatical errors are permitted. Successful completion earns a \$300 coupon. (5) For Deck Five, the student reads the root and affix(es) and identifies the target word. Group discussion is allowed. A correct response earns a \$150 coupon; challengers may attempt the task if the initial response is incorrect. If the card focuses on idiom(s), students are required to create a new sentence using the idiom(s) within 75 seconds, and the reward is a \$250 coupon. (6) For Deck Six, the student interprets the picture(s) on the card and answers the prompt. The student has 30 seconds to prepare and must speak for at least one minute. Successful completion earns a \$300 coupon.

Step 4: Awarding the Winner

After 20–25 minutes of play, the students with the highest total value of coupons are declared the winners. The teacher may prepare a small gift for each winner in advance.

4 Summary

I have argued in the preceding sections that students should avoid mechanical repetition in their vocabulary-learning journey. They need to understand the principle of deep processing and put it into practice in order to prolong the retention of newly learned words while also improving pronunciation. These empirical experiences motivated me to design the AWL self-learning materials. At the same time, no learner is an island, complete in itself; every language learner is a piece of the wider learning community,



Figure 5. A screenshot of sample coupons.

connected to the whole. Even self-disciplined and highly motivated students may struggle to make full use of vocabulary-learning materials when working entirely on their own. Therefore, I also designed an engaging activity to scaffold students' vocabulary acquisition. The card game provides students with more opportunities to focus on speaking, pronunciation, and communication. Through this process, newly learned vocabulary is more likely to be transformed from receptive knowledge (e.g., reading) into productive use (e.g., speaking).

Undoubtedly, the six AWL sublists provide students with only a limited number of lexical items. However, once students develop the habit of deep-processing new words in a principled sequence, as suggested by Bloom's Taxonomy (Bloom et al., 1956), this approach can evolve into a transferable learning strategy that subsequently facilitates vocabulary acquisition when learners face increasingly demanding learning needs. Empirical evidence supports

the necessity of this long-term perspective. For instance, non-native speakers typically acquire approximately 2,650 words within a single year when they make full use of an English-rich environment (Schmitt & McCarthy, 2002, p. 211). This figure far exceeds the current AWL learning load of our students—approximately 360 AWL words per year. If students are to achieve stronger academic performance, a substantially larger vocabulary size is therefore required. Achieving this goal necessitates the adoption of more systematic and scientifically grounded learning strategies. Deep processing could be one such strategy.

Looking ahead, considerable scope remains for further development in this area of scientific learning strategies. In the context of an AI-aided educational landscape and a generation of learners raised in a highly digitalized environment, it is reasonable to predict that the development of engaging, AI-supported vocabulary learning apps represents a meaningful and far-reaching next step.

References

- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., and Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals, by a committee of college and university examiners. Handbook I: Cognitive domain*. New York, NY: Longmans, Green.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671–684.
- Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 104(3), 268–294.
- Larsen-Freeman, D., & Long, M. H. (2000). *An introduction to second language acquisition research*. Beijing: Foreign Language Teaching and Research Press.
- Long, M. (1983). Does second language instruction make a difference? A review of the research. *TESOL Quarterly*, 17, 359–82.
- Richards, J. C. & Rodgers, T. S. (2014). *Approaches and methods in language teaching*. Beijing: Foreign Language Teaching and Research Press.
- Schmitt, N. & McCarthy, M. (2002). *Vocabulary: Description, acquisition and pedagogy*. Shanghai: Shanghai Foreign Language Education Press.
- Temur, M. (2021). The impact of awakening perception of learners' comprehension on English learning via regularly listening to songs in English. *International Journal of Progressive Education*, 17(1), 145–157.
- Tomlinson, C. A. (1999). *The differentiated classroom: Responding to the needs of all learners*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Wilkins, D. A. (1972). *Linguistics in language teaching*. Cambridge, MA: The MIT Press.
- Willis, D. (2009). *Rules, patterns and words: Grammar and lexis in English language teaching*. Cambridge: Cambridge University Press.

RESEARCH REPORT

Boosting Feedback Effectiveness: A Study on EAP Writing



Lu Zhang is teaching in Xi'an Jiaotong-Liverpool University (XJTLU) as a language lecturer. With an M.A. in TESOL from The University of Hong Kong, she has been teaching English for over 10 years to students in various age groups.

Lu.Zhang@xjtlu.edu.cn

Introduction

IN HIGHER EDUCATION contexts, it is an important stage for teachers to provide learners with feedback on their writing to monitor and assess their learning process, as well as to provide direction for improvement. However, the time-consuming feedback process can sometimes be unrewarding and may leave teachers frustrated when their feedback is not valued by their students (Winstone, Nash, Rowntre & Parker, 2017; Zhang & Hyland, 2022). Furthermore, the recent emergence of Artificial Intelligence (AI) has “promised” our students instant access to unlimited personalised machine-generated feedback on their writing. To make things even more complex, AI-produced drafts frequently surpass students’ own work, potentially replacing it to save their effort. Considering this, a new question arises: How can teachers improve students’ engagement with writing feedback and their ability to act on it, to protect our teaching quality and encourage greater student involvement throughout the writing process?

Literature review

Definitions of “Uptake” and “Engagement”

It is important to clearly define the two terms we will discuss first, as they are commonly used as indicators of feedback effectiveness: feedback engagement and feedback uptake (Zhang & Hyland, 2018, 2022; Dressler, Chu, Crossman & Hilman, 2019). The term “uptake” is relatively simple and focuses on the product, referring to learners’ behavioural responses to teacher feedback and how effectively they incorporate it into text revisions, that is, the extent to which feedback is used to improve their writing. In contrast, “engagement” is more about the process and is significantly more complex. It can be defined as “the extent students are invested or committed to their learning, embracing a complex of factors which can be seen in students’ responses to texts and their attitudes to writing and responding” (Zhang & Hyland, 2018, p. 91).

Feedback Engagement Model

Zhang and Hyland (2018, 2022) propose a model to measure the learner’s engagement level with teacher feedback, including three interlocking and overlapping dimensions:

affective, behavioural, and cognitive. Affective engagement involves learners' emotions and attitudes, such as their willingness to interact with the teacher and their satisfaction with the feedback they receive. Behavioural engagement is more concrete: the actions the learner takes after receiving feedback, such as revision actions and the time spent on revision. Cognitive engagement is less explicit and mainly includes the information processing in the learner's brain, such as choosing to use different strategies to analyse and address the issues pointed out in the feedback, such as asking peers for help, reorganising ideas, and elaborating on arguments.

Research on Writing Feedback Engagement and Uptake

To address low feedback engagement and uptake, researchers and practitioners have developed relevant frameworks and implemented various intervention activities with different focuses, drawing on Vygotsky's (1978) theory of Cognitive Development, which highlights social interaction and effective mediation within the Zone of Proximal Development (ZPD). Practically on interaction and mediation, Lee (2014) advocates the theory of mediated learning experience (MLE), arguing that MLE activities should be designed to make the writing tasks mainly include more process-oriented elements such as multiple-drafting, active student involvement, dialogic interactions, and teachers' focused and purposeful feedback. In a more systematic way, Maas (2017) promotes her framework called learner-driven feedback (LDF) which consists of the learners' feedback preferences about writing aspects and delivery modes (e.g. face-to-face, written, audio, video, etc.), as well as the learners' feedback requests. Moreover, Smith (2021) emphasises the important potential of personalised feedback for architecture students (who are one of the major subject groups in this study), considering the unique aspects of their discipline and the diverse methods of feedback delivery. His research suggests that exploring different approaches to feedback provision can greatly improve students' learning experiences, ultimately making education more effective and engaging. All these previous studies have laid the foundation for what we are discussing here regarding teacher feedback effectiveness.

My Research

Research Aims

This research study seeks to find out how to improve the effectiveness of teacher feedback. Specifically, the study aims to identify the more effective intervention activities and the more effectively-addressed writing aspects through this refined essay writing process by assessing student engagement with feedback and how well they incorporate (uptake) it. The following are the research questions:

RQ1: What specific intervention activities tend to be more effective in enhancing the students' feedback engagement and uptake?

RQ2: What aspects of the summative assessment (Situation-Problem-Solution (SPS) essay) can the feedback address more effectively?

Research Participants

The participants of this study are from my own teaching groups in a Year Two undergraduate English for Specific Academic Purposes (ESAP) module in Spring 2024, who majored in Architecture or Urban Planning and Design at a transnational, English-as-a-Medium-of-Instruction (EMI) university in the eastern part of the Chinese Mainland. The number of participants varied for different rounds of data collection, making the sample sizes for the three surveys 53, 49, and 26, respectively. In addition, 24 participants also provided their first and final essay drafts. All the participants were Chinese with Mandarin as their first language (L1).

Research Context

This Year Two ESAP module required four contact hours per week, divided into two sessions (seminars with a class size of 20 students per class group). By the end of that semester, the students were expected to achieve the English proficiency level of CEFR

B2+. For the writing segment of this module which is related to this research, students were required to write a 1000-word SPS essay on a given site, using sources and a proper referencing style, related to their field of study.

Intervention

To investigate the two research questions, an intervention was implemented throughout the SPS essay writing process. Details regarding the intervention, together with the schedule for the entire writing process, are listed in Table 1.

Timeline	Actions	Intervention?	Data Collected?
The Pre-Intervention Stage			
Week 1 Lesson 1	- Anonymous Survey 1 on participants' opinions and attitudes towards writing feedback in their previous learning experiences - Brief introduction to this research study	No	Yes: Survey data (Survey 1)
Week 1 Lesson 2	Explanation of the practical use and value of the SPS essay	Yes	No
The Intervention Stage			
Week 4	Training on individual outline tutorial	Yes	No
Week 5	- Individual outline tutorial - Anonymous Survey 2 on participants' willingness and satisfaction with the outline tutorial	Yes	Yes: Survey data (Survey 2)
Week 8 Lesson 1	Introduction to and use of the Feedback Preference Form (FPF)	Yes	No

Timeline	Actions	Intervention?	Data Collected?
Week 8 Lesson 2	- First draft submitted for the teacher's individual written feedback - Collection of relevant data	No	Yes: Survey data (FPF) + essay first drafts + individual written feedback
Weeks 8-9	Activities as whole-class feedback on common issues identified from the students' first drafts	Yes	No
Week 9	- Individual written feedback on the first draft with FPF returned to each student - Small-group tutorial on the first draft regarding individual written feedback	Yes	No
End of Week 9	Anonymous Survey 3 on participants' opinions and attitudes towards the intervention activities and the whole SPS writing process	No	Yes: Survey data (Survey 3)
Week 10	Final draft submitted for grading	No	No
End of the Semester	Collection of final drafts	No	Yes: Essay final drafts

Table 1. Data collection and intervention procedure

Research Instruments

This study employs three instruments to collect data: the Feedback Preference Form (FPF), three surveys, and two essay drafts for each participant.

The FPF was used in Week 8, right before the participants submitted their first draft for my individual written feedback. Data about their feedback preferences (in ranking) were collected, including their preferred delivery modes, as well as three essay aspects they preferred me to pay closer attention to when providing feedback, together with any additional requests (see Appendix 1 for the FPF). The purpose of this form is to collect data about how to conduct the subsequent feedback provision, to explore more about the participants' cognitive engagement, and to set comparison groups at a later stage.

The three surveys were implemented before, in the middle of, and right after the whole writing process, respectively. Survey 1 was conducted in Week 1, before the writing process or the intervention started, to investigate the participants' perceptions and attitudes towards teacher feedback based on their past experiences in general and their feedback expectations. Surveys 2 and 3 were used in Week 5 and Week 9, respectively, to collect data about the participants' perceptions and attitudes on the teacher feedback and the whole process in order to investigate the affective engagement.

Two essay drafts (first and final) were collected after the whole writing process was complete to assess the participants' behavioural engagement and feedback uptake level by tracking the textual changes from the first to the final draft based on the teacher feedback and quantifying the quality of each change. However, unlike the three-value system (0, 0.5, 1) used by Alharbi (2022) and Dressler et al. (2019) which cannot differentiate those more effective changes (addressing more than 50% of the issue identified by the feedback) from those less effective ones (such as those addressing only 20%), in this study, a more detailed four-value system is applied to code each textual change and assign a value to show the change quality according to the feedback: (a) 0 -- not addressed, (b1) 0.25 -- partially addressed ($0 < \text{change quality} < 50\%$), (b2) 0.75 -- partially addressed ($50\% \leq \text{change quality} < 100\%$), and (c) 1 -- fully addressed (100%).

Specifically, behavioural engagement, indicating the quantity of the student's textual changes based on the teacher's feedback, was measured via a percentage (termed as "overall feedback integration rate" later in this paper) by adding the total frequency of the feedback addressed (b1, b2, and c types) against the total frequency of feedback identified in the individual written teacher feedback. In contrast, feedback uptake level focuses on the quality of the feedback integration and is measured via another percentage (termed as "effective feedback integration rate") by assigning a respective value (0.25 for b1, 0.75 for b2, and 1 for c) to each textual change according to the extent to which it addresses the feedback, adding all the values, then dividing the number by the total frequency of all the feedback identified in the teacher's individual written feedback.

Results

This section will present the main findings related to each of the two research questions.

Firstly, feedback integration rates calculated from each participant's textual changes in their drafts were used to assess the overall effectiveness of this SPS writing process. As Figure 1 shows, the overall feedback integration rate of all the participants was 82.1%, which means the students responded to as much as 82.1% of my individual written feedback by making textual changes in their drafts. Regarding the quality of these changes based on the effective feedback integration rate, the participants successfully addressed 61.3% of all the issues flagged in my feedback. Specifically, when dividing all the participants into two subgroups (Group Y, N=14, in which participants used and submitted the FPF together with their first draft, and Group N, N=10, in which participants did not), the data shows that Group Y outscored Group N in both overall feedback integration rate (86.2% vs. 76.9%) and effective feedback integration rate (68.8% vs. 51.6%).

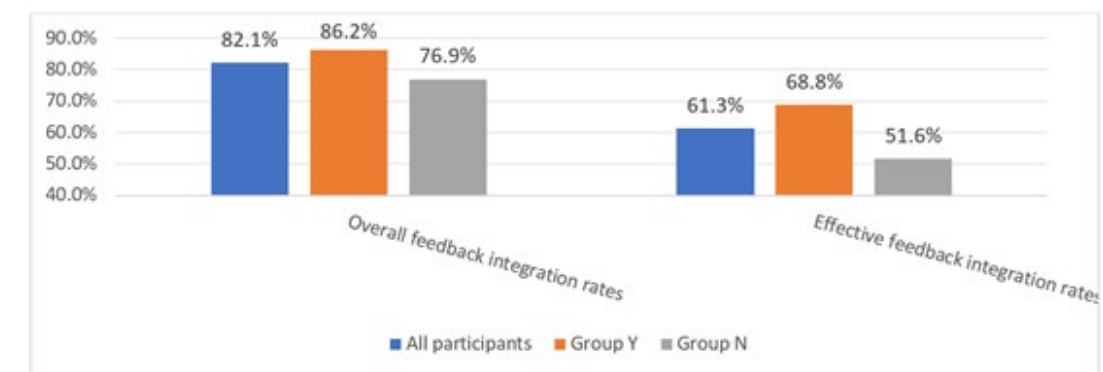


Figure 1. Overall vs. effective feedback integration rates for this SPS essay task

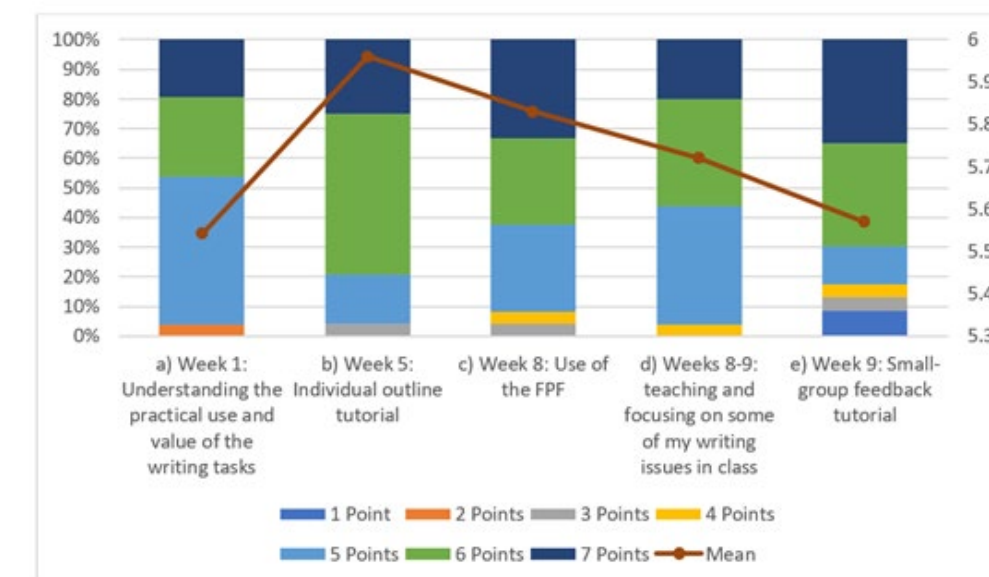


Figure 2. Effectiveness of specific intervention activities in 7-point Likert Scale (7 indicating most effective)

In terms of the effectiveness of different intervention activities, Figure 2 shows the relevant survey data on the participants' perceptions of the effectiveness of different intervention activities implemented at different times of this SPS writing process. Among the five specific intervention activities, the individual outline tutorial and the use of the FPF were the two intervention activities with the highest means (5.96 and 5.83, respectively) on a 7-point Likert Scale (7 indicating most effective). These two activities also had higher percentages of 6 and 7 points (80% and over 60% of the total number of participating students, respectively). One more activity that contained a higher percentage of 6 and 7 points combined was the small-group feedback tutorial. However, it is surprising that its mean score was very low, as the participants' opinions on this activity were quite polarised, and more students chose very negative options compared with other activities. Responses to the following open-ended question on the possible reasons also show that the students' perceptions on the effectiveness of small-group tutorials varied substantially, as among the 21 respondents who left valid comments, the positive features mentioned most was the oral and face-to-face feature making comprehension easier (four times), while the negative features mentioned most were tutorial duration too short (five times) and low efficiency (three times).

When turning to the seven essay aspects identified based on the official marking descriptor of this task, the feedback integration rates (in quantity and quality) also varied. Apart from Aspect "I. Topic selection and scope" in which I only wrote five pieces of feedback among the 24 first draft samples, I provided a total number of 241 pieces of individual written feedback across the other six aspects, with the number in each aspect ranging from 31 to 59 pieces. As Figure 3 shows, among these six aspects, the three aspects with the highest overall feedback integration rates (issues addressed in quantity) were "II. Idea development with clear logic" (92.5%), "V. Paragraph writing and cohesion" (89.2%), and "IV. Overall structure, cohesion, and thesis statement" (83.8%). Other aspects, including "III. Source use" (78.0%), "VII. Vocabulary" (77.4%), and "VI. Grammar" (73.0%), were less addressed. Interestingly, Group Y, in which all participants used the FPF, outscored Group N, in which students did not, in almost all these aspects, to different degrees.

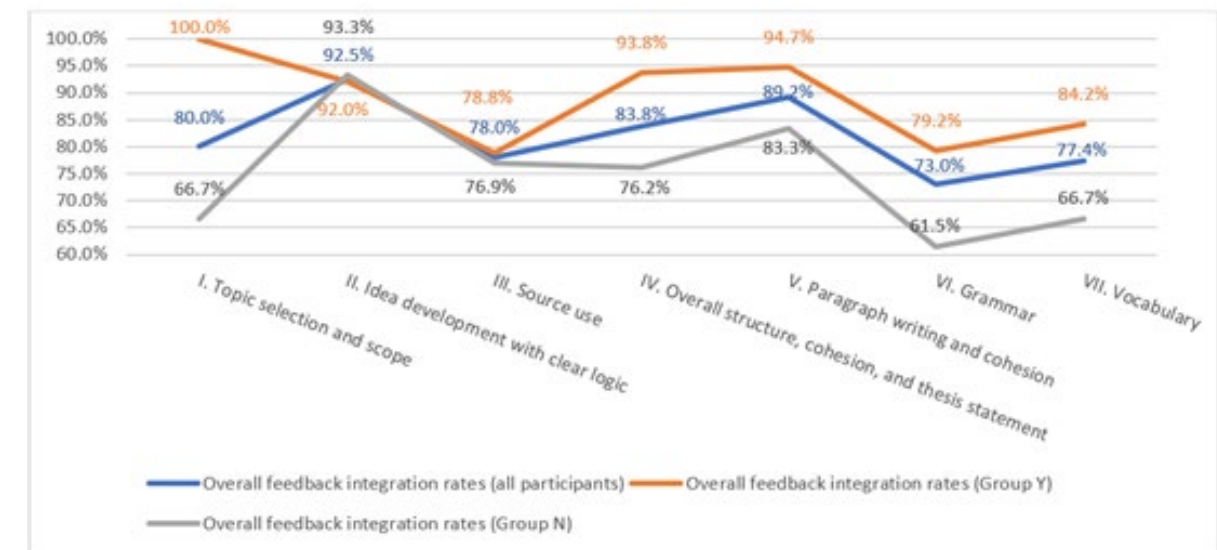


Figure 3. Overall feedback integration rates in seven essay aspects

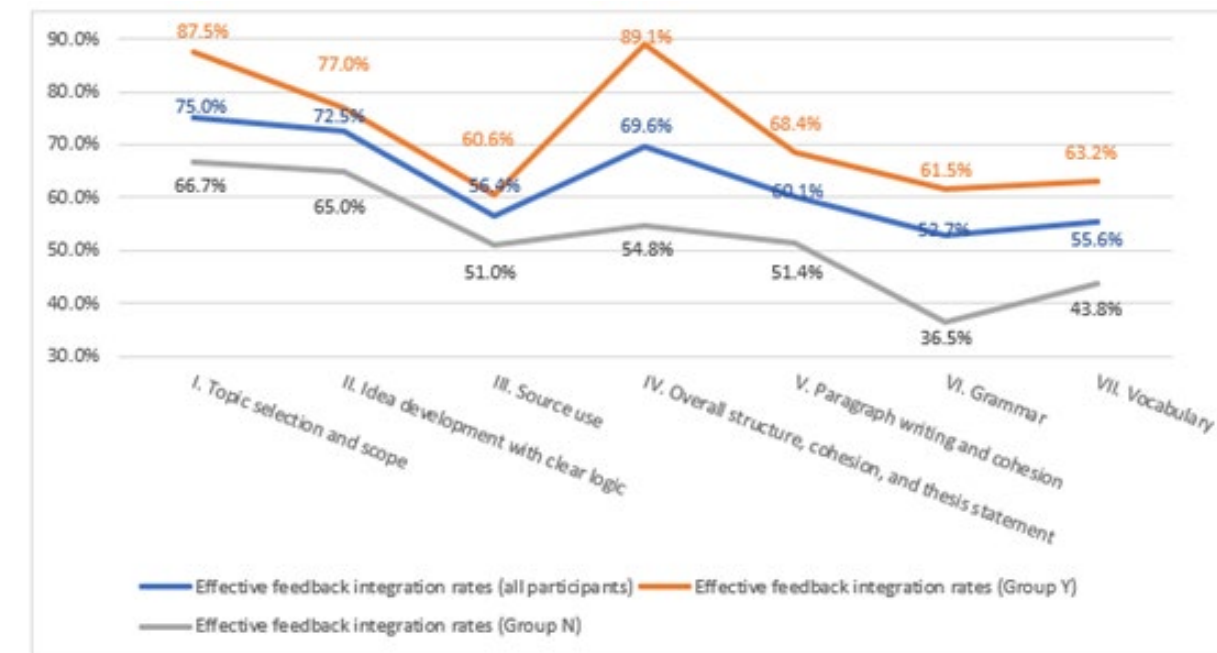


Figure 4. Effective feedback integration rates in seven essay aspects

Finally, Figure 4 on effective feedback integration rates (quality) across these same seven aspects also shows very similar results: Aspects II, IV, and V were the top three aspects most effectively addressed in quality by this intervention writing process (72.5%, 69.6%, and 60.1% respectively); Aspects III, VII, and VI least effectively addressed (56.4%, 55.6% and 52.7% respectively). Comparing Group Y with Group N, the former outscored the latter in every aspect, with the most significant difference in "IV. Overall structure, cohesion, and thesis statement" and "VI. Grammar".

Discussion

RQ1: What specific intervention activities tend to be more effective in enhancing the students' feedback engagement and uptake?

Based on the results, the effectiveness of different intervention activities varies.

The most effective feedback activities are (1) the outline tutorial conducted in the planning stage and (2) the FPF during the initial draft writing. As this writing task was new to most of my students, they saw the outline tutorial as a useful and timely opportunity to ask for more guidance and boost their writing confidence. This may be one of the reasons this activity was ranked top in both Surveys 1 and 3 and received very positive feedback in Survey 2. This implies that when teachers assign writing coursework, apart from the tutorials on the first draft at a later stage, they can consider using this outline tutorial as an additional activity (and maybe also as part of the assessment). The other very effective feedback activity, adapted from Maas's (2017) teaching design about LDF, is the use of the FPF. By using FPF, my students gained my help on the writing aspects that they actually preferred (and ideally via their preferred feedback modes). Although some of their preferences on feedback mode were not fully addressed due to relevant module requirements, their voices were heard, respected, and responded to; thus, a good rapport between us was established, which is essential to encourage them to engage with the feedback and the writing process. Furthermore, the FPF also guided my students to evaluate their own drafts by asking them to choose the writing aspects that they wanted their teacher to focus more on, which pushed them to use their cognitive strategies to assess the paper's strengths and weaknesses, making them value teacher feedback more. Therefore, I would suggest that more teachers use the FPF as it can provide students with more respect and communication opportunities, as well as guide them to conduct self-evaluation on their writing and practice learning autonomy.

On the other hand, it is worth discussing the effectiveness of the small-group tutorial. Although I grouped my students by their writing issues according to their individual written feedback, there were still times during these tutorials when I interacted with only one student, while the others sat there, listening (or maybe not) and waiting. This may be the reason why they did not see it as very effective. Teachers using this activity should be more

careful about student grouping, presentation of content (preparing group notes in bullet points), students' awareness of learning from each other, and time management so as to achieve higher effectiveness. Based on students' divergent opinions on different feedback modes, if more time or more feedback delivery modes are allowed, teachers can respect learners' individual needs and use their preferred modes, hopefully to boost the feedback engagement and uptake.

RQ2: What aspects of the summative assessment (Situation-Problem-Solution (SPS) essay) can the feedback address more effectively?

It was found that there are differences in feedback integration rates of the seven SPS essay aspects, which indicates the differences in these aspects addressed by the writing intervention. It seems that generally global issues (mainly Aspects II, IV, V) from the first draft were more effectively addressed than local issues (mainly Aspects VI, VII). One possible reason for this is that close correction and annotations on the essay drafts were not permitted for this assessment (only listing one or two examples of the error in the paragraph-long feedback at the end of the essay was allowed), making it more challenging for the students to track and fix these local issues. This may also indicate that, unlike addressing global issues, addressing local issues requires closer correction, which makes teacher feedback more time-consuming and less efficient than machine-generated versions. Another contributing factor may be my students' priorities on the global issues over the local issues, as Aspects II and IV were selected by most students in their FPFs. This might be because the students were used to using some technologies for grammar and vocabulary check, such as Grammarly, which they could access via their university student accounts. Global issues in writing, on the other hand, are much more complex and thus less effectively addressed by technologies, so the students value their teachers' feedback more.

It is a bit surprising that source use (Aspect III) was also among the lowest integration aspects, although more focus was given in classroom teaching upon my students' requests in Survey 2. This is probably due to its complexity caused by its high contextualisation and the high requirements of skill combination (i.e. reading, summarising, critical analysing, reorganising, integrating to the writer's own text, and referencing). Consequently, more focus could be provided on how to effectively address this aspect in academic writing.

Conclusion

In this research, I investigated the effectiveness of a refined writing process in terms of feedback engagement and feedback uptake, and identified more effective feedback activities and the writing aspects that were more efficiently addressed in an EAP writing task. With reference to the results, relevant pedagogical suggestions were also provided.

Looking ahead, the points discussed above merit further exploration, and more research with larger sample sizes is necessary to determine whether the findings from this study can be generalised, shedding more light on the true effectiveness of teacher writing feedback and how to improve it. This is particularly important as our learners now have access to extensive feedback from increasingly advanced AI applications. Writing feedback should not be a one-way product in writing; instead, it should be a process-oriented, interactive opportunity to develop and evaluate multiple academic skills, as well as to promote and celebrate interpersonal communication.

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

Data statement: The relevant data, including instruments (three questionnaire surveys and the FPF), survey and FPF data, and the participants' first and final drafts, have been uploaded to IRIS online data repository. All these data are available here: <https://doi.org/10.48316/K1er7-Z0wJ9>

References

Alharbi, M. A. (2022). Exploring the impact of teacher feedback modes and features on students' text revisions in writing. *Assessing Writing*, 52, 100610. <https://doi.org/10.1016/j.asw.2022.100610>

Dressler, R., Chu, M.-W., Crossman, K., & Hilman, B. (2019). Quantity and quality of uptake: Examining surface and meaning-level feedback provided by peers and an instructor in a graduate research course. *Assessing Writing*, 39, 14-24. <https://doi.org/10.1016/j.asw.2018.11.001>

Lee, I. (2014). Revisiting teacher feedback in EFL writing from sociocultural perspectives. *TESOL Quarterly*, 48(1), 201-213. <https://doi.org/10.1002/tesq.153>

Maas, C. (2017). Receptivity to learner-driven feedback in EAP. *ELT Journal*, 76(3), e8. <https://doi.org/10.1093/elt/ccw003>

Smith, C. (2021). How does the medium affect the message? Architecture students' perceptions of the relative utility of different feedback methods. *Assessment & Evaluation in Higher Education*, 46(1), 54-67. <https://doi.org/10.1080/02602938.2020.1733489>

Vygotsky, L. (1978). *Mind in society: the Development of Higher Psychological Processes*. Harvard University Press.

Winstone, N. E., Nash, R. A., Rowntree, J., & Parker, M. (2017). 'It'd be useful, but I wouldn't use it': barriers to university students' feedback seeking and recipience. *Studies in Higher Education*, 42(11), 2026-2041. <https://doi.org/10.1080/03075079.2015.1130032>

Zhang, Z. & Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing*, 36, 90-102. <https://doi.org/10.1016/j.asw.2018.02.004>

Zhang, Z. & Hyland, K. (2022). Fostering student engagement with feedback: An integrated approach. *Assessing Writing*, 51. <https://doi.org/10.1016/j.asw.2021.100586>

Appendix 1. Feedback Preference Form (FPF)

Additional: Research material on writing feedback material (Week 7): SPS First Draft Feedback Preference Form

Your Name: _____

Important note: This preference form is for a research project 'Action research on improving the writing feedback engagement and uptake' approved by XITLU Ethics Review Panel(s), and is **optional** for the students to use. You need to agree to participate in the research project in order to make his form effective.

1. Modes of feedback	Which type(s) of the following feedback do you want to receive for your SPS report first draft? (Please rank the top 3 that you like most: 1 meaning you prefer the most) (The teacher will try her best to provide you feedback in your preferred modes, but this is not guaranteed due to module and administration requirements.) _____ Individual feedback – written feedback (in sentences or short paragraphs) at the end of my report or separate for me to check on Learning Mall Core _____ Individual feedback – a checklist with short written feedback _____ Individual feedback – a teacher's audio recording _____ Individual feedback – an individual (private) conversation with teacher _____ Group feedback – small-group (3-5 students) feedback in class or a group tutorial with the teacher _____ Group feedback – whole-class feedback with some teaching/learning activities _____ Other (please specify) _____
2. Aspects of feedback to focus on	Which of the following aspects do you want your teacher to focus on* when giving feedback? (Please rank the top 3 aspects that you want your teacher to give feedback on most: 1 meaning the aspect in which you want feedback most) _____ Whether I have met the core task requirements _____ Whether my ideas are developed enough and with clear logic _____ Whether I have used sources appropriately and effectively _____ Whether my overall structure, cohesion, and thesis statement are satisfactory _____ Whether my individual paragraphs and paragraph cohesion are clear _____ Whether I have some vocabulary and spelling issues (including academic style of language) _____ Whether my grammar, sentence structure, and punctuation are satisfactory _____ Other (please specify) _____
3. Any other question(s) or request(s) that you may have	I want to ask (for):

* This means your teacher will pay particular attention to the aspects that you choose, but she may also give feedback on other aspects if necessary; this does NOT mean your teacher will only look at the aspects that you choose here.

Source: Maas, C. (2017). Receptivity to Learner-Driven Feedback in EAP. *ELT Journal*, 71(2), 127-140. <https://doi.org/10.1093/elt/ccw065>

Evaluating the Accuracy and Reliability of AI Detector Results: A Comparative Analysis



Mengqin Mao (Lexie) has 7 years of English teaching experience in XJTLU. She holds a Master's degree in TESOL (Teaching English to Speakers of Other Languages) from University College London. Her research interests include interactive classroom design and technology enhanced language learning. She investigates AI's pedagogical impact—exploring how AI can be utilized in language learning and ethical AI-human collaboration.

Mengqin.Mao@xjtlu.edu.cn



Xidan Li has 8 years of English teaching experience in Xi'an Jiaotong-Liverpool University in terms of English for Academic Purposes. She has been teaching students from a variety of majors in Year 1, Year 2 and Pre-sessionals with a Master degree in TESOL (Teaching English to Speakers of Other Languages) from University of Liverpool. Her primary research interests include technology-enhanced learning and teaching, and learner autonomy.

Xidan.Li@xjtlu.edu.cn

1. Introduction

THE RAPID INTEGRATION of generative Artificial Intelligence (AI) into academic writing has led to a significant challenge for academic integrity. As language teachers, we have witnessed AI tools reshaping the classroom and transforming every aspect of learning and teaching over the past several years, bringing forth concerns related to academic integrity, particularly when students start to use advanced AI tools such as ChatGPT to complete their essay writing. While advanced AI offers support for brainstorming, it can also cause potential plagiarism and cheating problems. As a result, some educational institutions are considering using AI detection software to identify AI-generated text, as these detectors can scan a paper and flag any text that might be machine-written.

However, both teachers and students may have concerns while using the AI detectors. Are these detectors actually accurate and reliable? Can we trust them? More specifically, how do the detectors distinguish between human and AI-generated text, especially when human intervention is involved?

2. Literature review

Can we trust the software designed to detect its use?

As generative AI becomes a common tool for students in academic writing, a critical challenge emerges: Can we trust the software designed to detect its use?

Technical Accuracy

A primary concern in the literature is the technical accuracy of detection algorithms. Early studies by Weber-Wulff et al. (2023) tested various commercial and open-source detectors and found that none was fully reliable. Accuracy rates varied dramatically, and all detectors share a weakness: they often incorrectly flag human-written text as AI-generated. This false-positive problem is especially unfair to non-native English speakers, whose writings are more likely to be misread by algorithms (Liang, Yuksekgonul, Mao,

Wu & Zou, , 2023). Such inaccuracies suggest that AI detectors are not perfect, and their mistakes may lead to real consequences for students. A wrong flag can lead to anxiety and frustration, which is harmful to students’ mental health.

Human Modification

Further research explored how human modification of AI text affects accuracy. Results consistently show that detectors are particularly vulnerable to human-modified text. With minimal human intervention, such as simple paraphrasing or editing, detector performance can degrade, leading to a substantial drop (Perkins, Roe, Postma, McGaughran & Hickerson, 2023; Sadasivan, Kumar, Balasubramanian, Wang, & Feizi, 2023; Weber-Wulff et al., 2023). It can be inferred that detectors are only looking for patterns in pure AI writing. Once a person changes words or sentences, the tools get confused and become unreliable. This creates a problematic grey zone where AI-assisted student work is the hardest to judge because superficial features of texts, such as sentence patterns and word choices, can be changed easily with minor editing.

Bridging the Gap

As most studies test raw accuracy on simple datasets, rather than the mixed, edited text students produce, there is little research on tools such as Turnitin and GPTZero under identical conditions. This study bridges the gap by testing how AI detectors work in practice. We tested Turnitin and GPTZero on 100% student-written essays and specially created samples that were lightly edited, heavily rewritten, or mixed with AI.

This study aims to compare two major AI detectors: Turnitin & GPTZero and answer the following three research questions.

- How accurate are the AI detectors: Turnitin & GPTZero?
- How consistent are the results obtained by Turnitin & GPTZero?
- How does manual editing affect the detection of ChatGPT-generated text?

We tested the tools on a set of sample academic writings that included pure human writing, pure AI writing, and several mixed versions in which AI texts were edited by us. The results provide a clear look at the current capabilities and limitations of the two widely

used detectors, fostering more responsible AI use among both students and teachers in academic settings.

3. Methodology

3.1 AI Detectors Tested

This small-scale study evaluates the accuracy and consistency of two publicly available AI detectors: Turnitin and GPTZero. The former is primarily used to detect plagiarism in educational institutions, with an AI detection function as an integrated feature. Most university lecturers have access to the software, and its AI function may serve as a grading reference. GPTZero was selected as a representative commercial AI detector because it claims to be the most accurate tool according to third-party benchmarking (Adam, Cui & Tian, 2025). Students tend to use commercial AI detectors such as GPTZero before submission since universities may restrict their access to Turnitin’s AI feature.

3.2 Data Collection

12 short writing samples on the same assignment prompt, ‘Evaluate the strengths and weaknesses of a technology’, were tested by the two AI detectors. The average word count was about 300 words. Only evaluation paragraphs were analyzed because they had the highest originality, which were stripped of diagrams and formulaic expressions. The samples include varying degrees of ChatGPT integration, ranging from 100% human-written to 100% ChatGPT-generated.

Sample Number	Description
Sample 1 & Sample 2	100% human-written
Sample 1a & Sample 2a	Main ideas are human-written, supporting ideas are ChatGPT-generated
Sample 3 & Sample 4	100% ChatGPT-generated
Sample 3a & Sample 4a	Sample 3 & Sample 4 paraphrased with synonyms by human
Sample 3b & Sample 4b	Sample 3 & Sample 4 heavily paraphrased by human using multiple techniques
Sample 3c	Mixed content: Sample 3 & Sample 3b combined
Sample 4c	Mixed content: Sample 4 & Sample 4b combined

Table 1. An overview of the samples used.

4. Results and Discussion

4.1 Accuracy of Turnitin and GPTZero

Turnitin	True Positive	True Negative	False Positive	False Negative	Missing CGPT Detection	False Accusation	Accuracy
Sample 1	24	N/A	N/A	0	N/A	0%	100%
Sample 2	17	N/A	N/A	0	N/A	0%	100%
Sample 1a	5	0	4	0	44.40%	0%	55.60%
Sample 2a	2	0	4	1	57.10%	14.30%	28.60%
Sample 3	N/A	11	0	N/A	0%	NA	100%
Sample 4	N/A	13	0	N/A	0%	NA	100%
Sample 3c	0	5	0	6	0%	54.50%	45.50%
Sample 4c	7	0	6	1	42.90%	7.10%	50%
Average					24.10%	12.65%	72%

Table 2. Turnitin Responses.

GPTZero	True Positive	True Negative	False Positive	False Negative	Missing CGPT Detection	False Accusation	Accuracy
Sample 1	24	N/A	N/A	0	N/A	0%	100%
Sample 2	17	N/A	N/A	0	N/A	0%	100%
Sample 1a	1	4	0	4	0%	44.40%	55.60%
Sample 2a	0	4	0	3	0%	42.90%	57.10%
Sample 3	N/A	11	0	N/A	0%	NA	100%
Sample 4	N/A	13	0	N/A	0%	NA	100%
Sample 3c	1	5	1	4	9.1%	36.40%	54.50%
Sample 4c	1	6	0	7	0%	50%	50%
Average					1.51%	28.95%	77%

Table 3. GPTZero Responses.

Note. Performance metrics calculated as follows: *Missing CGPT Detection=FN/(TP+TN+FP+FN). *False Accusation=FP/(TP+TN+FP+FN). *Accuracy=(TP+TN)/(TP+TN+FP+FN). *GPTZero likeliness lower than 0.03 is regarded as human writing. *Heavily paraphrased sentences are regarded as human-written in 3c and 4c.

Table 2 and Table 3 show the accuracy percentages for Turnitin and GPTZero, with details of successful, missed, and false detections. The numbers in the tables represent the smallest unit of detection: sentence. Accuracy percentage is calculated based on the ratio of correctly classified sentences to all sentences. According to the tables, GPTZero’s average accuracy is slightly higher than Turnitin’s (77% vs 72%), but this does not necessarily mean GPTZero is more trustworthy than Turnitin. We still need to take a close look at their performance across different scenarios and analyze the data drivers.

4.2 Key Similarities and Differences

Both Turnitin and GPTZero achieve 100% accuracy when testing the clear-cut samples (Samples 1, 2, 3, and 4), which are either entirely human-written or entirely ChatGPT-generated. Precise detection of purely ChatGPT-generated samples can be attributed to linguistic patterns such as the use of natural language, predictable word sequences and AI-preferred phrases (Barlow & Chen, 2024). However, the robustness of both Turnitin and GPTZero’s results decreases when human intervention is involved. The tables suggest that once human and AI contents are combined, the accuracy rate drops under 60%. This result aligns with the findings of Weber-Wulff et al. (2023) that most AI detectors have higher accuracy when dealing with human-written texts compared with human-edited AI-generated texts.

One important consideration to judge the AI technology applied in diagnostics is reliability in that diagnoses is closely related to the treatment of the patients and one mistakes could cause serious consequence. One notable example is IBM Watson for Oncology, which uses AI to provide evidence-based treatment recommendations. However, there have been reports of inconsistencies in its recommendations, which raised concerns about its reliability. In 2018, internal documents revealed that Watson had suggested “unsafe and incorrect” cancer treatments, underscoring the critical importance of rigorous testing and validation [4]. The care recommendations generated by AI technology combine with doctors’ experience will reduce the probability of misdiagnose. This collaborative approach minimizes the risk of misdiagnosis by cross-verifying AI insights with clinical expertise, ensuring more comprehensive and reliable patient assessments [5]. However, the AI technology is not reliable and needs to be improved for those diseases that have not enough information and for fully automatic diagnose.

Figure 1. Sample 2a’s Turnitin Result.

Note. *Sentences in black are human-written. *Sentences in red are ChatGPT-generated. *The sentence highlighted in blue is flagged as AI-generated by Turnitin.

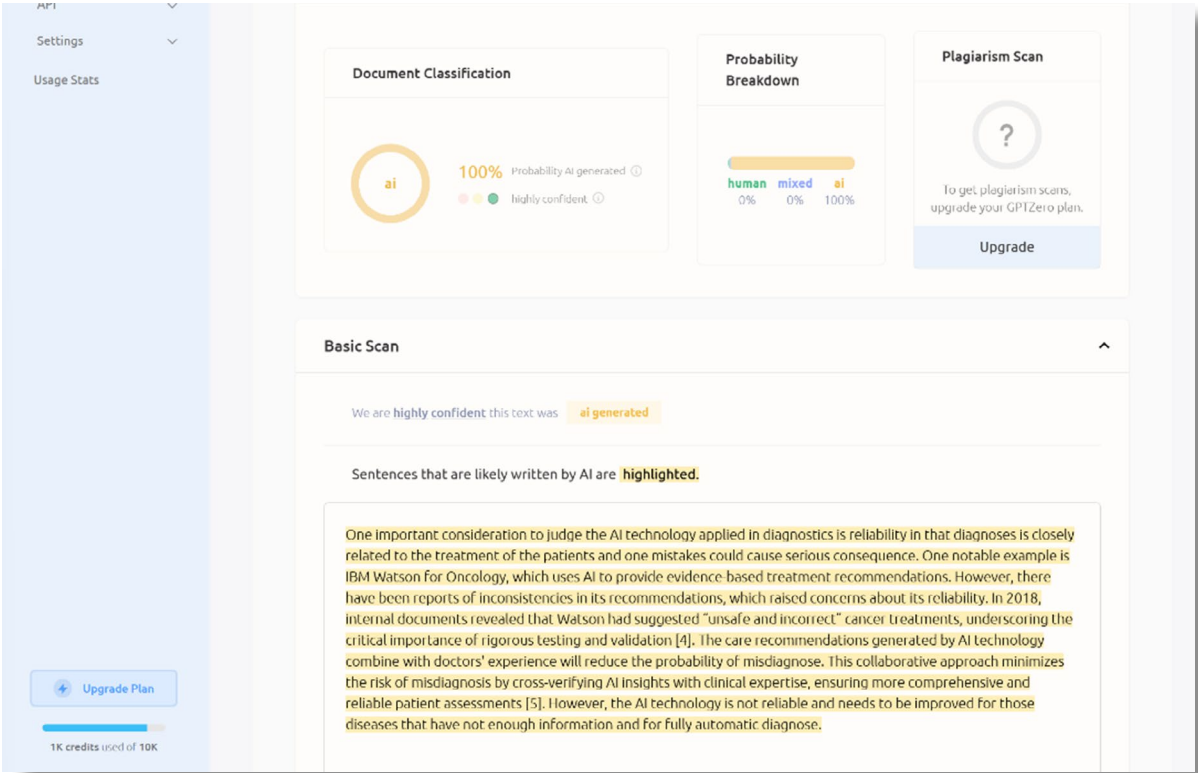


Figure 2. Sample 2a's GPTZero Result.

The major difference between the two AI detectors' judgment lies in the significant data variance regarding 'missing ChatGPT detection' and 'false accusation'. Overall, Turnitin has a high 'missing ChatGPT detection' percentage (24.1%), but comparatively low 'false accusation' percentage (12.65%). In contrast, GPTZero has a very low 'missing ChatGPT detection' percentage (1.51%), but a quite high 'false accusation' percentage (28.95%). In other words, GPTZero is more biased towards human-written texts, while Turnitin is more conservative. Take sample 2a (main ideas are human-written and supporting ideas are generated by ChatGPT) as an example, Figure 1 and Figure 2 show the responses from Turnitin and GPTZero. Turnitin only highlights 1 sentence as AI-generated, but GPTZero claims the whole paragraph as AI's work. The actual AI rate is only 51.6% as colour-coded in Figure 1. This false-detection pattern applies to the 2 detectors' responses in Sample 1a as well. One possible interpretation of this huge inconsistency is that, as a commercial AI detector, GPTZero tends to flag all the potential AI-generated texts because it is usually used for students to self-check their AI rate before submission. It does not hurt to flag any possible cases while proofreading. On the other hand, Turnitin is a plagiarism tool used by universities worldwide. When they develop the AI detection plug-in, it can be easily accessed by the markers and influence their academic judgment. They should be more cautious about false accusations, even with the compromise of a possible unfair advantage.

Sample 4: Despite its advancements, eye-tracking technology faces several accuracy-related challenges. External factors such as lighting conditions, reflections from glasses, and rapid head movements can interfere with the precision of eye-tracking data.

Sample 4a: Despite its **improvements**, eye-tracking technology **has many reliability-related problems**. External **elements** such as lighting **environment**, reflections from glasses, and **fast** head movements can **influence** the **accuracy** of eye-tracking data.

Figure 3. Example of Synonym Paraphrased Sentences.

Sample 4: In marketing, accurate eye-tracking can reveal what captures consumer interest, allowing for the optimization of advertisements and product placements.

Sample 4b: **For example**, in marketing, **what costumers are interested in can be accurately detected by eye-tracking technology**. **As a result**, the optimization of advertisements and product placements **can be realized**.

Figure 4. Example of Heavily Paraphrased Sentences.

4.3 Human Intervention

There are two types of human intervention in the tested samples: 1. ChatGPT was used to produce the whole text, but humans paraphrased the sentences with synonyms only (Samples 3a and 4a). 2. ChatGPT was used to produce the whole text, but humans heavily paraphrased the sentences by using multiple techniques such as adding cohesive devices, changing sentence structure, changing part of speech and using synonyms (Samples 3b and 4b). Samples 3c and 4c are designed to further challenge the AI detectors' ability as their components are more complicated (a combination of 3&3b and 4&4b). Figure 3 and Figure 4 demonstrate how the sentences are manually paraphrased.

Table 4 reflects the two AI detectors' classifications of the samples involving human intervention. Regarding Samples 3a and 4a, subsequent simple synonym paraphrasing does not change the detection result compared to Samples 3 and 4. Both Turnitin

AI Percentage	Sample 3a	Sample 4a	Sample 3b	Sample 4b	Sample 3c	Sample 4c
Turnitin	100%	100%	40%	0%	100%	5%
GPTZero	100%	100%	3%	2%	82%	82%

Table 4. Responses from Two AI Detectors Regarding the Likelihood of Human-Paraphrased Samples.

and GPTZero identify 3a and 4a as 100% generated by AI. Howard (1995) defines this type of writing, which changes a few phrases only without thoroughly re-expressing ideas, as patchwriting. This can be a good warning to the students who want to simply generate an essay by using AI and change a few words to avoid academic misconduct. However, if the student puts more effort into paraphrasing by adopting diverse paraphrasing techniques, AI detectors may consider the writing as human-produced. The AI likelihood of Samples 3b and 4b is nearly 0%, except for Turnitin's response to Sample 3b. The amount of heavily paraphrased sentences also plays a key role, as reflected in the AI probability of Samples 3c and 4c. When 100% AI-generated sentences are mixed with heavily paraphrased sentences, there is a big chance that the writing will be identified as highly likely to have been created by AI. Although teachers usually encourage students to write independently first, then use AI to proofread or polish the language use, it is very difficult to tell who owns the authorship when the process is reversed.

5. Conclusion

In conclusion, this comparative analysis evaluated the accuracy and consistency of Turnitin and GPTZero in detecting AI-generated text, particularly when modified by human paraphrasing. The findings show that both detectors achieve high accuracy (100%) on purely human-written or AI-generated texts. However, when it comes to human-AI mixed content, the accuracy falls below 60%. Another significant finding is the differing technical bias of the two tools. While Turnitin is more conservative, missing more AI text (24.1%) but making fewer false accusations, GPTZero prioritizes minimizing missed AI content but has a high false accusation rate (28.95%). This does not mean we can simply say one tool is better than the other, but it reflects their different intended uses: Turnitin is more suitable for formal assessment, while GPTZero can be used for self-checking. In addition, this study demonstrates that simply using synonyms for paraphrasing fails to evade detection, whereas using multiple paraphrasing techniques can make it far less likely for detectors to flag as AI-generated.

This study acknowledges several limitations related to the evolving AI technology and the limited sample size. Firstly, the research was conducted in June 2024. Given the rapid evolution of both generative AI models and detection algorithms, these specific statistics may not accurately represent the current capabilities of these tools. In addition, while this study provides insights by analyzing 12 discipline-specific writing samples, its scope is limited to examining only two AI detectors and a small sample size.

Therefore, it is recommended that educators and institutions interpret AI detector results with caution, especially for mixed-origin work and avoid using these results as the only proof of academic misconduct. Teachers and schools should not rely 100% on AI detection tools to determine whether a student used AI improperly. Instead, they should provide clear guidelines on acceptable AI use and emphasize the effectiveness of comprehensive paraphrasing by using multiple techniques. For students, this study sends a straightforward message that complete reliance on AI or superficial lexical modifications is not acceptable. To avoid detection, students have to earn authorship by demonstrating originality or by using different paraphrasing skills to recompose. In future research, more diverse datasets and updated detectors can be used to develop better guidelines on how to use them in academic settings.

References:

- Adam, A., Cui, A., & Tian, E. (2025, October 15). *GPTZero is officially the most accurate commercial AI detector*. Retrieved from: <https://gptzero.me/news/gptzero-accuracy-stats/>
- Barlow, A., & Chen, V. (2024, October 14). *How AI content detectors work to check for AI*. Retrieved from: <https://gptzero.me/news/how-ai-detectors-work/>
- Howard, R. M. (1995). Plagiarisms, authorships, and the academic death penalty. *College English*, 57(7), 788-806. <https://doi.org/10.2307/378403>
- Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. *Patterns*, 4(7), 100779. <https://doi.org/10.1016/j.patter.2023.100779>
- Perkins, M., Roe, J., Postma, D., McGaughan, J., & Hickerson, D. (2023). Detection of GPT-4 generated text in higher education: Combining academic judgement and software to identify generative AI tool misuse. *Journal of Academic Ethics*. 22(1), 1-25. doi:10.1007/s10805-023-09492-6
- Sadasivan, V. S., Kumar, A., Balasubramanian, S., Wang, W., & Feizi, S. (2023). Can AI-generated text be reliably detected? [Preprint] arXiv. <https://arxiv.org/abs/2303.11156>
- Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P., & Waddington, L. (2023). Testing of detection tools for AI-generated text. *International Journal of Educational Integrity*, 19, Article 26. <https://doi.org/10.1007/s40979-023-00146-z>

Appendix: Text Samples

Sample 1:

Smart technologies incorporated in cities bring numerous benefits that enhance the quality of life for citizens. First, in terms of transportation, smart technologies revolutionize traffic monitoring by utilizing sensor networks, GPS, and air quality sensors. With these technologies, real-time traffic information becomes readily available, enabling authorities and residents to make informed decisions about routes and travel times, leading to reduced traffic congestion. Moreover, in healthcare, IoT technologies improve patient monitoring and workflow efficiency in hospitals. IoT enables the tracking of patients, staff, and ambulances, ensuring timely and accurate medical care. It also facilitates data collection and sensing, leading to faster diagnoses and real-time information on patient health indicators. Despite the numerous advantages, the introduction of smart technologies in cities can exacerbate existing inequalities and create disadvantages for certain segments of the population. The case of individuals with disabilities highlights the potential exclusion of marginalized communities. For instance, inaccessible smart city services, such as digital payment-only parking systems, can hinder the participation of people with disabilities who face challenges in utilizing such technologies. This lack of consideration for diverse needs and limited accessibility underscores the importance of inclusive planning and implementation of smart technologies. Furthermore, relying solely on data-driven algorithms can lead to decisions that prioritize the interests of the affluent while neglecting the needs of the poor, minorities, or disabled individuals. The theoretical example of building a road instead of a subway demonstrates how decisions solely based on data might overlook the specific circumstances and needs of underprivileged citizens who rely on public transportation. In such cases, the moral dimensions and diverse needs of individuals should be considered to ensure equitable outcomes. To mitigate these disadvantages, inclusive planning and moral considerations measures must be prioritized in the implementation of smart technologies in cities.

Sample 1a:

Stability is an essential criterion for evaluating an automatic parking system. In general, an automatic parking system with high stability can work with a slight delay. In fact, there are many factors which can affect the stability of the system such as the vehicle’s velocity, controller and computing delay [12]. An example of this can be seen in Tesla’s Enhanced Autopilot system, which provides automatic parking capabilities. The vehicle’s velocity must be carefully controlled to ensure smooth parking maneuvers, and any sudden changes can impact system performance [13]. However, for a long term, with the development of computers, computers and sensors may process data quicker and time delay will be reduced greatly. This improvement will enhance the overall efficiency and reliability of automatic parking systems, making them more responsive and user-friendly. As a result, drivers will experience smoother and faster parking maneuvers, further increasing the adoption and trust in autonomous vehicle technologies. Despite the slight delay, from this perspective, this application is basically in line with this criterion.

Sample 2:

4.1 Reliability

One important consideration to judge the AI technology applied in diagnostics is reliability in that diagnoses is closely related to the treatment of the patients and one mistakes could cause serious consequence. An experiment shows that the accuracy of applying AI technology in diagnosing different types of diseases was around 80%. As for lung cancer, the accuracy was around 64% which is quite lower than other diseases [4]. The reason for this result could mainly because the lung cancer data that used in this experiment as the database for the system is not enough. Another research for an AI system Watson designed by IBM Company which is used to diagnose cancer indicates that the overall accuracy of the care recommendations for leukaemia cases was 82.6% [5]. Therefore, as assisted tools for doctors the diagnosis produced by AI technology in particular diseases is reliable. The care recommendations generated by AI technology combine with doctors’ experience will reduce the probability of misdiagnose. However, the AI technology is not reliable and needs to be improved for those diseases that have not enough information and for fully automatic diagnose.

4.2 Efficiency

Another criterion for evaluating the application of AI in medical diagnoses is efficiency in that this technology is design to assist doctors and simplify the process. The new technology ANN is considered to be more efficient and faster than the original algorithm which needs to input large amounts of knowledge to produce the results [2]. However, the training phase for ANN is time-consuming because it requires a large number of data sets to obtain useful conclusions and become efficient. In addition, it also takes time to find the suitable data sets [7].

AI technologies that used different algorithms are applied to diagnose different diseases in order to become more efficient. Indeed, this method works efficiently in its particular field [6]. However, the separate usage of different algorithms leads to a problem that they could not work interdisciplinary. For example, an AI technology that is used to diagnose heart diseases has no idea about cancer. It is not efficient for comprehensive diagnosis.

Sample 2a:

One important consideration to judge the AI technology applied in diagnostics is reliability in that diagnoses is closely related to the treatment of the patients and one mistakes could cause serious consequence. One notable example is IBM Watson for Oncology, which uses AI to provide evidence-based treatment recommendations. However, there have been reports of inconsistencies in its recommendations, which raised concerns about its reliability. In 2018, internal documents revealed that Watson had suggested “unsafe and incorrect” cancer treatments, underscoring the critical importance of rigorous testing and validation [4]. The care recommendations generated by AI technology combine with doctors’ experience will reduce the probability of misdiagnose. This collaborative approach minimizes the risk of misdiagnosis by cross-verifying AI insights with clinical expertise, ensuring more comprehensive and reliable patient assessments [5]. However, the AI technology is not reliable and needs to be improved for those diseases that have not enough information and for fully automatic diagnose.

Sample 3:

One of the most significant strengths of 3D printing is its unparalleled flexibility. This technology allows for the creation of highly customized and complex designs that would be difficult, if not impossible, to achieve with traditional manufacturing methods. Designers can easily modify digital models and print variations without the need for retooling or significant downtime. This flexibility is especially valuable in industries such as healthcare, where patient-specific medical devices and prosthetics can be produced with precision, and in aerospace, where complex geometries and lightweight structures are essential. Furthermore, 3D printing supports the use of a wide range of materials, from plastics to metals and even biocompatible substances, enabling diverse applications across various fields.

However, the flexibility of 3D printing also comes with certain weaknesses. One major limitation is the variability in material properties. While 3D printing can utilize different materials, the mechanical properties of printed items often do not match those produced by conventional methods, which can restrict their use in critical applications. Additionally, the flexibility in design and production can lead to longer print times and increased costs for complex or large-scale items. This can make 3D printing less efficient for mass production compared to traditional manufacturing methods. The need for specialized knowledge and skills to operate and maintain 3D printing equipment further adds to the challenges, potentially limiting its accessibility and widespread adoption.

Sample 3b

To start with, flexibility is one of the most important advantages of 3D printing as the personalized and complicated design provided by the technology is difficult for traditional method to achieve. One reason is it prevents retooling or downtime when designers need to edit digital models and print variations. Also, 3D printing provides precise production on specific medical devices and lightweight structure in healthcare and aerospace. The last point is that it is flexible in terms of materials. Various kinds of substances can be used such as plastics, metals and biocompatible materials, which leads to in different applications in a wide range of fields.

However, there are also disadvantages. The most significant one is the variability in material properties. Although 3D printing allows the flexibility of printed materials, the mismatch between the mechanical properties of those printed items and the ones produced by the old ways is likely to restrict the uses of 3D printing in certain applications. Another limitation is the lower efficiency it may bring compared to conventional way since the flexibility in design and production may cause longer time to print and more cost for mass production. Lastly, the accessibility widespread use can be a challenge without expertise. To operate and maintain 3D printing machines, professional knowledge and specialized skills are required.

Sample 3a

One of the most vital advantages of 3D printing is its excellent flexibility. This technology provides the initiation of highly personalized and complicated designs that would be challenging, if possible, to accomplish with conventional manufacturing methods. Designers can easily edit digital models and print changes without the need for recasting or significant downtime. This flexibility is particularly useful in fields like healthcare, where patient-specific medical equipment and limbs can be made with carefulness and in aerospace, where complex geometries and light structures are highlighted. Moreover, 3D printing accepts the use of many materials, from plastics to metals and even biocompatible elements, enabling various applications in different areas.

However, the flexibility of 3D printing also brings certain weaknesses. One main disadvantage is the variability in material properties. While 3D printing can use various substances, the mechanical properties of printed items usually do not match those produced in traditional ways, which can restrain their use in important applications. In addition, the flexibility in design and production can cause longer print times and increased expenses for large items. This can make 3D printing less constructive for mass production compared to traditional manufacturing methods. The need for technological information and competence to work and keep 3D printing equipment further leads to the difficulties, possibly constraining its availability and public attention.

Sample 3c:

To start with, flexibility is one of the most important advantages of 3D printing as the personalized and complicated design provided by the technology is difficult for traditional method to achieve. Designers can easily modify digital models and print variations without the need for retooling or significant downtime. Also, 3D printing provides precise production on specific medical devices and lightweight structure in healthcare and aerospace. Furthermore, 3D printing supports the use of a wide range of materials, from plastics to metals and even biocompatible substances, enabling diverse applications across various fields. However, there are also disadvantages. One major limitation is the variability in material properties. Although 3D printing allows the flexibility of printed materials, the mismatch between the mechanical properties of those printed items and the ones produced by the old ways is likely to restrict the uses of 3D printing in certain applications. Additionally, the flexibility in design and production can lead to longer print times and increased costs for complex or large-scale items. This can make 3D printing less efficient for mass production compared to traditional manufacturing methods. Lastly, the accessibility widespread use can be a challenge without expertise. To operate and maintain 3D printing machines, professional knowledge and specialized skills are required.

Sample 4:

Eye-tracking technology has made significant strides in accuracy, offering precise insights into human visual attention and behavior. Modern eye-tracking systems can detect and record eye movements with remarkable precision, often down to fractions of a degree. This high level of accuracy is invaluable in various fields, such as user experience (UX) research, psychology, and neurology, where understanding visual focus and attention is crucial. In marketing, accurate eye-tracking can reveal what captures consumer interest, allowing for the optimization of advertisements and product placements. Additionally, in assistive technology, eye-tracking enables individuals with disabilities to interact with computers and other devices effectively, enhancing their quality of life. The continuous improvements in hardware and algorithms have further minimized latency and increased the reliability of eye-tracking data.

Despite its advancements, eye-tracking technology faces several accuracy-related challenges. External factors such as lighting conditions, reflections from glasses, and rapid head movements can interfere with the precision of eye-tracking data. Calibration processes are often necessary to achieve optimal accuracy, which can be time-consuming and user-specific. Furthermore, the accuracy of eye-tracking can diminish over extended periods, requiring periodic recalibration. Individual differences in eye physiology, such as variations in pupil size and eye shape, can also affect the consistency of data. Additionally, while high-end eye-tracking systems are highly accurate, they can be prohibitively expensive, limiting accessibility for some researchers and institutions. These limitations highlight the need for ongoing advancements to enhance the robustness and affordability of eye-tracking technology.

Sample 4a:

Eye-tracking technology has made great progress in accuracy, providing accurate insights into human visual attention and behavior. Modern eye-tracking systems can notice eye movements with high precision, usually include small changes. This high degree of accuracy is valuable in different fields, such as user experience (UX) research, psychology, and neurology, where comprehending visual focus and attention is vital. In marketing, accurate eye-tracking can show what stimulates consumer interest, allowing for the boost of advertisements and product placements. Additionally, in assistive technology, eye-tracking allows users with disabilities to communicate with computers and other devices effectively, improving their quality of life. The unceasing improvements in hardware and algorithms have further decreased latency and raised the accuracy of eye-tracking data.

Despite its improvements, eye-tracking technology has many reliability-related problems. External elements such as lighting environment, reflections from glasses, and fast head movements can influence the accuracy of eye-tracking data. Calibration processes are often needed to gain maximum accuracy, which can be time-consuming and user-specific. Furthermore, the accuracy of eye-tracking can decrease over extended periods, needing regular recalibration. Individual variations in eye physiology, such as differences in pupil size and eye shape, can also influence the consistency of data. Additionally, while top eye-tracking systems are super accurate, they can be extremely pricy, restricting accessibility for some scholars and organizations. These limitations stress the need for continuous advancements to strengthen the robustness and affordability of eye-tracking technology.

Sample 4b:

One benefit brought by eye-tracking technology regarding accuracy is the precise insights into human visual attention and behavior. Even subtle eye movements can be captured precisely by modern eye-tracking systems. Many fields which are very demanding about understanding visual focus and attention can benefit from this highly accurate technology, including user experience (UX) research, psychology, and neurology. For example, in marketing, what costumers are interested in can be accurately detected by eye-tracking technology. As a result, the optimization of advertisements and product placements can be realized. In addition, in assistive technology, eye-tracking technology can assist the disabled to communicate with computers and other devices effectively, which improves their quality of life. With the continuously improved hardware and algorithms, the technology will have less latency and more reliability.

Apart from the advantages, eye-tracking technology also has many problems in terms of accuracy. The technology's accuracy may be influenced due to external factors such as lighting conditions, reflections from glasses, and rapid head movements. Moreover, it is time-consuming to adjust the calibration in order to optimize accuracy. Furthermore, the accuracy of eye-tracking may decrease with time, which requires regular recalibration. Individual data may also be inconsistent because of the differences in pupil size and eye shape. In addition, the top eye-tracking systems with high accuracy are extremely pricy, restricting accessibility for some scholars and organizations. The limitations mentioned above stress the need for continuous development to strengthen the robustness and affordability of eye-tracking technology.

Sample 4c:

Eye-tracking technology has made significant strides in accuracy, offering precise insights into human visual attention and behavior. Even subtle eye movements can be captured precisely by modern eye-tracking systems. This high level of accuracy is invaluable in various fields, such as user experience (UX) research, psychology, and neurology, where understanding visual focus and attention is crucial. For example, in marketing, what costumers are interested in can be accurately detected by eye-tracking technology. As a result, the optimization of advertisements and product placements can be realized. Additionally, in assistive technology, eye-tracking enables individuals with disabilities to interact with computers and other devices effectively, enhancing their quality of life. With the continuously improved hardware and algorithms, the technology will have less latency and more reliability.

Apart from the advantages, eye-tracking technology also has many problems in terms of accuracy. External factors such as lighting conditions, reflections from glasses, and rapid head movements can interfere with the precision of eye-tracking data. Moreover, it is time-consuming to adjust the calibration in order to optimize accuracy. Furthermore, the accuracy of eye-tracking can diminish over extended periods, requiring periodic recalibration. Individual data may also be inconsistent because of the differences in pupil size and eye shape. Additionally, while high-end eye-tracking systems are highly accurate, they can be prohibitively expensive, limiting accessibility for some researchers and institutions. The limitations mentioned above stress the need for continuous development to strengthen the robustness and affordability of eye-tracking technology.

IN PRACTICE

Teaching Students to Think **With** AI, Not **Through** It:

A Practical Gen-AI-Supported Writing Feedback Design for EAP Classrooms



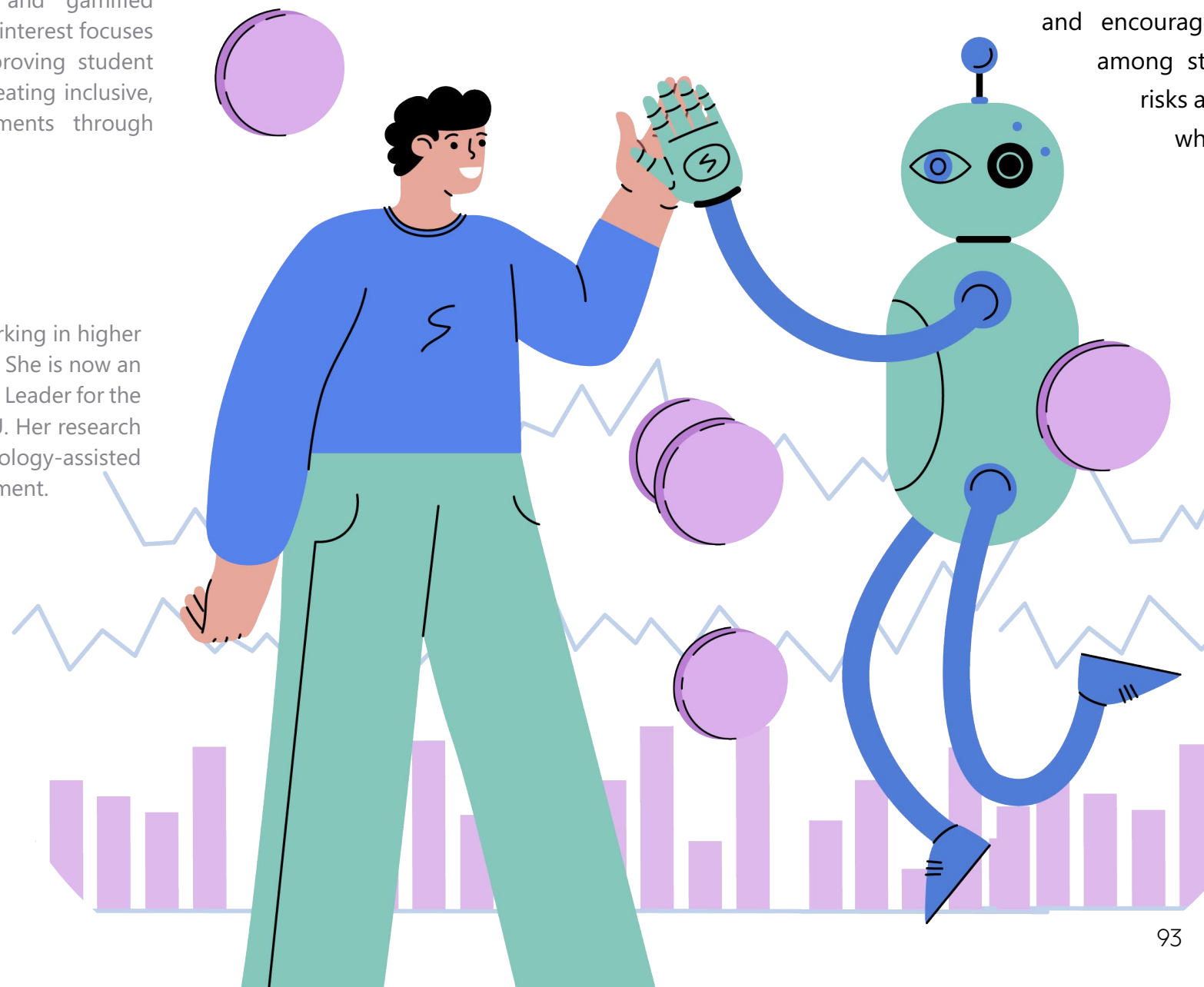
Bin Xiang is an Associate Language Lecturer at Xi'an Jiaotong-Liverpool University. Specializing in AI-enhanced and gamified language learning, her research interest focuses on generative AI's role in improving student writing, critical thinking, and creating inclusive, effective educational environments through technology.

Bin.Xiang@xjtlu.edu.cn



Jiashi Wang has been working in higher education for almost 10 years. She is now an EAP instructor and the Module Leader for the Y1 Advanced module in XJTLU. Her research interests lie primarily in technology-assisted teaching and student engagement.

Jiashi.Wang@xjtlu.edu.cn



From Tool Adoption to Pedagogical Design

THE RAPID UPTAKE OF generative artificial intelligence (Gen-AI) in tertiary education has sparked extensive discussion among language teachers and researchers, particularly regarding its implications for academic writing instruction. A growing body of research suggests that Gen-AI tools can support students' writing development by improving idea generation, language fluency, coherence, and organisation (Urzúa et al., 2025). For time-pressed teachers, these tools also appear to offer a partial solution to the long-standing challenge of providing detailed formative feedback on multiple drafts.

At the same time, serious concerns remain. Studies have warned that Gen-AI may generate inaccurate or fabricated content, fail to fully understand task context, oversimplify complex arguments, and encourage surface-level revision or overreliance among students (Kim, Detrick & Li, 2024). These risks are particularly problematic in EAP contexts, where writing development depends not only on linguistic accuracy but also on critical thinking, source use, and disciplinary awareness. As a result, an emerging consensus in the literature argues that the pedagogical value of Gen-AI lies not in what it produces, but in how students are taught to engage with it critically (Qian, 2025; Zhu, Wang, & Searsmith, 2025).

This article shares a teaching practice that responds to these challenges in a realistic EAP setting. Rather than positioning Gen-AI as a shortcut for revision or a replacement for teacher feedback, the approach treats AI as a cognitive partner that can support self-regu-

lated learning when carefully designed and mediated. The focus is on helping students learn to evaluate, justify, and reflect on AI-generated feedback, while also addressing the practical issue of feedback overload for teachers.

Teaching Context and Pedagogical Problem

The teaching practice was implemented with 101 Year-1 advanced students enrolled in an English for Academic Purposes course at Xi'an Jiaotong-Liverpool University (XJTLU), an English-medium instruction university in China. Most students had prior experience with AI tools, although this experience was informal and largely unguided. Some students already rely on AI to "polish" their writing, often accepting suggested revisions wholesale without questioning their relevance or accuracy.

From a teaching perspective, two interconnected problems emerged. First, students needed regular, detailed feedback to support their transition into academic writing at the university level, yet providing this feedback for multiple drafts across a large cohort was not sustainable. Second, although students were using Gen-AI, they often failed to engage with its feedback critically. Rather than reflecting on the AI's suggestions, many tended to accept or implement them

without question. This indicates a need not only to develop evaluation skills, but first to foster the essential habit of pausing to think twice. These challenges also raised a key pedagogical question: how could Gen-AI be integrated into writing instruction in a way that reduced teacher workload while simultaneously strengthening, rather than weakening, students' critical engagement with feedback?

Key Concepts: Gen-AI, Self-Regulation, and AI Literacy

To address this question, several key concepts informed both the design of the intervention and how it was presented to students. Gen-AI was defined in simple terms as a text-based tool capable of producing, revising, or commenting on writing in response to user prompts. Self-regulated learning was introduced as the process by which learners plan, monitor, and evaluate their own learning, particularly during complex tasks such as writing and revision (Zimmerman, 2002). Finally, critical AI literacy was framed as the ability to question AI output, recognise its limitations, and make informed decisions rather than treating AI suggestions as authoritative (Chan & Colloton, 2024).

These ideas were deliberately explained

in accessible language. Students were repeatedly reminded that effective academic writing involves making choices, and that AI, like teacher feedback or peer comments, represents only one perspective in the revision process.

Pedagogical Design: Applying the 6 Ps Framework

The instructional design followed the six-step pedagogical framework for supporting self-regulated learning in academic writing using text-based generative AI proposed by Kong, Lee, and Tsang (2024). The framework is composed of two parts. The first part is teacher-led preparation (P1-P3), which

includes planning content and structure, preparing and generating prompts for AI tools, and previewing and verifying tool's output. The second part is student-led engagement (P4-P6), encompassing producing writing using corrected output, peer review and proofreading, and portfolio-tracking and reflection (see Figure 1).

Part 1: Teacher-led Preparation

In the first phase, the teacher prepared sample paragraphs and essays that modelled academic writing conventions, including clear argumentation, appropriate register, and effective source integration. These samples functioned as reference points, allowing students to compare their own writing and AI-generated suggestions against established standards rather than relying on AI output

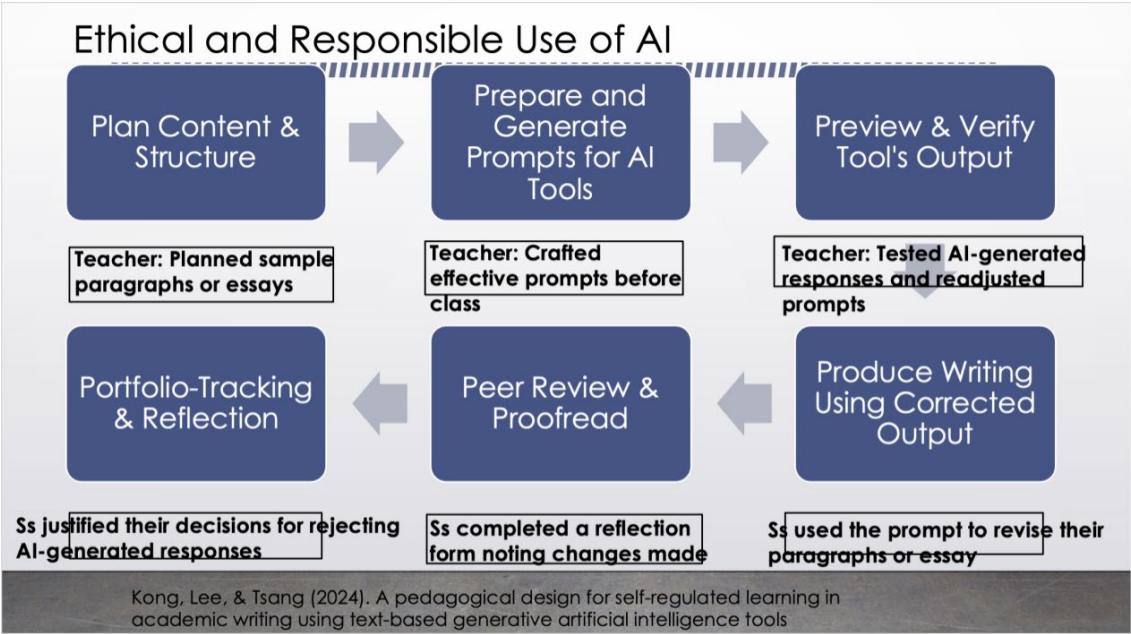


Figure 1. The six-step framework proposed by Kong et al. (2024)

alone.

The second phase focused on prompt design. Instead of asking students to “improve” their writing in general, the teacher developed and tested a set of targeted prompts aligned with task objectives, such as improving topic sentence clarity or academic register (see Figure 2). These prompts were trialled using sample student texts, and revisions were made to reduce vague or misleading AI responses. Although students were encouraged to experiment with their own prompts later, this initial scaffolding was

essential for setting expectations and ensuring quality interaction with the AI tools.

The third preparatory step involved previewing AI responses. By examining how the AI responded to different drafts and prompts, the teacher was able to anticipate common issues, such as over-generalisation or changes that altered meaning. This step increased teacher confidence in releasing responsibility to students and reduced the likelihood of students receiving unhelpful feedback.

Prompt: I’m a freshman in college learning EAP. Here are 10 academic style rules I need to follow:

	Academic style rules	Instead, academic writers SHOULD:
1	Do not use informal vocabulary (<i>kids, so on or a bit</i>) or abbreviations (<i>etc.</i>)	Use academic vocabulary.
2	Avoid vague or general terms (<i>things, stuff or good</i>).	Use topic-specific or academic vocabulary.
3	Do not state your opinion, personal feelings or be subjective (<i>I think..., I hope..., luckily</i>).	Try to be objective .
4	Avoid using personal pronouns (<i>I, our, you, etc.</i>) or imperatives – directly telling the reader what to do (<i>Stop polluting! Take this as an example!</i>).	Use third-person pronouns (<i>they, she</i>) or use passive voice.
5	Do not use contractions (<i>shouldn’t or won’t</i>).	Use the full terms (<i>should not or will not</i>).
6	Do not use <i>And, So or But</i> at the beginning of a sentence.	Use formal linking devices (<i>Additionally, However, or Therefore</i>).
7	Avoid using strong or absolute statements (<i>as we all know, obviously, or everybody knows</i>).	Use hedging language (<i>generally, research suggests, might</i>).
8	Do not ask rhetorical questions (<i>How can this be solved?</i>).	Use direct statements (<i>There are several ways to solve this problem.</i>).
9	Do not use proverbs (<i>every coin has two sides, practice makes perfect</i>) or clichés (<i>double-edged sword, with the development of...</i>).	Use direct statements (<i>However, there are also some disadvantages to</i>) and academic vocabulary.
10	Do not write integers between 1-9 (<i>1, 2, 3rd, etc.</i>) in paragraphs or essays.	Write the full word (<i>one, two, third</i>).

Please review my essay below and point out where I can improve according to these rules.

[insert your own paragraph below]

Figure 2. Teacher-tested prompts used by students to check academic register

Part 2: Student-Led Engagement

Once the groundwork had been laid, students took the lead. Over a twelve-week period, they completed three formative writing tasks: an argumentative paragraph, an argumentative essay, and a source-based paragraph. Before submission, students used XIPU AI, a university-funded Gen-AI tool, to revise their drafts using the teacher-provided prompts to start with. Importantly, students were allowed to interact freely with the AI, and examples were provided in class via PPT slides (see Figure 3), enabling students to ask follow-up questions and seek clarification when needed.

To deepen learning, peer discussion in class

was encouraged after revision. Informal conversations about AI suggestions often prompted students to articulate doubts, compare alternatives, and justify choices, thereby externalising their thinking. These discussions helped shift AI use from an individual, hidden activity to a shared learning experience.

The most critical element of the design was a structured reflection form, which functioned similarly to a writing portfolio. Students documented their original drafts, AI-generated suggestions, the revisions they adopted, and the suggestions they rejected. They were required to justify rejections and write brief reflection memos explaining what they had learned from the revision process. An example of good practice was shown to students (see Figure 4).

Free interactions allowed

***Note: If you want AI to explain certain feedback, you may send another prompt as follows:**

- “Please explain the revised topic sentence as follows. How is it different from mine? Why did you make that change?”
- Original: “A vegetarian diet may not be healthier than a diet includes meat as a diet includes meat includes overall nutrition.”
- Revised: “A vegetarian diet may not be as healthy as a diet that includes meat, primarily because meat provides essential nutrients that are often lacking in vegetarian diets.”

Feel free to interact with AI to help you understand and reflect on the feedback given.

Figure 3. Classroom examples illustrating free interaction with the AI

- Prompt: Revise the following paragraph in terms of organization (cohesion and coherence), idea development and language and point out how it can be improved: [essay here]

Original Paragraph:

A vegetarian diet may not be healthier than a diet includes meat as a diet includes meat includes overall nutrition. Meats has some important nutrition that vegetable does not include. For example, meat is mainly composed of protein. Not noly could protein strengthen your body and muscle but it could also balance blood press by highing it. That is why many vegetarian people are thinner and often feel powerless and headache. Besides, people are born with the desire for meat. Reseachers have shown that meat is with uncertain source that is essntial for nerves. The nerve without the uncertain source would become less active, which will make people's think slowly and easy to get angry.

Revised Paragraph:

A vegetarian diet may not be healthier than a diet that includes meat, as a diet with meat provides overall nutrition. Meat contains essential nutrients that vegetables do not. For example, meat is primarily composed of protein, which not only strengthens the body and muscles but also helps regulate blood pressure by increasing it. This explains why many vegetarians are thinner and often feel powerless and experience headaches. Additionally, humans are naturally inclined to desire meat. Researchers have shown that meat contains certain nutrients essential for nerve function. Without these nutrients, nerves become less active, leading to slower thinking and increased irritability.

Improvements:

1. **Organization (Cohesion and Coherence):**
 - Clarified the main argument in the first sentence.
 - Used transitional phrases like "For example," "Additionally," and "This explains why" to improve the flow between ideas.
2. **Idea Development:**
 - Clearly separated the discussion of protein benefits and the natural human inclination for meat.
 - Provided specific examples and explanations to support each point.
3. **Language:**
 - Corrected grammatical errors and awkward phrasing.
 - Used more precise and formal language to maintain an academic tone.

Detailed Breakdown:

1. **Introduction of Main Argument:**
 - **Original:** "A vegetarian diet may not be healthier than a diet includes meat as a diet includes meat includes overall nutrition."
 - **Revised:** "A vegetarian diet may not be healthier than a diet that includes meat, as a diet with meat provides overall nutrition."
2. **First Point (Nutritional Content):**
 - **Original:** "Meats has some important nutrition that vegetable does not include. For example, meat is mainly composed of protein."
 - **Revised:** "Meat contains essential nutrients that vegetables do not. For example, meat is primarily composed of protein."
3. **Benefits of Protein:**

Figure 4. Example of the reflection form for students to document their original drafts, AI-generated suggestions, adopted revisions, and rejected suggestions.

Further guidance on the use of AI

Good example

You reflection should look like this:

I do not accept the following suggestions:

Avoid personal pronouns: The original paragraph used "young adults" multiple times. The revised version replaced some instances with "individuals" to avoid repetition and maintain formality.

I think the concept "young adults" differs a lot from just "individuals", "young adults" are a group of people who have just walked over the threshold of the real society but perhaps not mature enough to handle all the problems they come into, this group is much more special than just "individuals". We are not supposed to deliberately avoid repetition sacrificing the accuracy of concepts.

Figure 5. Example of good practice in which a student justifies the rejection of AI-generated responses.

Highlight: Encouraging Critical Dialogue With AI

Students were taught how to engage in dialogue with the AI tool rather than treating it as a one-way editor. They were shown how to request explanations for specific revisions and compare original and revised versions to understand the rationale behind changes. This practice aligns with research suggesting that interactive engagement with AI supports deeper learning than passive acceptance of corrections (Qian, 2025; Zhu et al., 2025). To further support this process, examples of high-quality

student reflections from previous cohorts were shared (see Figure 5), demonstrating what thoughtful engagement with AI feedback looked like in practice. For example, the student argued that the original wording "young adults" should not be replaced by "individuals" because there is a clear difference between these two concepts, whereas the AI's suggestion sounded overly generalized. This explicit justification process encouraged metacognitive awareness and reinforced the idea that rejecting AI feedback can be an academically valid decision.

From a theoretical perspective, the design integrates principles of self-regulated learning, formative assessment, and critical AI literacy. Reflection forms and justification memos supported students' ability to monitor and evaluate their

writing decisions, while the structured use of AI feedback reframed revision as an active process of choice rather than error correction. For teachers, AI handled much of the surface-level feedback, allowing instructional effort to shift towards task design, modelling, and meta-feedback, which are more pedagogically impactful and sustainable.

Limitations and Implications

Despite its benefits, the approach has limitations. First, the initial investment in prompt design and material preparation is substantial, from experimenting with effective AI prompts to preparing in-class demonstration examples. To reduce the upfront burden, educators can build a shared library of proven prompts and adaptable templates with colleagues, and utilize AI tools to help generate and refine initial drafts of both prompts and instructional materials. In addition, not all students immediately value reflective writing. Some are reluctant to devote additional time to completing reflection forms and may even resort to minimal responses such as “I accept all the suggestions” to reduce their workload. To address this, future adaptations could simplify the reflection form or adjust its requirements.

A more concerning issue is that some students may still accept AI-generat-

ed feedback uncritically, especially when it appears fluent or authoritative. These students often reported not being confident enough to question the feedback. For example, if AI comments that an essay's topic sentences are weak or that it fails to use a wide range of cohesive devices, a student unfamiliar with these concepts may simply agree with the feedback rather than evaluate its validity. For those students who are unfamiliar with such concepts, teachers should set clear learning objectives and provide concept-checking tasks to reinforce fundamental understanding, such as identifying effective topic sentences or categorising cohesive devices in a sample text. For language-specific issues, tools such as Grammarly can be introduced to help students verify language accuracy independently.

Conclusion

Gen-AI does not eliminate the need for writing instruction; instead, it demands more thoughtful pedagogical design. When used uncritically, AI risks encouraging dependency and superficial revision. When embedded within a structured, reflective framework; however, it can support students in becoming more autonomous, critical, and self-regulated writers. By teaching students to think with AI rather than through it, EAP practitioners can harness the strengths of Gen-AI while preserving the core educational values of academic writing.

References

- Chan, C. K. Y. & Colloton, T. (2024). *Generative AI in Higher Education: A ChatGPT Effect*. Routledge. <https://doi.org/10.1080/1360080X.2024.2447091>
- Kim, J., Yu, S., Detrick, R., & Li, N. (2024). Exploring students' perspectives on generative AI-assisted academic writing. *Education and Information Technologies*. 30(1), 1265-1300. <https://doi.org/10.1007/s10639-024-12878-7>
- Kong, S.-C., Lee, J. C.-K., & Tsang, O. (2024). A pedagogical design for self-regulated learning in academic writing using text-based generative artificial intelligence tools: 6-P pedagogy of plan, prompt, preview, produce, peer-review, portfolio-tracking. *Research and Practice in Technology Enhanced Learning*, 19, 030. <https://doi.org/10.58459/rptel.2024.19030>
- Qian, Y. (2025). Pedagogical Applications of Generative AI in Higher Education: A Systematic Review of the Field. *TechTrends* 69, 1105-1120. <https://doi.org/10.1007/s11528-025-01100-1>
- Urzúa, C. A. C., Ranjan, R., Saavedra, E. E. M., Badilla-Quintana, M. G., Lepe-Martínez, N., & Philominraj, A. (2025). Effects of AI-Assisted Feedback via Generative Chat on Academic Writing in Higher Education Students: A Systematic Review of the Literature. *Education Sciences*, 15(10), 1396. <https://doi.org/10.3390/educsci15101396>
- Zhu, X., Wang, C., & Sears Smith, D. (2025). *Writing With Machines and Peers: Designing for Critical Engagement with Generative AI*. <https://doi.org/10.48550/arXiv.2511.15750>
- Zimmerman, B. J. (2002). Becoming a Self-Regulated Learner: An Overview. *Theory Into Practice*, 41(2), 64-70. https://doi.org/10.1207/s15430421tip4102_2

REVIEW

Gamification in Review Classes

Enhancing Student Engagement with

WAYGROUND



Ke Long is a Spanish Associate Language Lecturer from Modern Languages Center, Global Cultures and Languages Hub at Xi'an Jiaotong-Liverpool University. Her research interests are task-based language teaching and technology-enhanced learning and teaching.

Ke.Long@xjtlu.edu.cn

Background

I HAVE BEEN TEACHING beginner-level Spanish classes this semester (S1, AY25/26), covering five units in total. At the end of each unit, we have a review session. However, I have often found that students' interest and engagement drop significantly during review classes. As a result, the learning atmosphere in review sessions tends to be less engaging, and the attendance was not as good as in classes introducing new learning content.

Faced with this dilemma, I have been looking for a way to better engage students and avoid doing repetitive grammar and vocabulary exercises.

Introduction

Gamification refers to the application of principles and methods from games to non-game aspects of life (Kapp, 2012). Kapp (2012) also adds that in principle, any process or activity can be gamified, but it is most commonly used in business, marketing,

and education. Szabó and Kopinska (2023) pointed out that in recent years, gamification has become more and more important in foreign language teaching. It not only boosts motivation but also helps structure and organize learning processes. Among the gamification platforms, Kahoot, a game-based student response system, is frequently incorporated into classes to maintain student engagement (Hu, 2023). I have been using Kahoot for many years in my teaching. However, I often notice that, although students like Kahoot, sometimes they become bored with it because they use it too much in different courses. Additionally, because the speed of responses is very important in Kahoot, I found that students tended to focus more on speed, rather than carefully thinking about the question during the game. More importantly, I observed that when one student at the same table submitted the answer, sometimes others would ask for the answer without any engagement with the question.

Taking into consideration all these problems, I decided to look for a platform that could

maintain the gamification feature but place more emphasis on accuracy. The solution I found was Wayground. Wayground offers a wide variety of game modes and supports personalized learning.

Procedure

On the platform, teachers can create questions in multiple choice and fill-in the blank formats with a free account. In addition, they can choose between individual and team game modes. I prefer the individual mode as it guarantees active participation from every student. The game mode on Wayground is based on a certain topic, such as climbing a mountain or getting the most points in Strike and Shield (see figure 1). As Szabó and Kopinska (2023) argued, students' motivation is strongly influenced by the goal of catching up and learners can complete levels individually. Smirani and Yamani (2024) further commented that the most effective gamification techniques in education include stars, badges, progress bars, leaderboards, storylines and visuals.

These elements help motivate students, track progress, and make learning more interactive and enjoyable. The game

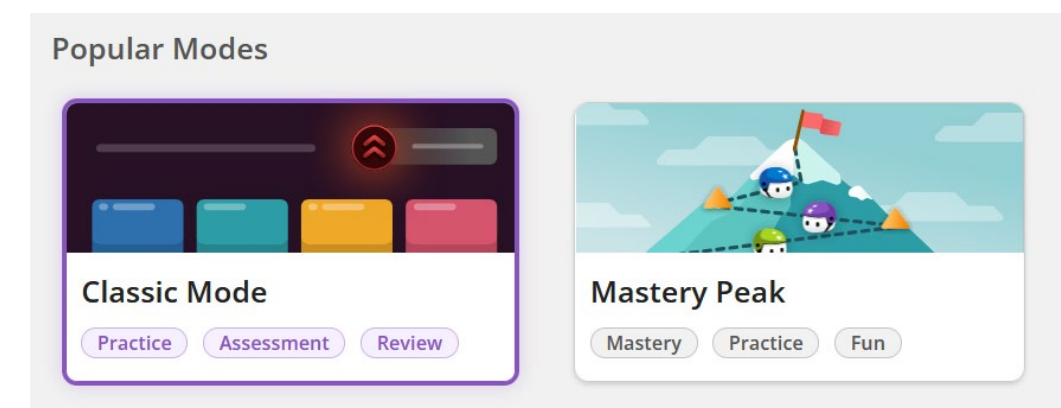


Figure 1 Popular modes on Wayground

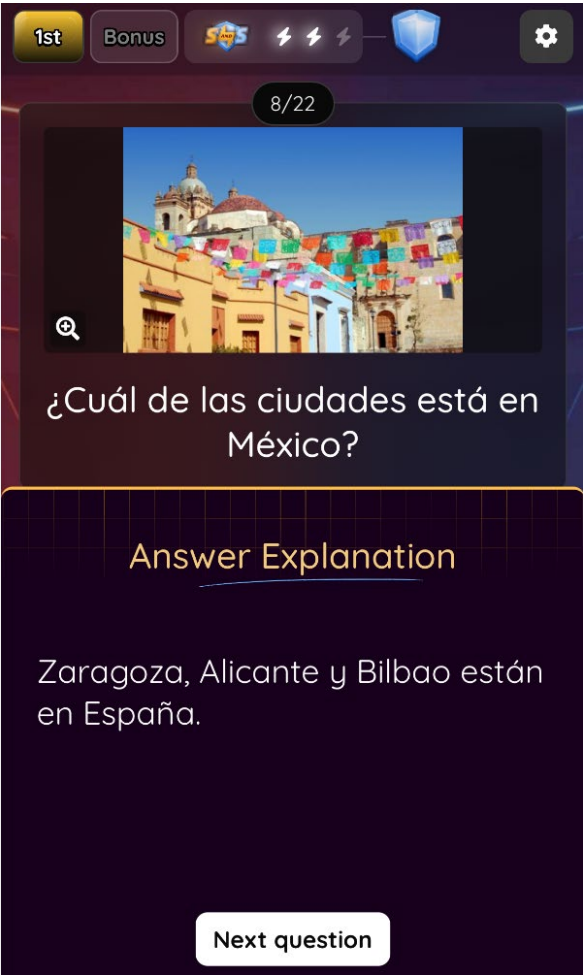


Figure 2 Feedback for question

mode and visual design of Wayground were found to be appealing to students in my class and were effective in engaging them.

During the activity, students were randomly given questions so that they could answer at their own pace without being influenced by others. The teacher can also choose whether to show the feedback for each question after it is answered. I always choose instant feedback so that students can immediately identify their mistakes (see figure 2). Additionally, the leaderboard is based on accuracy, that is, speed is not as important as accuracy (see figure 3). This is a significant feature on this platform since it encourages students to engage deeply with the questions rather than

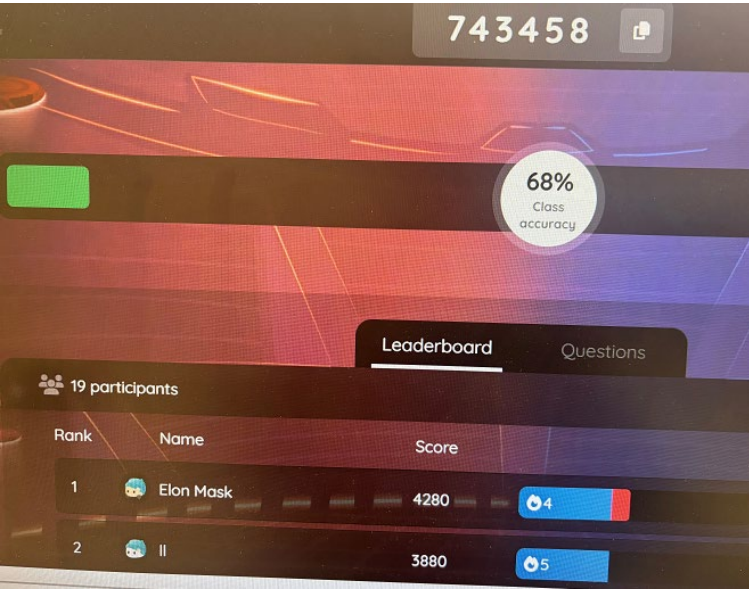


Figure 3 Leaderboard on Wayground

rushing through them. During the activity, when I walked around the classroom, students often asked me questions when they were unclear about the feedback. It demonstrated that they were thinking and engaged in this gamified activity.

Every student is eager to achieve a higher ranking and this is the motivation of game-based activities. Liu, Zhou, Li & Ye. (2022) indicated that while competition can boost motivation and participation, it may also produce stress and anxiety, particularly among students with lower proficiency levels. In Wayground, students were shown the leaderboard throughout the game and I encountered this problem with weaker students. Once in my class, a weaker student gave up when he saw that he was always at the bottom of the board. At that moment, I encouraged him to focus on the question and ignore the temporary leaderboard. I also helped him find the answer. By working at his own pace, he finally reached the peak. Therefore, my experience is that to ensure

students' participation in Wayground, it is very necessary that the teacher monitors the process and offers instant assistance when necessary.

On the teacher's end, I can monitor both accuracy and progress throughout the game. When students finish, especially those who always complete tasks more quickly, they can review all the questions at their own pace. In this way, they can customize their learning rhythm without having to wait until the end of the session. In addition to this, in my class, when all the students have finished, I also review all the questions together with them to ensure their understanding of these questions. The advantage of Wayground is that by looking at the accuracy of each question, the teacher can wisely select which questions to review. This helps to maintain students' attention in the general feedback stage.

References

Hu, Y. (2023). Kahoot! in the classroom: Examining the impact of a game-based student response system on pre-service teachers' academic achievement and perceptions. *Innovations in Education and Teaching International*, 61(5), 960-971. <https://doi.org/10.1080/14703297.2023.2250757>

Kapp, K. M. (2012). *The gamification of learning and instruction: Game-based methods and strategies for training and education*. Pfeiffer.

Liu, Y.-J., Zhou, Y.-G., Li, Q.-L., & Ye, X.-D. (2022). Impact study of the learning effects and motivation of competitive modes in gamified learning. *Sustainability*, 14(11), 6626. <https://doi.org/10.3390/su14116626>

Smirani, L., & Yamani, H. (2024). Analysing the impact of gamification techniques on enhancing learner engagement, motivation, and knowledge retention: A structural equation modelling approach. *Electronic Journal of E-Learning*, 22(9), 111-124. <https://doi.org/10.34190/ejel.22.9.3563>

Szabó, F., & Kopinska, M. (2023). Gamification in foreign language teaching: A conceptual introduction. *Hungarian Educational Research Journal*, 13(3), 418-428. <https://doi.org/10.1556/063.2023.00202>

Conclusion

Wayground is an engaging game-based learning platform and supports visual and gamified learning. In my review classes, it is helpful in alleviating the dilemma of repeating the grammar and vocabulary exercises. When I informally asked students about their thoughts on using it, students expressed that they found it fun and enjoyable, and were excited having an alternative gaming platform to Kahoot. More importantly, Wayground places more emphasis on accuracy, and students tend to focus more on finding the answers rather than on speed during the game. However, the teacher should pay more attention to weaker students' participation and provide instant assistance.

