Glass-Box Feedback: Turning AI Chatbots into Metacognitive Writing Partners for Year 5 Girls

Makiko Ryland

Kincoppal-Rose Bay School of the Sacred Heart, Sydney, Australia

Abstract

Many students approach writing as a task to complete rather than a process of drafting, reflecting and refining ideas. This action research project investigated how AI-generated feedback influenced Year 5 girls' metacognitive engagement and revision practices in writing within a primary school context in Sydney, Australia. With generative AI increasingly present in classrooms, there is a need to understand whether AI feedback can support revision without replacing students' thinking. The 10-week intervention involved designing and implementing a custom AI chatbot, Blue Bot. Blue Bot was co-constructed with students through shared success criteria, task-specific rubrics, and clear guardrails to shift AI from a "black box" to a "glass-box" tool. Students engaged in repeated drafting cycles: writing an initial draft, receiving rubric-aligned AI feedback, revising independently, and submitting a second draft for teacher assessment. A mixed-methods approach was used to collect data through student journals, pre- and post-intervention surveys, focus group interviews, chatbot interaction logs, teacher field notes, and rubric-scored writing samples with calculated revision gains. Findings indicate that, when explicitly scaffolded, AI feedback can strengthen evaluative judgement and support deeper revision beyond surface editing. However, the impact of AI feedback varied depending on students' perceptions of the chatbot, and some learners (including EALD and lower-achieving writers) required additional scaffolding to interpret and apply feedback. The study highlights the importance of transparent design, explicit teaching and "human in the loop" principles to ensure AI supports metacognitive growth and equitable access to revision improvement.

Glossary

Black Box AI: An AI system whose internal processes are not visible or understandable to users, making its outputs appear opaque or unexplained.

Chatbot: A conversational artificial intelligence tool designed to simulate dialogue and provide responses or feedback to users.

EALD (English as an Additional Language or Dialect): Students who speak a language or dialect other than Standard Australian English at home and may require additional language support.

Generative AI: Artificial intelligence systems capable of producing original content such as text, feedback, or responses based on user input.

Glass-Box AI: An AI system whose processes, purpose, and feedback mechanisms are transparent and understandable, allowing users to critically interpret and evaluate its outputs.

Metacognition: The ability to reflect on, monitor, and regulate one's own thinking and learning processes.

Metacognitive Engagement: Active reflection on learning processes, including planning, monitoring, evaluating, and revising work.

Glass-Box Feedback: Turning AI Chatbots into Metacognitive Writing Partners for Year 5 Girls

In many classrooms, students approach writing as a task to complete rather than a process of drafting, reflecting and refining ideas. Furthermore, students often struggle to revise their writing independently. Revision is not simply the correction of surface errors: effective revision requires metacognitive engagement, which involves students planning, monitoring, and evaluating their thinking. Metacognition helps students understand and use feedback to improve their work. When students engage metacognitively with feedback, they move beyond passive correction and become active participants in their learning. However, students often lack opportunities to engage deeply with feedback and revision.

The recent emergence of generative artificial intelligence (AI) tools has created new opportunities to support students' learning. AI chatbots can provide immediate, personalised feedback, creating opportunities for students to engage in iterative revision. Unlike traditional teacher feedback, which is often delayed, AI feedback allows students to revise their work in real time. This immediacy has the potential to support students in reflecting on feedback and actively improving their writing. To this end, this study examined the question: *How does AI-generated feedback influence Year 5 girls' metacognitive engagement and revision practices in writing?*

Action research was an appropriate methodology for this project as it enables teachers to investigate and improve their own practice. Following Mertler's (2020) action research model, this study involved implementing a targeted intervention and examining its impact on student learning. This cyclical process allowed for reflection on how AI-generated feedback influenced students' engagement with revision. By embedding this intervention within authentic classroom practice, the study aimed to generate practical insights that could inform teaching and learning.

This research contributes to growing discussions about the role of AI in education. While AI tools are increasingly used in classrooms, less is known about their impact on metacognitive engagement and revision skills, particularly in primary school contexts. Understanding how AI-generated feedback influences students' revision behaviours may help educators use AI in ways that support learning rather than replace thinking. It is hoped that this study will provide practical insights into how AI can be used to support students in becoming more reflective, independent writers.

Literature Review

Writing is a complex and cognitively demanding process that plays a central role in students' learning across all areas of schooling (UNESCO, 2017). While writing outcomes remain a concern in Australian and international contexts (ACARA, 2021; Bañales et al., 2020; Thomas, 2020), improving writing is not solely a matter of correcting surface errors or raising test scores. Meaningful improvement depends on students' ability to engage metacognitively with their work; to plan, monitor, evaluate, and revise their ideas in response to feedback. Writing, therefore, is not merely a product, but a process of thinking and rethinking (Villalón et al., 2015). Developing students' capacity to regulate revision is critical. As digital tools increasingly shape educational environments, understanding how emerging technologies influence students' metacognitive engagement with revision has become an urgent area of inquiry.

Metacognition, commonly understood as thinking about one's own thinking, plays a central role in effective writing development (Mahdavi, 2014). In the context of writing, metacognitive engagement involves planning ideas, monitoring clarity and coherence, evaluating the effectiveness of language choices, and making deliberate revisions to strengthen meaning. Self-regulated writers actively interpret feedback, identify gaps between their current performance and desired standards, and decide strategically how to improve

(Nicol & Macfarlane-Dick, 2006). Revision is a complex evaluative process requiring judgement, reflection, and decision-making (López et al., 2021). Research distinguishes between reproductive conceptions of writing, where students focus on reproducing information with minimal revision, and epistemic conceptions, where writing is understood as a tool for refining ideas and deepening understanding (Villalón et al., 2015). Developing students' metacognitive capacity to engage in this deeper form of revision is essential if feedback is to translate into meaningful improvement.

Feedback is widely recognised as one of the most powerful influences on learning, yet its impact depends on how students interpret and use it (Hattie & Timperley, 2007). Effective feedback does more than identify errors; it clarifies standards, prompts reflection, and supports learners to close the gap between current and desired performance (Nicol & Macfarlane-Dick, 2006). Crucially, feedback only becomes formative when students actively engage with it, questioning suggestions, evaluating their relevance, and making purposeful revisions (Black & Wiliam, 1998). Without this metacognitive engagement, students may engage passively with feedback, limiting its potential to act as a catalyst for improvement.

In writing instruction, iterative drafting cycles combined with timely, specific feedback have been shown to strengthen students' revision practices (Black & Wiliam, 1998; López et al., 2021). However, the effectiveness of feedback is shaped by students' writing beliefs, confidence, and perceptions of whether feedback is supportive, rather than judgemental. These dynamics are particularly significant in girls' education, where affective factors such as perfectionism and feedback sensitivity may influence how revision opportunities are taken up (Madigan, 2019).

Within girls' education, engagement with feedback and revision is shaped not only by cognitive skills but also by affective factors. Research suggests that girls are more likely to hold epistemic conceptions of writing, viewing it as a tool for refining ideas and deepening

understanding (Villalón et al., 2015). These dynamics are particularly significant, as affective factors such as perfectionism and feedback sensitivity may influence how revision opportunities are taken up (Madigan, 2019). Emerging research also suggests that engagement with AI technologies may be gendered, with some studies reporting a predominantly male user base (Cann, 2025) highlighting the need to support girls to engage confidently and critically with AI.

When revision is perceived as exposing weakness rather than fostering growth, students may prioritise error avoidance over substantive improvement, giving feedback particular emotional weight. Brandmo and Gamlem (2025) note that students' perceptions of feedback influence whether it is taken up productively or experienced as threatening. In girls' classrooms, vague, overly critical, or public feedback may undermine confidence, whereas constructive, dialogic and autonomy-supportive feedback can strengthen motivation and ownership. These conditions may significantly shape girls' metacognitive engagement with revision.

Against this backdrop, the emergence of AI-generated feedback introduces both opportunities and risks for students' engagement with revision. While generative AI can provide immediate, individualised feedback that may feel less evaluative, its impact depends on how it is used. The concept of cognitive offloading or outsourcing of mental processes to external tools (Risko & Gilbert, 2016), is central to understanding this dynamic. Recent research distinguishes between beneficial offloading, which supports learning, and detrimental offloading, where learners bypass the cognitive effort required for deep understanding (Lodge & Loble, 2026). Evidence also suggests that unstructured AI use often trends toward this latter form, creating a "performance paradox" in which task completion improves while long-term learning is diminished (Lodge & Loble). This paradox is associated with reduced metacognitive engagement, as students may rely on AI outputs rather than

actively evaluating and revising their work. Consequently, the pedagogical value of AI-generated feedback lies not in its immediacy, but in how it is used to support active reflection and maintain learner agency.

The rapid development of artificial intelligence has introduced new possibilities for generating feedback at scale and in real time. Across studies, AI-generated feedback has been associated with increased revision activity, improved writing performance, and heightened engagement, when embedded within structured instructional designs. Chan et al. (2024) found that AI-generated comments led to measurable improvements in essay quality, although students perceived the feedback as less empathetic than teacher responses. Similarly, Meyer et al. (2024) report significant gains in text revision and positive learning emotions when AI feedback was paired with clear performance indicators, suggesting that structured scaffolding enhances its impact. However, emerging evidence also suggests that these benefits are not evenly distributed. Lower-performing students and English as an Additional Language or Dialect (EALD) learners may experience reduced motivational, emotional, and metacognitive benefits when AI feedback is linguistically complex, insufficiently scaffolded, or difficult to interpret independently (Shahini, 2025; Steinbach et al., 2025~~; Shahini, 2025~~).

At the primary level, Chang and Chow (2024) argue that AI is most effective when applied to surface-level features such as grammar and organisation, allowing teachers to focus on higher-order conceptual development. Dazzeo (2024) further demonstrates that involving students in co-constructing rubrics and critically reflecting on AI feedback can strengthen evaluative judgement and digital literacy. A recent meta-analysis by Deng et al. (2024) concludes that while generative AI can enhance revision activity and performance outcomes, its influence on deeper metacognitive regulation remains uneven and contingent on pedagogical framing and scaffolding.

Nonetheless, challenges persist. Current AI models struggle to capture the full nuance of human language, particularly creativity, cultural context, and emotional depth. Ethical concerns include bias, misinformation and student data privacy. Without addressing these issues, large-scale adoption risks diminishing learning quality, exacerbating inequalities, and raising integrity concerns (Deng et al., 2024). Historical parallels remind us that such anxieties are not new: Socrates, for example, feared writing would erode wisdom (Norman, n.d.), and contemporary debates warn of "cognitive debt" associated with AI use (Kosmyna et al., 2024). Mollick (2025) counters that AI's impact depends on how it is deployed—it can either dull or enhance thinking. Effective design should therefore prioritise nudging critical reflection, maintaining the "human in the loop" principle to ensure AI complements rather than replaces human judgment (Cipolla & Lenhart, 2024, p.7).

For girls, the stakes are particularly significant. AI feedback intersects with patterns of self-efficacy, feedback sensitivity, and perfectionism. It can support more sophisticated revision by building on their comparatively strong writing self-concept. AI chatbots also shift feedback from a one-way process to a more dialogic interaction, enabling students to question, clarify, and refine ideas through iterative revision (Xu et al., 2025). This process supports epistemic conceptions of writing as a tool for learning and reorganising thought (Villalón et al., 2015), rather than providing a unidirectional process of cognitive offloading. The judgement-free nature of AI aligns with girls' preferences for constructive input (Henderson et al., 2025), easing perfectionist pressures and supporting greater agency over the pace and direction of learning.

Despite this potential, empirical evidence in primary, all-girls' contexts remain limited. Most research to date has focused on older students or higher education, leaving limited understanding of how primary-aged girls interpret, question, and regulate AI-generated feedback during revision (Almohesh, 2024; Deng et al., 2024). Addressing this gap

is critical to understanding whether AI feedback can meaningfully support equity and engagement in early writing education. Within the context of girls' education, this issue takes on added significance: without careful implementation, generative AI risks compounding existing gendered inequities by further disenfranchising girls in the digital space. Girls' schools therefore play a vital role in equipping students with the confidence, digital literacy, and critical awareness needed to navigate these technologies, transforming a potential threat into an opportunity for empowerment.

In summary, writing development depends not only on instructional quality but on students' metacognitive engagement with feedback and revision. Meaningful improvement occurs when learners actively interpret, evaluate, and act on feedback rather than receiving it passively. In girls' education, these processes are shaped by affective factors such as perfectionism, feedback sensitivity, and self-efficacy, highlighting the importance of psychologically safe, dialogic feedback environments. Emerging evidence suggests that AI-generated feedback may support iterative revision and engagement when embedded within structured pedagogical approaches. However, its impact on deeper metacognitive regulation and revision ownership remains uneven and under-examined, particularly in primary, all-girls' contexts. This review, therefore, positions AI not as a replacement for teacher feedback, but as a tool with potential to support self-regulated writing, warranting further investigation into how Year 5 girls engage with, and regulate, AI-supported revision.

Research Context

Kincoppal-Rose Bay School of the Sacred Heart (KRB) is a Catholic independent school in Sydney, Australia, offering education from Early Learning to Year 12. Grounded in the Sacred Heart tradition, the school emphasises holistic education, with a strong focus on learner attributes including metacognition, critical thinking, and reflective practice.

This study was conducted with 21 Year 5 girls in my class. Participants were selected through convenience sampling, as they were my current class, providing a relevant context for exploring how AI-generated feedback influences writing revision. As their classroom teacher, I held an existing pedagogical relationship with participants, which supported ongoing observation and data collection.

Ethical approval was obtained through school-based processes, including consultation with the IT department and leadership team to secure permission for the use of PlayLab (n.d.), and the development of a formal parent consent process. Participants were assured that their involvement was voluntary and that all data would remain confidential. To protect anonymity, identifying details were removed in all reporting.

Research Action

The intervention in this action research project involved the development and integration of a customised AI chatbot, Blue Bot, into my Year 5 classroom. This chatbot was built using Playlab.ai (n.d.), a platform that enables teachers to design tailored AI tools for specific classroom contexts without requiring coding knowledge. The purpose of this intervention was to provide students with immediate, structured feedback to support metacognitive engagement and strengthen their ability to revise their writing independently.

Prior to implementation, students participated in introductory lessons on artificial intelligence, focusing on safe, ethical use and the role of feedback and revision in improving writing. To foster ownership, students voted to name the chatbot Blue Bot, reinforcing its role as a collaborative classroom tool. Students were also involved in its design by co-constructing success criteria, examining task-specific marking rubrics, and contributing to the guardrails that defined how the chatbot would respond. These rubrics, which varied depending on the writing task, were embedded into the chatbot's backend to ensure feedback aligned with classroom expectations. This transparent design supported a shift from perceiving AI as a

“black box” to a “glass box,” making feedback processes visible and understandable (Tabib & Alrabeei, 2024).

During writing lessons, students engaged in structured drafting and revision cycles. Each student completed a first draft in response to a short writing task, supported by a rubric tailored to that specific task. These tasks included narrative, persuasive, and descriptive writing activities, providing regular opportunities for feedback and revision. Students submitted their first draft to the teacher, who assessed it using the task-specific rubric to establish a baseline measure of performance.

Students then submitted their draft to Blue Bot, which provided feedback aligned with the same rubric. The chatbot followed a consistent workflow, first providing “two stars and a wish” feedback to identify strengths and an area for improvement. It then highlighted areas relating to spelling, grammar, and clarity. Blue Bot was configured with strict guardrails to prevent answer generation or direct correction. Instead, it prompted students to reflect on feedback and make revisions independently. Students engaged in an iterative feedback loop, revising their work and resubmitting where necessary to further refine their writing. Following revision, students submitted a second draft to the teacher, which was assessed using the same task-specific rubric. This enabled direct comparison between drafts and allowed for measurement of revision quality and improvement. This structured cycle of drafting, AI-generated feedback and revision positioned feedback as an active, reflective process rather than a passive one.

The integration of Blue Bot aimed to enhance students’ metacognitive engagement by supporting them to interpret feedback, monitor their progress, and independently implement revisions. By embedding structured, rubric-aligned AI feedback within classroom writing practice, the intervention encouraged students to take greater ownership of the revision process and develop greater awareness of how feedback could be used to improve their

writing. This approach aligns with action research principles by implementing and evaluating a targeted intervention within authentic classroom practice.

Data Collection

This study used a mixed-methods approach to investigate how a custom-built AI chatbot influenced the metacognitive engagement and revision practices of Year 5 girls. Following Mertler's (2020) action research framework, qualitative and quantitative data were collected to build a comprehensive understanding of how the tool supported drafting and improvement. Multiple data sources were used to polyangulate findings and strengthen credibility.

Student journals were used across the cycle to document how students engaged with feedback, revised their writing, and applied strategies prompted by the chatbot. Reflective prompts encouraged students to comment on how AI input shaped their editing decisions. Pre- and post- intervention surveys captured shifts in revision behaviours and writing practices. The pre-survey examined baseline understandings of revision, strategies for improving writing, and current feedback habits. The post-survey measured how these processes changed, particularly in relation to how the chatbot influenced students' ability to refine and improve their writing. Comparing the two sets of data highlighted changes in the girls' metacognitive awareness and engagement with revision.

Quantitative data came from writing samples across multiple tasks. Each piece was scored using a rubric for the first draft, the revised draft, and a calculated "revision score" (the difference between the first and second drafts), indicating improvement. Tracking these scores over time provided insight into how writing quality developed and how effectively students applied feedback. Changes between drafts also served as an observable indicator of students' metacognitive engagement, revealing the extent to which they reflected on feedback, evaluated their work, and made purposeful revisions to improve their writing.

Additional qualitative evidence came from PlayLab feedback logs and semi-structured group interviews. The logs recorded how students interacted with the chatbot and the types of revisions they attempted. Group interviews explored the chatbot's usefulness, how it supported writing improvement, and challenges with its use. These discussions were recorded, transcribed, and coded to identify patterns in writing decisions, revision strategies, and the role of AI. Teacher field notes further documented classroom interactions and observable changes in drafting practices.

Data Analysis

Data were analysed using Mertler's (2020) inductive sequence of organise, describe, and interpret. Survey responses, rubric scores, and work samples were organised into spreadsheets to track class and individual trends. Journal responses, interview transcripts, feedback logs, and field notes were collated for qualitative analysis.

The data were coded to identify recurring themes related to metacognitive engagement and revision practices. Coding supported the organisation of concepts such as revision depth, feedback interpretation, student decision-making, autonomy, and the transfer of feedback across drafts. Changes in rubric scores and written revisions were analysed alongside qualitative data to examine how students reflected on feedback, monitored their progress and made purposeful revisions. Interpretation involved linking these themes to observable patterns in students' revision behaviours and their evolving ability to engage reflectively with the writing process.

Polyangulation occurred by cross-referencing journal responses, survey results, writing samples, interview data, feedback logs, and teacher observations. This strengthened validity and ensured interpretations reflected students' metacognitive engagement and revision behaviours, rather than relying on a single data source. While some findings aligned with existing literature on feedback and revision, the analysis also revealed unexpected

patterns, including variation in students' critical engagement with AI feedback, differing levels of revision independence, mixed attitudes towards AI, and uneven accessibility, particularly for lower-ability and EALD students.

Discussion of Findings

This section synthesises the key themes identified through analysis of data collected across the action research cycle. Presented as thematic statements, the findings respond directly to the research question: How does AI-generated feedback influence Year 5 girls' metacognitive engagement and revision practices in writing? My analysis indicated that AI-generated feedback can strengthen metacognitive engagement when students are supported to interpret and evaluate it critically. However, its effectiveness appeared to vary depending on students' perceptions of the chatbot and their ability to independently interpret feedback. These patterns are explored across the following themes.

AI-Generated Feedback Can Support Year 5 Girls' Metacognitive and Critical Engagement with Writing

When intentionally implemented, AI-generated feedback often functioned as a thinking partner rather than a shortcut, supporting students' evaluative judgement and metacognitive regulation. Evidence from student journals and focus groups suggests that involvement in shaping how the chatbot generated feedback positioned AI as a system open to scrutiny rather than unquestioned authority, encouraging critical engagement with suggestions (Tabib & Alrabeei, 2024).

Journal analysis provided clear evidence that students increasingly evaluated AI feedback rather than accepting it uncritically. Several students identified limitations in chatbot suggestions and exercised judgement in deciding whether revisions were appropriate. One student explained that while the chatbot identified useful corrections, it also suggested changes that were "not actually there," requiring her to determine whether the feedback was

valid. Another described rejecting suggestions that encouraged narrative features in an information report, explaining that she “put matters in my own hands with my opinion.” Chatbot interaction logs further reinforced this pattern of critical engagement. One student questioned the chatbot’s feedback structure, asking, “why do you say that, you usually just say the first paragraph,” while later commenting “I have to go because your wishes are too long,” demonstrating awareness of feedback quality and usability. These responses demonstrate emerging evaluative judgement, a key component of metacognitive regulation (Ojeda-Ramirez et al., 2023), as students actively monitored and evaluated feedback rather than applying it automatically.

Focus group discussions further revealed a shift from surface-level editing toward deeper conceptual revision. Students described moving beyond correcting punctuation and spelling to improving clarity, organisation, and meaning. One student explained that prior to using the chatbot, editing focused primarily on surface features, whereas AI feedback prompted her to reconsider whether her ideas “made sense” to the reader. Others described how the chatbot’s step-by-step feedback helped them understand not only what to change, but why changes were necessary. This suggests that AI feedback supported deeper engagement with the revision process by making underlying writing decisions more visible and discussable.

Survey data reinforced the qualitative findings. The number of students able to clearly articulate how they improved their writing increased from 9 to 16 out of 21 following the intervention. This increase suggests improved metacognitive awareness of revision strategies, rather than improvement in writing quality alone. Taken together, evidence from journals, focus groups, and surveys indicates that AI-generated feedback strengthened students’ metacognitive engagement when positioned as a tool for reflection and evaluation rather than passive correction.

Year 5 Girls' Perceptions of AI Shape Their Engagement, Responsibility, and Revision Behaviours

Students' perceptions of the chatbot played a critical role in determining how they engaged with its feedback. This is particularly significant in girls' education contexts, where gender has been shown to influence students' familiarity and confidence with AI tools (Ofosu-Ampong, 2023).

Analysis of journals, focus groups, and chatbot interaction logs revealed substantial variation in how students perceived the chatbot. Some students demonstrated critical distance, recognising its limitations. One student noted that "seeing as it is a robot, this tool isn't always that accurate," indicating awareness that AI feedback required evaluation. Others anthropomorphised the chatbot, describing it as "like having an online teacher," suggesting a more relational interpretation of the tool.

These perceptions directly influenced revision behaviour. Students who viewed the chatbot as a credible instructional partner engaged more deeply with feedback, reflected in higher revision scores between first and second drafts. In contrast, students who perceived the chatbot as "just a robot" were more likely to disengage, dismiss feedback, or seek quick answers without meaningful revision. Notably, one higher-performing student who repeatedly asked the chatbot to "just tell me the answers" demonstrated a lower revision score than expected, suggesting that perceptions of the tool may have influenced engagement, independent of ability. Chatbot interaction logs provided further evidence of this variation. Some students engaged in reflective dialogue, asking clarifying questions and refining their work iteratively.

Focus group discussions reinforced this pattern. Students who described the chatbot as supportive reported feeling more confident revising independently, whereas those who perceived it as impersonal reported lower engagement. Several students also articulated the

importance of respectful interaction, demonstrating growing awareness of responsible AI use. These findings suggest that anthropomorphism influenced not only how students perceived the chatbot, but how actively they engaged in metacognitive revision.

Taken together, these findings align with Hasan (2024) and Placani (2024), who argue that anthropomorphism shapes perceptions of trust and authority in AI systems. Within this classroom context, students' perceptions of AI influenced whether feedback functioned as a tool for metacognitive reflection or a shortcut. This highlights the importance of explicitly teaching students how to critically interpret AI feedback, ensuring that AI supports evaluative thinking rather than passive reliance.

AI Feedback is Not Equally Accessible for All Year 5 Girls and Requires Differentiated Scaffolding

Consistent with existing research showing uneven accessibility of AI feedback for some learners (Shahini, 2025; Steinbach et al., 2025), similar patterns emerged in this study. Analysis of journals, teacher observations, focus groups, and markbook data indicated that some students required additional scaffolding to interpret and apply feedback independently.

Student reflections provided clear evidence of comprehension barriers. One EALD student wrote, "I think some times BlueBot give me a long sentences and I can't really understand what that mean so I need to think about it and asked my teacher Miss Ryland" (Student 21, journal). Teacher observations reinforced this interpretation, as the student frequently paused during revision and asked for clarification or assistance interpreting feedback before making revisions. These behaviours indicate that cognitive load and linguistic complexity limited independent engagement with AI feedback. In response, a personalised chatbot was developed with simplified language, shorter responses, and a focus on a single revision goal. When other students inadvertently accessed this version, they reported greater clarity and usability. Focus group participants described the feedback as

"shorter" and easier to remember, suggesting that reduced complexity supported independent revision. This shows that modifying feedback structure improved accessibility and engagement.

Quantitative markbook data further supported findings. Students receiving C or D grades showed the lowest revision gains, ranging from 6.43% to 9.55%, compared to higher gains among more proficient writers. These patterns suggest that some students struggled to independently interpret and apply feedback, rather than lacking motivation to revise. Similar findings have been reported in prior research, which shows that learners with lower proficiency benefit most from explicit scaffolding when using AI feedback (Yeung, 2025; Zhu et al., 2025).

Student journals, teacher observations, and focus group responses showed that some learners struggled to interpret AI feedback independently, particularly when responses were linguistically complex or overly detailed. This was reflected in lower revision gains in markbook data among these students, indicating that AI feedback was not equally accessible to all learners. Without appropriate scaffolding, AI risks reinforcing existing disparities by benefiting students already equipped with strong metacognitive and linguistic skills. However, when feedback was simplified and scaffolded, students demonstrated greater independence and engagement in revision. These findings highlight that without deliberate differentiation and teacher mediation, AI feedback may privilege more proficient learners, while leaving support and EALD students reliant on additional human scaffolding to access its benefits.

Conclusion

This project investigated how AI-generated feedback influenced Year 5 girls' metacognitive engagement and revision practices in writing. The findings suggest that AI feedback can strengthen students' metacognitive awareness and revision skills when it is

intentionally designed and embedded within structured drafting and feedback cycles. Rather than functioning as a shortcut, the chatbot supported students in evaluating their writing, reflecting on feedback, and making purposeful revisions. Many students demonstrated increased evaluative judgement, improving clarity, organisation and meaning. They were also better able to articulate how and why their writing improved, indicating greater metacognitive engagement.

However, these benefits were not automatic. Students' perceptions of the chatbot influenced how they engaged with feedback, and some learners required additional scaffolding to interpret and apply suggestions independently. These findings highlight the importance of positioning AI as a tool for reflection rather than correction. Transparent "glass-box" design, explicit instruction and teacher mediation were critical in supporting meaningful engagement and equitable access.

Future refinements should focus on simplifying and differentiating feedback, strengthening students' AI literacy, and ensuring alignment between AI and teacher guidance. When implemented thoughtfully, AI-generated feedback can support students in becoming more reflective, independent writers with greater ownership over their revision process.

Reflection

Engaging in this action research process has been one of the most professionally rewarding experiences of my teaching career. One of the greatest highlights was witnessing my Year 5 girls embrace the project with such enthusiasm and openness. I have always valued reading their writing, but this process provided a deeper insight into how they think, reflect, and respond to feedback. Their honesty and thoughtfulness were striking. The care with which they evaluated chatbot feedback was mirrored in the way they offered feedback to one another and to me. This experience reinforced the importance of centring student voice and

has made me a more reflective and responsive practitioner. It reminded me that meaningful change in teaching begins with listening closely to the learners themselves.

This project also transformed my own relationship with artificial intelligence. While I once viewed tools like ChatGPT with uncritical optimism, this process helped me develop a more nuanced understanding. Much like my students' shift from viewing AI as a "black box" to a "glass box," I came to recognise that AI is not inherently transformative on its own. Its value lies in how it is used to support thinking, not replace it. Most importantly, this experience reaffirmed that human relationships, trust, and understanding remain at the heart of meaningful learning, particularly in girls' education.

I am deeply grateful to the ICGS and GARC community, especially my research advisor Leanne Horwitz, for her consistent support and insight. Working alongside Team Horwitz has been intellectually stimulating and deeply affirming. I would also like to thank my incredible colleagues, Jess Palmer and Bec Kinnane, whose willingness to listen, challenge my thinking, and offer support made this journey possible. My sincere thanks go to my school-based mentor, Trinity Hastie, whose brilliance and generosity have profoundly shaped my thinking. I am also grateful to Kincoppal-Rose Bay School of the Sacred Heart and my Principal, Erica Thomas, for providing me with this extraordinary opportunity and for supporting innovation in teaching and learning.

Most importantly, I extend my deepest thanks to the 21 girls of 5 Blue. Their enthusiasm and willingness to engage so thoughtfully with this project were the foundation of its success. I have learnt so much from them, and this experience has reaffirmed my belief in their capacity to think critically. I have no doubt they will go on to question, shape and redefine the world around them.

References

Almohesh, A. R. I. (2024). AI application (ChatGPT) and Saudi Arabian primary school students' autonomy in online classes: Exploring students and teachers' perceptions. *International Review of Research in Open and Distance Learning, 25*(3), 1–18. https://doi.org/10.19173/irrodl.v25i3.7641

Australian Curriculum, Assessment and Reporting Authority. (2021). *National assessment program: Literacy and numeracy*. https://nap.edu.au/docs/default-source/default-document-library/2021-naplan-national-report.pdf

Bañales, G., Ahumada, S., Graham, S., Puente, A., Guajardo, M., & Muñoz, I. (2020). Teaching writing in grades 4–6 in urban schools in Chile: A national survey. *Reading and Writing, 33*(10), 2661–2696. https://doi.org/10.1007/s11145-020-10055-z

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Brandmo, C., & Gamlem, S. M. (2025). Students' perceptions and outcome of teacher feedback: A systematic review. *Frontiers in Education, 10*, 1572950. https://doi.org/10.3389/feduc.2025.1572950

Cann, G. (2025, January 7). *Users of AI chatbot companions say their relationships are more than 'clickbait' — but views are mixed on their benefits*. ABC News. https://www.abc.net.au/news/2025-01-07/ai-chatbot-husbando-waifu-relationships romance-psychology/104701556

Chan, S., Lo, N., & Wong, A. (2024). Generative AI and essay writing: Impacts of automated feedback on revision performance and engagement. *rEFLections, 31*(3), 1249–1284. https://doi.org/10.61508/refl.v31i3.277514

Chang, Q., & Chow, Z. (2024). The potential and implications of AI-generated feedback for

primary school composition writing. *ASEAN Journal of Applied Languages, 3*(1), 17–43. https://ejournal.maal.org.my/asjal/article/view/17

Cipolla, B., & Lenhart, A. (2024). *Generative AI in K–12 education: Challenges and opportunities*. Common Sense Media. In C. Dare & J. Robb (Eds.), *Commonsensemedia* (pp. 1–63). https://www.commonsensemedia.org/sites/default/files/research/report/generative-ai-in-k-12-education-white-paper-updated-aug-2024-final-2.pdf

Common Sense Media. (2025, July 16). *Nearly 3 in 4 teens have used AI companions, new national survey finds*. https://www.commonsensemedia.org/press-releases/nearly-3-in-4-teens-have-used-ai-companions-new-national-survey-finds

Dazzeo, R. (2024). AI-enhanced writing self-assessment: Empowering student revision with AI tools. *Journal of Technology-Integrated Lessons and Teaching, 3*(2), 80–85. https://doi.org/10.13001/jtilt.v3i2.9119

Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2024). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education, 105*, 105224. https://doi.org/10.1016/j.compedu.2024.105224

Hasan, A. (2024, August 20). *Are you anthropomorphizing AI?* Blog of the APA. https://blog.apaonline.org/2024/08/20/are-you-anthropomorphizing-ai-2/

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/003465430298487

Henderson, M., Bearman, M., Chung, J., Fawns, T., Buckingham Shum, S., Matthews, K. E.,

- & de Mello Heredia, J. (2025). Comparing generative AI and teacher feedback: Student perceptions of usefulness and trustworthiness. *Assessment & Evaluation in Higher Education, 1–16.* https://doi.org/10.1080/02602938.2025.2502582

Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X.-H., Beresnitzky, A. V., Braunstein, I., & Maes, P. (2024). *Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for an essay writing task.* MIT Media Lab. https://www.brainonllm.com/

Lodge, J. M., & Loble, L. (2026). *Artificial intelligence, cognitive offloading and implications for education*. University of Technology Sydney. https://doi.org/10.71741/4pyxmbnjaq.31302475

López, P., Torrance, M., Rijlaarsdam, G., & Fidalgo, R. (2021). Evaluating effects of different forms of revision instruction in upper-primary students. *Reading and Writing, 34*(7), 1741–1767. https://doi.org/10.1007/s11145-020-10025-5

Madigan, D. J. (2019). A meta-analysis of perfectionism and academic achievement. *Educational Psychology Review, 31*(4), 967–989. https://doi.org/10.1007/s10648-019-09484-2

Mahdavi, M. (2014). An overview: Metacognition in education. *International Journal of Multidisciplinary and current research*, *2*(3), 529-535.

Mertler, C. (2020). *Action research: Improving schools and empowering educators* (6th ed.). SAGE Publications.

Meyer, J., Jansen, T., Schiller, R., Liebenow, L. W., Steinbach, M., Horbach, A., & Fleckenstein, J. (2024). Using LLMs to bring evidence-based feedback into the classroom: AI-generated feedback increases secondary students' text revision, motivation, and positive emotions. *Computers and Education: Artificial Intelligence, 6,* 100199. https://doi.org/10.1016/j.caeai.2023.100199

Mollick, E. (2025, July 8). Against brain damage. *One Useful Thing.*

https://www.oneusefulthing.org/p/against-brain-damage

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education, 31*(2), 199–218. https://doi.org/10.1080/03075070600572090

Norman, J. M. (n.d.). *Socrates on the invention of writing and the relationship of writing to memory*. History of Information. https://www.historyofinformation.com/detail.php?id=3439

Ofosu-Ampong, K. (2023). Gender differences in perception of artificial intelligence-based tools. *Journal of Digital Art & Humanities, 4*(2). https://doi.org/10.33847/2712-8149.4.2_6

Ojeda-Ramirez, S., Rismanchian, S., & Doroudi, S. (2023). Learning about AI to learn about learning: Artificial intelligence as a tool for metacognitive reflection. *International Journal of Artificial Intelligence in Education*. https://doi.org/10.35542/osf.io/64ekv

Placani, A. (2024). Anthropomorphism in AI: Hype and fallacy. *AI and Ethics, 4*, 691–698. https://doi.org/10.1007/s43681-024-00419-4

Playlab. (n.d.). *Where educators do their best work with AI.* Playlab AI. https://www.playlab.ai

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in cognitive sciences, 20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002

Shahini, A. (2025). Emotional dimensions of feedback: How AI and human responses shape ESL learning outcomes. *Ampersand, 100235*. https://doi.org/10.1016/j.amper.2025.100235

Steinbach, M., Fleckenstein, J., Kuklick, L., & Meyer, J. (2025). (De)motivating zero performing students with negative feedback: Does the salience of performance information matter? *Journal of Computer Assisted Learning, 41*(4), e70070. https://doi.org/10.1111/jcal.70070

Tabib, F. M., & Alrabeei, M. M. (2024). Can guided ChatGPT use enhance students' cognitive and metacognitive skills? In R. Huang, J. Chen, & N.-S. Chen (Eds.), *Reimagining education: Innovations in teaching and learning with generative AI* (pp. 153–169). Springer. https://doi.org/10.1007/978-3-031-52280-2_10

Thomas, D. P. (2020). Rapid decline and gender disparities in the NAPLAN writing data. *The Australian Educational Researcher*, *47*(5), 777-796.

UNESCO. (2017). *Reading the past, writing the future: Fifty years of promoting literacy*. https://unesdoc.unesco.org/ark:/48223/pf0000247563

Villalón, R., Mateos, M., & Cuevas, I. (2015). High school boys' and girls' writing conceptions and writing self-efficacy beliefs: What is their role in writing performance? *Educational Psychology*, *35*(6), 653–674. https://doi.org/10.1080/01443410.2013.836157

Xu, X., Wang, H., & Li, Y. (2025). Enhancing self-regulated learning and learning experience in generative AI environments: The critical role of metacognitive support. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.13599

Yeung, S. (2025). University students' engagement with generative AI-supported automated writing evaluation (AWE) feedback. *Journal of Second Language Writing, 68*, 101203. https://doi.org/10.1016/j.jslw.2025.101203

Zhu, Q., Wang, Z., & Koedinger, K. R. (2025). How students use AI feedback matters: Differential effects of optional and enforced AI assistance. *arXiv*. https://arxiv.org/abs/2505.08672