

Practice Without Pressure: Using a Teacher-Trained AI Agent to Build Exam Writing Skills in a Year 12 Girls' VCE Visual Arts Classroom

Perri Winter

Presbyterian Ladies' College, Melbourne, Australia

Abstract

Over many years of teaching senior art, I noticed a persistent gap: students demonstrating genuine artistic sophistication in studio work were frequently unable to reproduce the same depth of thinking in written exam responses. This action research study investigated how a teacher-trained AI feedback agent, “Agent Art,” shaped ten Year 12 girls' independent exam preparation and analytical writing at Presbyterian Ladies' College (PLC) in Melbourne. I designed Agent Art as an exam practice partner trained on subject-specific content, rubrics, and Victorian Curriculum and Assessment Authority (VCAA) assessment criteria language. Using a mixed-methods, qualitative-led approach encompassing AI conversation logs, pre- and post-intervention surveys, a focus group, individual conversations, practice exam artefacts, and a reflexive researcher journal, I applied triangulation and analysed data inductively through open, axial, and selective coding. I identified four themes as answers to the research question: girls engaged with Agent Art because they trusted the teacher behind it; they used it as a low-stakes space for checking and rehearsing rather than for high-stakes judgement; reading the conversation logs reshaped my teaching and in turn how girls prepared; and when the tool produced generic or inaccurate output, girls disengaged unless they had been taught to evaluate feedback critically. The findings suggest that a teacher-trained AI agent can meaningfully support girls' analytical writing when implemented with visible teacher oversight, relational trust, and explicit teaching of critical feedback literacy.

Glossary

Polyangulation: A qualitative enrichment strategy in action research in which data from multiple sources are cross-referenced to deepen understanding, surface contradictions, and strengthen the trustworthiness of interpretations, rather than to prove results statistically (Mertler, 2024).

Member Checking: A trustworthiness strategy in which the researcher shares preliminary interpretations with participants to verify that their perspectives have been accurately represented. In this study, member checking involved returning emerging themes to students in individual conversations and in the focus group.

Critical Feedback Literacy: The capacity to evaluate, interrogate, and selectively act on feedback received, including feedback generated by automated or AI systems. Critical feedback literacy involves recognising both the value and the limitations of a feedback source and making informed decisions about which guidance to trust and apply.

Practice Without Pressure: Using a Teacher-Trained AI Agent to Build Exam Writing Skills in a Year 12 Girls' VCE Visual Arts Classroom

There is a particular frustration many visual arts teachers will recognise. A student sits down at her folio and talks about her work with fluency, insight, and genuine feeling. Then she picks up a pen in an exam and writes something generic, rote learned, or half the length it needs to be. Over many years of teaching, I have watched this gap between girls' artistic intelligence and their written exam performance repeat itself, not because they lacked knowledge, but because they lacked the practice, confidence, and space to rehearse the specific analytical writing the exam demands.

The above challenge has been compounded by the social dynamics of a girls' classroom. Many of my students were reluctant to ask what they perceived as a silly question in front of their peers. This pattern is well documented: adolescent girls frequently exhibit perfectionism, a reluctance to take intellectual risks, and a heightened awareness of social judgement when seeking help (Booth & Nolen, 2012; Matthews, 2020). The girls were more comfortable seeking reassurance quietly, at the end of a lesson, when my time was already spoken for. The result was that feedback, one of the most powerful levers for learning (Education Endowment Foundation, 2021), reached them too infrequently and too late.

Generative artificial intelligence (GenAI) offers a compelling response to the disconnect between girls' artistic thinking and writing. AI tools can deliver feedback that is instant, scalable, and available outside school hours. Yet, the literature is equally clear about the risks: AI that answers for students, rather than scaffolding their thinking, can foster what Fan et al. (2025) term "metacognitive laziness," a reduction in the effortful cognitive work that produces durable learning. The question, therefore, is not whether AI should be used for feedback, but how it should be designed and implemented to genuinely serve girls as learners.

My intervention, designed and built in-house, was Agent Art: a teacher-trained AI feedback agent built on Google's Gemini large language model (LLM), and configured with VCE Art Making and Exhibiting content, VCAA rubrics, sample responses, and my pedagogical knowledge of the class. Agent Art functioned as both a first-responder during class when I was working with another student and as an independent study partner between lessons, generating practice exam questions and providing structured feedback. The agent was positioned not as an answer-generator but as a scaffold for analytical thinking.

The research question guiding this study was: *How does a teacher-trained AI feedback agent shape girls' independent exam preparation and analytical writing in Year 12 VCE Art Making and Exhibiting?*

Literature Review

The case for AI-assisted feedback rests on established evidence about what makes feedback effective. The Education Endowment Foundation (EEF, 2021) defines high-quality feedback as timely, specific, and actionable, and emphasises that students must be given opportunity to act on it. As the EEF report states, “feedback should focus on moving learning forward, targeting the specific learning gap identified by the teacher, and ensuring that a pupil improves” (p. 21).

The persistent challenge in providing feedback, however, is time. Delivering feedback at the frequency and specificity every learner needs is resource-intensive, particularly in the final year of secondary school. This challenge is especially acute in girls' classrooms. Matthews (2020) documents that girls place particular weight on their relationships with teachers and often seek teacher approval as part of how they learn, which concentrates feedback demand on the teacher rather than distributing it across self-assessment or peer review.

GenAI tools address the teacher-time constraint by making feedback abundant, instant, and repeatable (Lindsay et al., 2025). In theory, a girl can attempt a practice question, receive targeted feedback, revise, and attempt again, a cycle prohibitively time-intensive if dependent solely on teacher availability. This feedback-abundance argument is complicated, however, by evidence about what students do with AI-generated responses. Fan et al. (2025) demonstrate that students using GenAI can produce higher-quality written work while simultaneously showing poorer knowledge retention and fewer self-correcting behaviours, a pattern they describe as metacognitive laziness, warning that “AI technologies such as ChatGPT may promote learners’ dependence on technology and potentially trigger metacognitive laziness” (p. 489). Darvishi et al. (2024) reinforce this concern, arguing that the design and implementation of an AI tool determine whether it enhances or erodes student agency. This matters particularly for girls. The research literature on girls’ education consistently emphasises the importance of voice, agency, and authentic authorship (Matthews, 2020); an AI tool that substitutes for student thinking risks undermining the very capacities single-sex education is designed to build. Therefore, while AI systems that scaffold reflection and prompt critical thinking can strengthen self-regulation, those that supply answers risk producing dependency.

Self-regulated learning is central here. Li et al. (2023) demonstrate that adaptive scaffolding, responsive to the individual learner's real-time actions, is more effective than fixed scaffolding in promoting the planning and monitoring behaviours that characterise strategic learners. A tool calibrated to a specific student context is more likely to foster genuine self-regulation than a generic one. Laak and Aru (2025) extend this argument, calling for AI tools that develop agency, critical thinking, and enduring learning skills rather than narrow performance metrics.

Trust is equally important in the provision of feedback, and it is particularly salient in a girls' school. Research on automated feedback shows that students' willingness to engage with AI feedback is strongly influenced by their trust in its source (Khosravi et al., 2025). This trust is magnified for girls, who place significant value on the teacher-student relationship. Matthews (2020) notes that “our culture works hard to rob girls and women of our sense of agency” (para. 5), which makes the teacher’s role in validating and making visible each girl’s thinking a central part of how they learn. Furthermore, Lindsay et al. (2025) found that students perceive co-produced feedback, where a human teacher is involved in the development of an AI tool, as more genuine and reliable, and engage with it more deeply. This co-production logic underpinned the design rationale for Agent Art: by embedding my pedagogical knowledge, vocabulary, and assessment criteria into the tool's training, I aimed to create a system girls would experience as an extension of my teaching rather than an impersonal external service.

Specific dynamics apply in a girls' education context that reach beyond trust. Booth and Nolen (2012) provide empirical evidence that girls’ risk-avoiding behaviours are not fixed traits but environmentally produced. In a controlled experiment across single-sex and coeducational schools, they found that “girls in an all-girls group or attending a single-sex school were more likely than their coed counterparts to choose a real-stakes gamble” (p. F56), suggesting that observed gender differences reflect social learning rather than inherent disposition. This is encouraging. If the reluctance I have observed is environmentally shaped, a private, low-stakes space where risk-taking carries no social cost might shift behaviour. The literature thus produces a clear design rationale: an AI agent is most likely to support independent study and analytical writing when it is teacher-trained to build trust, designed to scaffold thinking rather than supply answers, and implemented with enough teacher visibility that girls feel accountable and supported.

Research Context

This study was conducted at Presbyterian Ladies' College (PLC), Melbourne, Australia, a K to 12 independent single-sex girls' school. The participating class comprised ten Year 12 students enrolled in VCE Art Making and Exhibiting, a subject in which students develop a visual arts journal, written assessments, and finished artworks, and sit a one-and-a-half-hour written examination. The examination requires sustained analytical writing under timed conditions, which was the area of student performance that motivated this study.

The discipline-specific demands of VCE Art Making and Exhibiting exam writing shaped the intervention's design. The exam requires students to write analytically about artworks they have not previously studied, to apply curatorial language with precision, and to construct arguments within tight time constraints. These demands are genre-specific: the skills required are a specialised form of art-critical discourse that most students can articulate verbally before they can produce it reliably in writing. A feedback tool trained to recognise and reinforce this register, and to redirect students when responses drift into description, is better positioned to scaffold improvement than a generic writing assistant.

The intervention ran from 1 September 2025 through the VCAA examination in November 2025, with a post-intervention survey administered in late December 2025. Baseline data were established in July 2025 using the 2024 VCAA practice examination. The class met for regular timetabled lessons throughout Terms 3 and 4 (July to November).

As classroom teacher and researcher, I occupied a dual role. I maintained a reflexive journal to document decisions, note changes in student behaviour, and interrogate my assumptions. I was conscious that my presence as researcher introduced the possibility that the girls would respond in ways they believed would please me, a risk heightened in a girls' school setting where the literature confirms students' strong inclination to please their teachers (Matthews, 2020). I addressed this by framing instruments to invite honest critique,

structuring the focus group to include direct challenge questions, and consulting my research advisor throughout.

Informed consent was obtained from parents or guardians, and student assent from each participant. Students were advised participation was voluntary and would have no bearing on academic assessment. All student names have been replaced with Pokémon character pseudonyms, and all conversation log data were de-identified prior to analysis. Data are stored securely on school-encrypted servers. The study was assessed as low-risk.

The Action

Agent Art was designed to address two interconnected problems: the scarcity of timely, individualised feedback in exam preparation, and the reluctance of girls in this setting to seek help in ways that exposed their uncertainty to others. The agent's core purpose was to serve as a private, practice-focused partner available during class when I was occupied with another student and also be available between lessons when girls chose to practise independently.

My original plan had been to build ten individual agents, one for each student, personalised to her artform and artists of influence. In practice, the complexity of maintaining ten distinct tools alongside full teaching responsibilities made that approach unworkable. Instead, I built a small suite of targeted agents serving the whole class: a vocabulary agent focused on VCE Art Making and Exhibiting terminology; several practice exam agents trained on specific question types, marking language, and sample responses; and a shared Google NotebookLM loaded with the study design, VCAA marking guides, sample responses from the 2024 exam, content on the artists studied, and approaches to curatorial, conservation, and art-critical writing.

Student onboarding occurred in early September 2025 and began with something more personal than a technical demonstration. Before introducing the agent, I gave each student a

physical Pokémon character card (see Appendix A). I chose each Pokémon deliberately, based on two years of knowing each girl, and rewrote each card description slightly so that, while it read as a description of the Pokémon, the student would recognise herself in it. I also made sticker sets for each student so they could use their character across their materials. The purpose was not decorative. I wanted each student to know, before she encountered the AI tool, that her pseudonym had been chosen with care and attention, and that she was known. This choice was grounded in what the literature tells us about girls: relationships are central to how they engage with their teachers and their learning, and feeling seen by the teacher is part of feeling safe enough to take academic risks (Matthews, 2020).

In the class introduction, I demonstrated Agent Art's function and explained its limitations: the agent could generate practice questions and give feedback, but girls should treat its feedback as a first response to consider carefully, not as a final verdict. I modelled a full practice exchange, submitting a response, reading the feedback, and deciding which elements to act on. Students were provided with a written guide to accessing the agent independently and were encouraged to use it in study periods and at home.

During lessons, Agent Art served as a first-responder. When students were working on written tasks and I was conferencing with another, they were encouraged to trial their response with the agent rather than waiting. Between lessons, the agent generated fresh practice exam questions, provided feedback, and offered vocabulary support. The agent was deliberately designed using a Socratic approach: it would never state a correct answer directly but would prompt students to consider what the question required, what criteria applied, and how their response could be refined.

Teacher oversight was maintained throughout. I reviewed conversation logs regularly, both to monitor whether the agent's feedback was accurate and to gain pedagogical intelligence about individual students. Where I identified problems, I adjusted the training

materials. This “human-in-the-loop” mechanism was not merely a safeguard; it became one of the most informative data sources in the study.

Data Collection

Data collection followed a mixed-methods, qualitative-led approach, consistent with action research methodology (Mertler, 2024). In line with Mertler's (2024) principle of polyangulation, the deliberate cross-referencing of diverse data sources to deepen understanding and surface contradictions, I collected data across six categories.

Written artefacts formed the foundation of the dataset. These included students' responses to the 2024 VCAA practice examination in July 2025, which were used as a baseline measure of analytical writing; weekly practice question responses completed during Term 3; a teacher-and-AI-assessed PLC practice examination after students had finished working with the agent and immediately prior to their external exam; and students' final VCAA examination scripts. These artefacts allowed for descriptive comparison of pre- and post-intervention analytical quality, attending to shifts in writing register, use of subject vocabulary, and the structure of argument.

Anonymised AI conversation logs documented every interaction between students and Agent Art across the intervention period. They provided process evidence that surveys and interviews could not: the questions students asked, how they responded to feedback, whether they revised their answers, and how their approach evolved. The logs also gave me insight into each individual student's thinking that would not otherwise have been visible.

Pre- and post-intervention surveys gathered students' self-reported confidence levels, attitudes towards AI feedback, and perceptions of the agent's usefulness. The post-intervention survey, administered in late December 2025, included open-response questions that explicitly invited critique. A focus group was conducted after the agent intervention concluded, and I held brief individual conversations with students throughout as a form of

member checking. My reflexive journal, maintained throughout, recorded observation notes, decisions, doubts, and moments of surprise. Discussions with my research advisor provided critical distance and an external perspective.

Data Analysis

Data analysis followed Mertler's (2024) three-stage inductive process of organising, describing, and interpreting. I sorted all data sources by time period, student activity type, and source type, with conversation logs organised by pseudonym and interaction type. Open coding generated initial categories including trust language, first-responder use, low-stakes checking, iterative revision, and non-use explanations. Axial coding identified relationships between categories, notably the link between knowing the teacher was monitoring the tool and willingness to seek feedback independently. Selective coding produced the four themes presented below.

Conversation logs were used to test whether survey perceptions were reflected in observable behaviour, a central polyangulation move that helped me avoid over-relying on self-report. Disconfirming evidence was retained deliberately. Students who did not engage consistently with Agent Art, and their explanations for non-use, were treated as equally significant data and informed the conditions-for-success analysis in my study's conclusions. Discussions with my research advisor at each stage provided an external check on the plausibility of my interpretations.

Discussion of Findings

Analysis of the full dataset produced four themes that together answer the research question. Each is examined here in relation to the data and the literature, with disconfirming cases acknowledged.

A Teacher-Trained Agent Shaped Girls' Exam Preparation Because They Trusted the Teacher Behind It

Teacher Involvement Enhances Girls' Trust and Engagement in AI Use

The most consistent finding across every data source was that girls' willingness to use Agent Art depended less on the quality of the AI system itself and more on their trust in me as the teacher who had built and was monitoring it. One student in the focus group explained she was “probably more likely to use [it] ... because I know you’re checking whether we’re on track.” Another described the experience as “very comforting and relieving ... a safe space ... non-judgmental ... it’s always there to help you.” This is not language about a technical tool; it is language about a relationship. The agent was trusted because I was trusted. This aligns with what the literature tells us about girls as learners: Matthews (2020) documents that girls place particular weight on their relationships with teachers, and that their learning behaviours are shaped by the desire to be seen and supported.

The accountability dimension was equally important. Students knew I read the logs, and this shaped how they used the tool. This is consistent with Lindsay et al.’s (2025) finding that students engage more deeply with AI feedback when they perceive a human teacher involved in its development and oversight; in my classroom, teacher involvement was explicit, repeated, and visible.

The trust finding connected directly to the social dynamics of a girls’ school. Several girls in the focus group described feeling guilty about emailing me for help. One said simply, “I feel bad for sending an email,” noting that “teachers are our only source of feedback.” Another, reflecting on a particularly tiring afternoon session, said “my brain is a bit pooped anyway ... I’m kind of glad that we did this instead.” Agent Art reduced the guilt of help-seeking by providing a first step that did not impose on my time. Relief is a meaningful emotional outcome in a Year 12 context, and it connects to what Matthews (2020) describes

as girls' social risk-aversion in help-seeking: the agent offered a private alternative that did not require the girls to expose uncertainty to peers or to take up teacher time they already felt protective of.

Trust was not universal from the outset. When asked whether they would trust the agent if it had not been built and monitored by me, several students indicated low confidence. Capsakid continued using an alternative study resource for part of the intervention, believing the agent was still in development. Clefairy's trajectory, however, illustrates a different pattern: she engaged hesitantly at first, then became one of the most active and purposeful users, asking highly specific questions in the weeks before the examination. Her post-survey comment that the agent "helped me easily find specific information about the conservation of Atem's specific medium" captures the moment her use shifted from exploratory to purposeful. Her journey from uncertainty to advocacy was clear evidence that initial reluctance does not preclude eventual deep engagement. Taken together, this theme answers the research question directly: the agent shaped girls' exam preparation because the teacher-trained design transformed AI feedback into a trusted extension of the teacher-student relationship.

Agent Art Shaped Girls' Writing Rehearsal as a Low-Stakes Space for Checking, Not for High-Stakes Assessment

Girls' Confidence in AI Decreases as the Stakes Get Higher

A central finding was that girls made a clear and consistent distinction between what they used Agent Art for and what they wanted a teacher for. For low-stakes purposes, confirming understanding, checking vocabulary, trying out a topic sentence, they were comfortable using the agent independently. For high-stakes tasks, in particular teacher-marked practice examinations, they wanted me. This distinction was observable in the logs, consistent across survey responses, and confirmed in the focus group.

The conversation logs showed that the most common starting intent was test-style or question-based prompting: students initiating exchanges with phrases such as “test me,” “give me a question,” or “ask me about.” The next most common intents were definition-seeking and vocabulary confirmation, followed by feedback requests on drafted responses. What the logs did not show was girls using Agent Art to receive and simply accept a final verdict on examination-standard work. One student stated the position bluntly: “I feel like I would try it out just to see what it says ... but I wouldn’t like put my own trust into it.” Another summarised the shared view: “The feedback part of AI is not trustworthy for marking of exams.” This pattern is consistent with Matthews’s (2020) account of girls as learners who value the relational authority of teacher judgement and calibrate carefully where they place their trust in a source of feedback.

Bellossom described her post-intervention study approach as doing an entire paper, marking it with the agent, then redoing the paper and fixing all responses based on Agent Art’s feedback. This is retrieval practice in its fullest form: active generation, external feedback, correction, and repetition. Another student captured a related dynamic by describing the agent’s value as offering “an infinite amount of questions.” The word “infinite” points to something real: for the first time, practice felt genuinely unlimited, without the social cost of asking the teacher again. The girls who used the agent this way showed the most coherent development in analytical writing across the practice and final examination period. This is consistent with what Darvishi et al. (2024) describe as the conditions under which AI enhances rather than erodes agency: when the tool scaffolds thinking rather than completing it on the student’s behalf.

Not all students used the agent consistently. Lillipup was candid in her post-survey that she did not use it because she did not study in general. Glimmet articulated a considered position: she did not enjoy chat-based AI and retained information best through handwritten,

traditional study methods. Glimmet’s survey comment that the tool “gather[s] information from lots of different resources using relevant areas” acknowledged its usefulness, but her preferred mode of learning remained handwritten and deliberate. This is a legitimate learning preference, not a failure of engagement. In answer to the research question, Agent Art shaped analytical writing rehearsal most powerfully when girls used it as an unlimited, low-stakes practice partner, while the teacher remained the trusted assessor for high-stakes judgement.

Reading the Conversation Logs Changed My Teaching, Which in Turn Shaped Girls’ Exam Preparation

Girls’ AI Use Shapes Teacher Pedagogy

One of the most significant and least anticipated outcomes was what I learned as a teacher from reading the conversation logs. The logs did not simply tell me whether students were using the agent; they showed me how individual girls were thinking about the subject: which artists a student kept returning to, which questions she could not frame clearly, which aspects of the rubric she was misapplying. What I saw in the logs changed what I taught next, which in turn shaped how the girls prepared for the exam. In this sense the logs functioned as a feedback loop within the intervention itself, not merely a record of it.

Reading the logs changed my teaching in specific ways. When several students repeatedly asked the agent about the same conservation concepts, I prioritised those in the next lesson. When one student’s log showed a pattern of asking for sample answers rather than attempting her own response first, I addressed that directly. When I saw another student working through practice questions at ten o’clock at night, I understood something about her study habits and anxiety that I had not seen in class and adjusted my support accordingly. Survey feedback confirmed that this support was felt: Espeon wrote that the tool “answers my question directly and is very fast, saved my time of scrolling through all the information.” Another student described the accessibility of information through the agent as “very

comforting and relieving,” language consistent with what we know about girls’ preference for learning environments that feel safe and relational (Matthews, 2020).

Khosravi et al. (2025) argue that learning analytics can give teachers insights unavailable through conventional observation. In my case, the logs served exactly that function, not as surveillance but as formative information I used to adjust my teaching. The girls knew I read them; they also knew the purpose was to help me help them better. The answer to the research question here is indirect but important: Agent Art shaped girls’ exam preparation partly by shaping me. The teacher who emerged from reading the logs taught differently, and the girls prepared differently as a result.

Generic or Unreliable Feedback Disengaged Girls Unless They Had Been Taught to Evaluate It Critically

Girls’ Critical Feedback Literacy Affects Engagement With AI

The fourth theme addresses the tension beneath all the others: girls wanted the benefits of instant feedback, but they were acutely aware that generative AI systems can produce inaccurate or unhelpful responses. When Agent Art performed well, students used it willingly. When it produced generic output or failed to connect to VCAA-specific expectations, students disengaged quickly.

Students described the tool as “hit and miss” and cited “connection error” as a practical barrier. More substantively, several girls worried about being “led in the wrong direction,” a concern that was, in some instances, warranted. Espeon was direct in her survey: “there is one question I asked and it gave me a wrong answer.” Furthermore, Ninetales noted that the tool “sometimes ... gives illogical answers,” and Wishcash observed that “the answer is [a] bit too long,” a small but telling sign that generic output, even when technically accurate, can overwhelm rather than support. My reflexive journal records occasions during September and October 2025 when the agent’s feedback was too generic or slightly

misaligned with the current VCAA rubric emphasis. This connects to what we know about girls as learners: their tendency toward perfectionism and desire to get things right makes them particularly sensitive to imprecise feedback, and imprecision can trigger disengagement rather than correction (Matthews, 2020).

The girls who continued to use the tool productively despite unreliable output had developed a practice of cross-referencing: using the agent for an initial structure, then checking feedback against my teaching materials and VCAA sample answers before deciding what to act on. This is what the literature describes as critical feedback literacy, the capacity to evaluate feedback rather than simply trust or distrust it (Lindsay et al., 2025). Yu and Guo (2023) argue that because generative AI models are opaque, students cannot read the quality of feedback from its surface; in an examination context, this makes critical literacy essential. Clefairy gave a particularly measured account in her survey, suggesting the agent was useful for “auto-correcting our answers, suggesting answers if a student is stuck,” but adding that she would “probably not use it for creativity and ideas because I think these should be generated by the human mind.” This is critical feedback literacy in action: a student who understands precisely where AI adds value and where it does not. Several girls who encountered unreliable output disengaged entirely rather than interrogating it. In answer to the research question: Agent Art shaped girls’ exam preparation positively only when they had been equipped to evaluate its output. When they had not, unreliability caused disengagement. Critical feedback literacy must be explicitly taught.

Conclusions

This study investigated how a teacher-trained AI feedback agent shaped girls’ independent exam preparation and analytical writing in Year 12 VCE Art Making and Exhibiting. The evidence indicates that Agent Art shaped both study behaviours and writing quality in meaningful, conditional, and sometimes surprising ways. The clearest way to

characterise my findings is through highlighting a distinction between what the agent did and what the teacher did: Agent Art equipped the girls with practice, feedback, and differentiated support, while the teacher-student relationship remained the element that empowered them. Both mattered, but they were not interchangeable.

Girls engaged with the agent primarily because they trusted the teacher behind it, a finding consistent with what we know about girls as learners and their attachment to relational authority (Matthews, 2020). That trust rested on three conditions: they knew I had trained the tool on our specific content and rubrics; they knew I was reading the logs; and the tool's questions and feedback felt consistent with the exams they already knew. When any one of those conditions was absent, engagement dropped. A teacher-trained AI agent is not a standalone tool; it is an extension of the teacher-student relationship.

Girls used Agent Art most purposefully for low-stakes checking and rehearsing and continued to prefer the teacher for high-stakes assessment judgement. This distinction reflects how girls calibrate trust in sources of feedback (Matthews, 2020) and is pedagogically healthy. It suggests the agent found its appropriate role as a first layer of support rather than a replacement for teacher feedback.

Reading the conversation logs gave me pedagogical intelligence about individual girls that conventional classroom observation alone could not afford. The logs did not simply record how the girls used the agent; they reshaped how I taught in ways that fed back into how the girls prepared for the exam. This human-in-the-loop mechanism may be the most underappreciated element of a teacher-trained AI intervention.

When the tool failed or produced generic feedback, students who had developed the capacity to evaluate AI feedback critically continued to benefit; students who had not, disengaged. Critical feedback literacy, therefore, must be explicitly taught. It is the skill that

makes the difference between a student who uses an imperfect tool productively and one who is defeated by it.

Implications for Practice

For teachers considering a similar intervention in a girls' classroom, this first cycle suggests four practical commitments are worth making from the outset:

- Make the teacher visible in the tool. Train the agent on your own rubrics, vocabulary, and sample responses, and tell the girls you have done so; trust in the teacher behind the tool is what girls act on, not trust in the tool itself.
- Read the conversation logs as teaching data, not as surveillance. The logs surface what is difficult, what is avoided, and what is misunderstood, and they let the next lesson respond directly.
- Teach critical feedback literacy as a concrete exam skill before the tool is used at scale. Model cross-referencing against rubrics and sample answers so that unreliable output becomes a prompt for interrogation rather than a reason to disengage.
- Position the agent as a low-stakes rehearsal space, not a marker. Girls calibrate carefully where they place trust, and preserving the teacher as the high-stakes assessor protects the relational authority that gives the agent its credibility in the first place.

Several limitations shape how these findings should be read. The study was situated in a single cohort of ten students in an independent girls' school with strong device access and curriculum support. As teacher and researcher, I was not a disinterested observer; despite polyangulation, reflexive journaling, and critical-friend consultation, the risk of confirmation bias could not be fully eliminated. Furthermore, the study could not establish causal relationships between Agent Art use and assessment outcomes.

In the next cycle, I will communicate access and availability to all students more clearly from the outset; introduce a dedicated early session teaching critical feedback literacy

as a concrete exam skill; develop contingency protocols for when the tool is unavailable; and work more deliberately to engage students whose preferred study methods are not agent-based.

Reflection

I came into this project thinking of myself primarily as a curriculum designer. I left it thinking of myself as a researcher. That shift, from building a tool to reading the conversation logs and asking what they actually revealed about how each girl was thinking, has changed how I teach. The logs were the most unexpected gift of this study: they gave me a window into student reasoning that observation alone had never afforded, and they fed directly back into the lessons and the support I offered each girl in the lead-up to the examination. I suspect that habit, of treating the logs as teaching data, will continue to shape how I think about feedback, accountability, and the particular learning dynamics of the girls I work with.

The finding that surprised me most was not about the AI. It was about trust. I had designed Agent Art to be technically accurate and exam-relevant. What I had not fully anticipated was that none of that would matter much unless the girls trusted me behind it. Hearing a student describe the agent as a safe space was both moving and instructive. The affective dimension of learning is always present, even when the tool appears purely functional.

I am grateful to my research advisor, Karen Lewis for asking the difficult questions and for pushing me to sit with the evidence that did not fit my preferred interpretation. The difference between action research and advocacy is the willingness to report what you actually found, including the limitations and the failures. I have tried to honour that difference in this report.

To Glimmet, Combee, Capsakid, Bellossom, Clefairy, Lillipup, Dragonair, Wishcash, Espeon, and Ninetales: thank you for your honesty in the surveys, in the focus group, and in

the quiet conversations at the end of lessons when you told me what was actually working and what was not. Your voices are the most important data in this study. I hope the next version of Agent Art is better because of what you were willing to tell me.

Finally, thank you to ICGS and the GARC community for creating a structure in which a classroom teacher can conduct this kind of rigorous, practitioner-led inquiry, share it with peers, and believe that it matters beyond her own classroom.

References

- Booth, A. L., & Nolen, P. J. (2012). Gender differences in risk behaviour: Does nurture matter? *The Economic Journal*, 122(558), F56–F78. <https://doi.org/10.1111/j.1468-0297.2011.02480.x>
- Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D., & Siemens, G. (2024). Impact of AI assistance on student agency. *Computers and Education*, 210, 104967. <https://www.sciencedirect.com/science/article/pii/S0360131523002440>
- Education Endowment Foundation. (2021). *Teacher feedback to improve pupil learning*. Education Endowment Foundation. <https://educationendowmentfoundation.org.uk/education-evidence/guidance-reports/feedback>
- Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*, 56(2), 489–530. <https://doi.org/10.1111/bjet.13544>
- Khosravi, H., Shibani, A., Jovanovic, J., Pardos, Z. A., & Yan, L. (2025). Generative AI and learning analytics: Pushing boundaries, preserving principles. *Journal of Learning Analytics*, 12(1), 1–11. <https://doi.org/10.18608/jla.2025.8961>
- Laak, K. J., & Aru, J. (2025). *AI and personalized learning: Bridging the gap with modern educational goals* [Preprint]. <https://doi.org/10.48550/arXiv.2404.02798>
- Li, T., Fan, Y., Tan, Y., Wang, Y., Singh, S., Li, X., & Gašević, D. (2023). Analytics of self-regulated learning scaffolding: Effects on learning processes. *Frontiers in Psychology*, 14, 1206696. <https://doi.org/10.3389/fpsyg.2023.1206696>

Lindsay, E. D., Zhang, M., Johri, A., & Bjerva, J. (2025). *The responsible development of automated student feedback with generative AI* [Preprint]. arXiv.

<https://doi.org/10.48550/arXiv.2308.15334>

Matthews, D. (2020, January 14). Girls, confidence problems, and peer pressure. *Psychology Today*. [https://www.psychologytoday.com/au/blog/going-beyond-intelligence/](https://www.psychologytoday.com/au/blog/going-beyond-intelligence/202001/girls-confidence-problems-and-peer-pressure)

[202001/girls-confidence-problems-and-peer-pressure](https://www.psychologytoday.com/au/blog/going-beyond-intelligence/202001/girls-confidence-problems-and-peer-pressure)

Mertler, C. A. (2024). *Action research: Improving schools and empowering educators* (7th ed.). SAGE Publications.

Yu, H., & Guo, Y. (2023). Generative artificial intelligence empowers educational reform: Current status, issues, and prospects. *Frontiers in Education*, 8, 1183162.

<https://doi.org/10.3389/feduc.2023.1183162>

Appendix

Pokémon Character Cards and Sticker Sheets (Onboarding Materials, September 2025)

The following images show the personalised Pokémon character cards and sticker sheets prepared for each student prior to Agent Art onboarding in September 2025. Each card bears the student's Pokémon pseudonym, an illustration, a short power statement, and a description lightly adapted so that, while it read as a description of the Pokémon, the student would recognise something true about herself in it. The sticker sheets allowed each student to carry her character across her study materials. These onboarding materials supported the relational trust-building strategy described in The Action section, and are grounded in what the literature tells us about girls as learners: that relationships are central to how they engage with their teachers and their learning, and that feeling seen by the teacher is part of feeling safe enough to take academic risks (Matthews, 2020).

Bellossom

Power: Spreads joy and positivity wherever it goes.

Bellossom is known for its vibrant petals that dance in the sunlight. It brings joy, warmth, and positive energy to any group, turning even dull days into something bright.



Capsakid –

Power: Radiates bold confidence and energy.

Capsakid's fiery nature and lively spirit make it a bold presence. It thrives on challenges and has courage far beyond its size, inspiring others to be brave too.



Combee

Power: Works together to build and create sweetness.

Combee is a symbol of cooperation and hard work. Each member contributes to the hive, bringing sweetness and strength to the team's efforts.



Dragonair -

Power: Commands respect with quiet elegance.

Dragonair is a graceful and mysterious Pokémon, said to control the weather. It carries a sense of untapped greatness, inspiring those around it.



Espeon -

Power: Senses needs and reacts instantly.

vibrant petals that dance in the sunlight. It brings joy, warmth, and positive energy to any group, turning even dull days into something bright.



Glimmet -

Power: Endures challenges with quiet strength.

Glimmet thrives under pressure, drawing strength from its resilience. It works steadily without needing the spotlight, yet always delivers results.



Lillipup -

Power: Always loyal and ready to help.

Lillipup is brave and dependable, always looking out for its friends. Its loyalty builds trust, making it a treasured companion.



Ninetales -

Power: Inspires awe with wisdom and grace.

Ninetales is majestic and admired for its presence. Legends say it can live for a thousand years, carrying the wisdom of ages.



Whiscash -

Power: Calm until it unleashes its power..

Whiscash is peaceful and easygoing, yet when needed, it can call upon immense strength to protect and guide its surroundings.



Clefairy -

Power: Spreads wonder and light-hearted joy..

Clefairy is rare and adored, known for its magical nature. Its curiosity and charm make it a delight to those lucky enough to meet it.

