Measured, Not Heard: AI-Generated Feedback and Year 6 Girls' Self-Perception as Public Speakers

James Porter

Wimbledon High School GDST, London, United Kingdom

Abstract

This action research study explored how Year 6 girls interpret and respond to AI-generated feedback on public speaking, and what impact this had on their positive self-perception as speakers. Conducted over a thirteen-week autumn term in a London girls' junior school, the inquiry was embedded within an existing oracy curriculum and centred on three speaking tasks supported by Microsoft Speaker Progress. A mixed-methods design combined three-timepoint self-perception questionnaires, AI-generated metric reports and stored recordings, think-aloud protocols during rehearsal, focus groups conducted before, during, and after the intervention, teacher field notes, and reflexive journalling. Rather than producing a straightforward narrative of confidence gain or loss, the findings present a more complex and ethically significant picture. AI feedback did not operate as a discrete intervention acting uniformly upon pupils; instead, it was encountered within a wider feedback ecology in which it interacted with teacher feedback, peer responses, prior experience, emotional safety, and pupils' own developing self-judgements. Within this ecology, AI feedback sometimes supported performance, sometimes reduced emotional risk, and sometimes introduced uncertainty, meaning its influence on self-perception was partial, conditional, and relational. Importantly, the quantitative and qualitative strands did not converge neatly: questionnaire items linked to understanding, preparation, and knowing how to improve showed clearer movement than affective items relating to confidence and nervousness. Read as analytically productive, this divergence helps explain why clarity and competence did not reliably translate into confidence, and why shifts in self-perception were not always captured by quantitative measures alone.

Measured, Not Heard: AI-Generated Feedback and Year 6 Girls'

Self-Perception as Public Speakers

To speak confidently in public is not a politically or socially neutral skill. For girls and women, voice is bound with identity, agency, and the ability to be heard in societies where power is unevenly distributed (Ahmed, 2017; hooks, 1989). In the wake of #MeToo and the persistence of online and offline misogyny, strengthening girls' capacity to trust and project their voices has taken on renewed urgency (Banet-Weiser, 2018; Jane, 2017). In girls' schools, this urgency becomes a pedagogical concern: how we teach speaking, give feedback, and how pupils understand themselves as speakers has repercussions beyond the classroom.

The rapid arrival of artificial intelligence feedback systems into schools has only amplified this urgency. AI, often framed as objective and neutral, is largely developed within a heavily gendered industry whose assumptions and biases seep into the technologies it produces (Benjamin, 2019; Noble, 2018). This raises ethical questions about what it means to be evaluated and "heard" by such systems, particularly when feedback functions as recognition as well as instruction. If voice develops relationally, then the introduction of an algorithmic listener into a child's speaking life is not simply a technical innovation; it represents a rift in the classroom fabric of speaking and being heard.

When the International Coalition of Girls' Schools (ICGS) announced its research theme on navigating the AI frontier in girls' schools, I found myself reflecting on a persistent pattern in my own practice. Through leading "Speak Up! Year 6," a cross-sector public speaking programme, I noticed a recurring tension: some girls delivered fluent, well-received speeches yet continued to describe themselves as lacking confidence. In a small number of cases, girls began and ended the course rating their confidence as zero, despite positive feedback and objective success. This disjunction between competence and self-perception prompted a question: what was it about my feedback that meant some girls could perform well yet still not see themselves as confident speakers?

The ICGS theme led me to consider whether AI-generated feedback might be received differently from teacher or peer responses. Because such systems are often perceived as emotionally detached and criteria-driven, they may feel less socially entangled. It seemed possible that this detachment could reduce the interpersonal risk of teacher feedback. At the same time, introducing AI into such a relational domain raised ethical questions about credibility, trust, and recognition. This led to the research question: How does AI-generated feedback impact Year 6 girls' positive self-perception as public speakers?

This study, therefore, explored AI-generated feedback not as a stand-alone solution, but within a wider classroom ecology in which teacher feedback, peer response, prior experience, emotional safety, and pupils' developing self-judgements interact. The focus was not simply whether AI could improve performance, but how it shaped girls' relationship with evaluation and with themselves as speakers. It was, therefore, appropriate to approach this as an action research cycle, since the question emerged from a lived problem of practice and required inquiry embedded in real classroom conditions (Mertler, 2020).

Literature Review

The centrality of voice in education has long been recognised. Dewey (1916) framed communication and participation as essential to democracy, arguing that education is not merely about the transmission of knowledge but about the cultivation of citizens capable of dialogue.

Yet research and lived experience show that girls' confidence in this domain is precarious. Gilligan's *In a Different Voice* (1982) famously identified how girls, as they grow, can experience the silencing of authentic expression, learning to doubt what they know and fear that their words will not be heard. This "dissociative split" (Gilligan, p. xxi) between experience and articulation names the tension I observed earlier, helping to explain why girls who demonstrate competence in oracy may nevertheless describe themselves as lacking confidence. Crucially, voice, for Gilligan, is not simply sound, but relation: "to have a voice is to be human [...] but speaking depends on listening and being heard" (p. xvi).

Large-scale studies illustrate that competence and confidence in oracy often diverge for girls. Clark (2011), reporting on a survey of more than 6,800 primary and secondary pupils for the National Literacy Trust, found that despite school-aged girls valuing communication skills more highly, they reported lower confidence than boys, which Clark attributed to perfectionism and self-criticism. More recent studies reinforce this concern, showing that girls undervalue their ability and are less likely to volunteer to speak in public despite equivalent competence (De Paola et al., 2021; Girlguiding, 2024; Speakers Trust, 2023).

This lack of confidence does not vanish with age: Female university students report higher oral communication apprehension than male peers despite comparable performance, and young women's confidence in expressing political opinions remains lower than men's, with fear of saying the wrong thing a significant barrier (Best et al., 2023; Loureiro et al., 2020).

Taken together, these findings suggest that confidence is relationally mediated and not simply the outcome of skill. Strengthening girls' self-perception, therefore, requires more than technical instruction; it demands environments that validate, nurture, and listen to their voices.

Feedback is a key mechanism for shaping that validation, yet its impact is uneven. Hattie and Timperley (2007) stress that it works best when it is specific and process-focused, while Black and Wiliam (2009) emphasise its formative, iterative role in learning. However, these accounts can still frame feedback as a technical instrument, aimed at closing a gap between current and desired attainment. Winstone and Carless (2019) extend this by distinguishing feedback as information from feedback as a learning-focused process of sense-making, judgement, emotional management and uptake. For girls, reception is more complex still. Gilligan (1982) explains why praise can sound hollow: if it does not resonate as authentic, it can deepen doubt rather than reduce it. Dweck (2006) likewise found that praise, when focused solely on ability, may reinforce self-doubt, underscoring that affirmation can backfire when it is disconnected from effort or strategy. Thus, feedback's effectiveness lies not only in its accuracy but in its authenticity.

In this context, the entry of AI into classrooms appears both promising and troubling. Proponents highlight AI's capacity for immediacy, consistency, and scalability. Abdallah (2024) found that self-regulated learning with AI chatbots enhanced reflective writing and self-expression among EFL student teachers, while Dennis (2024) reported that AI-powered speech recognition improved pronunciation and was valued for its instant feedback. Moreover, Sayed et al. (2024) associated AI-mediated oral testing with gains in well-being, autonomy, and buoyancy. However, while Zheng et al. (2024) found that automated assessment in English public speaking could function as an efficient and apparently objective formative tool, they found no significant impact on learners' self-efficacy, engagement, or competence, and noted that peer assessment produced higher social engagement.

Yet objectivity is a slippery claim (Benjamin, 2019; Noble, 2018). AI systems are created within gendered cultural and industrial contexts that shape what is noticed, measured, and rewarded (Guzmán et al., 2024; Wajcman & Young, 2023). Values of speed, competition, and detachment may clash with the relational validation girls often need to feel heard. Gilligan's (1982) warning that "speaking depends on listening" (p. xvi) thus remains pertinent: AI may provide feedback, but if it lacks resonance, it risks replicating the very silencing it seeks to redress.

A phenomenological lens shifts attention from what feedback says to how it is experienced. Feedback becomes an encounter through which identity is affirmed or denied; when it fails to resonate, it can trigger "emotional backwash" (Winstone & Carless, 2019, p. 158), where negative affect undermines uptake even when feedback is otherwise sound.

The literature converges on three points. First, girls' self-perceptions of confidence are not direct reflections of ability but are mediated by recognition and relational experience. Second, feedback plays a crucial role in aligning competence with confidence, yet it is effective only when perceived as authentic. Third, AI feedback holds promise in shaping not only performance but affective outcomes. But contradictions remain. Feedback is widely assumed to be transformative, yet for many girls it is not. AI is framed as objective, yet its lack of human recognition may limit its impact, and its design is rooted in masculinised cultures that risk encoding bias. Most importantly,

the current evidence base focuses on older or second-language learners. Little is known about how primary-aged girls experience AI feedback in public speaking, nor whether it can strengthen positive self-perception in ways that matter for their future confidence and agency.

This study addresses that gap by examining how Year 6 girls interpret AI-generated feedback on their public speaking and what this means for their positive self-perception as speakers. In a cultural moment where girls' voices are too often diminished, ensuring that girls see themselves as speakers who deserve to be heard is an urgent task for educators. To nurture girls' confidence in their voice affirms their place not only in the classroom, but as leaders, creators and active citizens.

Research Context

Wimbledon High Junior School is an academically high-achieving independent girls' day school in south-west London, with strong emphasis on playful scholarship, intellectual risk-taking, dialogic learning and oracy.

Participants were 25 Year 6 pupils aged 10–11 from one class. The parallel class followed the same curriculum and rehearsal opportunities but did not participate in formal data collection.

I occupied a dual role as both teacher and researcher. This required reflexivity, as my existing relationships with pupils and investment in the curriculum may have shaped how they responded, particularly during focus group discussions. A reflexive journal was maintained to document assumptions and support interpretive transparency.

Parental consent was obtained on an opt-in basis. Pupils were given an age-appropriate explanation of the study and were reminded that their contributions to focus groups and reflective discussions were voluntary. All data were anonymised and identifying details removed.

The Action

In designing this study, I aimed to understand how the girls experienced the unfamiliar presence of AI in a space that, until then, had always been human. I was interested not only in whether AI feedback could sharpen performance, but in what it might do to the relational fabric of the classroom.

The project ran over a thirteen-week autumn term within the Year 6 Socratic discussion and critical thinking curriculum—"Experientia." Rather than creating a separate intervention, the action was embedded in classroom practice. The cycle unfolded through three speaking tasks: an initial presentation, a mid-unit debate, and a final prepared speech. Each followed the same pattern: pupils drafted and rehearsed, recorded practice versions using Speaker Progress within Microsoft Teams, reviewed the automated feedback, refined their delivery, and then performed live.

Public speaking was taught explicitly through rhetoric, argument construction, delivery skills and dialogic engagement. At the outset, criteria for effective public speaking were developed collaboratively with the class. Pupils identified the features they associated with effective speaking, and these were aligned with key elements of Microsoft Speaker Progress feedback, including pace, clarity, filler words, pitch variation, and audience engagement. This ensured coherence across peer, teacher, and AI-generated feedback.

Throughout rehearsal, pupils received feedback from three sources: structured peer assessment, formative teacher feedback, and AI-generated feedback. These were positioned as complementary rather than hierarchical, and pupils were encouraged to compare them and reflect on how they influenced delivery.

As none of the pupils had previously used AI in classroom learning, the introduction of Speaker Progress was itself part of the action. The novelty of automated evaluation, the presentation of quantifiable metrics, and the absence of interpersonal cues shaped how the tool was encountered.

Following each speaking task, pupils completed written reflections and engaged in guided discussion about how different forms of feedback affected their confidence, motivation, and sense of themselves as speakers.

Data Collection

I used a mixed-methods approach that combined quantitative data with qualitative insight. Following Mertler's (2020) concept of "polyangulation," (p. 27) I drew on multiple sources:

questionnaires; focus groups; observations and field notes; automated feedback reports, metrics, and recordings from Microsoft Speaker Progress; and think-aloud protocols recorded during feedback and rehearsal sessions. Together, these created a record of how feedback was interpreted, discussed, and internalised.

I began with a shared exploration of what artificial intelligence is and what it might mean for a machine to “listen.” This established the conceptual baseline for the project. Each pupil then completed a short, age-appropriate self-perception questionnaire on baseline confidence and nervousness. Using simple statements and a five-point Likert scale, it traced within-cohort change. The questionnaire was repeated at the mid-point and again at the end of the intervention, providing a quantitative benchmark against which qualitative data could be read.

As the project unfolded, the girls delivered short speeches that were analysed by Speaker Progress, which generated feedback and metrics on pace, pitch variation, filler words, and body language. Alongside these AI-generated reports, I observed performances and recorded the small human details the algorithm could not quantify, such as subtle shifts in tone and posture.

Focus groups took place before, during, and after the intervention. A semi-structured, conversational format ensured inclusive dialogue and gave girls space to lead discussion about how feedback sounded and made them feel. They compared their reception of the AI’s feedback with mine and their peers’ and were invited to explain differences in interpretation and perceived intent. Each discussion was recorded and transcribed verbatim.

Data Analysis

For analysis, I followed Mertler’s (2020) sequence of organising, describing, interpreting, and reporting, alongside an inductive, phenomenological approach that looked for patterns in what the girls said about confidence, perception, and recognition. I read and re-read transcripts and notes, highlighting recurring words and images, then coded and grouped these into broader themes.

I then compared the qualitative data with the questionnaire results and Speaker Progress indicators, looking for alignment and contradiction. Regular discussion with my mentor helped keep

my interpretations honest. Collecting and analysing these data felt, at its core, like an act of listening; to the girls, to myself, to the technology, and to the relationship forming between us.

Discussion of Findings

The findings are presented thematically, foregrounding pupils' voices and drawing together quantitative and qualitative data to illuminate how AI-generated feedback was experienced and interpreted in practice.

AI-Generated Clarity Strengthens Performance, But Positive Self-Perception Develops Through Recognition Rather Than Criteria Alone

Across the study, AI-generated feedback was associated with observable improvements in performance. In rehearsal, girls visibly slowed down, lifted their heads, lowered their shoulders, gestured and smiled more, and paused instead of using filler words. These shifts were later observed in live classroom speeches, suggesting that the gains extended beyond interaction with the tool itself, even if some speed and filler words returned under performance pressure. The tool's metric data made performance criteria explicit and actionable, and pupils increasingly used these metrics when evaluating speeches.

Questionnaire data reflect this pattern, with items related to knowing what to work on showing consistent positive movement for eighteen participants. In focus group discussions, girls repeatedly valued the clarity and specificity of the tool's metrics because they made improvement concrete. One pupil explained why she liked the *pace* metric: "because if you speak too fast people can't understand you," while another valued the tool because "it tells you what you actually do, like filler words."

This clarity supported performance. Pupils made deliberate adjustments when speaking to the AI, often replaying speeches, commenting that they had been "too fast" or said "uhm" too often, and then refining delivery accordingly. One reflected, "I talk slower to Speaker Progress," linking this change to the experience of speaking being "less nerve wracking." Another explained, "It made me speak slower because I had to think about it." These comments, together with my classroom

observations, align with feedback literature that emphasises the importance of task-level clarity for improvement (Hattie & Timperley, 2007). Read through Winstone and Carless's (2019) learning-focused model, the metrics only became meaningful when pupils made sense of them, rehearsed differently, and carried altered behaviours into later performances. The improvement was therefore not produced by AI alone, but by the interaction between visible criteria, repeated rehearsal, and pupil uptake.

However, increased clarity and improved performance did not reliably strengthen girls' positive self-perception as speakers. Across the questionnaires, affective items relating to confidence and nervousness did not show a clear overall positive shift, even as pupils became clearer about what effective speaking required. This pattern was mirrored in pupil voice. As one girl put it, "I still get nervous even if I know what I'm meant to do." Here, competence and confidence remained distinct, echoing Zheng et al.'s (2024) finding that automated assessment may support public speaking development without necessarily producing significant shifts in self-efficacy.

This resonates with Gilligan's account (1982) of the dissociation between knowing and speaking, particularly for girls, whose confidence is often shaped relationally rather than procedurally. Pupils themselves articulated what was missing. While AI feedback made performance legible, it did not always feel personally affirming. One girl captured this distinction succinctly: "It tells you what you did, but it doesn't really know you." In this analysis, positive self-perception did not develop through criteria alone, but through experiences of recognition—being known, received, and understood as a speaker. Where feedback functioned primarily as instruction, performance improved, but confidence shifted only tentatively; where feedback functioned as acknowledgement, pupils were more likely to describe feeling seen and valued.

AI-Generated Feedback Supports Girls' Positive Self-Perception When It Preserves Agency and Choice

A second pattern concerned the role of agency in shaping how AI feedback was experienced. Over the course of the study, questionnaire data indicated that many pupils came to feel more

confident about *how to improve* than about speaking itself. Qualitative data help explain this distinction by showing that AI feedback was most supportive of self-perception when it increased pupils' sense of control over evaluation.

Girls consistently engaged with AI feedback selectively rather than deferentially. "I trust the filler words," one pupil stated, while another reflected, "Maybe you have to take it [pronunciation] with a pinch of salt." Pupils tested the system, noticed inconsistencies, and made judgements about usefulness. This agency was visible in the classroom: girls compared metrics against their own perceptions, used each other to count filler words, tested the pace metric by deliberately speeding up, tried different accents against the pronunciation metric, and dismissed outputs they judged unreliable, while continuing to act on the feedback they judged credible. These responses reflect the active role that Winstone and Carless (2019) argue is central to feedback literacy: they were not passively receiving feedback, but making sense of it, judging it, and deciding what deserved uptake.

Agency was particularly evident in how girls talked about correction. One pupil explained why AI feedback felt acceptable: "It doesn't like shout at you to say it correctly; it just tells you at the end, and you can choose to ignore it." Another summarised the experience simply: "I just take it better when I speak to a screen." These accounts echo Abdallah's (2024) study of AI-supported self-regulated learning, in which learner agency helped students take ownership of improvement. Here, however, girls regulated not only how they spoke, but how far they allowed the feedback to shape their self-judgement.

AI strengthened aspects of positive self-perception not by offering reassurance, but by reducing the subtle coercive pressures associated with relational feedback, a dynamic particularly significant in a girls' school context, where the wish to maintain connection may shape how evaluation is received (Gilligan, 1982). Once pupils could choose what to keep, discount or ignore, they often became louder, slower, less fidgety, and more physically settled in rehearsal.

Later, when girls could choose which aspects of the AI feedback to disable, questionnaire data, observations and pupil talk suggested they felt less anxious about both the feedback and their

own ability. Pupils were less likely to question or reject the metrics once they had removed categories they had identified as unreliable or unhelpful. Notably, six pupils became more nervous from the pre- to mid-point questionnaire, but five of these six then stabilised or became less nervous after they had switched off specific metrics. This broadly echoes Sayed et al.'s (2024) finding that AI supported autonomy and wellbeing by allowing learners to practise and receive feedback at their own pace. However, my data suggest that, for these girls, control over the terms of evaluation mattered as much as control over the pace. Increased control appeared to support a steadier sense of self when being evaluated.

Girls' Emotional Safety in Speaking is Situational and Relational, Not a Property of AI-Generated Feedback Itself

Girls' accounts revealed that AI feedback often reduced emotional risk, particularly in early or private practice, and this was reflected in behaviour: as soon as the task began, many moved away from peers to find quieter spaces in the classroom or corridor in which to rehearse. Several pupils described feeling less anxious when speaking to the AI tool, especially when contrasted with teacher or peer feedback that carried interpersonal expectations. Across focus groups, teachers and friends were described by several pupils as likely to "sugar coat" feedback "to be nice," which was understood as kind but also emotionally loaded. Against this backdrop, AI feedback was experienced as safer precisely because it was perceived as less personally invested. One girl reflected: "mmm … you're kind of more confident to a computer almost, because you're not speaking to someone, so they're not going to criticise you or hurt your feelings … the computer is not going to hurt your feelings because it's not a living thing—it's going to tell you the truth." This resonates with studies suggesting that AI-mediated practice can reduce anxiety by creating a less socially exposed rehearsal space (Abdallah, 2024; Sayed et al., 2024).

Yet pupils' accounts also show that emotional safety was not an inherent property of AI feedback. Zheng et al. (2024) found that automated assessment functioned as a useful formative tool, but that peer assessment produced higher social engagement. This distinction was visible in my

own data. The absence of social presence made AI feedback feel safer in rehearsal because it reduced the immediate risk of embarrassment, judgement, or disappointing a teacher. However, that same absence limited the relational cues through which pupils could experience themselves as heard. In this sense, AI's perceived objectivity was understood less as infallibility than as emotional detachment: feedback that felt less socially charged, but also less personally affirming.

Reduced emotional risk supported rehearsal; I observed girls restarting more readily, working with greater focus and determination, and even telling peers to "go away" when interrupted. Pupils linked feeling less nervous to concrete changes in delivery, such as slowing down or concentrating more carefully. In my observations, pupils were more willing to initiate additional rehearsal, adjust phrasing, and experiment with delivery when practising with the tool than when speaking live; with peers, silliness and giggles more often returned, along with faster speech and more filler words.

At the same time, pupils articulated the relational limits of AI-mediated speaking. One girl offered a striking distinction: "When you're speaking to a camera … you're kind of speaking at it, but to a person, to it." This formulation captures a central tension in the findings. Speaking to AI reduced exposure, but it also altered the relational quality of speaking itself. The absence of tone, response, and mutual presence made it harder for girls to experience themselves as being heard. One pupil suggested that the tool would feel more comforting if it were "a character," indicating a desire for relational cues that could be interpreted emotionally. This was echoed in reflections that "with AI you can't hear how they're saying it," highlighting how interpretability shaped whether feedback felt emotionally safe.

Questionnaire responses mirrored this complexity. Affective items did not show a cohort-wide shift towards greater calm; instead, pupils' trajectories remained heterogeneous across the cycle. For two pupils, moments of doubt about the tool's accuracy appeared to coincide with increased nervousness, suggesting that emotional safety could be disrupted when feedback felt less reliable or harder to interpret.

This helps explain why reduced nervousness did not reliably translate into increased confidence. While AI removed some interpersonal threat, it also removed relational cues through which girls experience recognition. Zheng et al.'s (2024) finding that automated assessment cannot substitute for the social engagement of peer assessment sharpens this point: the girls did not need feedback only as measurement, but as a felt response. In Gilligan's (1982) terms, a girl's voice develops in relationship; her confidence as a public speaker depended not only on reduced fear or increased skill, but on speaking to someone capable of hearing her.

Negotiating Algorithmic Authority: Misrecognition, Trust and Girls' Self-Perception

Finally, a fourth theme concerned how pupils negotiated algorithmic authority when AI feedback appeared to misrecognise their speech. This was most evident in pronunciation, where the tool's judgements sometimes conflicted with pupils' own embodied knowledge of what they had said. These moments were noticed and named but rarely contested. One pupil remarked, "It said I pronounced a bunch of things wrong—which I didn't. I don't understand how you can say 'this' wrong." Another responded pragmatically: "It might hear us wrong." Rather than rejecting the feedback outright, the girls filtered it: "I still looked at the other bits." In my observations, these moments were often handled through quick rationalisations or subtle dismissal, allowing pupils to continue without publicly challenging the tool. This reflects both the active sense-making central to feedback literacy (Winstone & Carless, 2019) and, more specifically, the pressure girls can feel to preserve relational harmony rather than openly challenge perceived authority (Gilligan, 1982). When feedback was clearly misrecognised, girls would mutter "well that's just wrong" or "how can it say that," yet contestation remained quiet, and no girls openly challenged it with me.

These responses help explain why confidence measures in the questionnaires did not show sharp declines. Misrecognition rarely produced overt discouragement. Instead, it introduced epistemic doubt—uncertainty about whose judgement to trust. As one pupil observed, "AI can make you second-guess yourself if it says something you think is wrong." Accent-related concerns sharpened this risk, particularly when pupils wondered whether the AI was "making fun of my

accent.” This distinction is significant because, while AI-powered speech recognition has been shown to support pronunciation, its effectiveness depends on system precision, learner context, and adaptability to diverse speakers (Dennis, 2024). In this study, a misrecognised word carried affective as well as technical implications. For some girls, it became a moment in which they had to decide whether the problem lay in the tool or in themselves.

This also helps interpret the mixed trajectories in affective questionnaire items across the cycle. Two of the seven pupils who began the study with low nervousness later reported an increase, while three of the seven who began more nervous described feeling less nervous by the end. While these shifts cannot be attributed to AI alone, they suggest that emotional responses were negotiated differently depending on prior self-perception and on whether feedback was experienced as credible. In this sense, misrecognition does not simply “lower confidence”; it can alter the conditions under which pupils trust their own perceptions, particularly when the tool’s judgement conflicts with their lived sense of self. The girls’ responses were therefore not passive acceptance but emerging evaluative judgement: they were learning how to use feedback while also appraising the credibility of an apparently objective system. This extends Winstone and Carless’s (2019) account of feedback as sense-making by showing that, in an AI-mediated context, uptake also required pupils to judge the authority of the feedback tool itself.

The girls’ management of misrecognition resonates with Gilligan’s (1982) account of self-silencing. It also connects with wider critiques of AI neutrality, which caution that systems often acquire authority through the appearance of objectivity, even when they foreground particular ways of seeing or judging (Benjamin, 2019; Noble, 2018). Within the feedback ecology of the classroom, girls’ capacity to rationalise inaccuracies may have protected them from immediate negative affect, helping account for the lack of sharp declines in confidence scores. At the same time, this quiet accommodation raises ethical questions about the longer-term impact of feedback systems that misrecognise speakers without offering avenues for dialogue or repair. The concern, then, is not that

AI simply damaged confidence, but that it sometimes shifted the burden of interpretation onto the child.

Conclusion

This study set out to explore how AI-generated feedback impacted Year 6 girls' positive self-perception as public speakers. The findings suggest a more cautious, but ultimately more revealing, conclusion. AI feedback supported performance. What it did not consistently do was transform how girls understood themselves as speakers.

Positive self-perception arose from a feedback ecology in which AI, teacher response, peer reflection, and pupils' own judgements interacted. Within this ecology, confidence was shaped less by clarity alone and more by acknowledgement, interpretability, and trust. AI reduced some forms of social exposure. But it could not generate the relational cues through which speakers experience themselves as heard.

Nor was its authority straightforward. Girls valued its directness and specificity, yet they also tested it, filtered it, and quietly negotiated its inaccuracies. Misrecognition did not produce dramatic collapse in confidence; instead, it introduced subtler forms of epistemic doubt. When a system contradicts a pupil's lived sense of her own speech, the question is not simply whether she corrects the error, but whose judgement she learns to privilege. This tension shaped how girls came to trust—or doubt—their own self-perception as speakers.

These findings complicate both optimistic and alarmist narratives: AI neither silenced nor automatically empowered the girls, but illuminated the conditions under which confidence grows. It made authority more ambiguous. It exposed how easily feedback can shift from support to surveillance, and how carefully pupils manage that tension.

What emerges, then, is not a verdict on AI, but a reframing of the problem. Confidence in speaking is not secured through metrics alone. It develops when competence, agency, and recognition converge. AI can clarify the craft of speaking. But it cannot, by itself, create the experience of being received.

The deeper task, then, is to ensure that speaking is met with recognition rather than measurement. The work of nurturing girls' voices will always remain relational. While AI can sit within that ecology, it cannot do the interpersonal work through which a girl comes to believe that her voice matters. Without human recognition, improvements remain technical; with it, voice can become identity. The question, therefore, is not whether AI belongs in girls' education, but how we shape the ecology in which it operates: to use clarity in the service of courage, and to ensure that measurement never becomes a substitute for listening.

Reflection Statement

There have been many moments of professional growth during this project, but the most significant has been the sustained opportunity to listen carefully to how my Year 6 girls interpret being evaluated. Hearing them articulate what they trusted, what unsettled them, and what made them feel heard was both illuminating and humbling. The research process slowed down my assumptions about feedback and allowed space for girls to name distinctions that are often invisible in classroom practice: the difference between clarity and confidence; between correction and judgement; between being assessed and being recognised.

This inquiry has strengthened my conviction that voice development in girls cannot be reduced to technical proficiency. Public speaking is relational work. Confidence does not grow solely from clearer criteria or improved metrics, but from experiences of credibility, agency, and acknowledgement. AI-generated feedback, while powerful as a performance tool, revealed its limits most clearly when it could not provide recognition. That insight will continue to shape how I design feedback environments in the future. Technology can support rehearsal and reduce certain emotional risks, but it cannot replace the human conditions under which girls come to see themselves as speakers who deserve to be heard.

I am deeply grateful to Karen Lewis, my research adviser, for her rigorous challenge and steady guidance throughout the process. My thanks also go to Debbie Hill for leading and sustaining this work across the GDST and for championing practitioner research with such clarity of purpose. I

am grateful to Harriet Fryer, Head of Juniors, for her encouragement and belief in the value of this inquiry, and to Sarah Chittenden, Deputy Head, for her logistical wizardry and for working with me to make space for this project alongside an already full professional life.

Most of all, I owe thanks to Kate Cheetham, my mentor, who has been present at every stage of this journey. Her thoughtful encouragement has shaped this project, and indeed my practice, in ways both visible and invisible.

Finally, I must thank the Year 6 girls who participated. Their honesty, perceptiveness, and willingness to engage critically with unfamiliar technology made this research possible. I began this project wondering whether AI might help girls trust their voices more readily. I conclude it more convinced than ever that girls' confidence grows when we listen to them seriously — and when we design classrooms, technologies, and curricula that recognise them as thoughtful speakers in their own right.

References

Abdallah, M. M. S. (2024). Using self-regulated learning supported by artificial intelligence (AI) chatbots to develop EFL student teachers' self-expression and reflective writing skills. *Academic Journal of Faculty of Education, Assiut University 40*(9) (Sept.), 1–50.

Ahmed, S. (2017). *Living a feminist life*. Duke University Press. https://doi.org/10.2307/j.ctv11g9836

Banet-Weiser, S. (2018). *Empowered: Popular feminism and popular misogyny*. https://doi.org/10.2307/j.ctv11316rx

Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new jim code*. Polity.

Best, E., Clark, C., & Picton, I. (2023). *Confident young adults: Communication and social, public, and civic engagement*. National Literacy Trust. https://literacytrust.org.uk/research-services/research-reports/confident-young-adults-communication-and-social-public-and-civic-engagement/

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, *21*(1), 5–31. https://doi.org/10.1007/s11092-008-9068-5

Clark, C. (2011). *Boys, girls, and communication: Their views, confidence, and why these skills matter*. National Literacy Trust.

De Paola, M., Lombardo, R., Pupo, V., & Scoppa, V. (2021). Do women shy away from public speaking? A field experiment. *Labour Economics*, *70*(626), 102001. https://doi.org/10.1016/j.labeco.2021.102001

Dennis, N. K. (2024). Using AI-powered speech recognition technology to improve English pronunciation and speaking skills. *IAFOR Journal of Education*, *12*(2), 107–126. https://doi.org/10.22492/ije.12.2.05

Dewey, J. (1916). *Democracy and education: An introduction to the philosophy of education.* Macmillan.

Dweck, C. S. (2006). *Mindset: The new psychology of success*. Random House.

Gilligan, C. (1982). *In a different voice: Psychological theory and women's development*. Harvard University Press.

Girlguiding. (2024). *Girls' attitudes survey 2024*. https://www.girlguiding.org.uk/globalassets/docs-and-resources/research-and-campaigns/girls-attitudes-survey-2024.pdf

Guzmán, E., Fischer, R. A.-L., & Kok, J. (2024). Mind the gap: Gender, micro-inequities and barriers in software development. *Empirical Software Engineering*, 29, Article 17. https://doi.org/10.1007/s10664-023-10379-8

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

hooks, bell. (1989). *Talking back: Thinking feminist, thinking black*. South End Press.

Jane, E. A. (2017). *Misogyny online: A short (and brutish) history*. Sage Publications Ltd.

Loureiro, M., Loureiro, N., & Silva, R. (2020). Differences of gender in oral and written communication apprehension of university students. *Education Sciences*, *10*(12), 379. https://doi.org/10.3390/educsci10120379

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

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. University Press.

Pitt, E., & Norton, L. (2017). "Now that's the feedback I want!" Students' reactions to feedback on graded work and what they do with it. *Assessment & Evaluation in Higher Education*, *42*(4), 499–516. https://doi.org/10.1080/02602938.2016.1142500

Sayed, B. T., Zein, Alkhayyat, A., Iroda Adhamova, & Teferi, H. (2024). To be with artificial intelligence in oral test or not to be: A probe into the traces of success in speaking skill, psychological well-being, autonomy, and academic buoyancy. *Language Testing in Asia*, *14*(1). https://doi.org/10.1186/s40468-024-00321-0

Speakers Trust. (2023, April 12). *Almost half of girls fear public speaking*. Speakers Trust. https://speakerstrust.org/almost-half-of-girls-fear-public-speaking/

Wajcman, J., & Young, E. (2023). Feminism confronts AI: The gender relations of digitalisation. In J. Browne, S. Cave, E. Drage, & K. McInerney (Eds.), *Feminist AI: Critical perspectives on algorithms, data, and intelligent machines* (pp. 47–64). Oxford University Press. https://doi.org/10.1093/oso/9780192889898.003.0004

Winstone, N., & Carless, D. (2019). *Designing effective feedback processes in higher education: A learning-focused approach*. Routledge.

Zheng, C., Chen, X., Zhang, H., & Chai, C. S (2024). Automated versus peer assessment: Effects on learners' English public speaking. *Language Learning & Technology*, *28*(2), 210–228. https://doi.org/10.64152/10125/73578