Exploring Conceptual and Critical Thinking: Using GenAI to Enhance Year 9 Girls' Understanding of Scientific Models

Linda Zhe Jue Chui

MLC School Burwood, Sydney Australia

Abstract

This action research investigated the integration of generative artificial intelligence (GenAI) to support the development of conceptual and critical thinking in Year 9 (14–15-year-old) girls through scientific modelling. In school science classrooms, models are frequently used to introduce new content and abstract concepts; however, within the constraints of the densely packed New South Wales curriculum, in-class opportunities for explicit evaluation and critiques of models are often compromised. Consequently, students may continue to perceive science as a subject reliant on rote learning rather than requiring deep conceptual understanding. Over a 12-week period, a range of scientific models was introduced as course content progressed. These included physical models constructed by students, GenAI-generated models produced using CanvaAI, and evaluative tasks supported by both CanvaAI and Microsoft Copilot. Students were required to compare, critique, and refine these models, and, in some instances, receive written feedback from Copilot on evaluative modelling responses. Findings indicate growth in girls' conceptual understanding, particularly through the comparison of GenAI-generated models with physical representations. Increased evidence of active learning was also observed during these modelling and evaluation tasks. However, for some students, limitations in their scientific knowledge contributed to instances of metacognitive laziness, particularly when tasks required higher-order evaluative judgement while comparing their own work with GenAI outputs.

Overall, students were able to pause and review GenAI outputs with greater precautions. They also included more appropriate details when answering scientific modelling questions that required their applications of evaluative skills, demonstrating improved critical thinking skills.

Glossary

GenAI: Generative artificial intelligence

Conceptual Thinking: The process where students actively and critically reflect on their understanding of a concept (Solstad et al., 2024)

Critical Thinking: The cognitive skill that enables students to evaluate evidence, question assumptions, and make reasoned judgements rather than relying on rote memorisation

Metacognitive Laziness: Term to describe the lack of critical engagement, whereby students cognitively offload problem-solving processes due to limited foundational knowledge

Molymod: A chemistry modelling kit

Paired T-Test: A statistical test used to analyse numerical data (nominal data such as test scores) when samples are dependent, for instance, it compares the means scores of the same group at different times.

Exploring Conceptual and Critical Thinking: Using GenAI to Enhance Year 9 Girls' Understanding of Scientific Models

My research question states: How does using GenAI image creation and modelling enhance girls' conceptual and critical thinking in a Year 9 Science class? Statisticians Box and Draper (1987) state that "all models are wrong, but some are useful" (p. 424). In science, concepts are often taught through modelling, as they are often abstract and not directly observable; they are tools for science communication and therefore have the tendency to evolve as new scientific knowledge arises. However, students often approach learning science with rote learning; therefore, it is important for teachers to emphasise the limits of scientific modelling to communicate scientific concepts correctly in absolute terms.

Another reason to implement this action research was to build girls' confidence in attempting tasks that are challenging and open-ended. There is anecdotal evidence from my experience of teaching girls that some usually want to know the correct answer and will not attempt questions because they do not want to write down the incorrect response; usually giving the reason, "I don't want to write the wrong answer because It'll ruin my book," or "It's too hard." Research suggests, this trait of perfectionism is more prominent in Year 9, particularly in gifted girls, and more observable in senior grades (Siegle & Schuler, 2000).

Utilising GenAI to generate and evaluate scientific models not only saves class time, but it provides more opportunities for girls to focus on image evaluation rather than image generation, bypassing the trap of perfectionism and prolonging active learning and potentially resulting in more effective conceptual understanding.

Literature Review

Since the release of ChatGPT in November 2022, generative artificial intelligence (GenAI) has transformed the world of education rapidly by providing tools for content generation, personalised feedback, and adaptive learning experiences (Kasneci et al., 2023). From my perspective as a science educator, utilising this tool in my teaching has been intriguing, especially with a goal of benefiting students' understanding of difficult scientific concepts.

Scientific models can be defined as representations of ideas, objects, processes, or systems that are employed to describe and explain phenomena, as well as to predict future observations (Science Learning Hub, 2011). Within science education, models serve as a central foundation for conceptual understanding and critical thinking. Scientific models also extend beyond static diagrams and exact replicas to include experiments, mathematical representations such as graphs, physical models, and dynamic computer simulations, all of which aim to help students to visualise and conceptualise abstract or unobservable scientific phenomena and or ideas (Treagust et al., 2002). However, research conducted by Tsai (2004) highlights that students often hold counterintuitive perceptions, assuming that science learning primarily involves passive knowledge acquisition; for example, memorising content and practising tutorial problems for the purpose of passing tests. In addition, while models are formally embedded in the curriculum, students may not always recognise them as practical tools actively utilised by teachers to support the communication and development of scientific concepts in everyday lessons (Treagust et al., 2002). This perception limits students' ability to use models to actively interrogate and refine conceptual understanding. Thus, integrating modelling into regular classroom practice is critical for developing deeper scientific reasoning, as modeling competence plays an integral part in scientific literacy (Chiu & Lin, 2019).

The use of scientific models is also a crucial element in enabling students to develop and apply conceptual thinking skills. Conceptual thinking refers to students actively and critically reflecting on their understanding of a concept (Solstad et al., 2024). The ability to conceptualise, therefore, requires a prior grasp of the concept, enabling learners to evaluate and refine their understanding. Pedagogical approaches that foster conceptual thinking are most effective after explicit teaching has occurred; however, this does not imply that explicit instruction renders students passive. Rather, explicit teaching can still involve active engagement, but it is distinct from the deeper reflective processes characteristic of conceptual thinking.

By contrast, non-conceptual thinking describes a surface-level approach to learning in which students focus primarily on producing correct answers, often in response to teacher prompts (Vinner, 1997). Although sometimes viewed negatively, non-conceptual thinking is not entirely absent from effective classrooms. Instead, it is common that both types of thinking, or the complete absence of thinking, are present in any classroom depending on factors such as the time of day, students' cognitive load, or ability to concentrate (Singh, 2025). These contextual influences shape the extent to which learners can engage in conceptual thinking and active learning.

Critical thinking is another fundamental skill in science learning, as it enables students to evaluate evidence, question assumptions, and make reasoned judgements, rather than relying on rote memorisation. Developing critical thinking requires students to actively take ownership of scientific ideas by reconstructing concepts, internalising meaning, and articulating their understanding through explanation and communication. This process reflects the central role of conceptual reconstruction and model-based reasoning in science education (García-Carmona, 2023; Treagust et al., 2002).

According to Wilson et al. (2016), a gender gap is evident in the Australian Science Olympiad Examination for Physics, with male students outperforming female students on items assessing projectile motion, two-dimensional motion, forces, and tasks requiring the interpretation of diagrams. These findings suggest that male and female students may process scientific models and visual-spatial information differently, which in turn can contribute to differences in their conceptual thinking and therefore their understanding of scientific concepts.

Using GenAI to convert text to image using prompts may be one way to bridge the gender gap in girls' conceptual thinking. For example, using computer simulation to teach chemistry (electrolysis) narrowed such a gender gap in a 2021 study (Oladejo et al., 2021). Navigating GenAI in model construction and evaluation allows students to focus their cognitive load on analysing models through AI prompts rather than the busy work of constructing a physical model. The intention of my action research, therefore, was to reduce "busy time" and increase time for conceptual thinking to take place.

Recent neuroscientific evidence indicates that students completing essay-writing tasks without GenAI support demonstrate stronger and more widely distributed brain connectivity, suggesting greater cognitive engagement (Kosmyna et al., 2025). In contrast, Kestin et al. (2025) found that GenAI tutoring could outperform traditional in-class active learning, highlighting its potential to enhance immediate instructional effectiveness. However, this outperformance may be short-term as suggested by Bastani et al. (2024) that while GenAI may boost short-term performance, overreliance could dampen long-term retention. This pattern again aligns with the findings from the latest 2026 OECD report (Organisation for Economic Co-operation and Development, 2026) and research by Ahern (2025) and Fan et al. (2024), which emphasises that GenAI tools can improve students' immediate task performance; however, such gains do not always translate into deeper learning. The OECD report also warns that cognitive offloading to GenAI may increase the risk of metacognitive laziness and disengagement, potentially undermining long-term skill development.

At present, the long-term pedagogical value of GenAI remains inconclusive, particularly regarding its ability to sustainably foster critical and conceptual thinking or to engage more girls to select a science subject in their senior years (Years 11 and 12) and to pursue a career in STEM beyond school. In 2025, NSW Higher School Certificate (HSC) enrolment data (NSW Education Standards Authority [NESA], 2025) showed that female students comprised a majority of biology enrolments (64%), were close to gender parity in chemistry (49%), and remained substantially underrepresented in physics (24%). These enrolment patterns, particularly the persistent underrepresentation of girls in physics, underscore the need to investigate if and how GenAI-supported pedagogies might influence girls' conceptual engagement, subject selection, and confidence in choosing senior science pathways.

To address the persistent underrepresentation of women in STEM fields globally and within Australia, research by Mackenzie et al. (2021) indicated that girls who endorsed a growth mindset and developed stronger self-efficacy demonstrated greater confidence in their ability to succeed in science and were, therefore, more likely to continue studying physics and chemistry in the senior years of high school, increasing the likelihood of pursuing STEM-related pathways beyond school. Supporting the development of students' self-efficacy and growth mindsets is also a priority within MLC School and, therefore, represents an area that may be positively influenced through the implementation of this action research project.

Research Context

MLC School is an independent all-girls school located in the inner west suburb of Sydney, Australia. MLC school used to be called Methodist Ladies' College in 1914, before adopting the name MLC School in 1977 following the formation of the Uniting Church. It accepts girls from Pre-Kindergarten (3 years old) to Year 12 (18 years old). It was established in 1886, with the intention of preparing girls for university honours at the University of Sydney. It was the first school in New South Wales to prepare girls to study physics at a tertiary level.

My class consisted of 23 Year 9 girls aged 14–15 years old at a stage where studying science is mandatory. In NSW high schools, the Stage 4 (Years 7 and 8) and Stage 5 (Years 9 and 10) Science courses include content in biology, chemistry, physics, earth and environmental science, and data science. Students are exposed to all the above-mentioned disciplines prior to their subject selection in the senior years (Years 11 and 12). In the senior years, the school offers both the IB (International Baccalaureate's Diploma Programme) and the HSC (Higher School Certificate) courses.

I chose Year 9 girls (14–15 years old) for this action research as this age represents a critical period for engaging and encouraging students to continue studying science in the senior years. Research by DeWitt et al. (2014) indicates that while children typically report positive attitudes towards science at around age 10 years old, interest and aspirations to pursue science subjects declined markedly by age 14 years old. This decline is also reflected in senior secondary enrolment patterns in the 2025 NSW HSC courses (NESA, 2025).

Both the participants and their parents were provided with details about the research, along with the consent forms to support informed participation in the study. Permission forms for data collections were provided to all parents of my students immediately after the introduction of the action research, highlighting that each student's name would remain anonymous.

The Action

My action utilised GenAI in my lessons to assist girls' conceptual and critical thinking through scientific modelling. CanvaAI was chosen as a GenAI platform in the first part of the action research. Girls first received explicit teacher instructions and participated in a guided group practical using Molymods (chemical modelling kit) to model the polymerisation of monomers. Each student then independently translated her conceptual understanding of polymerisation into a written text prompt, which was uploaded to CanvaAI to generate a corresponding visual model. Individually, students evaluated the accuracy and effectiveness of the GenAI-generated model by comparing it with their practical model and iteratively refining their text prompts to produce visual representations that more closely aligned with, or more effectively communicated, the polymerisation process. This approach extended traditional Molymod-based instruction by integrating multimodal learning, requiring students to translate conceptual understanding across physical, textual, and visual representations.

Lastly, girls were explicitly introduced to a different scientific concept—the natural greenhouse effect—through conducting an experimental model. They then needed to insert this concept into Copilot through a text prompt on the scientific method and see if the results aligned with their expectations. Individually, girls evaluated the effectiveness of the practical in enriching their conceptual understanding of the science of the natural greenhouse effect; their evaluation was also compared with Copilot's evaluation to exercise both their conceptual and critical understanding. During the implementation of this action research, girls' access to Copilot was inconsistent due to age-based licensing requirements and legislative timing; therefore, students without access were paired with peers who had access to ensure participation for all students.

Following the implementation of the GenAI evaluation activities, girls completed an "examination style" question in class as evidence of learning. This task enabled me to provide timely, targeted feedback on students' responses. In addition, their written responses were submitted to ChatGPT, and the GenAI generated feedback was printed and provided to students. Students then compared teacher feedback with GenAI feedback to identify strengths, areas for improvement, and refinements needed in their original responses; thus, supporting iterative improvement and metacognitive engagement with feedback. After this activity, each student reattempted the same question and peer marking was conducted to exercise girls' critical thinking skills.

Data Collection

To investigate conceptual thinking in students via GenAI image modelling and written prompts, I employed a mixed-methods approach from Mertler's (2020) action research framework, combining qualitative and quantitative data to explore students' development on conceptual understanding and critical thinking. Baseline data were gathered, and mid- and end-of-term surveys were conducted to gauge girls' own perceptions on the effectiveness of GenAI in assisting them with their conceptual understanding in science throughout the progress of the action research.

I used the following data collection methods:

1. Baseline data were collected via a Microsoft Form to gain a perspective on students' confidence on learnt scientific concepts and their usage of GenAI.

2. Teacher-led activity that investigated students’ understanding by comparing the qualitative differences between a practical model to two GenAI models they generated in class via text prompts. In conjunction with this evidence, a classroom observation was recorded by another teacher to ensure a different perspective on students’ engagement and interaction with the activity. Observation notes from this lesson were valuable as it added additional dimensions to observe students’ behaviour and engagement.
3. Quantitative data from one question in 3 separate assessment tasks (see Appendix A) were collected after the implementation of GenAI to assist students’ development of conceptual thinking through modelling. These data were compared with two questions in previous assessment tasks, which also accessed students’ understanding in scientific modelling.
4. A mid-term survey was conducted to assess students’ confidence in using GenAI in scientific modelling.
5. Additionally, an interview was conducted with two students who volunteered themselves to provide overall feedback of their experience with this project.
6. An end-of-term survey was also given out via Microsoft Forms to gather students’ reflections at the end of the project.

Data Analysis

Inductive analysis of qualitative data was utilised to analyse my qualitative data, with the three steps of organisation, description and interpretation (Mertler, 2020). Baseline data and mid-term survey questions both contained qualitative and quantitative data, enabling polyangulation. Common themes identified in the qualitative data were organised to compare girls’ progress and their perception on conceptual understanding before and after the implementation of GenAI in their learning. The quantitative data collected within the student surveys provided insights on students' perception of their own confidence (e.g. 4 out of 5) in utilising GenAI and later, if their conceptual understanding in scientific modelling had improved from the implementation of AI, as this allowed me to cross-reference with students' reflections to further support my findings.

The organisation of students' responses for my second piece of evidence also involved sorting through common themes and choice of words when reading girls' evaluations and responses. This qualitative approach was also cross-referenced with data collection method 3's quantitative analysis, involving two paired t-tests to analyse students' improvement from three questions relating to scientific modelling, one conducted in Term 1, another in Term 2, and lastly Term 3, mainly to assess if there was any statistical significance in students' improvement from the implementation of GenAI. Results of the two paired t-tests are found in Appendix B. Although action research emphasises qualitative inquiry, the paired-samples t-test was selected Aas an appropriate complementary method to examine within student change, aligning with the cyclical and improvement-focused nature of the research and polyangulate my findings.

Lastly, girls' reattempts of a written response (an exam-style modelling question) were analysed following feedback from GenAI, peers, and the teacher to determine changes in students' evaluative reasoning, critical thinking and scientific literacy. Improvements in these areas were used as indicators of increased confidence and deeper consolidation of conceptual understanding and critical thinking of the introduced scientific model.

Discussion of Results

The following key findings were identified after data were analysed in the action research process.

Using AI Models Improves Girls' Growth in Conceptual Understanding of Introduced Scientific Models

End of term survey data supported this finding, with students providing specific examples of how GenAI supported their understanding of scientific concepts. One student noted that "AI helped visualise models better - help break down complex topics into digestible chunks of information", and another stated "GenAI helped by giving me more ideas, and fact-checking my answers."

This was further reinforced by the end of term video recording of student reflections. One student mentioned that "Using it [GenAI] in class has definitely helped me understand the concepts that we have being working through … I've realised like I've kind of learnt how to write prompts that AI can understand."

Taken together, the quantitative data also suggest that students' ability to evaluate scientific models improved across the year, particularly when compared to their initial engagement with modelling tasks early in the course. This improvement may reflect the cumulative effect of repeated exposure to modelling activities, including GenAI-supported visual modelling, which provided students with multiple opportunities to generate, critique, and refine representations of scientific ideas.

Whilst the quantitative analysis between Task 1 and Task 4 did not show statistical significance, I noticed an interesting pattern when marking the modelling response for Task 4. Although the task invited students to evaluate physical models in general, most students chose to focus on evaluating a familiar model, most commonly the polymerisation model shown in the stimulus material, rather than discussing broader advantages and limitations of physical models as a category. This tendency suggests increasing confidence and familiarity with seen models when engaging in evaluative reasoning. At the same time, it highlights an area for further development, and that is to support students to interpret task requirements carefully and to generalise evaluative criteria beyond familiar examples.

Using GenAI in Scientific Modelling Improves Girls' Self Efficacy in Scientific Conceptual Development and Critical Thinking

Evidence of increased self-efficacy was observed both within and beyond the classroom. One student independently explored additional scientific models related to the development of atomic theory outside of class time, driven by personal interest. She later provided detailed feedback on the accuracy and validity of multiple models generated by Copilot, demonstrating growing confidence and evaluative reasoning when engaging critically with scientific models. Similarly, several students continued to refine and modify text prompts in GenAI tools after completing the mandatory activities, expressing curiosity about whether the accuracy of their virtual AI-generated models could be improved. This voluntary extension of learning indicates heightened curiosity, ownership of learning, and engagement in active learning processes, which are all essential characteristics to strengthen their conceptual understanding and critical thinking abilities.

Classroom observation notes from my colleague further supported this finding, noting that "intellectual risk-taking was excellent in this activity and failure wasn't personal. They felt empowered by what they knew over AI." Students all completed the set work in the lesson, and they also started experimenting with CanvaAI to generate more images of the polymer despite only having to complete two attempts. This reinforced my finding that GenAI-supported modelling creates a low-stakes learning environment in which students were more willing to experiment, critique, and refine scientific ideas, thus leading to higher self-efficacy in students.

This increase in self-efficacy among girls at this stage of learning is significant, as confidence in one's scientific capability has been shown to influence subject persistence (student's continuation on a subject when it becomes optional) and potentially future STEM participation (Brage-del-Río et al., 2025). While small-scale, this approach may contribute to reducing gender disparities in senior STEM subject selection (Years 11 and 12) and in longer-term STEM career pathways (Brage-del-Río et al., 2025).

GenAI-Assisted Modelling and Evaluation Fosters Girls' Active Learning

Active learning is another critical element in students' conceptual development (Olimpo & Esparza, 2020; Sundstrom et al., 2025). GenAI modelling supports active learning by providing near real-time feedback in response to students' text prompts, capturing students' attention within a short timeframe. It also allows students to co-create the content, which reinforces students' ownership to both the content and the conceptual knowledge gained from the activities. Several students commented positively on the efficiency of model generation when using CanvaAI and Copilot, noting that it made visual models quickly. This efficiency enabled students to generate, review, and evaluate multiple models within a short timeframe, fostering deeper consideration of each model's accuracy and limitations rather than focusing on a single static representation.

One student further reflected that the GenAI-generated model prompted her to draw on prior scientific knowledge to identify inaccuracies and evaluate the model's validity. Her original text prompt began with "Could you please give me an image of polymerisation", which subsequently developed into: “Could you please generate an image of a model of polymerisation. Initially there are 3 monomers of ethane and each molecule has 2 carbons that are double bonded … there are six carbons and 12 hydrogens.” Although scientific errors are evident in this student’s prompt, this second text provided significantly more detail and therefore she was able to guide GenAI to produce a visual product that is more closely aligned with the physical model. Figure 1 below shows the visual product resulting from the first prompt, Figure 2 is the product from the second and it is more closely aligned with Figure 3, which students all built during their practical.

Figure 1 - Image of polymerisation from first prompt

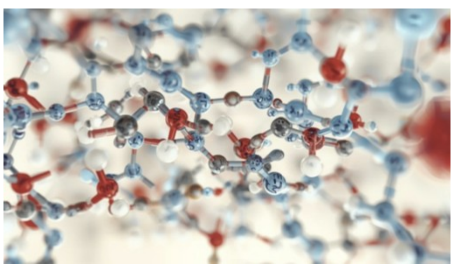


Figure 2- Polymerisation image from second prompt

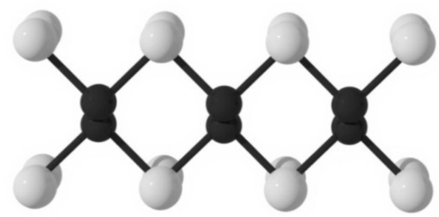
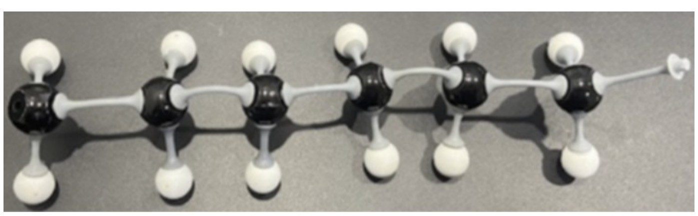


Figure 3 - Physical model of polymerisation built by student



This response indicates a shift from passive acceptance of information to critical engagement with scientific representations, a key feature of active learning that supports the development of conceptual thinking and understanding (Sahito et al., 2025).

Classroom observation notes from my colleague also highlighted that “the GenAI modelling was introduced in an explicit way.” To illustrate, CanvaAI was projected on the classroom screen while I input a text prompt describing the polymerisation process. When the first generated image did not align with the physical model of the polymer, a second prompt was revised. This explicit modelling process demonstrated to students how scientific models are iteratively constructed and refined. This explicit teaching of modelling tasks also encouraged students to reflect on the status of their own knowledge, and the need to revisit their first prompt, and develop an understanding of how scientific models are developed and evaluated at their stage (Year 9), aligning with research on effective modelling instruction (Chiu & Lin, 2019).

GenAI has Limited Impact on Girls' Critical Thinking

Some students reported that the GenAI-generated visual models were overwhelming at times. Observational notes indicated that repeated refinement of text prompts to generate multiple models enabled students to "evaluate the difference between the two" and "sparked interesting discussion and curiosity," particularly among those who wanted the GenAI models to be more scientifically accurate.

However, due to students' limited Stage 5 (Years 9 and 10) content knowledge, some learners expressed cognitive overload during the task. Comments such as "Can we just learn about science now?" and "Okay, I'm kinda traumatised now … also I'm kinda confused" suggest that, for these students, the complexity of evaluating multiple imperfect models hindered rather than supported their conceptual understanding at that stage of learning.

Additionally, some students did not engage as deeply with the visual modelling task or the practical model evaluation task. Some of their evaluative responses were also brief and lacked clarity, often focusing on superficial visual features rather than scientific accuracy. For example, comments such as "different colours, wrong colours, not labelled" and "highly inaccurate, gibberish language" were common, despite language not being an intended output of the task. This indicates that these students may have struggled to distinguish between identifying qualitative differences and linking them to conceptual criteria when evaluating models, highlighting the need for additional scaffolding to support evaluative reasoning.

It was also evident that some girls displayed signs of metacognitive laziness during the practical model-evaluation task. For example, several students did not recognise that the practical model described by the GenAI differed from the model they had implemented themselves. This suggests limited cognitive engagement in comparing similarities and differences between the GenAI-generated evaluation and their own evaluative judgments, indicating that critical thinking is hindered. As suggested by Ahern (2025), metacognitive laziness refers to a lack of critical engagement, whereby students cognitively offload problem-solving processes due to limited foundational knowledge. In this context, reliance on GenAI appeared to have promoted learner dependence on technology and potentially triggered metacognitive laziness in their learning (Fan et al., 2024).

Girls Place Greater Value on Teacher and Peer Feedback Than on GenAI When Evaluating Scientific Models

Evaluation of models is a critical component of the modelling process (Chiu & Lin, 2019). Accordingly, effective and timely feedback plays a central role in supporting students' growth and conceptual development (Hattie & Timperley, 2007). While GenAI is capable of providing immediate feedback, the feedback received by students was often generic. This is likely due to students' limitations to craft precise text input prompts that would enable GenAI models to generate personalised and diagnostically useful feedback.

Evidence from the teacher–student interview at the end of this action research further highlighted this limitation. Two students who volunteered for this interview reported that they "trust teacher's feedback, it provides more details, and teacher knows more on what might be pseudoscience," and that "AI often misses a lot of aspects; it often gives me really high marks but without super detailed feedback." These perceptions suggest that, although GenAI offers efficiency and immediacy with feedback, it currently lacks the depth, epistemic judgement, and contextual awareness required for high-quality formative assessment.

This gap highlights an opportunity to embed more structured reflections within classroom practice, allowing students to engage in the critical evaluation of feedback from multiple sources, including GenAI, the teacher and peers. This aligns with Hattie and Timperley's (2007) assertion that effective feedback requires learners to actively engage with and make sense of feedback, rather than passively receive it. Student voice further reinforced this need, with one participant suggesting in the End of Term Survey that we should "maybe compare AI and human feedback and see which is more accurate.", indicating an emerging awareness of the importance of assessing the validity and reliability of different feedback sources.

Conclusion

This action research explored how GenAI could be used to support Year 9 girls' conceptual and critical thinking through scientific modelling and evaluation. Guided by a mixed-methods approach (Mertler, 2020), it focused on GenAI-assisted visual modelling, written prompts, and GenAI's teacher and peer feedback and their impact on girls' conceptual understanding, evaluative reasoning, confidence, and engagement within a Stage 5 science classroom.

My findings suggest that, when embedded within explicit teacher's instruction, GenAI can support girls' conceptual thinking by providing accessible visual representations and opportunities to compare, critique, and refine scientific models. This is also reinforced by

Mandouit and Hattie's (2023) findings, suggesting that AI tools can provide timely and effective feedback that fosters metacognitive engagement, with immediate visual outputs enabling students to evaluate the clarity and precision of their scientific descriptions.

Many girls demonstrated increasing confidence in engaging with scientific representations and greater willingness to articulate evaluative judgements, particularly when working with familiar models introduced earlier in the course.

The benefits of GenAI however were not experienced uniformly. A few girls also experienced cognitive overload or engaged at a surface level when evaluating multiple inaccurate models, particularly when their scientific knowledge was still developing. In these cases, reliance on GenAI outputs appeared to limit comparison and evaluative reasoning, indicating instances of metacognitive laziness. These findings underscore the importance of careful scaffolding, explicit evaluative criteria, and teacher intervention to ensure GenAI supports girls' thinking rather than undermines it.

Girls also emphasised the importance of teacher and peer feedback in supporting critical thinking. While GenAI provided timely feedback, girls reported greater trust in teacher and peer feedback due to their specificity, curriculum alignment, and capacity to identify misconceptions and/or pseudoscientific reasoning. This highlights the relational and epistemic dimensions of feedback that are particularly valued by girls (Carroll & Park, 2024) and reinforces the central role of the teacher in guiding evaluative judgement in science.

This action research is limited in scale (23 participants); however, the insights gained have informed refinements to my pedagogical practice, including increased scaffolding for evaluative reasoning, explicit support for prompt constructions, and the deliberate positioning of GenAI as a comparative tool rather than a direct source of knowledge.

For this research to extend further, I will focus on managing cognitive load during modelling tasks and exploring hybrid feedback approaches that combine teacher expertise with scaffolded GenAI support. Through these refinements, this work aims to continue improving practice in ways that support girls' conceptual and critical understanding, confidence, and sustained engagement with science learning.

Reflection

This action research project provided a valuable opportunity to reflect more deeply on my pedagogy, engage with current educational research, and identify areas for improvement and innovation within my classroom practice. Through this process, I have become more intentional and reflective in my teaching.

Generative AI represents an important tool for students to explore as they prepare for future learning and careers. However, the development of critical and conceptual thinking remains essential if students are to use these technologies meaningfully and responsibly. While I do not consider myself an expert in Generative AI, I hope that this project may encourage other educators to explore this emerging area and experiment with the tools available to them in ways that enhance student learning.

While my data collection mainly consisted of student surveys and classroom observations, incorporating more data from video interviews and audio recordings would have provided richer insights into students' experiences, allowing for deeper exploration of student thinking to inform future iterations of this project.

I would like to sincerely thank my research advisor, Debbie Hill, for her tireless support and guidance throughout this project, from refining the research question to reviewing the final draft of my report. I am also grateful to my in-school mentor, Toby Duncan, for his guidance and the many valuable conversations we shared during the course of this action research. I would also like to thank my Principal, Lisa Maloney, for supporting my participation in this research project. Finally, I extend my thanks to my Year 9 Science class for their enthusiasm, engagement, and willingness to participate in this study.

References

Ahern, S. J. (2025). What is lazy metacognition and what can we do about it? *Journal of Learning Development in Higher Education*, *37*. https://doi.org/10.47408/jldhe.vi37.1713

Bastani, H., Bastani, O., Alp Sungu, Ge, H., Özge Kabakcı, & Mariman, R. (2024). Generative AI can harm learning. *The Wharton School Research Paper*. https://doi.org/10.2139/ssrn.4895486

Box, G., & Draper, N. R. (1987). *Empirical model-building and response surfaces*. John Wiley & Sons.

Brage-del-Río, M., Martin-Núñez, J. L., & Pablo-Lerchundi, I. (2025). Educational strategies to reduce the gender gap in the self-efficacy of high school students in stem teaching. *Frontiers in Education*, *10*. https://doi.org/10.3389/feduc.2025.1553001

Carroll, G., & Park, S. (2024). Towards expansive model-based teaching: A systematic synthesis of modelling pedagogies in science education literature. *Studies in Science Education*, *62*(1), 1–39. https://doi.org/10.1080/03057267.2024.2417157

Chiu, M.-H., & Lin, J.-W. (2019). Modeling competence in science education. *Disciplinary and Interdisciplinary Science Education Research*, *1*(12). https://doi.org/10.1186/s43031-019-0012-y

DeWitt, J., Archer, L., & Osborne, J. (2014). Science-related aspirations across the primary–secondary divide: Evidence from two surveys in England. *International Journal of Science Education*, *36*(10), 1609–1629. https://doi.org/10.1080/09500693.2013.871659

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Dragan Gašević. (2024). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*, *56*(2). https://doi.org/10.1111/bjet.13544

García-Carmona, A. (2023). Scientific thinking and critical thinking in science education. *Science & Education*, *34*. https://doi.org/10.1007/s11191-023-00460-5

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://www.jstor.org/stable/4624888

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., & Stadler, M. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*(102274). https://doi.org/10.1016/j.lindif.2023.102274

Kestin, G., Miller, K., Klales, A., Milbourne, T., & Ponti, G. (2025). AI tutoring outperforms in-class active learning: An RCT introducing a novel research-based design in an authentic educational setting. *Scientific Reports*, *15*(1), 1–10. https://doi.org/10.1038/s41598-025-97652-6

Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X.-H., Beresnitzky, A. V., Braunstein, I., & Maes, P. (2025). Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task. *ArXiv*. https://arxiv.org/abs/2506.08872

Mackenzie, E., Berger, N., & Holmes, K. (2021). Predicting adolescent girls' intentions to study science in senior high school. *Issues in Educational Research*, *31*(2), 574–585.

https://www.researchgate.net/publication/352538662_Predicting_adolescent_girls'_intentions_to_study_science_in_senior_high_school

Mandouit, L., & Hattie, J. (2023). Revisiting "The power of feedback" from the perspective of the learner. *Learning and Instruction*, *84*(84). https://doi.org/10.1016/j.learninstruc.2022.101718

Mertler, C. (2020). *Action Research Improving Schools and Empowering Educators*. SAGE.

NSW Education Standards Authority (NESA). (2025). *2025 HSC course enrolments*. NSW Government. https://www.nsw.gov.au/education-and-training/nesa/hsc/facts-and-figures/enrolment-data/courses

Organisation for Economic Development (2026). *OECD digital education outlook 2026: Exploring effective uses of generative AI in education*. https://doi.org/10.1787/062a7394-en

Oladejo, A. I., Nwaboku, N. C., Okebukola, P. A., & Ademola, I. A. (2021). Gender difference in students' performance in chemistry – can computer simulation bridge the gap? *Research in Science & Technological Education*, *41*(3), *1*–20. https://doi.org/10.1080/02635143.2021.1981280

Olimpo, J. T., & Esparza, D. (2020). Active learning and conceptual understanding in biology. In: J.J. Mintzes & E.M. Walter (Eds.). *Active learning in college science.* (pp. 43–57). Springer. https://doi.org/10.1007/978-3-030-33600-4_4

Sahito, Z.H., Khoso, F.J., & Phulpoto, J. (2025). The effectiveness of active learning strategies in enhancing student engagement and academic performance. *Journal of Social Sciences Review*, *5*(1), 110–127. https://doi.org/10.62843/jssr.v5i1.471

Science Learning Hub. (2011, May). *Scientific modelling*. Science Learning Hub. https://www.sciencelearn.org.nz/resources/575-scientific-modelling

Siegle, D., & Schuler, P. A. (2000). Perfectionism differences in gifted middle school students. *Roeper Review*, *23*(1), 39–44. https://doi.org/10.1080/02783190009554060

Singh, J. S. (2025). Correlation between cognitive load, learning strategies, and academic success in STEM subjects. *International Journal for Research Publication and Seminar*, *16*(4), 35–44. https://doi.org/10.36676/jrps.v16.i4.319

Solstad, T., Eivind Kaspersen, & Eggen, M. (2024). Eye movements in conceptual and non-conceptual thinking. *Educational Studies in Mathematics*, *118*, 323–338. https://doi.org/10.1007/s10649-024-10358-w

Sundstrom, M., Gambrell, J., Green, C., Traxle, A. L., & Brewe, E. (2025). Relative benefits of different active learning methods to conceptual physics learning. *ArXiv)*. https://doi.org/10.48550/arxiv.2505.04577

Treagust, D. F., Chittleborough, G., & Mamiala, T. L. (2002). Students' understanding of the role of scientific models in learning science. *International Journal of Science Education*, *24*(4), 357–368. https://doi.org/10.1080/09500690110066485

Tsai, C. (2004). Conceptions of learning science among high school students in Taiwan: A phenomenographic analysis. *International Journal of Science Education*, *26*(14), 1733–1750. https://doi.org/10.1080/0950069042000230776

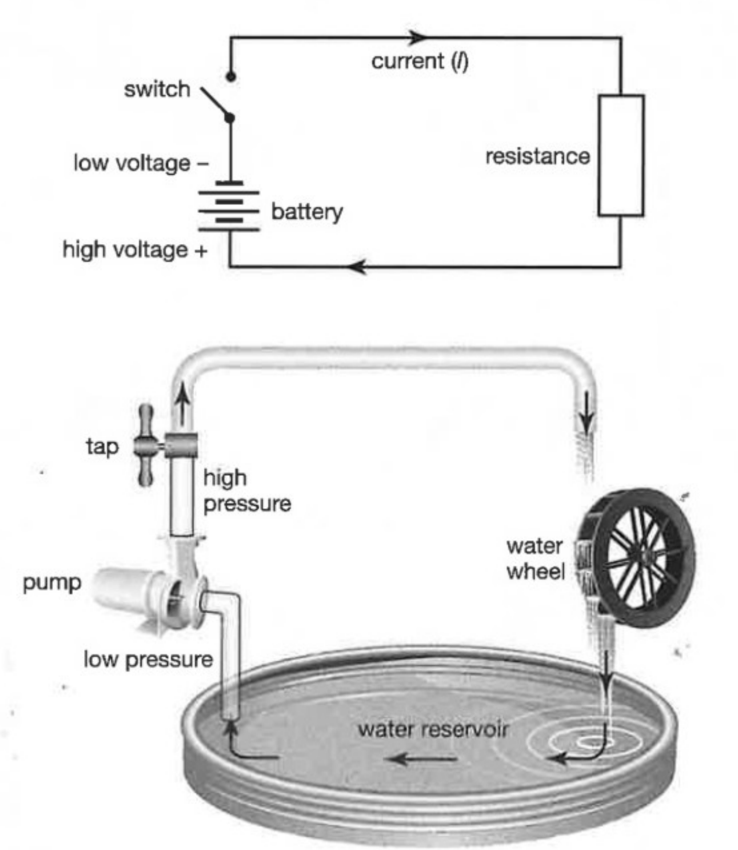
Vinner, S. (1997). The pseduo-conceptual and the pseudo-analytical thought processes in mathematics learning. *Educational Studies in Mathematics*, *34*(2), 97–129. https://doi.org/10.1023/a:1002998529016

Wilson, K., Low, D., Verdon, M., & Verdon, A. (2016). Differences in gender performance on competitive physics selection tests. *Physical Review Physics Education Research*, *12*(2). https://doi.org/10.1103/physrevphyseducres.12.020111

Appendix A - Scientific Models From Assessment Tasks

Term 1 - Water Wheel Model

7. Scientists use models to help them understand what is going on. An *analogy* is a model that compares something that is difficult to understand with something that is easy to understand. In a similar way, a simple electric circuit can be compared with the water 'circuit' that runs a water wheel or fountain in a garden pond. The figure below compares the two.



a) Identify **one** features that make this a good analogy. **1**

____

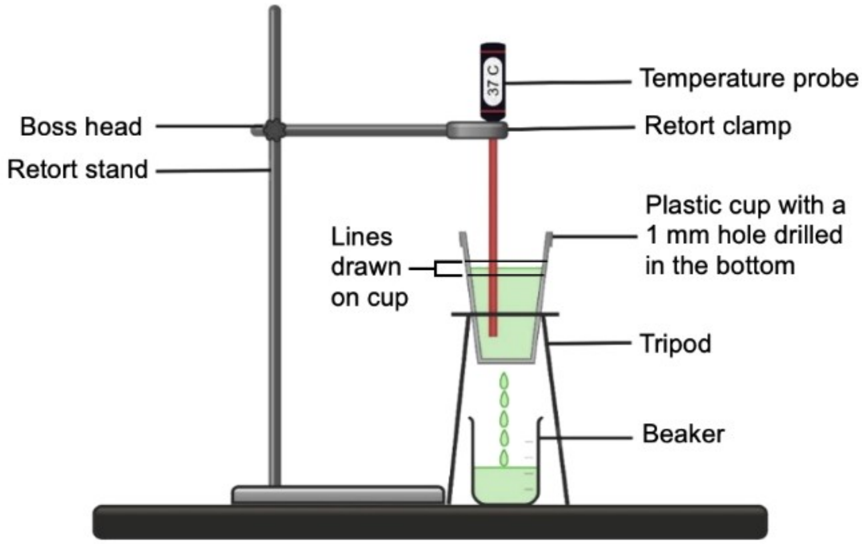
b) Identify **one** feature that makes this a poor analogy. **1**

____

____

Term 2 - Homeostasis Model

5. During this topic you completed a practical task to model homeostasis (apparatus shown below). Water lost was replaced by adding hot or cold water using a pipette.



a. **Describe** one **positive** aspect of the model in representing the maintenance of a constant body temperature. Explain why this is a strength of the model. **2**

____

____

____

____

b. **Describe** one **limitation** of the model in representing the maintenance of a constant body temperature. Explain why this is a weakness of the model. **2**

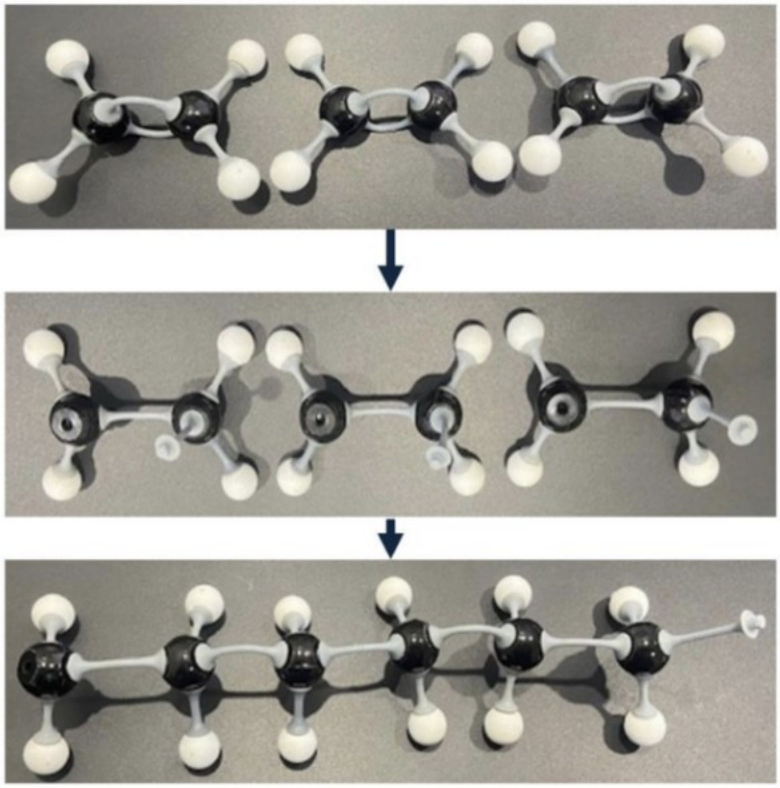
____

____

____

Term 4 - Physical Model

11. Scientific models are often used to represent concepts in science. For example, using the Molymod model kits to model the polymerisation of ethene into polythene.



Evaluate (make a judgement) the advantages and limitations of using physical models (e.g. molymod) to represent scientific concepts. **4**

Appendix B – Paired T-Test

Figure 1: Paired T-Test Results

T Test: Paired Two Sample for Means			T4VsT1
	0	0	
Mean	0.5	0.30434783	
Variance	0.07386364	0.08498024	
Observations	23	23	
Pearson Correlation	0.0717155		
Hypothesized Mean Difference	0		
df	22		
t Stat	2.44332749		
P(T<=t) one-tail	0.01151918		
t Critical one-tail	2.07387306		
P(T<=t) two-tail	0.02303835		
t Critical two-tail	2.40547274		

T Test: Paired Two Sample for Means			T4vsT2
	0	0	
Mean	0.5	0.33695652	
Variance	0.07386364	0.08868577	
Observations	23	23	
Pearson Correlation	-0.1755032		
Hypothesized Mean Difference	0		
df	22		
t Stat	1.78936054		
P(T<=t) one-tail	0.04366714		
t Critical one-tail	2.07387306		
P(T<=t) two-tail	**0.08733428**		
t Critical two-tail	2.40547274		

The statistical analysis calculated a p-value of 0.023 (<0.05) for the comparative analysis between the mean score in the first task and the final task, which suggests a statistically significant improvement from Task 1(Term 1) to Task 4 (Term 4). While the mean performance also increased from Task 2 (Term 2) to Task 4, the calculated p-value is 0.087(>0.05), which indicates the mean difference between tasks was not statistically significant, indicating some improvement in the mean scores within each students’ performance but not all.