

From Text to Symphony: A Longitudinal Study of GenAI Integration in a Group Project Assessment

YU Wing Yan ¹

¹ School of Optometry, The Hong Kong Polytechnic University, Hong Kong, China

Abstract

This longitudinal mixed-methods study examined the evolution of student engagement with generative artificial intelligence (GenAI) in a health science general education course at The Hong Kong Polytechnic University across seven cohorts (2022–2025, $n = 398$). Drawing on 74 GenAI usage logs and associated reflections, course-end surveys ($n=185$), and focus groups, it tracked a shift from text-based exploration (e.g., ChatGPT for brainstorming) to multimodal orchestration (e.g., Suno AI, Perplexity for music, verification, and video). Surveys revealed consistently high perceptions of GenAI fostering critical thinking ($M=4.6–4.8/5$) and tool proficiency enhancement (Mean = $4.5–4.8/5$). Qualitative data highlighted persistent challenges including hallucinations requiring human validation, assessment anxiety over rubrics, preferences for human versus AI voices, and a “paywall barrier” limiting premium multimodal access, exacerbating equity divides. These findings advocate redefining AI literacy as workflow management, urging institutions to provide premium AI tools and ungraded practice spaces to ensure equitable assessments for all students.

Keywords: GenAI assessment, AI literacy, health science education, critical thinking, multimodal learning

1. Introduction

The public release of Large Language Models (LLMs) and Generative AI (GenAI) tools in late 2022 precipitated a paradigm shift in higher education, particularly within the biomedical and health sciences fields. Tools such as ChatGPT demonstrated the capability to pass standardised medical licensing examinations and generate complex scientific discourse, challenging traditional modes of assessment and curriculum delivery (Dave et al., 2023; Kung et al., 2023). While initial institutional responses fluctuated between prohibition and cautious observation, the consensus has increasingly shifted toward integration. Educators are now tasked with defining how these technologies can be leveraged to enhance digital literacy while maintaining academic rigor (Cheng et al., 2025; Masters, 2023). However, the landscape of GenAI is evolving rapidly.

While literature focused predominantly on text-based applications, such as automated essay writing and summarisation (Cheng et al., 2025), the proliferation of multimodal models capable of generating images, audio, and video has expanded the utility of AI in science education (Artsi et al., 2025; Lu et al., 2024). This is particularly relevant in our health science general education subject, where the visualisation of creative ideas is as critical as their textual description.

Despite the potential benefits, the integration of GenAI introduces significant pedagogical and ethical challenges. A primary concern is the phenomenon of hallucinations, where AI models generate plausible but factually incorrect scientific data (Cheng et al., 2025). This necessitates a pedagogical pivot from content generation to content validation, emphasising the critical role of human involvement in the workflow. Furthermore, as GenAI tools bifurcate into free-tier and premium-subscription models, there is a growing concern regarding equity (Francis et al., 2024). The “digital divide” in education may no longer be defined solely by access to hardware, but by access to superior algorithmic reasoning and multimodal capabilities. Current research often captures a snapshot of AI adoption at a single point in time. There is a paucity of longitudinal studies that track how student engagement with these tools evolves over consecutive cohorts as the technology matures. This study addressed that gap by analysing data from a health science general education course across seven cohorts (2022–2023 Summer Term, 2023–24 Semester 1, 2 and Summer Term, and 2024–25 Semester 1, 2 and Summer Term). The research focused on examining how students’ GenAI adoption, perceptions, and challenges evolved over time. The study also explored the implications for assessment redesign in higher education.

To investigate these dynamics, this study posed the following four research questions:

- **Question 1:** How did students’ GenAI usage patterns and purposes change across the seven cohorts (2023–2025)?
- **Question 2:** How did students’ perceptions of GenAI (e.g., trust, authenticity, usefulness) evolve over time?
- **Question 3:** What challenges of GenAI use did students report, and how did these vary by cohort?
- **Question 4:** What are the implications of these shifts for designing assessments in GenAI-rich learning environments?

2. Methodology

2.1. Study Context and Pedagogical Design

This longitudinal study was conducted within a 3-credit, 13-week General Education course classified under the Science, Technology, and Environment cluster at The Hong Kong Polytechnic University. The course was designed for undergraduate students from diverse disciplinary backgrounds, aiming to equip them with essential knowledge about eye health and the skills to debunk common health myths. The pedagogical framework utilised a blended-learning approach. Students engaged with online resources including a custom Massive Open Online Course (MOOC) on visual functions and a 3D anatomy application, alongside synchronous face-to-face mini-lectures, tutorials, and project discussions. To achieve the course learning outcomes, which emphasised the critical evaluation of information sources and the strategy to debunk health myths, the major group assessment strategy was redesigned to mandate the integration of GenAI tools since the 2022-23 Summer Term. Students were tasked with using advanced technologies to research, develop, and present educational videos or slide shows. This assignment required students to act as “human-in-the-loop” editors, using AI-driven research tools, video editing software, and the like to communicate complex scientific ideas while strictly validating the accuracy of the AI-generated content.

2.2. Participants and Data Collection

Data were collected longitudinally across seven consecutive cohorts, spanning 24 months from August 2023 to August 2025 (2022-2023 Summer Term, 2023–24 Semester 1, 2 and Summer Term, and 2024-25 Semester 1, 2 and Summer Term). The total study population consisted of 398 students. A mixed-methods approach was employed to capture both the quantitative and qualitative trends in students’ use of GenAI.

2.3. Data Sources and Instruments

Three primary data sources were utilised to address the research questions:

1. **GenAI Usage Logs and Reflections:** To track usage patterns and workflow shifts, student groups were required to submit detailed “GenAI Usage Logs and reflections” alongside their final projects. These logs documented the specific tools used, purpose of usage, reasons for choosing one tool over others, perceived usefulness, and challenges experienced. A total of 74 usage logs and associated reflective reports were collected and analysed.
2. **Course-End Survey:** A quantitative survey was administered at the conclusion of each cohort to gauge student perceptions of AI efficacy.

A total of 185 students responded to the survey. The instrument utilised a 5-point Likert scale to measure agreement with the following statements:

- “The use of GenAI tools promoted me to evaluate information from different sources and develop critical thinking skills.”
- “The course developed my ability to use GenAI tools effectively.”

3. **Focus Group Interviews:** To obtain deeper insights into the challenges and ethical considerations faced by students, two separate semi-structured focus group interviews were conducted. These sessions took place with students from the 2022-23 Summer Term and the 2023-24 Summer Term cohorts, allowing for a comparative qualitative analysis of student experiences as GenAI technologies matured.

3. Results and Discussion

3.1. Quantitative Perceptions of Skill Development

Quantitative data from the course-end surveys ($n = 185$) indicated a consistently high perception of GenAI's utility throughout the 24-month study period. Student agreement with the statement that GenAI tools promoted information evaluation and critical thinking remained robust across all seven cohorts, ranging from a mean of 4.6 to 4.8 (SD 0.4–0.7). Similarly, students consistently reported that the course effectively developed their ability to use GenAI tools, with a mean of 4.5 to 4.8 (SD 0.4–0.7). Notably, despite the rapid introduction of more complex multimodal models in the later cohorts (2024–2025), the scores did not fluctuate significantly. This stability suggested that the pedagogical scaffolding, specifically the requirement to critique and validate AI outputs, remained effective regardless of the specific generations of tools available.

3.2. Evolution of Usage Patterns from Textual Reliance to Multimodal Orchestration

Analysis of the 74 GenAI Usage Logs and associated reflections with focus group interviews revealed a distinct evolutionary trajectory in how students engaged with these technologies over the two-year period.

- **Phase 1: Textual Exploration (2022-2023 Summer Term – Sem 1 2023-24 Semester 1):** During the first two cohorts, students primarily utilised GenAI for textual tasks. Logs were dominated by ChatGPT and Grammarly. In the August 2023 focus groups, students described a linear workflow of using tools such as Bing to summarise technical articles they found too complex and using ChatGPT to brainstorm ideas. However, scepticism was evident. Students from mathematical disciplines

noted that early GenAI tools were not “beneficial or efficient” for their specific needs, and others expressed uncertainty regarding how the use of these tools would impact their grades.

- **Phase 2: Visual and Auditory Experimentation (2023-24 Semester 2 – 2023-2024 Summer Term):** As the study progressed, usage shifted towards visual and auditory experimentation. Tools such as Canva, Midjourney, CapCut, and Poe appeared frequently. Students moved beyond text generation to creating “digital assets.”
- **Phase 3: Specialised Orchestration (2024-25 Semester 1 – 2024-25 Summer Term):** The final cohorts demonstrated a sophisticated “orchestration” of specialised tools. The student reflections revealed that this phase was characterised by a strategic division of labour within groups. One student handling video editing, another prompting for voiceovers, and another verifying information. Logs showed the use of Perplexity and Scispace for literature verification, DeepSeek for reasoning, and Suno AI for music generation. Students reported this specialisation significantly improved group efficiency compared to traditional workflows.

3.3. Students’ Perceptions and Perceived Challenges on GenAI Use in the Assessment

Qualitative reflections and focus group feedback underscored that while work efficiency increased over the 24 months, the cognitive load shifted from content creation to validation, aesthetic refinement, and navigating assessment anxiety.

Human vs AI Voice Outputs

While GenAI generally enhanced workflows, the 2024 focus groups highlighted a discrepancy in students’ perception on using GenAI to generate video narrations. Students who were less confident in English welcomed AI voiceover. They shared that less time was needed for practising the scripts. However, students with a stronger command of English noted that recording their own voiceovers was often more efficient and emotionally resonant than engineering prompts for AI narration. They described AI-generated voices as “too generic” and lacking the necessary emotion for drama storytelling.

Assessment Anxiety and Rubrics

A persistent theme across both years was anxiety regarding assessment. In 2023, students were “worried how the use of GenAI tools would affect the grade.” By 2024, this evolved into a specific critique of the grading rubrics.

Students suggested that while GenAI use should be a compulsory requirement, it should not be a graded component of the rubric, fearing that using “too little” or “too much” AI would be penalised arbitrarily.

The Equity of GenAI Tool Access (The Paywall Barrier)

The 2024–2025 data highlighted a growing barrier regarding GenAI tool access. Focus group participants explicitly expressed that advanced video and image generation tools required paid credits. Some groups abandoned ideas (e.g., full AI video generation) because it consumed too many credits. This validated concerns regarding the “digital divide”. Groups willing to pay for premium subscriptions had access to superior multimodal capabilities.

Hallucinations and Accuracy

Across all cohorts, the hallucination of facts remained a primary challenge. Students noted that while ChatGPT provided a “fuller picture” on information, it required cross-referencing. This confirms that students successfully adopted the role of “human-in-the-loop” editors.

3.4. Discussion

The longitudinal data suggested that student engagement with GenAI matured from passive consumption of text to active curation of multimedia. The stability of the quantitative results implied that the core competency learned was “AI Literacy”, which is the ability to evaluate, integrate, and validate outputs. The focus group insights revealed a critical divergence in discipline-specific utility. While GenAI was efficient for literature review, it was viewed as less “friendly” for design disciplines due to the lack of precise aesthetic control. Furthermore, the students’ request to decouple GenAI usage from grading rubrics suggested a desire for a “safe sandbox” environment where experimentation was encouraged without the fear of academic penalty. Future curricula must address the “paywall barrier” issue by considering institutional bulk subscriptions or tokens, as suggested by students. Additionally, assessment design should pivot away from grading the quantity of AI use and focus on the quality of the critique and verification process, acknowledging that for some tasks, human intelligence and output remains the gold standard.

4. Implications

This longitudinal study concluded that student engagement with GenAI was dynamic, evolving from passive textual reliance to the sophisticated orchestration of multimodal tools. This trajectory necessitated a theoretical shift in defining AI literacy, moving beyond basic prompt engineering to emphasise

workflow orchestration and the critical role of the “human-in-the-loop” as the ultimate validator of scientific accuracy. Practically, these findings challenge global assessment paradigms. The documented student anxiety regarding grading implies that GenAI integration should probably function as a pedagogical “sandbox”, which can be a mandatory but non-graded prerequisite, rather than a rigid rubric criterion, thereby fostering experimentation without fear of penalty. Furthermore, the emergence of a “paywall barrier” highlights a critical equity issue relevant to the global higher education sector. As premium models offer superior multimodal capabilities, universities should consider centralised provisioning of tools to prevent a deepened digital divide, ensuring that academic success remains a product of intellectual capability rather than financial access to advanced algorithms.

Acknowledgments

The author thanks the School of Optometry and Educational Development Centre at The Hong Kong Polytechnic University for their support in assessment redesign.

Declaration of Use of AI in the Writing Process

The author(s) used Gemini 3 Pro and Perplexity-Sonar-Pro during the preparation of this work to summarise the findings and draft the abstract. After utilising the tool/service, the author(s) thoroughly reviewed and edited the content as necessary and assumed full responsibility for the publication’s content.

References

- Artsi, Y., Sorin, V., Glicksberg, B., Korfiatis, P., Thomas, D. C., Nadkarni, G. N., & Klang, E. (2025). AI-generated videos in medical education: systematic review. *BMJ Open Qual*, 14(4). <https://doi.org/10.1136/bmj-joq-2025-003704>
- Cheng, A., Calhoun, A., & Reedy, G. (2025). Artificial intelligence-assisted academic writing: recommendations for ethical use. *Adv Simul (Lond)*, 10(1), 22. <https://doi.org/10.1186/s41077-025-00350-6>
- Dave, T., Athaluri, S. A., & Singh, S. (2023). ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. *Front Artif Intell*, 6, 1169595. <https://doi.org/10.3389/frai.2023.1169595>
- Francis, N. J., Jones, S., & Smith, D. P. (2024). Generative AI in Higher Education: Balancing Innovation and Integrity. *Br J Biomed Sci*, 81, 14048. <https://doi.org/10.3389/bjbs.2024.14048>

- Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepano, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., Maningo, J., & Tseng, V. (2023). Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digit Health*, 2(2), e0000198. <https://doi.org/10.1371/journal.pdig.0000198>
- Lu, M. Y., Chen, B., Williamson, D. F. K., Chen, R. J., Zhao, M., Chow, A. K., Ikemura, K., Kim, A., Pouli, D., Patel, A., Soliman, A., Chen, C., Ding, T., Wang, J. J., Gerber, G., Liang, I., Le, L. P., Parwani, A. V., Weishaupt, L. L., & Mahmood, F. (2024). A multimodal generative AI copilot for human pathology. *Nature*, 634(8033), 466-473. <https://doi.org/10.1038/s41586-024-07618-3>
- Masters, K. (2023). Ethical use of Artificial Intelligence in Health Professions Education: AMEE Guide No. 158. *Med Teach*, 45(6), 574-584. <https://doi.org/10.1080/0142159X.2023.2186203>