Brought to you by the UGC-funded 8-university project:
Exemplary learning, teaching and assessment practices in UGC-funded institutions in the age of GenAI

Collection of Selected Papers from the International Conference on GenAI and Pedagogical Innovations 2026

International Conference on GenAI and Pedagogical Innovations in Higher Education (GaPI)

GaPI Conference

Dates: 20 – 22 May 2026

Venue: The Hong Kong Polytechnic University

The project is led by The Hong Kong Polytechnic University 香港理工大學
with members from
Lingnan University 嶺南大學
City University of Hong Kong 香港城市大學
Hong Kong Baptist University 香港浸會大學
The Chinese University of Hong Kong 香港中文大學
The Education University of Hong Kong 香港教育大學
The Hong Kong University of Science and Technology 香港科技大學
The University of Hong Kong 香港大學

General Information

•

Copyright © 2026

Copyright of the papers is retained by the authors. No part of this collection may be reproduced by any process without prior written permission of the copyright holders.

ISBN 978-988-19170-0-3 (PDF)

Published in Hong Kong, China; by the project entitled "Exemplary learning, teaching and assessment practices in UGC-funded institutions in the age of GenAI [ELTA G-AI]", funded by the Fund for Innovative Technology-in-Education (FITE) - Inter-institutional Collaborative Activities (IICA) from the Hong Kong University Grants Committee (UGC).

Title: Collection of Selected Papers from the International Conference on GenAI and Pedagogical Innovations 2026 / edited by Julia Chen, Alle Leung, Elza Tsang, Angela Ng, Jacky Chau, Roy Kam, Mitesh Patel, Dawn Lo, Barbara Tam, Leo Chon, Kevinia Cheung, Eunice Tang and Kelly Ho.

Identifiers: ISBN 9789881917003 (PDF) | Handle http://hdl.handle.net/10397/121738

Produced by

Educational Development Centre

The Hong Kong Polytechnic University

Citation (APA Style): Chen, J., Leung, A., Tsang, E., Ng, A., Chau, J., Kam, R., Patel, M., Lo, D., Tam, B., Chon, L., Cheung, K., Tang, E., & Ho, K.(Eds.). (2026). Collection of Selected Papers from the International Conference on GenAI and Pedagogical Innovations 2026. http://hdl.handle.net/10397/121738

Contents

Introduction to the GaPI Conference Collection 1

Welcome Message to the GaPI Conference from the Organising Committee 8

Foreword to the GaPI Conference from the 8-university Project Team 10

GaPI Conference Organising Committee, Local and International Advisors 11

Section One: Papers from Keynote Speakers 12

Section Two: AI Tools and Pedagogy 38

Section Three: Learning Analytics 118

Section Four: Incorporating GenAI in Learning, Teaching and Assessment 126

Section Five: Teacher Professional Development 384

Section Six: Student AI Literacy, AI-Readiness, and Cognitive Development 433

Section Seven: Hybrid (Human-AI) Intelligence 504

Section Eight: Technology, Curriculum Innovation, and Sustainability in Higher Education 530

Section Nine: Smart Learning Environments 580

Contributors 585

Keyword index 607

Introduction to the GaPI Conference Collection

•

1. The birth of this Collection

The idea for the International Conference on GenAI and Pedagogical Innovations in Higher Education 2026 (**GaPI)** emerged three years ago in mid-2023, at a time when generative artificial intelligence (GenAI) was not yet the everyday presence in education that it is now. A group of educational developers from the eight public universities in Hong Kong had already begun to recognise the profound changes that GenAI would bring to higher education. They saw that these technologies would not simply provide new tools for teaching and learning, but would challenge some of the sector's most fundamental assumptions about knowledge production, assessment, authorship, feedback, academic integrity, professional preparation, and the role of human judgement in education.

The educational developers from the eight universities decided to submit an Inter-institutional Collaborative Activity (IICA)[1] proposal for the Hong Kong University Grants Committee (UGC)'s Fund for Innovative Technology-in-Education (FITE) 2023-26. The project aimed to be a shared cross-institutional effort to support universities in responding thoughtfully and proactively to the GenAI era. The pace of technological change was, and remains, extraordinary. Tools that generate text, code, images, music, video, simulations, feedback, and interactive dialogue are now widely accessible to students and teachers, often outside institutional platforms and formal learning environments. This creates new possibilities for personalised learning, formative feedback, multimodal expression, simulation-based practice, and inclusive access to support. At the same time, it raises urgent concerns about over-reliance, cognitive offloading, academic integrity, misinformation, hallucination, privacy, bias, authorship, and equity of access.

Successfully held from 20 to 22 May 2026 at The Hong Kong Polytechnic University, the GaPI Conference and this Collection of selected papers serve as timely platforms for moving beyond immediate reactions to GenAI. Rather than focusing only on whether AI should be allowed or prohibited, the

[1] The FITE-IICA 2023-26 project was led by The Hong Kong Polytechnic University, and comprised City University of Hong Kong, Hong Kong Baptist University, Lingnan University, The Chinese University of Hong Kong, The Education University of Hong Kong, The Hong Kong University of Science and Technology, and The University of Hong Kong.

Conference invited educators, researchers, students, professional developers, and institutional leaders from around the globe to ask deeper pedagogical questions: What kinds of learning matter most when AI can generate fluent outputs? How should assessments be redesigned when the production of text or code is no longer reliable evidence of learning? What forms of AI literacy do students and teachers need? How can GenAI be used to support, rather than weaken, critical thinking, creativity, disciplinary judgement, ethical awareness, and human agency? How can higher education prepare graduates not only to use AI, but to work responsibly, reflectively, and creatively with it?

These questions are not unique to Hong Kong. They are being asked across the world. Yet Hong Kong's higher education sector has been particularly well positioned to contribute to this global conversation. The close collaboration among the eight UGC-funded universities in the geographically compact city of Hong Kong, the active involvement of educational development units in the local higher education sector and beyond, and the willingness of teachers across disciplines to redesign practice have created a rich environment for experimentation, reflection, and knowledge sharing. The GaPI Conference brought together this local momentum with international perspectives, creating a space for rigorous exchange across disciplines, institutions, regions, and countries.

This Collection grew out of that shared endeavour. It stands as a tangible tes-tament to the effort and insights that presenters brought to their presentations at GaPI. It brings together selected papers, posters, and workshops presented at the Conference, documenting how higher education communities are engaging with GenAI not as a passing technological trend, but as a catalyst for rethinking pedagogy. The contributions show that the most promising responses to GenAI are not tool-driven but learning-driven. Across the papers, AI is treated vari-ously as a tutor, simulator, writing partner, feedback provider, grading assistant, research aid, design collaborator, learning analytics resource, and professional development catalyst. Yet the central message is consistent: GenAI must be em-bedded within sound pedagogical design, ethical practice, disciplinary judge-ment, and human-centred education.

2. Overview of the papers in this Collection

The Collection is organised into nine sections that follow the main themes of the GaPI Conference, and reflect the breadth of work presented at the GaPI Conference 2026 and the many ways in which GenAI is transforming higher education.

Section One: Papers from Keynote Speakers opens the Collection by framing some of the most pressing conceptual and institutional challenges facing higher education. The first keynote paper by Jason Lodge examines assessment reform in the age of GenAI, arguing that no simple or definitive solution exists and proposing a pragmatic "Two-Lane" model that distinguishes between secure assessment for assurance of learning and AI-integrated assessment for authentic professional practice. The second keynote paper by Abram Anders focuses on custom AI assistants and strategic cognitive partnership, showing how students can learn to design AI tools that support rather than replace their thinking. The third keynote paper by Allison Littlejohn and her collaborators Francisco Duran del Fierro and Koula Charitonos introduces the idea of "academic meta-work", highlighting the often invisible labour educators undertake as AI reshapes teaching, assessment, ethical practice, professional identity, and student relationships. Together, these keynote papers establish a central theme of the Collection: GenAI requires not merely technical adaptation, but deep rethinking of educational work, expertise, and agency. The editors would like to express our special thanks to Professor Littlejohn for contributing her paper to this Collection despite circumstances having prevented her from delivering her keynote address at the Conference. We are also grateful to Professor Diana Laurillard for kindly stepping in at short notice to deliver a keynote address online. Given the limited timeframe, Professor Laurillard was not invited to contribute a written paper to this Collection.

Section Two: AI Tools and Pedagogy showcases practical and discipline-specific uses of AI tools in teaching and learning. Several papers demonstrate how AI can support communication, language learning, presentation skills, and disciplinary creativity. The NinjOrAItor project illustrates a hybrid intelligence model for English oral presentation training, combining AI feedback with self- and peer assessment. A study on English connected speech processes shows how AI-enhanced pronunciation training can support Chinese EFL learners. Other contributions explore custom AI agents for design entrepreneurship, AI-mediated copywriting for EFL learners, and a "Learn by Design" philosophy that uses AI to foster metacognitive reflection, critical interrogation, and empathetic simulation. Additional papers examine video-based learning for technological social responsibility, virtual exchange in the AI era, AI avatars for flipped classroom materials, workflow automation using n8n, AI agents as teaching sidekicks, and AI-based negotiation simulators. Collectively, this section demonstrates that effective AI pedagogy requires carefully designed interaction among learners, teachers, tools, and disciplinary purposes.

Section Three: Learning Analytics focuses on how data from AI-mediated learning environments can support deeper understanding of learning processes. The paper on classroom AI-avatar dialogue uses natural language processing to examine semantic alignment, lexical overlap, and interaction patterns between students and AI avatars. It shows that AI may sound fluent while still missing students' intended meanings, and it offers educators practical ways to interpret AI-supported classroom dialogue beyond surface-level fluency. This work points to the growing importance of learning analytics in evaluating the quality of human-AI interaction, not merely its frequency or novelty.

Section Four: Incorporating GenAI in Learning, Teaching and Assessment is the largest section of the Collection and reflects the centrality of assessment and curriculum redesign in the GenAI era. The papers cover a wide range of disciplines, including language education, nursing, anatomy, chemistry, science, design, occupational therapy, journalism, finance, business, geomatics, programming, mathematics, and postgraduate interdisciplinary education. Several studies examine GenAI-supported language learning, including IELTS speaking practice, academic writing, EFL copywriting, peer and AI collaboration in writing, and process-based writing assessment. Others explore GenAI-supported clinical and professional simulations, such as virtual patients in nursing and occupational therapy, AI-powered clinical role-play, and business negotiation simulators. A number of papers focus on assessment integrity and grading, including AI-assisted chemistry grading, automated essay scoring, transparent rubrics, SmartFeedback for data analysis reports, human-AI grading comparisons, and authentic assessment design. This section also includes studies of students' perceptions of GenAI use in coursework, GenAI in science and anatomy education, AI-supported equity valuation, journalism writing workflows, collaborative problem-solving with AI teammates, and interdisciplinary postgraduate learning. Across this section, the emphasis is not on replacing human judgement, but on designing assessment systems in which AI use is visible, bounded, meaningful, and educationally productive.

Section Five: Teacher Professional Development turns attention to the educators who must lead pedagogical change. Papers in this section examine professional development programmes, teacher AI literacy, STEAM competency, new academics' adoption of GenAI, and discipline-sensitive faculty development. One paper reports on a large-scale professional development programme at Temasek Polytechnic that uses a human-in-the-loop approach to help educators reclaim agency in the age of AI. Another develops and pilots instruments for evaluating university teachers' AI literacy and STEAM competency, identifying gaps between motivation and practical confidence. A study of new academics at PolyU

highlights both adoption patterns and barriers, including lack of knowledge, concerns about accuracy, cost, and impact on student learning. Another paper compares STEM and non-STEM faculty perspectives on GenAI, showing how disciplinary epistemologies shape attitudes, assessment design, and professional development needs. The section also includes an institutional-level account of reintroducing flipped learning supported by GenAI, and the development of an AI Interaction Knowledge scale for higher education teachers. These contributions make clear that AI transformation depends on teacher learning, institutional support, and communities of practice, not simply on access to tools.

Section Six: Student AI Literacy, AI-Readiness, and Cognitive Development explores how students develop the knowledge, dispositions, and cognitive capacities needed to engage with AI responsibly and effectively. One longitudinal study examines reciprocal relationships between positive youth development and AI literacy among Filipino adolescents, showing how moral, relational, and digital competencies may co-develop. Another study investigates AI interaction, AI cognition, and creative thinking among university students in China, demonstrating the role of positive AI cognition in supporting creativity. Other papers explore programmable robots for authentic AI literacy learning, the OpenImpactLab platform for project-based AI readiness and social impact, critical reading in the AI era, GenAI use in media design education, AI dependency and AI literacy among Hong Kong university students, and a 5P framework for scaffolding metacognitive engagement with GenAI. A recurring theme is that AI literacy cannot be reduced to technical skill. It also includes ethical awareness, critical evaluation, metacognition, self-regulation, creativity, resilience, and the ability to avoid unproductive dependence on AI.

Section Seven: Hybrid Human-AI Intelligence examines how humans and AI can work together in educationally meaningful ways. The papers in this section consider emerging topics such as co-agency, collaboration, and the emotional dimensions of human-AI interaction. A study on AI-assisted writing feedback shows that stronger and weaker learners engage with AI feedback differently, underscoring the need for differentiated hybrid pedagogy. Another conceptual paper introduces the AI-Augmented Creative Resilience Framework, addressing Creative Displacement Anxiety and showing how learners may move from fear of replacement to adaptive co-creation with AI. A workshop on GenAI and AI agents further explores how educators can use AI tools to augment teaching, reduce workload, and prepare students for AI-mediated workplaces. This section highlights a crucial insight: hybrid intelligence is not simply about combining human and machine outputs; it is about designing relationships in which human judgement, creativity, emotion, and responsibility remain central.

Section Eight: Technology, Curriculum Innovation, and Sustainability in Higher Education extends the discussion to immersive learning, curriculum reform, and institutional change. Papers in this section examine holographic projection for crystal structure learning, institutional GenAI-in-assessment reform through Kotter's change model, AI-powered metaverse environments for food science education, smart teaching supported by knowledge graphs and large models, GenAI-enabled semi-flipped classes and project-based design thinking, and technology-enhanced pre-service learning for service-learning in nursing homes. These contributions show that GenAI is part of a broader ecology of educational technologies, including metaverse platforms, virtual reality, holographic projection, AI agents, and smart learning systems. They also demonstrate that successful innovation requires more than experimentation at the course level; it requires institutional strategies, leadership, professional development, and alignment with broader educational goals.

Section Nine: Smart Learning Environments concludes the Collection by looking toward future learning spaces. The Math City workshop introduces an inquiry-based metaverse platform for mathematics education, integrating gamification, digital avatars, authentic assessment, and GenAI tools. It exemplifies how smart learning environments can make abstract concepts more visible, interactive, and engaging, while also fostering collaboration, inquiry, and 21st-century skills. This final section reminds us that the future of AI-enhanced education will not be confined to chatbots or text generation, but will increasingly involve immersive, adaptive, and interactive learning spaces.

Taken together, the papers in this Collection offer a rich and timely snapshot of how higher education is responding to GenAI as of the first half of 2026. They provide empirical studies, conceptual frameworks, design cases, professional development models, workshops, and institutional reflections. They also span a wide range of contexts and disciplines, from chemistry, engineering, health sciences, and geomatics to language education, design, journalism, business, mathematics, and teacher development. This diversity is one of GaPI Collection's greatest strengths. It shows that there is no single 'GenAI pedagogy'; rather, there are many situated, discipline-sensitive, ethically informed ways of designing learning with AI.

Several cross-cutting themes emerge. First, assessment must shift from product-only evaluation toward process, judgement, reflection, authenticity, and evidence of learning. Second, AI literacy must include critical, ethical, emotional, and metacognitive dimensions, not merely operational skill. Third, educators need sustained professional development and institutional support if they are

to integrate AI meaningfully. Fourth, students require guidance to use AI as a partner in learning without surrendering agency, creativity, or responsibility. Finally, human judgement remains indispensable. Across the Collection, the most compelling examples of GenAI use are those in which AI amplifies human learning, supports collaboration, makes thinking visible, and creates opportunities for reflection, rather than replacing the intellectual and relational work of education.

Encouragingly, this conversation has already extended beyond this Collection. At the time of publication, a paper emerging from the Teacher Panel at the GaPI Conference has been published in an academic journal. That paper further develops the panel's discussion at GaPI by proposing a role-based framework for future faculty development. This development reinforces this Collection's emphasis on sustained professional support for educators; and the paper can be found at https://journal.hep.com.cn/fde/EN/10.1007/s44366-027-0117-3.

With these ongoing developments in mind, we hope this GaPI 2026 Collection will serve as a resource for educators, re-searchers, academic developers, institutional leaders, and policymakers seeking to respond to GenAI with both imagination and care. The papers gathered here do not offer final answers. Instead, they offer tested practices, emerging frameworks, critical questions, and shared commitments. In doing so, they contribute to a larger international effort to shape higher education that is AI-informed, ethically grounded, pedagogically rigorous, and deeply human-centred.

Julia CHEN and Jacky CHAU
The Hong Kong Polytechnic University

Welcome Message to the GaPI Conference from the Organising Committee

•

A warm welcome to GaPI: International Conference on GenAI and Pedagogical Innovations in Higher Education! It is our distinct honour and pleasure to welcome you to the Conference, and to the vibrant city of Hong Kong for those of you who can come in person.

As we stand at the threshold of a new era in education, the GaPI Conference serves as a vital platform for dialogue, innovation, and reflection. The rapid evolution of AI presents us with unprecedented opportunities and complex challenges, and there is no better time for the global academic community to gather and discuss the future of teaching and learning.

We are truly overwhelmed by the global interest in this event. We received over 180 submissions for paper presentations, interactive workshops and poster presentations at the Conference. After a double-blind review process, and many went on to a triple review, over 100 of the submissions have been selected for presentation at the Conference and around 70 are included in the Conference Collection. We are very pleased that our presenters at GaPI come from more than 20 countries and regions, including Australia; Canada; Chinese Mainland; Croatia; Greece; Hong Kong, China; India; Japan; Macau, China; Morocco; New Zealand; Norway; Philippines; Russia; Singapore; Sweden; Taiwan, China; Thailand; United Kingdom; United States and Vietnam. This diversity of perspectives ensures that our discussions will be rich, inclusive, and globally relevant. Under the overarching theme of GenAI and pedagogical innovation, our programme is structured around 10 critical sub-themes that reflect the current landscape of the field. For details, please refer to our conference website at https://events.polyu.edu.hk/gapi2026.

Behind the organisation of the Conference is the staunch support of an eight-university project team whose project has been funded by the Hong Kong University Grants Committee (UGC)'s Fund for Innovative Technology-in-Education (FITE) Fund. We are grateful to the UGC for the grant that has allowed the eight universities to co-explore good practices in embracing AI in education. The eight universities include (according to the order on the UGC homepage): City University of Hong Kong, Hong Kong Baptist University, Lingnan University, The Chinese University of Hong Kong, The Education

University of Hong Kong, The Hong Kong Polytechnic University, The Hong Kong University of Science and Technology and The University of Hong Kong.

Our deepest gratitude goes to our keynote speakers and plenary panel speakers, whose visionary insights provide the foundation for our sessions. We also wish to thank our pre-conference workshop facilitators and parallel session speakers for sharing their specialised expertise and practical strategies. The vibrancy of the GaPI conference is further enhanced by our poster presenters and interactive workshop presenters, whose creative approaches and hands-on demonstrations offer a glimpse into the cutting edge of pedagogical experimentation.

Beyond the academic sessions, we believe in the power of cultural exchange to foster community. We would like to extend a special ‘thank you’ to our two cultural event performers for their breathtaking contribuons, and to the Culture Promotion and Events Office for their invaluable assistance in connecting us with the instrumental quintet and for the use of the beautiful venue: the Jockey Club Auditorium at The Hong Kong Polytechnic University. Our sincere thanks also go to our sponsors. We encourage all attendees to visit the sponsor exhibitions to explore the latest tools and services supportng the educational sector.

Finally, a conference of this scale would be impossible without the tireless work of our project staff, especially our capable and thorough project associate, Miss Alle Leung; as well as our professional, technical, administrative colleagues and student volunteers who have worked meticulously behind the scenes to ensure every detail is in place.

As you engage in the many sessions at GaPI, we hope you find inspiration, forge new partnerships, and leave with a renewed sense of purpose. Thank you for being part of this journey to redefine the intersection of technology and humanity in the classroom.

Welcome to the GaPI Conference!

The Organising Committee of the GaPI Conference

Foreword from the Project Team

On behalf of the Project Team, it is our great honor to welcome you to the International Conference on Gen AI and Pedagogical Innovations in Higher Education. This conference represents the culmination of collaboration among eight universities in Hong Kong, united by a shared vision to respond to the rapid development of generative AI and its profound impact on higher education.

Throughout this journey, our team has worked to transform pedagogies, curriculum, assessment, and student development to meet students' evolving learning needs, while thoughtfully incorporating and leveraging AI. Our goal has always been to make learning meaningful, effective, and future-ready. The path has been fascinating, urgent, and challenging, given the rapid pace of technological change and the diverse views among professionals regarding AI in education.

We believe this conference is not only a platform for academic exchange but also an opportunity to foster academia–industry collaboration, ensuring authentic learning experiences that prepare students for real-world challenges. Two major themes under the project—promoting technological social responsibility and safeguarding academic integrity—have guided our work and remain central to the discussions we hope to inspire here.

This conference is the fruit of hard work, continuous exchange, and learning through trial and error. We are delighted to share insights with participants worldwide. We hope all will treasure the dialogue and enjoy the fruitful exchange this conference promises.

— The Project Team

Organising Committee

Chair: Julia Chen

Event Implementation Lead: Alle Leung

Paper Review Sub-committee: Barbara Tam, Dawn Lo, Roy Kam, Mitesh Patel, Leo Chon, Kevinia Cheung, Alle Leung, Julia Chen

Collection Sub-committee: Leo Chon, Alle Leung, Julia Chen

PR & Liaison Sub-committee: Barbara Tam, Alle Leung

Programme Sub-committee: Barbara Tam, Dawn Lo, Leo Chon, Alle Leung, Kelly Ho, Julia Chen

Technical Sub-committee: Raymond Ho, Mag Xu, Ethan Hung, Birley Chan (Graphic Design), Alle Leung (Website Design)

Logistics Sub-committee: CK Chu, Crissy Chan, Winnie Tam

Local Advisors

City University of Hong Kong: Prof. Kenneth Lo

Hong Kong Baptist University: Dr Theresa Kwong, Dr Louis Cai

Lingnan University: Prof. Frankie Lam

The Chinese University of Hong Kong: Prof. Cecilia Chun, Mr Andy Wong

The Education University of Hong Kong: Dr Gary Cheng, Ms Crystal Luo

The Hong Kong University of Science and Technology: Dr Sean McMinn

The University of Hong Kong: Prof. Cecilia Chan

International Advisors

Prof. Simon Bates, University of British Columbia, Canada

Dr Claire Gordon, London School of Economics and Political Science, UK

Prof. Marcelo Milrad, Linnaeus University, Sweden

Prof. Sun Hua, Peking University, China

Prof. Danping Wang, University of Auckland, New Zealand

1 Papers from Keynote Speakers

Beyond the Gaze of the Guru: Why Nobody Has the Answers to Assessment Reform

Jason M. LODGE [1]
[1] Professor of Educational Psychology and Director of the Learning, Instruction, and Technology Lab in the School of Education, The University of Queensland

Abstract

The rapid proliferation of generative AI has exposed a cavernous gap between educational research and the functional needs of policy and practice. While self-proclaimed gurus claim definitive solutions, the reality is a messy landscape of complex compromises. This paper argues that higher education must move past the paralysis of 'admiring the problem' and embrace a heuristic approach to assessment reform in the short term. By redefining expertise as the ability to navigate deep domain knowledge while being entangled with technology, we can develop more resilient frameworks over longer time horizons. Using the 'Two-Lane' assessment model as a basis, I propose that, in the short term, we prioritise the assurance of learning over the impossible quest for a perfect, AI-proof curriculum and assessment. Beyond this, the challenge for all stakeholders is to find ways to close the gap between research, policy, and practice. This effort will require deep collaboration between stakeholder groups.

Keywords: Assessment reform, Generative AI, Higher education, Expertise

1. Introduction

The current state of assessment across educational contexts globally is complicated and messy. There are currently no definitive answers on how to adapt educational assessment for the age of AI. As described by Corbin et al. (2025), this reform is a 'wicked problem' that requires complex solutions. However, this has not stopped many purported solutions from emerging that purport to provide the answer to the problem of 'cheating' with AI.

For the past eighteen months, I have been looking at the issues of AI and assessment across sectors and countries, and the most striking takeaway is this: nobody has the answers. Despite the deluge of self-appointed social media gurus and educational consultants claiming they have cracked the code, we are all operating in an environment of deep uncertainty. This uncertainty is exacerbated by the gap between research and policy, which is not just vast;

it is expanding rapidly. This has been a consistent problem with the impact of technology on education (Thompson & Lodge, 2020). Changes in policy and practice lag far behind developments in research and theory.

When it comes to adapting education to AI, this gap is a significant problem. AI use is already widespread, yet its applications and ethics remain highly variable (OECD, 2026). Educational institutions and sectors need to make decisions now and are doing so. These policy decisions are often binary, messy decisions about purchasing, resourcing, and curriculum. The guardrails and frameworks proposed by many are insufficient to support these decisions, as they often provide a level of certainty about aspects of AI use that does not reflect the state of the research literature. Given the early stages of the research endeavour, the findings to date are not sufficiently translated into practice in real educational environments.

This level of uncertainty means the confidence placed in the solutions being offered to educators and educational institutions is unwarranted, to say the least. Consider the social media analogy: educational systems do not need a nuanced 50-page white paper on the sociological benefits of algorithmic connection or an oversimplified framework to know that continuing to allow 11-year-olds access to social media is potentially harmful and should be controlled. It is a weight-of-evidence decision. Assessment policy in light of AI requires the same level of pragmatic decisiveness. Given the current uncertainty about AI and its ongoing development, this is a significant dilemma. Accepting simplified solutions (such as 'traffic light' frameworks) is not an appropriate path forward. The question is then: how to continue the process of adapting in the face of such high levels of uncertainty?

2. Methodology

This work synthesises policy analysis and cross-sectoral observations across global higher education systems. It uses a heuristic framework to evaluate how we define and credential expertise in environments where human intelligence is increasingly distributed across and entangled with machine agents. The synthesis provided here, triangulated with the emerging research literature, enables the projection of possible approaches to closing the gap between research, policy, and practice in real educational environments.

3. Results and Discussion

The results of this cross-sector analysis suggest that current assessment regimes in higher education are built on a shaky foundation. Education systems have been confusing knowing students have done the assessment with knowing

students have done the learning (Winstone et al., 2026). Prior to the emergence of generative AI, contract cheating (see Eaton et al., 2022) highlighted the growing problems with assessment. AI has simply accelerated this corruption of traditional assessment practices. In this context, students' inappropriate use of AI to complete assessment tasks is understandable; these tasks have drifted too far from the learning processes and have become increasingly performative.

Concerningly, generative AI enables performance in the complete absence of actual learning (Yan et al., 2025). The nature of the human-AI relationship creates further challenges in understanding appropriate forms of assessment. AI simulations of agency look like human output, but they are not. Categorising AI as an agent is often nonsensical, but we must do it anyway to make choices. Emerging data (e.g., Lan et al., 2026) suggest that many students, and more broadly people, treat generative AI tools as collaborators, even though they do not appear to have legitimate agency. Many unanswered questions remain about the relationships between students and AI as a distributed cognitive system. While theories and metaphors abound, the actual mechanisms of these emerging relationships remain ambiguous.

The acute threat to academic integrity created by the emergence of generative AI necessitates a response, and educational institutions and systems are responding with vigour (e.g., Lodge et al., 2025). However, the level of uncertainty surrounding human-AI interaction makes it difficult to create precise, nuanced approaches. The acuteness of the problem necessitates unambiguous frameworks that sacrifice much of the nuance needed to adapt to the threat of inappropriate AI use to academic integrity.

One of the most prominent models is The University of Sydney's 'Two-Lane' model (Bridgeman et al., 2024):

- **Lane 1:** Assessment that is secure, often invigilated, and designed to assure that the individual has actually internalised the learning.
- **Lane 2:** Assessment that is integrated with AI, reflecting the reality of a world where expertise is entangled with technology.

Rather than being overly prescriptive, the Two-Lane model provides a straightforward heuristic for decision-making. To make sense of what belongs in each lane, it is critical to examine the knowledge and skills essential in the age of AI. Assessment does not exist in a vacuum; it is closely tied to issues related to curriculum, teaching, and, perhaps most importantly, to questions about what the assessment purports to assess.

One of the critical roles of higher education is to support students to become experts in their discipline or profession. The role of an expert in this new reality we find ourselves requires:

- Deep, domain-specific knowledge schemas;
- An ability to use that knowledge to make high-stakes decisions;
- The capacity to navigate new knowledge within their domain; and
- An ability to work within a distributed intelligence network without losing their own agency.

So, questions about the nature of assessment tasks become more complicated when seen in the light of these capacities. Does the task require knowledge acquisition? Can it be assessed securely? Is it essential for the domain? These issues do not lend themselves to simple solutions, and this is the longer-term work required to build from here. For example, exams are not the solution; they are a compromise, one of many compromises that educators, institutions, and education sectors are making to maintain the integrity of degrees.

4. Implications

The reality of higher education means that practitioners and policymakers need to make clear decisions regarding student learning and assessment. Reiterating that the problem of assessment reform is deeply complex does not help the educator designing a syllabus for 2,000 students on Monday. Oversimplified solutions are also not adequate. While the complexity of emerging human-machine systems becomes untangled over time, educators and educational institutions need to stop waiting for the research community to develop a solution and start building heuristics that enable action, such as the Two-Lane model.

Ultimately, the job of higher education is to produce experts and citizens who can navigate a world of synthetic language without losing their own agency within complex human-machine systems. The path forward is to accept that it will be some time before we understand the human-machine entanglement. Developing this understanding and creating solutions will require a collaborative effort across stakeholder groups, moving towards more nuanced and fit-for-purpose assessment regimes for the age of AI. That process is underway in some contexts, for example through the Castlereagh Statement (Castlereagh Summit, 2026) in Australia.

In the meantime, the Two-Lane model is the best available heuristic decision-making tool.

Ethical Declaration Statement

As the project relies on policy analysis and cross-sectoral observations, human research ethical approval was not required.

Statement

The author(s) used Gemini (V. 3) during the preparation of this work to edit the paper for clarity. After using the tool/service, the author(s) thoroughly reviewed and edited the content as necessary and assumed full responsibility for the publication's content.

References

Bridgeman, A., Liu, D., & Weeks, R. (2024). *Program level assessment design and the two-lane approach*. University of Sydney. https://educational-innovation.sydney.edu.au/teaching@sydney/program-level-assessment-two-lane/

Castlereagh Summit (2026). *The Castlereagh Statement: A cross-sector call to action on Australian education and training in the age of AI*. Sydney, Australia. https://castlereagh.ai/

Corbin, T., Bearman, M., Boud, D., & Dawson, P. (2025). The wicked problem of AI and assessment. *Assessment & Evaluation in Higher Education*, 1–17. https://doi.org/10.1080/02602938.2025.2553340

Eaton, S. E., Curtis, G. J., Stoesz, B. M., Clare, J., Rundle, K., & Seeland, J. (Eds.). (2022). *Contract cheating in higher education: Global perspectives on theory, practice, and policy*. Springer Nature.

Lan, Y., Liu, S., & Chen, H. (2026). The effects of responsiveness, perceived warmth, and anthropomorphism on university students' use of conversational AI for learning support: a chain mediation analysis based on S-O-R framework. *Frontiers in Psychology*, 17, 1769964. https://doi.org/10.3389/fpsyg.2026.1769964

Lodge, J. M., Bearman, M., Dawson, P., Gniel, H., Harper, R., Liu, D., McLean, J., Ucnik, L., & Associates (2025). *Enacting assessment reform in a time of artificial intelligence*. Tertiary Education Quality and Standards Agency, Australian Government.

OECD. (2026). *OECD Digital Education Outlook 2026: AI and the future of education*. https://www.oecd.org/en/publications/oecd-digital-education-outlook-2026_062a7394-en.html

Thompson, K., & Lodge, J. M. (2020). 2020 vision: What happens next in education technology research in Australia. *Australasian Journal of Educational Technology*, 36(4), 1–8. https://doi.org/10.14742/ajet.6593

Winstone, N. E., Gravett, K., & Elkington, S. (2026). Black Box Assessment: rethinking integrity and learning for a time of Generative AI. *Assessment & Evaluation in Higher Education*, 1–18. https://doi.org/10.1080/02602938.2026.2661365

Yan, L., Greiff, S., Lodge, J. M., & Gašević, D. (2025). Distinguishing performance gains from learning when using generative AI. *Nature Reviews Psychology*, 4(7), 435–436. https://doi.org/10.1038/s44159-025-00467-5

Designing Custom AI Assistants to Practice Strategic Cognitive Partnership

Abram ANDERS [1]
[1] Jonathan Wickert Professor of Innovation and Associate Director of the Student Innovation Center, Iowa State University, United States

Abstract

Recent research identifies a U-shaped relationship between AI use and learning: minimal or piecemeal delegation degrades performance, while strategic cognitive offloading paired with cognitive vigilance enables transformative learning (Wang & Zhang, 2026). What pedagogical approaches can move students into the productive zone? This paper argues that designing custom AI assistants—through prompts and curated knowledge bases that students iterate against their own evaluative criteria—is a *writing-to-engage* activity that promotes both cognitive pathways simultaneously. Structured around a Plan–Iterate–Evaluate cycle, the assignment compresses strategic offloading into an explicitly scaffolded process: students externalise disciplinary knowledge, plan the interaction structure, refine the assistant against self-set standards, and use the completed artifact for tasks they care about. Drawing on threshold concepts research and prior empirical work on AI literacy development, the paper presents the assignment design and discusses implications for promoting AI fluency across the curriculum.

Keywords: generative AI literacy, cognitive offloading, writing pedagogy, threshold concepts, human–AI collaboration

Introduction

Recent research finds that the same generative AI tools can erode skills, leave students worse off than if no tool were used at all, or produce gains in performance and learning. Wang and Zhang (2026) recently mapped this divergence, finding a U-shaped relationship between cognitive offloading and transformative learning across 912 students in three cultural contexts. Partial, uncritical delegation—what might be called the "muddy middle" of AI use—produced the worst outcomes; deeper offloading combined with cognitive vigilance produced the strongest results. The two pathways activate together in pedagogical partnerships, enabling deeper offloading to free cognitive resources for higher-order reflection rather than displacing it.

This finding sharpens a question that AI integration in higher education has not yet answered: what assignments teach students to occupy the productive zone? Students readily learn that they should be responsible users of AI. The harder work is developmental: reorienting how they conceive of AI itself, from a tool that completes tasks to a partner whose work depends on what the student supplies.

Threshold concepts research has documented this shift in writing with AI contexts as three interlinked moves: treating engagement as experimental rather than transactional; using situated expertise to direct and evaluate outputs; and maintaining agency through deliberate process choices (Anders & Dux Speltz, forthcoming). This paper argues that designing custom AI assistants, through prompts and curated knowledge bases, is a *writing-to-engage* activity (Palmquist, 2020) ideally suited to promoting both vigilance and strategic cognitive offloading simultaneously. Building such an assistant requires students to externalise disciplinary knowledge, plan how it should be used, and refine the result against criteria they have set. The artifact then operationalises that prior thought, allowing students to offload routine cognitive labour on later tasks while retaining authorship over the choices that matter.

The sections that follow describe the assignment's design, the cognitive moves it produces, and its implications for course design.

Methodology

This paper presents a pedagogical design grounded in recent empirical work on AI literacy and learning. Three findings from prior research orient the design. First, Wang and Zhang's (2026) dual-pathway model shows that effective human–AI partnerships activate cognitive vigilance and cognitive offloading together; the productive zone is not less offloading but *strategic* offloading paired with critical evaluation. Second, three threshold concepts identified in students' reflections on AI-assisted writing (experimentation, expertise, and agency) name the developmental shifts that productive collaboration requires (Anders & Dux Speltz, forthcoming). Third, the Plan–Iterate–Evaluate cycle, derived from self-regulated learning theory and applied to AI literacy development, gives operational form to those shifts: planning activates situated knowledge and sets evaluative criteria; iteration tests prompts and outputs against those criteria; evaluation reflects on both products and processes to inform future use (Anders & Dux Speltz, 2025).

The pedagogical design synthesises these findings into a three-phase assignment, framed as a *writing-to-engage* task (Palmquist, 2020), in which

students build a custom AI assistant for a learning or communication problem they care about. Prior pedagogical work established the value of teaching explicit prompting strategies for self-directed AI use (Mollick & Mollick, 2023); the present design extends that work by treating prompting and knowledge curation as the interrelated halves of a single planning practice.

In the **Plan** phase, students assemble the assistant's knowledge base and write its instructions. The knowledge base is a structured set of domain content, conventions, examples, and constraints that the assistant will draw on. The instructions specify the assistant's role, the tasks it will support, the process it should follow, and the constraints it must observe. Writing both forces students to make explicit what they would otherwise leave tacit: what counts as essential disciplinary knowledge for the task, the standards by which good output should be judged, and the division of labour between student and assistant.

In the **Iterate** phase, students test the assistant against tasks drawn from their own work, refine the knowledge base and instructions in response to outputs that fall short, and document what they learn from each cycle. The phase makes the experimental cycle explicit: each iteration reveals a misalignment between assistant behaviour and student standards, which leads to a targeted revision.

In the **Evaluate** phase, students use the completed assistant for a learning or communication task they care about, assess outputs against their rhetorical goals, and reflect on what the build experience taught them. The transition from build to use is the move where the assistant becomes a tool for thinking and communicating rather than a target of construction; evaluation now encompasses both how well the work meets the student's purposes and what the build reveals about effective collaboration with AI.

The course context is an undergraduate Artificial Intelligence and Writing course; empirical patterns from prior iterations of related assignments (Anders & Dux Speltz, 2025) provide the illustrative student practices discussed in the next section.

Results and Discussion

The assignment's three phases (Plan, Iterate, Evaluate) produce the developmental shifts that Wang and Zhang (2026) describe as the productive zone of the U-curve. Each phase activates the three threshold concepts of writing with AI in different proportions, and each contributes to the joint development of cognitive vigilance and the capacity for productive cognitive

offloading (Anders & Dux Speltz, forthcoming).

Plan: Students Articulate Situated Expertise and Rhetorical Choices

Compiling a knowledge base requires students to decide what their discipline considers essential and what constitutes success in a given domain; writing the instructions requires them to specify the assistant's role and the division of labour with the assistant. Anders and Dux Speltz (2025) document the same pattern in their planning-phase analysis: successful students "activate domain knowledge" before engaging the tool, drawing on disciplinary expertise to identify both the application and the criteria for judging it. Knowles (2024) names this work as *rhetorical load sharing*: writers retain the rhetorical labour that defines authorship while delegating routine execution. The Plan phase activates expertise most strongly, while agency emerges in choices about delegation and experimentation enters as anticipation that prompts will need testing.

Iterate: Students Enact Experimentation as a Method

Testing the assistant, refining the knowledge base, and revising instructions in response to outputs that fall short enacts the experimental threshold concept directly (Anders & Dux Speltz, forthcoming). Outputs that look fluent but miss the disciplinary mark become visible as misalignments between the assistant's behaviour and the student's own standards. This work develops what Bearman and Ajjawi (2023) call *evaluative judgment*: the capacity to assess output quality against disciplinary standards a student has articulated rather than against an external rubric. Anders and Dux Speltz (2025) found that students described iterative development as the most valuable practice they developed in similar contexts. The Iterate phase activates experimentation directly, while expertise sustains evaluation of outputs and agency continues to refine through decisions about what to delegate.

Evaluate: Students Use the Assistant and Reflect on the Work

When students apply the completed assistant to a task they care about, the assignment shifts from build-target to working tool. The cognitive cost of specifying conventions and criteria has been invested in advance, so the student can offload routine labour while retaining authorship of the choices that matter. This is the configuration Wang and Zhang's (2026) U-curve names as productive: high offloading paired with retained vigilance. Reflection on outputs and on the assistant design itself closes the cycle, generating insights that inform future work. The Evaluate phase activates agency in its fullest form, because the work is now the student's own; experimentation continues as questions about what to refine; expertise is sharpened by what use reveals.

Table 1: Phases of the custom AI assistant assignment and the threshold concepts each activates

Phase	Student Activity	Threshold Concepts Activated
Plan	Curate a knowledge base; write instructions specifying the assistant's role, scope, and standards.	Expertise: Articulating disciplinary knowledge and quality criteria. Agency: Deciding what to delegate and what to keep. Experimentation: Anticipating that prompts will require testing.
Iterate	Test outputs against tasks; refine knowledge and instructions in response to misalignments.	Experimentation: Plan-and-test cycles enacted directly. Expertise: Applied to evaluation of outputs. Agency: Adjusting delegation in light of what works.
Evaluate	Use the completed assistant on learning or communication tasks; reflect on outputs and on the build process.	Agency: Retaining authorship of the work the assistant supports. Experimentation: Outputs generate questions for future builds. Expertise: Standards refined in light of use.

Implementation has become more flexible since earlier versions of this assignment. The simplest form is a prompt and pasted knowledge base used in a chat interface, which any student can produce in a single class session. A mid-tier option is a custom chatbot built in a no-code platform such as a ChatGPT GPT or a Microsoft Copilot agent, which gives the assistant a persistent identity and a dedicated interface. Another approach is an on-demand reusable skill, such as a Claude skill or a Gemini gem, which can be called within other tasks rather than visited as a separate tool. The pedagogical principle holds across these implementations: each requires the same Plan, Iterate, and Evaluate moves. What changes across implementations is portability and reach.

Implications

The argument has consequences for course design, theory, and programme-level work in higher education.

For Course Design: The assignment functions as an anchor for AI fluency across the curriculum. Many AI literacy interventions teach skills in isolation: prompt engineering as one module, evaluation as another, ethical use as a third. Building a custom AI assistant integrates these moves in a single sustained activity. The assignment's content is whatever knowledge the discipline already

considers essential for the task, which makes the design portable to disciplinary instructors without specialist training.

For Theory: The design illustrates a productive convergence. Wang and Zhang's (2026) dual-pathway finding, the threshold concepts identified in writing with AI contexts (Anders & Dux Speltz, forthcoming), and Knowles's (2024) account of rhetorical load sharing all describe the same developmental territory from different vantage points. Each names a version of the same shift: from passive offloading toward an integrated practice in which delegation and critical oversight depend on each other. The custom-assistant assignment offers a single pedagogical move that produces all three.

For Programme-Level Work: The assignment supports an integration model that does not require AI to become a separate subject. Because the assistant is built around disciplinary knowledge and tasks, the work belongs in disciplinary courses, with the writing programme or AI literacy initiative providing scaffolding and faculty development. This pattern matches what Writing Across the Curriculum programmes have done well for decades and offers a near-term path for institutions that want to develop AI fluency without building it from scratch.

Two questions remain open. Does the experience of building one assistant change how students approach AI tools they have not built? Do the cognitive moves the build produces persist when students later work without the artifact? These are areas that it will be valuable for future work to address across course, programme, and disciplinary contexts.

Ethical Declaration Statement

This paper presents a pedagogical design rationale and does not report new primary research with human subjects. Empirical findings cited from prior published work were approved under the institutional review board protocols of the original studies.

Declaration of Use of AI in the Writing Process

The author used Claude (Anthropic) during the preparation of this work to support review of source material, drafting, and editorial revision. After using the tool, the author thoroughly reviewed and edited the content as necessary and assumes full responsibility for the study content.

References

Anders, A. D., & Dux Speltz, E. (2025). Developing generative AI literacies through self-regulated learning: A human-centred approach. *Computers and Education: Artificial Intelligence, 9*, 100482. https://doi.org/10.1016/j.caeai.2025.100482

Anders, A. D., & Dux Speltz, E. (forthcoming). Threshold concepts for writing with AI: Experimentation, expertise, agency. *Computers and Composition*.

Bearman, M., & Ajjawi, R. (2023). Learning to work with the black box: Pedagogy for a world with artificial intelligence. *British Journal of Educational Technology, 54*(5), bjet.13337. https://doi.org/10.1111/bjet.13337

Knowles, A. M. (2024). Machine-in-the-loop writing: Optimising the rhetorical load. *Computers and Composition, 71*, 102826. https://doi.org/10.1016/j.compcom.2024.102826

Mollick, E. R., & Mollick, L. (2023). *Assigning AI: Seven Approaches for Students, with Prompts*. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4475995

Palmquist, M. (2020). A Middle Way for WAC: Writing to Engage. *The WAC Journal, 31*(1), 7–22. https://doi.org/10.37514/WAC-J.2020.31.1.01

Wang, S., & Zhang, H. (2026). Pedagogical partnerships with generative AI in higher education: How dual cognitive pathways paradoxically enable transformative learning. *International Journal of Educational Technology in Higher Education, 23*(1), 11. https://doi.org/10.1186/s41239-026-00585-x

Beyond AI and Education: The Emergence of Academic Meta-Work

Allison LITTLEJOHN [1], Francisco DURÁN DEL FIERRO [1], Koula CHARITONOS [2]
[1] UCL Knowledge Lab, University College London, UK
[2] Institute of Educational Technology, The Open University, UK

Abstract

This paper considers how the work practices of university educators are changing as AI is used in teaching and learning. Based on a Rapid Assessment Review, the areas of educators' core work affected are: 1) the adaptation of practices in different settings; 2) the fostering of equitable learning environments; 3) the development of expertise; and 4) ethics of care and ethical practices. The disruption of these work practices alters the relationships between educators and students. In this paper, we propose that the reconfiguration of these areas of work is linked to the emergence of forms of meta-work—"work that enables work". Taking an expansive view of meta-work, we posit that it offers ways to 'scaffold' educators' learning about new ways of working, opening up evolving learning opportunities as work transforms. We conclude that the visibility and invisibility of meta-work need to be negotiated continually as a way to guide university policy and practice.

Keywords: AI and Education; professional learning; meta-work

1. AI and the Emergence of Labour

1.1 A State of 'Becoming'

Artificial intelligence is permeating education systems worldwide, reshaping teaching, learning, and assessment. Research shows that most university educators now use AI tools in their practice (Wang et al., 2026). Even those educators who choose not to use AI in their teaching still encounter AI use in their work because of the large number of students choosing to use AI to carry out coursework. A Higher Education Policy Unit study showed that almost all students in UK Higher Education use AI (Stephenson & Armstrong, 2026). Most students (68%) think AI skills are essential for their future, yet only 48% believe that university education helps them to develop AI skills, and only 38% say universities provide them with access to AI tools.

Despite this enthusiasm, educator resistance to AI use may be well-founded: early studies show that the use of AI, particularly Generative AI (GenAI), may reduce students' cognitive effort and impair long-term retention (Barcaui, 2025), demonstrating that AI tools should be used with care based on evidence. Tech vendors promote the use of AI tools in education to support innovation, learning at scale, personalisation, and efficiency. However, a UNESCO report found many claims are unevidenced, with the pace of development prioritised over an evidence-based approach to implementation (Holmes and Miao, 2023). This gap between tech development and education practice may account for university educators reporting an increasing sense of detachment from the core aspects of teaching when AI is used (Williamson et al., 2023).

While AI can create meaningful benefits for educators, for example by automating administrative tasks, educators also face significant challenges due to AI disrupting normative ways of working and generating new forms of labour. Examples include ways of verifying that coursework has been carried out by a student; universities not supporting equitable access to AI tools, which means educators have to take inequity into account; and educators being expected to support learners in using AI and keep up-to-date with rapid advances in AI (see, for example, Zawacki-Richter et al., 2019; Lee, 2021; Kuka, 2022). This additional labour takes time, resets student-educator relationships, and redefines professional identity.

There is growing evidence that, while AI transforms work processes, it also intensifies work, placing demands on professionals (Ranganathan and Ye, 2026). At the organisational level, university leaders may view digital transformation processes, such as the implementation of AI tools, as linear, strategic exercises. By contrast, at the practice level, transformation is experienced as an iterative process comprised of continual encounters with what professionals cannot 'yet determine' because it has not 'yet become' or may never 'become definite' (Charitonos et al., 2026). Transformation brings about ruptures with the 'norms' of the past, generating forms of work that establish new routines, values, and identities. These encounters lead to forms of 'meta-work', defined as 'work that enables work' (Aroles et al., 2023). Meta-work can disrupt and unsettle normative ways of working during periods of transformation in 'unforeseen and unfolding' ways (Nerland et al., 2025). At the same time, it can create the conditions needed to rethink work practices, values, and identities in ways that engender learning and development (Charitonos et al., 2026).

This paper considers how the work practices of educators are changing as AI is used in teaching and learning. We consider emerging and not 'yet determined' forms of meta-work and how these open opportunities for professional learning. We begin by introducing the concepts of 'epistemic culture' and 'agency' to conceptualise the transformation of work and learning. We then explore forms of meta-work generated through the use of AI. We end by discussing how universities may create the conditions needed for meta-work to lead to learning and positive change.

1.2 Epistemic Culture and Agency

Agency is the response of professionals to changes in knowledge, practice, and processes at work (Cetina, 2005). Drawing on Eteläpelto et al. (2014) and others, agency refers to the capacity of professionals to influence their work, construct meaning, and enact values within shifting structures. Introducing AI tools into university teaching engenders new knowledge regimes, material practices, and organisational expectations that reshape educators' work. Educators will respond to these changes by mobilising their agency.

Cetina's (1999) concept of *epistemic cultures*—the unique, localised practices, tools, social structures, and ways of reasoning and communicating within different research fields that shape how they produce, validate, and understand knowledge—offers a way to consider how agency might be enacted in different settings. As AI mediates increasing portions of educational activity, it alters existing epistemic cultures by introducing machine-generated judgments, automated instructional sequencing, algorithmic analytics, and 'data-driven' predictions. These changes impact epistemic cultures and disrupt power relations across disciplines, generating opportunities and barriers. Educators' response to transformation is mediated by their epistemic culture and how well new knowledge, practice, and values align with their professional identity.

For example, when automated systems conflict with educators' professional norms, giving authority to the methods and values of data scientists and technical developers, this can lead to certain forms of *epistemic injustice*, understood as a form of unfairness where a person is harmed in their capacity as a knower or expert (Fricker, 2007; Grasswick, 2017). These injustices may lead to all kinds of ramifications, such as data scientists holding positions of power in educational systems or student behaviours being interpreted incorrectly, thus harming the expertise of university educators. These effects may lead to agentic responses from educators, such as educator workarounds or diverse forms of refusal.

By engaging proactively and creatively in transformation processes, individuals and communities may reshape their knowledge, practice, values, and professional identity, and this reshaping can be seen, under certain circumstances, as a form of 'meta-work' (Duran del Fierro et al., 2024). Meta-work offers ways to 'scaffold' learning and new ways of working in contexts of transformation and uncertainty. Meta-work (as a form of scaffolding) is not static; instead, it changes continuously as work transforms. Thinking about digital transformation through the lens of meta-work and epistemic culture foregrounds cultural practices, emphasising that AI use extends beyond organisational and technological considerations. Different types of meta-work, therefore, can be considered as evolving cultural practices that support new ways of working. This paper examines what types of meta-work may be created that support new ways of working with AI in teaching and learning in universities. This was examined through a Rapid Evidence Assessment.

2. Method: Rapid Evidence Assessment

A Rapid Evidence Assessment (REA) was carried out to explore the emerging literature on how university educators use artificial intelligence (AI) in teaching and learning. Relevant studies were identified through searches of the academic database Google Scholar using the keywords 'AI in university learning and teaching' and 'artificial intelligence in higher education teaching practice'. Inclusion criteria focused on peer-reviewed publications between 2020 and the present day examining the application of AI within university teaching and learning. Studies were screened for relevance, and areas of educators' work critically affected by the use (or non-use) of AI were identified and grouped as:

1. The adaptation of practices in different settings;
2. The fostering of equitable learning environments;
3. The development of expertise;
4. Ethics of care and ethical practices, drawing on the work of Nerland et al. (2025). These groups are discussed in the next section.

3. Results and Discussion: Meta-work, Culture and Agency

The Rapid Evidence Assessment identified areas of educators' work that are critically affected by the use of AI. The ways that these transformations generate forms of meta-work that impact epistemic culture, influencing responses and forms of agency, are discussed in this section.

3.1 Adaptations of Practices in Different Settings

There is evidence that educators and students need to adapt systemic work practices when AI is used for learning. This work includes the redesign

of learning activities and assessments. There is evidence that both formative and summative assessments are being carried out by students using GenAI tools. In some cases, students generate assignment outputs with limited cognitive engagement and learning (Richardson et al., 2021). This problem is causing some educators to adapt how they design assessments. Adaptation of assessment practice expands beyond altering assessment design; it includes advising students how AI should be used to support their learning and revising the ways student work is monitored. These changes in practice redefine the relationship between the student and the educator.

Some educators are also adapting the ways they design learning activities by integrating GenAI tools such as chatbots, adaptive systems, feedback systems, or automated writing tools. Rapid progress in the development of AI tools can make it difficult for them to stay well-informed, which means their work is in constant flux. Working in uncertain circumstances may undermine the notion of the educator as the 'expert'. Uncertainty may be stimulating for some educators and a concern for others (Littlejohn, 2022), which means that educators' agentic responses will differ across different contexts.

Another core academic practice that needs to be adapted is the interpretation of guidelines on the use of AI in learning and teaching. Rapid development of AI tools places educators in the position of having to interpret then re-interpret guidelines. The different interpretations mean that varied instantiations of institutional guidance will emerge, as educators share and use what 'makes sense to them' in their disciplinary or unit context. Variations will proliferate as educators discuss whether they use and how they interpret guidelines and as they create versions (e.g., slide decks) to share with students in lessons and tutorials. Although educators have always had to define the conditions under which students work, these actions and decisions are seen as meta-work generated in response to uncertainties that arise with the use of AI in education. These adaptations in practice are critical for equitable practices and learning environments.

3.2 Fostering Equitable Learning Environments

The use of AI places educators in the position of having to deal with equity concerns to ensure learning environments are non-discriminatory. This may include scrutinising students' access to AI tools, supporting their AI literacy, and guiding how they use AI for learning. Universities that do not provide students with access to AI tools place responsibility on educators to take into consideration disparities in access, differences in digital literacy, and the potential for AI to amplify bias. Access to digital tools is normally the remit

of a university IT department to level out advantages afforded to students who can pay subscription fees for premium versions of AI tools. An opportunity is opened up to reflect on and learn about equity issues, enabling educators to form a wider understanding of their students.

Educators also support students' AI literacy, for example by teaching how to write AI prompts, how to critique AI outputs using fact-checking methods, and how to refine their queries, enabling effective student–AI interaction (Denny et al., 2023). To be able to do this, the educators themselves have to develop these AI literacy skills. This involves additional labour such as learning about prompt writing, understanding how different GenAI tools respond to different types of prompts. The rapid development of AI and GenAI adds to the complexity of staying up-to-date with AI literacy.

Educators working in universities that do not provide institutional guidance on the use of AI have to guide students on the acceptable use of AI (Nazaretsky et al., 2022). This work involves defining what forms of algorithmic collaboration are tolerable, when disclosure of AI use is required, and whether misuse will be reported. Defining AI guidelines is difficult in complex learning contexts, extending beyond routine academic work where guidance on academic work would be provided by a pan-university unit, such as the Registry Office. Setting boundaries and rules and implementing these can be considered a new form of labour for educators that opens up ways to reflect on and learn about interactions with students. However, educators under time pressure may choose not to deal with these equity issues, placing students at risk.

3.3 The Development of Expertise

As they use AI in their teaching, educators develop expertise around AI capabilities and limitations. Unlike teaching technologies that have evolved slowly over time (e.g., Learning Management Systems), generative AI evolves rapidly, requiring educators to engage in continual, rapid upskilling (Ma & Groß, 2024). This places educators in a position where their knowledge and practice may be seen as continually becoming out-of-date.

As identified by Ng and colleagues (2021), they may have to engage in sustained effort to know how AI works: by understanding the technologies behind the different products and services that are used; to use and apply AI in teaching, including their pedagogical implications; and to create AI and use it ethically, which refers to not only designing and making AI technologies but being able to identify the different social, political, and environmental

implications associated with AI products and services. Arguably, the latter is the most difficult dimension because AI is not a simple tool existing in isolation, but it is part of a wider opaque ecosystem embedded in platforms, systems, and services (Pangrazio, 2026).

When using AI, particularly Generative AI (GenAI), outputs are generated in ways that appear realistic; for example, GenAI might indicate they are 'thinking' and 'writing' answers. Nevertheless, they sometimes generate inaccurate, fabricated outputs, shifting the burden of verification onto students and also educators. Winograd (2022) warns of this tendency of GenAI tools to output 'stochastic parroting', emphasising that continuous verification of outputs is needed. In response, educators report spending extended hours verifying and fact-checking student outputs. Although forms of investigative work aimed at verifying attribution has always formed part of academic practice, the use of AI makes it much more difficult for educators to know who produced assignment outputs, placing educators in a liminal space between certainty and uncertainty and generating meta-work.

Verifying student outputs intensifies educators' labour as forms of 'policing' and 'detecting'. This may lead to 'penalising' or 'banning' students in ways that are at odds with the culture, norms, and values educators aspire to. In the UK, most universities have not yet implemented guidance on the use of AI (Stephenson & Armstrong, 2026). This precariousness creates a contested space where educators are unsure how to respond to students' non-declared use of AI, building tensions, eroding trust, and expanding labour. These tensions may drive educator responses in different ways; some educators may feel compelled to verify and fact-check student work, while others may feel unsure how to make clear judgements and put forward accusations to students. In all these situations, educators are confronted with possibilities for action that may or may not happen. These create opportunities to learn experientially about AI and student engagement, and to develop new forms of relationships with their students through ongoing considerations necessitating them to work out what may be the most appropriate response to the challenges AI invokes in educational settings.

3.4 Ethics of Care and Ethical Practices

The use of AI by students necessitates an ethics of care to make sure students safely use AI tools. Educators raise students' awareness of the personal consequences of the use of AI tools, guiding them on issues such as data privacy, intellectual property, bias, and the environmental impact of data use (Okulicz-Kozaryn et al., 2026). A report on AI governance initiatives by

the European Commission (2022) emphasises institutional responsibility in communicating these risks. However, since few universities have introduced guidance, the burden of responsibility is on educators, who may not be knowledgeable about—or experienced in—the use of AI or are still working out themselves how to use AI.

Educators invest effort in maintaining ethical practices by guiding students on how to make transparent and disclose their use of AI. This requires ongoing work that involves continually reviewing GenAI guidance, re-interpreting and re-framing guidance, communicating with other colleagues, communicating with students, making judgements beyond the quality of the work produced by a student, to include how that work was produced and whether AI has been used using agreed rubrics (Okulicz-Kozaryn et al., 2026). Some educators will take action to address this uncertainty by engaging in deliberate actions to resolve the tension, engaging thus in forms of meta-work which has the potential to lead to learning. Others may be concerned they are unable to judge academic rigour and may deal with the uncertainty by resisting having to set boundaries.

Maintaining ethical practice means that educators need to make decisions about whether the degree of assistance used by a student constitutes misconduct. However, use of AI, especially GenAI, can make these judgements difficult for educators. GenAI creates text, images, videos, sounds, and citations that seem authentic and plausible, but there are not reliable tools to detect attribution, thus making academic judgements (Njee, 2025). This means educators have to engage in new forms of interpretive academic labour that can be demanding and that constitute meta-work. Some educators may choose to respond by avoiding the risk of making the wrong judgement, flagging assignments to be considered by university-wide panels such as academic integrity committees so the educator avoids having to take responsibility. Others may feel uncomfortable with making judgements and choose not to engage with this situation.

In all these situations, what is becoming increasingly difficult for educators is to find time, space, and emotional energy to consider how AI shapes what is becoming of their profession and how it is changing individual practices and behaviours. Therefore, reconfiguration of work practices and the emergence of related forms of meta-work have implications for educator development, as discussed in the next section.

4. Implications: Meta-work as an Enabler for Educator Learning and Development

This paper highlights areas of university educators' core work critically affected by the adoption of AI in teaching and learning and are undergoing change. Within these core areas, new forms of work—or meta-work—are created when AI is used by educators and/or students. Meta-work is a specific form of work which embodies emerging forms of professional expertise that are critical for digital transformation.

What distinguishes meta-work from routine academic labour is not simply that it is more work, but that the work is still in a 'state of becoming', generally is unacknowledged, and operates at a meta-level under which teaching and learning occur. The use of AI in teaching and learning generates forms of meta-work that includes tasks that are broadly linked to rethinking assessment and learning activity design; the development of new forms of expertise and professional judgement in AI-mediated practice; the fostering of inclusive and equitable learning environments; and the application of a duty of care towards students and ethics of AI-related practice.

All these forms of work create opportunities for learning and professional development. While meta-work scaffolds opportunities for learning and development, it also creates a state of flux, challenging educator expertise as they work within a 'state of becoming'. Some forms of meta-work will eventually be stabilised and become routine work. However, meta-work may also lead to tensions between educators' work 'now' and what came 'before' that may (or may not) align with their identity, culture, and values, which may lead to various modes of resistance. Educators may find forms of meta-work exhausting, especially when they stimulate the erosion of traditional expertise and an increase in emotional labour associated with uncertainty.

This signals a call to action in identifying and acknowledging diverse forms of meta-work to inform future educational policy and practice. More research is needed to systematically examine forms of meta-work in relation to core areas of educators' work as AI is embedded within university learning and teaching.

Ethical Declaration

This project was part of a wider stimulus study entitled: *Professional Learning in the Multi-layeredness of Digital Transformation: Synthesising contributions to further developing a collaborative research agenda*. Funded by the University of Oslo (2024-25) with partners at University College London

(UK), The Open University (UK), University of Regensburg (Germany), and University of Gothenburg (Sweden).

References

Aroles, J., Bonneau, C., & Bhankaraully, S. (2023). Conceptualising 'meta-work' in the context of continuous, global mobility: The case of digital nomadism. *Work, Employment and Society*, 37(5), 1261-1278.

Barcaui, A. (2025). ChatGPT as a cognitive crutch: Evidence from a randomized controlled trial on knowledge retention. *Social Sciences & Humanities Open*, 12, 102287.

Cetina, K. K. (2007). Culture in global knowledge societies: Knowledge cultures and epistemic cultures. *Interdisciplinary Science Reviews*, 32(4), 361-375.

Charitonos, K., Duran del Fierro, F., & Littlejohn, A. (2026). Organising Work and Learning: The Mediating Role of Meta-Work in the Digital Transformation of Work. *Submitted to JVET special issue 'VET in the era of digital transformation: New forms of work and their implications for vocational learning'*, March 2026.

Denny, P., Leinonen, J., Prather, J., Luxton-Reilly, A., Amarouche, T., Becker, B. A., & Reeves, B. N. (2023). Promptly: Using prompt problems to teach learners how to effectively utilize AI code generators. *arXiv preprint arXiv:2307.16364*.

Eteläpelto, A., Vähäsantanen, K., Hökkä, P., & Paloniemi, S. (2014). Identity and agency in professional learning. In *International Handbook of Research in Professional and Practice-Based Learning* (pp. 645–672).

Fricker, M. (2007). *Epistemic injustice: power and the ethics of knowing*. Oxford University Press.

Grasswick, H. (2017). Epistemic injustice in science. In *The Routledge handbook of epistemic injustice*, Routledge, Taylor & Francis Group.

Hartley, et al. (2023). SKA and the paradigm shift in scientific investigation.

Holmes, W., & Miao, F. (2023). *Guidance for generative AI in education and research*. UNESCO Publishing. https://unesdoc.unesco.org/ark:/48223/pf0000386693

Kuka, L., Hörmann, C., & Sabitzer, B. (2022). Teaching and learning with AI in higher education: A scoping review. *Learning with Technologies and Technologies in Learning: Experience, Trends and Challenges in Higher Education*, 551-571.

Lee, J., Wu, A. S., Li, D., & Kulasegaram, K. (2021). Artificial intelligence in undergraduate medical education: a scoping review. *Academic Medicine*, 96(Supplement_2), S62-S70.

Littlejohn, A. (2020). Seeking and sending signals: Remodelling teaching practice during the Covid-19 crisis. *ACCESS: Contemporary Issues in Education*, 40(1), 56–62.

Littlejohn, A. (2023). Transforming educators' practice: How university educators learned to teach online from home during the Covid-19 pandemic. *Higher Education Research & Development*, 42(2), 366–381.

Mah, D. K., & Groß, N. (2024). Artificial intelligence in higher education: exploring faculty use, self-efficacy, distinct profiles, and professional development needs. *International Journal of Educational Technology in Higher Education*, 21(1), 58.

Nazaretsky, T., Ariely, M., Cukurova, M., & Alexandron, G. (2022). Teachers' trust in AI-powered educational technology and a professional development program to improve it. *British Journal of Educational Technology*, 53(4), 914-932.

Nerland, M., Mäkitalo, Å., Brandenberger, I., & Sadorge, C. (2025). Learning through metawork in digitalised healthcare: The case of monitoring patient falls. In *EARLI 2025 conference*, Graz, Austria.

Njee, L. C. (2025). *Online Professors' Experience With Students Misusing Artificial Intelligence (AI) in Higher Education: An Exploratory Case Study* (Doctoral dissertation, National University).

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2, 100041.

Okulicz-Kozaryn, W., Artyukhov, A., & Artyukhova, N. (2026). Will AI Replace Us? Changing the University Teacher Role. *Societies*, 16(1), 32.

Pangrazio, L. (2026). The (im) possibility of AI literacy. *Learning, Media and Technology*, 51(1), 1-7.

Scaife, A. (2020). Data models and scientific discovery in the era of the SKA.

Stephenson, R., & Armstrong, C. (2026). *Student generative AI survey 2026*. Higher Education Policy Institute: London, UK. https://www.hepi.ac.uk/reports/student-generative-ai-survey-2026/

Wang, Y., Liu, C., & Tu, Y. F. (2021). Factors affecting the adoption of AI-based applications in higher education. *Educational Technology & Society*, 24(3), 116-129.

Williamson, B. (2019). New power networks in educational technology. *Learning, Media and Technology*, 44(4), 395–398.

Williamson, B., Komljenovic, J., & Gulson, K. N. (2023). Introduction: Digitalisation of education in the era of algorithms, automation and artificial intelligence. In *World yearbook of education 2024* (pp. 1-19). Routledge.

Winograd, A. (2022). Loose-lipped large language models spill your secrets: The privacy implications of large language models. *Harv. JL & Tech.*, 36, 615.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators? *International Journal of Educational Technology in Higher Education*, 16(1), 39.

2 AI Tools and Pedagogy

Transforming Pedagogy with AI Tools: Impacts, Challenges, and Opportunities in Higher Education

Chi Ho CHAN [1], Christy CHAN [2], Julia CHEN [3], Pauli LAI [1]
[1] Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Hong Kong, China
[2] Chan Feng Men-ling Chan Shuk-lin Language Centre, City University of Hong Kong, Hong Kong, China
[3] Educational Development Centre, The Hong Kong Polytechnic University, Hong Kong, China

Abstract

Traditional pedagogy positions teachers as the primary source of knowledge, with educators providing academic guidance on English oral presentations—including both content and paralinguistic features—across university disciplines. However, subject-matter instructors often lack expertise in English language delivery, while large class sizes prevent the provision of timely individual feedback. Advances in technology enable a pedagogical shift to fill this gap. Rather than waiting for teacher feedback, students can leverage AI for immediate learning support. However, AI has limitations regarding nuanced aspects of performance. To complement AI feedback, we implemented **NinjOrAItor**—an AI platform for English oral presentations featuring hybrid intelligence that combines self- and peer-assessment. This hybrid intelligence pedagogy integrates AI capabilities with human insight. The platform addresses key challenges: teacher expertise limitations, scalability issues in large classes, and the need for timely feedback. By integrating AI assessment with structured human intelligence, *NinjOrAItor* demonstrated improved student presentation performance while maintaining pedagogical quality.

Keywords: AI, English oral presentations, hybrid intelligence, pedagogy, self-assessment, peer assessment

1. Introduction

Empirical evidence demonstrates that technological tools consistently enhance student achievement, engagement, and teaching effectiveness across educational levels and disciplines (Hu, 2024). In English oral presentation instruction, where students traditionally rely on delayed teacher feedback, technology offers timely and comprehensive assessment that can accelerate skill development and promote interactive learning experiences.

English serves as the *lingua franca* in tertiary education (Alhasnawi, 2021). Universities now require English oral presentations across disciplines, accompanied by teacher assessment and feedback. However, this traditional approach faces significant learning barriers: subject-matter instructors often lack specialised English language expertise to deliver targeted linguistic feedback, while large classes limit timely individual assessment, which is crucial for language skill development. Although AI platforms can provide faster feedback, current AI systems still struggle to evaluate nuanced presentation qualities, leaving a gap between technological potential and pedagogical needs.

To bridge this gap and optimise learning outcomes, we developed **NinjOrAItor**, an AI platform that employs hybrid intelligence—combining automated feedback with human input and structured learning resources. *NinjOrAItor* is grounded in three pedagogical principles essential for the effective development of English oral presentation skills: active learning, self-regulated learning (SRL), and human-AI hybridity. Details are as follows:

1. **Active Learning:** Emphasises student-centred, iterative practice and engagement over passive lecture-based instruction, enabling deeper skill acquisition through repeated performance, feedback, and refinement (Kozanitis, 2023).
2. **Self-Regulated Learning (SRL):** When properly designed, SRL consistently enhances students' learning outcomes and academic performance, as evidenced by multiple meta-analyses (Panadero, 2017).
3. **Human-AI Hybridity:** Leverages the complementary strengths of AI and human expertise. AI provides scalable, immediate analysis of technical and perceptual elements (e.g., pronunciation, fillers, and delivery), while humans contribute contextual judgment and nuanced decision-making (Holstein et al., 2020).

Together, these principles demand frequent practice cycles supported by timely, multi-layered feedback—conditions that traditional large-classroom instruction struggles to meet. *NinjOrAItor* operationalises them by integrating technology-enhanced active practice, AI-driven immediate feedback with self- and peer-assessment tools, and structured human oversight, creating a scalable environment that accelerates skill development while preserving pedagogical depth.

NinjOrAItor integrates Azure OpenAI GPT-4o and Azure OpenAI GPT-4 to deliver automated rubric-based feedback on eight dimensions: clarity of view, content, eye contact, facial expression, hand gestures, PowerPoint visu-

als, pronunciation, and vocal fillers. The system processes approximately 40 five-minute video submissions per hour to enable rapid feedback cycles while teachers retain final oversight for contextual judgment. Students access the platform through institutional email accounts. They upload presentation videos, receiving both AI and human feedback in the process.

The hybrid feedback loop works as shown in **Figure 1**. Students first receive immediate AI-generated comments. They then complete self-assessment by comparing their own reflections against the AI feedback, followed by peer assessment on linguistic and paralinguistic aspects, as indicated in **Figures 2**, **3**, and **4**.

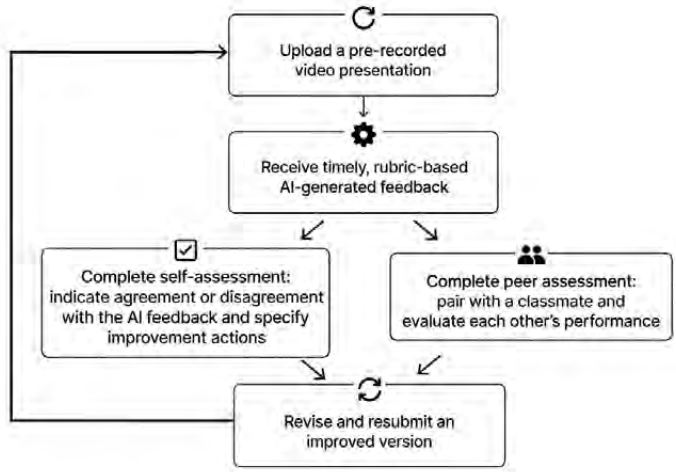


Figure 1: Hybrid feedback loop in *NinjOrAItor*

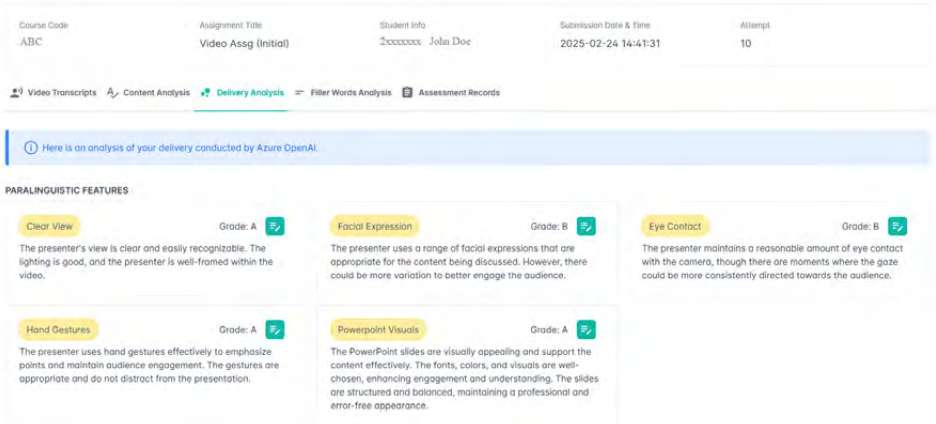


Figure 2: AI feedback interface

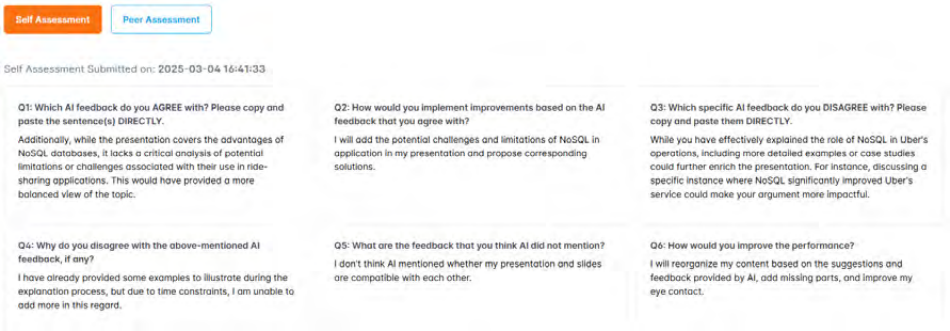


Figure 3: Self-assessment interface

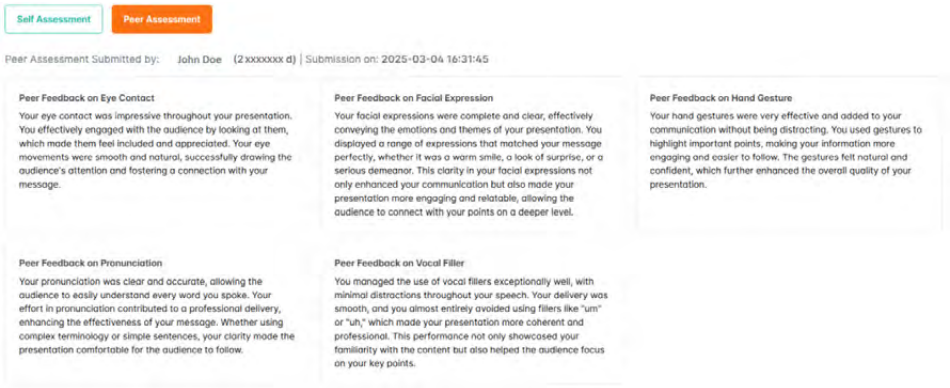


Figure 4: Peer-assessment interface

By prompting students to critique AI-generated feedback through structured self- and peer-assessment, *NinjOrAItor cultivates evaluative judgement*—the capacity to make informed decisions about the quality of one's own and others' work (Tai et al., 2018). This active process aligns with constructivist principles in language pedagogy, as students reconcile automated outputs with their own reflections and peer perspectives instead of receiving feedback passively. Such engagement fosters deeper metacognition and autonomy compared with the passive consumption of AI feedback.

2. Methodology

The study adopted a mixed-methods design (Creswell & Plano Clark, 2018) to evaluate *NinjOrAItor*'s impact on pedagogical practice and student learning in English oral presentations. Our dataset came from a database subject involving 171 students; over 98% of whom use English as an Additional Language (EAL). They were required to submit presentation videos through *NinjOrAItor*.

Quantitatively, pre- and post-surveys were administered to measure students' understanding of rubric criteria, confidence, and the use of rubrics for self-learning. Responses were coded on a 5-point Likert scale from 1 (Strongly Disagree) to 5 (Strongly Agree). These were analysed using weighted means (Σ [response value × frequency] / total responses) to capture both the direction and magnitude of change. The survey instrument employed the same eight rubric components as those used in *NinjOrAItor*, with no additional differential weighting applied beyond the 5-point Likert scale itself. A total of 104 students completed the pre-survey, while 59 filled in the post-survey. Using this approach, the study observed improvements in students' understanding of rubric criteria and their use of rubrics for self-learning after interacting with *NinjOrAItor*.

Qualitatively, semi-structured interviews with students examined user experience, perceived specificity and actionability of AI feedback, the value of learning units, and the complementarity of self/peer assessment. Survey, performance, and interview data were collected after students received immediate, rubric-aligned AI feedback and completed self/peer assessments embedded in course workflows.

3. Results and Discussion

The video marks were calculated based on eight rubric criteria with the following weightings: Content (35%), Clarity of View, Eye Contact, Facial Expression, Pronunciation, Vocal Fillers, and PowerPoint Visuals (10% each), and Hand Gestures (5%).

Students showed improved performance in English oral presentations after using *NinjOrAItor*. Among the 161 students who submitted both video assignments, average scores rose by 4.80% from M = 81.07 (SD = 4.41) on the first assignment to M = 84.96 (SD = 3.74) on the second. A paired-samples t-test confirmed the improvement was statistically significant, $t(160) = -13.01$, $p < 0.001$ (two-tailed).

Students received hybrid feedback from *NinjOrAItor* on their first submission and then revised and resubmitted an improved version as the second assignment. At the end of the semester, post-survey results (compared with pre-survey baseline) showed improved attitudes across most assessment criteria. The largest gains occurred in students' perceptions of vocal fillers (weighted mean increased by 16.67% from a pre-survey baseline of 3.33 to a post-survey value of 3.88), followed by paralinguistic techniques (weighted mean increased by 7.17% from a pre-survey baseline of 3.61 to a post-survey value of 3.86).

However, attitudes toward pronunciation assessment showed a modest decline (with the weighted mean decreased by 0.61% from a pre-survey baseline of 3.62 to a post-survey value of 3.59).

Students also reported improved understanding of rubric criteria and an enhanced ability to apply rubrics for self-directed learning. Perceptions of *NinjOrAItor* were mixed: while appreciation for the platform's feedback features increased (weighted mean up by 15.38%), confidence in its assessment accuracy decreased (weighted mean down by 8.60%), primarily attributed to technical limitations in hand gesture detection. Meanwhile, in the post-survey, 71.19% of students indicated that the platform was helpful for self-improvement.

Interviews with five consenting students revealed generally positive perceptions of *NinjOrAItor*, with participants describing it as intuitive and user-friendly. Students reported effective use of rubrics in preparing their presentations and consistent improvements in vocal filler reduction, pronunciation, and speech pacing after receiving AI feedback. The hybrid intelligence approach provided comprehensive guidance that helped students identify specific areas for improvement and enhance confidence in presentation preparation and delivery. One student described a complementary peer-AI loop: peers use AI's rubric-based overview (e.g., on eye contact) as a foundation but added contextual depth, such as content simplification for non-experts, creating a collaborative revision process.

Despite these benefits, AI limitations in visual analysis (hand gestures and slide content) and occasional content feedback inaccuracies were noted as areas needing refinement. Overall, students appreciated the objective baseline and structured support, particularly for inexperienced presenters, with rubrics and Learning Units facilitating targeted skill development. Instructors valued *NinjOrAItor*'s vocal filler detection for identifying expressions like "so," "like," and "uh." However, occasional misclassifications (e.g., flagging "but" as a filler) require human oversight. The platform enhanced grading efficiency while maintaining assessment quality by providing AI evaluations as an objective starting point. Educators retained full authority to adjust scores and feedback for specific pedagogical contexts, creating an optimal human-AI partnership that leverages automation while preserving personalised instruction.

This study has several limitations. Although 171 students were enrolled, only 104 respondents (60.82%) completed the pre-survey and 59 respondents (34.50%) completed the post-survey, resulting in a 43.27% dropout rate. This substantial dropout may introduce survivorship bias, as more motivated or en-

gaged students were more likely to persist with both surveys and video assignments. Consequently, the observed improvements in presentation scores and perceptions may be positively skewed.

4. Implications

The hybrid model implemented in *NinjOrAItor* resolves the core weaknesses identified in the literature and observed in our data (see Table 1). Students found the platform effective for independent presentation skill development, yet the slight decline in confidence in AI evaluation highlights the continued importance of human judgment. Through integrated self- and peer-assessment features, students learn to critically evaluate AI feedback while learning collaboratively from peers. These findings align with the theoretical advantages of hybrid intelligence outlined earlier.

Table 1 compares the four main feedback models currently used in English oral presentation instruction. It summarises the performance outcomes of four primary feedback models: instructor-only, automated (AI-only), peer/self, and the hybrid model. Since performance gains for the AI-only and teacher-only models could not be quantified in this study, they are described qualitatively as 'moderate' based on prior literature and observed classroom practices (Kaliisa et al., 2025).

Table 1: Comparative Analysis of Feedback Models in English Oral Presentation Instruction

Feedback Model	Timeliness of feedback	Scalability	Nuance in assessment	Performance gains observed	Data Source
Instructor-only	No	No	Yes	Moderate	Kaliisa et al. (2025); classroom observations
AI-only	Yes	Yes	No	Moderate	Kaliisa et al. (2025)
Peer/Self only	Partially	Partially	Partially	Moderate	Kaliisa et al. (2025)
Hybrid (NinjOrAItor)	Yes (AI + selective human)	Yes (AI + selective human)	Yes (AI + human contextual judgment)	High (+4.80% overall score, $p < 0.001$)	Present study (empirical, paired t-test)

Note: Non-NinjOrAItor rows draw on Kaliisa et al. (2025); NinjOrAItor results are from the present study.

The hybrid model implemented in *NinjOrAItor* (rightmost column) resolves the core weaknesses identified in the literature and observed in our data. While AI-only feedback delivers the timeliness and scalability demanded by large EAL classes (171 students, >98% EAL), it cannot provide the nuanced judgment on content relevance, cultural pragmatics, or motivational support

that human instructors supply. By embedding immediate AI analysis within a human-feedback loop, students not only improved presentation scores by 4.80% ($t(160) = -13.01$, $p < 0.001$) but also developed metacognitive skills through actively evaluating AI output.

Acknowledgments

This research was supported by the TDG Inter-institutional Collaborative Activities (IICA) 2022–25 grant through a partnership between The Hong Kong Polytechnic University (PolyU) and City University of Hong Kong (CityUHK). The authors extend their gratitude to the English instructors at CityUHK and the technical development team at PolyU, whose collaborative efforts have been instrumental in advancing this study on hybrid intelligence in English language education.

Ethical Declaration Statement

The project has been approved by the Human Subjects Ethics Sub-committee (HSESC) (or its Delegate) of The Hong Kong Polytechnic University (HSESC Reference Number: HSEARS20231206007).

Declaration of Use of AI in the Writing Process

The authors used Grammarly during the preparation of this work to check grammar and the use of English. 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

Alhasnawi, S. (2021). English as an academic lingua franca: Discourse hybridity and meaning multiplicity in an international Anglophone HE institution. *Journal of English as a Lingua Franca, 10*(1), 31–58. https://doi.org/10.1515/jelf-2021-2054

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE.

Holstein, K., Aleven, V., & Rummel, N. (2020). A conceptual framework for human–AI hybrid adaptivity in education. In I. I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin, & E. Millán (Eds.), *Artificial Intelligence in Education (AIED 2020)* (Lecture Notes in Computer Science, Vol. 12163, pp. 240–254). Springer. https://doi.org/10.1007/978-3-030-52237-7_20

Hu, S. (2024). The effect of artificial intelligence-assisted personalised learning on student learning outcomes: A meta-analysis based on 31 empirical research papers. *Science Insights Education Frontiers, 24*(1), 3873–3894.

Kaliisa, R., Misiejuk, K., López-Pernas, S., & Saqr, M. (2025). How does artificial intelligence compare to human feedback? A meta-analysis of performance, feedback perception, and learning dispositions. *Educational Psychology, 46*(1), 80–111. https://doi.org/10.1080/01443410.2025.2553639

Kozanitis, A., & Nenciovici, L. (2023). Effect of active learning versus traditional lecturing on the learning achievement of college students in humanities and social sciences: A meta-analysis. *Higher Education, 86*(6), 1377–1394. https://doi.org/10.1007/s10734-022-00977-8

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology, 8*, Article 422. https://doi.org/10.3389/fpsyg.2017.00422

Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education, 76*(3), 467–481. https://doi.org/10.1007/s10734-017-0220-3

Technology-Enhanced Training in English Connected Speech Processes Among Chinese EFL Learners

Hsueh-Chu CHEN [1] & Xiaona ZHOU [1]
[1] Department of Linguistics and Modern Language Studies, The Education University of Hong Kong, Hong Kong, China

Abstract

Connected speech processes (CSPs), referring to sound adjustments such as deletion and reduction, are commonly used by native speakers to ease articulation in natural speech. However, they are rarely employed by Chinese learners of English (EFL), partly due to a lack of instruction and practice. This study explored an online training programme on six types of CSPs—linking, deletion, insertion, modification, reduction, and multiple—adopting a technology-enhanced approach. The research investigated the training effects on Chinese EFL learners' CSP production through a pre-test–training–post-test design. Thirty participants were involved: half were assigned to an experimental group (EG) receiving eight sessions of CSP training, while the other half served as a control group (CG) receiving general drills. Results showed that the EG's CSP production was significantly improved. Generalisation of the CSP training was also observed. These findings suggest the pedagogical benefits of technology-enhanced CSP training.

Keywords: English connected speech processes; connected speech training; Chinese EFL learners; Artificial Intelligence; Doubao

1. Introduction

Chinese learners of English as a foreign language (EFL) often report that the sounds acquired in classroom settings differ noticeably from those encountered in real-life communication. Such differences are largely attributed to the influence of connected speech processes (CSPs). CSPs can affect not only specific sounds at the segmental level but also the overall rhythm and fluency of speech at the suprasegmental level. According to features of sound adjustments at word boundaries, CSPs are categorised into linking, deletion, insertion, modification, reduction, and multiple (Alameen & Levis, 2015). Although CSPs are commonly employed by native speakers to facilitate articulation, Chinese learners are found to rarely produce them (e.g., Wong et al., 2021). Despite their importance, instruction on CSPs is largely lacking, and research on CSP training remains limited.

In recent decades, Artificial Intelligence (AI) tools have been gradually integrated into pronunciation instruction and have been proven effective in various aspects of pronunciation. However, their integration into the training of CSPs has yet to be fully explored. This study, therefore, aims to develop a technology-enhanced CSP training programme and to investigate its effectiveness in improving Chinese EFL learners' production of CSPs. Accordingly, this study seeks to address the following research questions (RQs):

1. To what extent is the CSP training effective in improving English CSP production by Chinese EFL learners?
2. What are Chinese EFL learners' attitudes towards the CSP training?

2. Methodology

Participants

This study employed mixed methods to investigate Chinese EFL learners' difficulties with and the training effects on the production of CSPs. A total of 30 Chinese EFL learners from Northern China, with Northern Mandarin as their first language (L1), participated in this study. They were all university students from non-English majors, aged between 18 and 25, and reported intermediate English proficiency based on their self-reported scores from recognised English examinations.

They were randomly assigned to one experimental group (EG) and one control group (CG), with 15 participants in each. The EG received CSP training, while the CG received general drills. Each participant completed production pre- and post-tests before and after the training. Participants from the EG also completed post-training interviews regarding training effectiveness after completion of the post-test.

Stimuli

During the production tests, participants read aloud a set of 42 sentences adapted from various teaching materials and studies related to CSPs. These sentences elicited 72 tokens in total, covering six types of CSPs (12 tokens for each type). Of the 12 tokens, six were used as demonstration examples during the training, while the remaining six were not.

The target types of CSPs examined in this study included consonant-vowel (CV) linking, deletion, vowel-vowel (VV) linking, assimilation, vowel reduction, and multiple. The selection of tokens was based on the occurrence of the target CSP in the speech of at least three out of five native speakers.

Treatment

The CSP training included eight 30-minute sessions delivered online twice a week via Qualtrics. Each session consisted of three major components: explicit instruction, perceptual practice, and production practice.

- **Explicit Instruction:** This section focused primarily on CSP rules delivered through a 10-minute instructional video with an AI-generated voiceover.
- **Perceptual Practice:** This included four tasks: dictation, identification, discrimination, and shadowing, each using either authentic or AI-generated audio as task items. For the first three tasks, instant corrective feedback was provided.
- **Production Practice:** This component included both controlled and spontaneous oral tasks. In controlled practice, participants were instructed to read aloud a set of collocations, idioms, sentences, or short paragraphs containing the target CSPs. In the free speech task, participants were asked to use target CSPs while giving a short talk on one of several daily topics and to receive feedback from a Generative AI (GenAI) tool.

The GenAI tool used in this study was **Doubao** (https://www.doubao.com), a chatbot with capabilities similar to other popular tools such as ChatGPT, Gemini, and Copilot. Similar to these tools, Doubao offers powerful natural language processing and interaction features. Two core functions of Doubao were used during the production practice:

1. A speech-to-text function, which promptly recognised and transcribed participants' spoken input.
2. Providing feedback based on participants' prompts in an interactive, conversational format.

To facilitate optimal interaction and ensure consistent, high-quality feedback, sample prompts were provided to EG participants.

Data Analysis

All speech data were analysed sentence by sentence, focusing on the occurrence of CSPs. After completion of the perceptual analysis, an independent samples t-test was conducted to compare gain scores between the CG and the EG. Training transfer was then examined across different types of CSPs and between trained and untrained instances. Interview data were transcribed and analysed thematically to explore feedback on training effectiveness and the integration of AI technology.

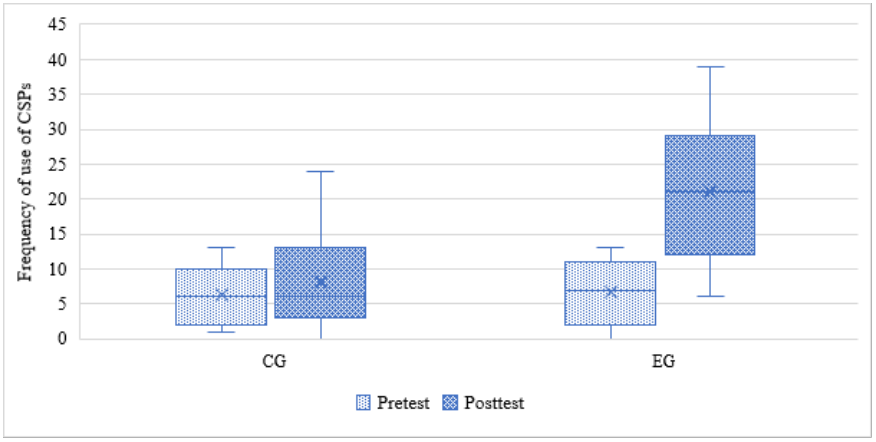
3. Results and Discussion

Effects of CSP Training on Production

To answer RQ1 regarding how the training affected the production of CSPs among Chinese EFL learners, the mean occurrence of CSP production for the CG and EG in the pre- and post-tests were compared (see Figure 1). Levene's test indicated that equal variances were not significant (p = .844), which indicated that the assumption of homogeneity of variance was met.

The t-test results revealed a significant difference in gain scores between the CG (M = 1.20, SD = 2.77) and the EG (M = 19.35, SD = 7.36), $t(28)$ = -8.94, p < .001 (one-tailed), with Cohen's d = -3.26. The EG showed significantly greater improvement in the use of CSPs in reading-aloud speech, with a large effect size.

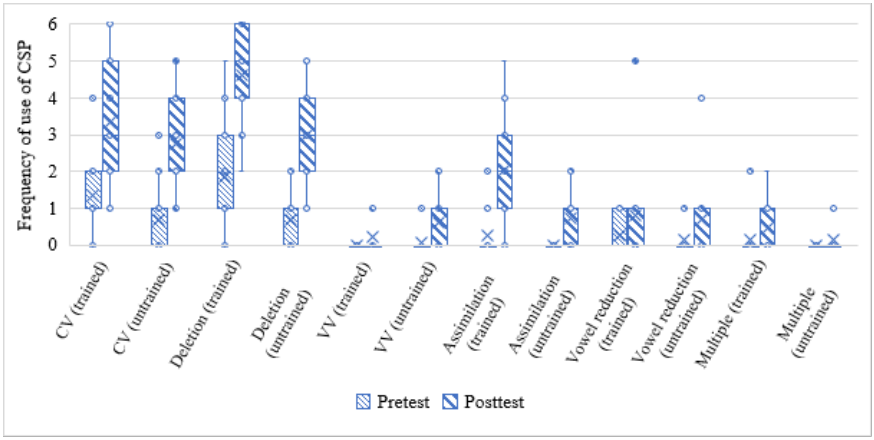
Figure 1: Performance of overall CSP production by CG and EG in the pre- and post-tests



With respect to the generalisation of training effects, participants' performance on trained and untrained instances between pre- and post-tests was also analysed. As illustrated in Figure 2, both trained and untrained instances demonstrated improvement in the production of six types of CSPs from pre- to post-test. However, the gains in production of VV linking (e.g. 'fair‿enough' /f ɪ ər ɪ ˈnʌf/, 'two(w)eggs' /tuːwegz/), vowel reduction (e.g. 'for' /fə/ in 'for several weeks'), and multiple processes (e.g. 'heard‿him' /hɜːd ɪ m) were relatively limited compared to the clearer improvements observed in CV linking (e.g. 'moved‿around' /muːvdəra ʊ nd/), deletion (e.g. 'past president' / p ɑ ːsprez.ənt /), and assimilation (e.g. 'baked you' / be ɪ kdʒuː/).

This pattern indicated that the training was effective in improving the production of certain CSPs. The limited improvement in the production of VV linking, vowel reduction, and multiple processes aligns with studies reporting asymmetric effects of training across different CSPs (e.g., Ting & Kuo, 2012).

Figure 2: Performance of use of each CSP in trained and untrained instances by EG



Chinese EFL Learners' Attitudes Towards the CSP Training

To answer RQ2, participants' attitudes towards the CSP training were evaluated through a semi-structured interview. Their feedback was largely positive. All participants agreed that the training helped them understand the importance of CSPs and expand their knowledge of CSPs in a systematic way. They reported being more attentive to CSPs when listening to and speaking English.

The participants also evaluated the usefulness of each component in improving their use of CSPs. All were considered helpful, with 11 participants identifying high-variability perceptual practice as the most beneficial due to its dual purpose. As one participant described, it was a form of 'learning through practice'.

The integration of AI technology into the training components elicited nuanced responses. Twelve out of 15 participants perceived the AI-generated voiceover as acceptable and did not consider the teacher's voice to be necessary for instructional videos. Interestingly, they did not realise that some audio files used in the perception task were generated by AI until the interview.

In terms of the use of GenAI in production practice, twelve out of 15 participants showed willingness to continue using Doubao to support their

self-directed learning. Over half of them indicated that Doubao's speech-to-text function motivated them to practice pronunciation more regularly; nevertheless, they also expressed concerns about the accuracy of the feedback it provided, particularly when there were inconsistencies with the instructional content presented during the training. Two participants even viewed verifying the feedback as a form of practice. For instances where they felt confident, they considered it beneficial. However, when some instances were unfamiliar or less mastered, they struggled to assess the accuracy and highlighted the need for teacher guidance. This concern is echoed in Mompean's (2024) study, which indicated that GenAI tools may provide inaccurate feedback, particularly regarding pronunciation accuracy.

Overall, the positive feedback of the training aligns with previous studies suggesting that well-designed instruction on CSPs could raise learners' awareness and enhance their ability in the use of CSPs (e.g., Momen & Pilus, 2022).

4. Implications

The study examined the effectiveness of the CSP training on Chinese EFL learners' CSP production. The results revealed limited use of CSPs and identified frequently used patterns for each CSP by Chinese EFL learners in connected speech. The results also confirmed the positive effects of the training on learners' CSP production. These results provided pedagogical implications for both pre- and in-service teachers, particularly in the Chinese context.

Pronunciation instruction should include focused training on CSPs for ESL/EFL learners. The AI-enhanced CSP training framework used in this study can serve as a practical model for integrating such instruction. It is important for teachers to introduce learners to the basic features of CSPs and to explicitly address inappropriate strategies that learners often adopt when encountering CSPs.

Different AI tools can serve multiple training purposes. For example, text-to-speech AI can be used for narration in instructional videos, generation of perceptual practice items, and imitation samples in production practice. GenAI tools are also useful for independent practice, though the accuracy of their feedback on CSPs remains a concern. This suggests a need for hybrid or flipped classroom models that include teacher-led feedback and in-class interaction, and also calls for more attention to the iteration of GenAI tools to enhance the accuracy of feedback on pronunciation and speaking.

These implications may also extend to other linguistic contexts or international EFL settings. The framework can be adapted by selecting pronunciation targets that are difficult for the local learner group and by tailoring AI prompts and practice materials to their learning needs and classroom constraints.

Acknowledgments

This research forms part of the first author's thesis. The first author gratefully acknowledges the guidance and support provided by their supervisor throughout the project.

Ethical Declaration Statement

The project has been approved by the Human Research Ethics Committee of The Education University of Hong Kong (Reference Number: Ref. no. 2024-2025-0508).

Declaration of Use of AI in the Writing Process

Doubao was used to support participants' production practice in the treatment phase. The authors used ChatGPT-4 during the preparation of this work to enhance clarity, grammar, and style. After using the tool, the authors thoroughly reviewed and edited the content as necessary and assumed full responsibility for the publication's content.

References

Alameen, G., & Levis, J. M. (2015). Connected speech. In J. M. Levis & M. Reed (Eds.), *The handbook of English pronunciation* (First edition) (pp. 157–174). Wiley Blackwell.

Momen, N., & Pilus, Z. (2022). The effects of connected speech instruction on second or foreign language learners' perceptive skills and connected speech production: A systematic review of the literature (2000-2021). *International Journal of Learning, Teaching and Educational Research, 21*(9), 389–414. https://doi.org/10.26803/ijlter.21.9.22

Mompean, J. A. (2024). ChatGPT for L2 pronunciation teaching and learning. *ELT Journal, 78*(4), 423–434. https://doi.org/10.1093/elt/ccae050

Ting, W. Y., & Kuo, F. L. (2012). Messages behind the unheard sounds: Crossing the word boundaries through songs. *NCUE Journal of Humanities, 5*, 75-92. https://doi.org/10.29602/XLZY.201203.0005

Wong, S. W. L., Dealey, J., Leung, V. W. H., & Mok, P. P. K. (2021). Production of English connected speech processes: an assessment of Cantonese ESL learners' difficulties obtaining native-like speech. *The Language Learning Journal, 49*(5), 581-596. https://doi.org/10.1080/09571736.2019.1642372

Co-Evolving Design Entrepreneurship: An AI Agent for New Venture Creation

Luke Edward Eric FEAST [1]
[1] School of Design, The Hong Kong Polytechnic University, Hong Kong, China

Abstract

Hong Kong requires its design graduates to combine creative problem-solving with entrepreneurial thinking to drive innovation. Traditional design education often lacks practical guidance on entrepreneurship, creating a significant challenge for students starting new ventures. This ongoing project addresses this gap by developing a conversational AI agent (chatbot) that integrates expert entrepreneurs' cognitive patterns with design theory. The AI agent's effectiveness will be rigorously tested via a three-arm randomised controlled trial comparing a control group, a standard foundational AI agent, and the expertise-enhanced AI agent. The anticipated outcome is a scalable pedagogical tool that significantly improves the quality of design students' decisions in venture creation, thereby supporting Hong Kong's innovation economy.

Keywords: Design Cognition, Effectual Entrepreneurship, Conversational AI, Pedagogy, Venture Creation

1. Introduction

To build an innovative economy, Hong Kong needs graduates who can successfully merge creative design with entrepreneurial action. A critical challenge is the disconnect between traditional design education, which emphasises creative problem-solving, and the adaptive, resource-based thinking required to create new ventures. While classical design cognition provides powerful accounts of creative problem-solving through mechanisms such as problem-solution co-evolution (Dorst & Cross, 2001), it often underspecifies the situational factors—such as finance and labour—that are central to entrepreneurial practice. Conversely, entrepreneurship research offers robust models for action under uncertainty, such as effectuation (Sarasvathy, 2022), but pays limited attention to the creative process of designing new products.

This paper details an ongoing research project that aims to bridge this theoretical-practical gap by integrating designerly and entrepreneurial thinking into a unified framework. The project's ultimate contribution to higher education is the development and validation of an expertise-enhanced conversational

AI agent. This tool is intended to provide design students with just-in-time, customised guidance for new venture creation, effectively scaling the cognitive expertise of successful entrepreneurs (Christensen et al., 2023).

Our goal is to answer the question: *Does a conversational AI agent trained on expert entrepreneurial transcripts improve design students' ability to create new ventures?* This project represents a pedagogical innovation with the potential to transform entrepreneurship education from mentor-dependent, resource-intensive teaching to scalable, AI-mediated cognitive scaffolding that maintains the richness of expert guidance while enabling personalised learning at scale.

2. Methodology

The project employs a three-phase methodology that progresses from theory development to empirical evidence and, finally, to technological application.

2.1. Theory Development and Operationalisation

First, we have developed a theoretical framework that integrates design cognition, specifically problem-solution co-evolution (Dorst & Cross, 2001), with effectual entrepreneurship (Sarasvathy, 2022; Sarasvathy & Venkataraman, 2021). This framework explains how entrepreneurs actively reshape the business context by leveraging resources, reframing problems, and committing to ventures despite uncertainty. The theory will be operationalised into a rigorous coding manual that translates theoretical concepts into observable behaviours for analysis.

2.2. Comparative Cognitive Study

Second, the main study will compare the cognitive and decision-making patterns of 20 expert entrepreneurs and 20 experienced designers. This sample size is consistent with established standards in think-aloud protocol research (Cross, 2023). A concurrent think-aloud protocol will be used, in which participants verbalise their thoughts while solving standardised, realistic venture-creation scenarios. Audio recordings will be transcribed and anonymised, with two independent coders annotating the transcripts using the developed coding manual. MANOVA will identify significant quantitative differences between the groups' cognitive patterns. The outcome will be a high-quality dataset of annotated transcripts and validated cognitive profiles that differentiate entrepreneurial thinking.

2.3. AI Agent Development and Testing

Third, the cognitive insights from the comparative study will be used to build and test the AI agent. The conversational AI agent will be built on a state-of-the-art foundation model and will use Retrieval-Augmented Generation (RAG). The annotated expert transcripts will form a domain corpus, embedding expert reasoning patterns into a vector store, allowing the agent to supplement its responses with contextualised retrievals of specialist thinking.

The agent will be evaluated through a three-arm randomised controlled trial involving 120 design students (40 per condition), stratified by prior experience:

(a) **Control:** Participants use worksheets and standardised guidelines;
(b) **Baseline:** Participants interact with a standard foundational AI agent;
(c) **Treatment:** Participants interact with the expertise-enhanced AI agent.

The primary outcome measure is a Decision Quality Score (DQS), determined by an expert panel rating written responses to entrepreneurial scenarios on a 7-point scale using standardised rubrics. Crucially, the trial includes a dedicated transfer phase designed to assess authentic learning.

2.4. Transfer Phase: Design and Rationale

A fundamental methodological concern in AI-assisted learning research is distinguishing genuine cognitive development from performance gains attributable solely to the availability of tools. To address this, the RCT incorporates a transfer phase in which all participants complete novel entrepreneurial scenarios without any AI assistance. This phase occurs after the main intervention period and presents participants with scenarios that are structurally similar but contextually distinct from training materials.

Authentic learning will be evidenced by Treatment group participants maintaining significantly higher DQS scores during this unassisted phase compared to the Control and Baseline groups. Such maintenance would indicate that expert cognitive patterns have been internalised rather than merely accessed. Conversely, if the treatment group scores decline to Baseline levels during transfer, this would suggest the AI functioned primarily as a performance scaffold rather than a learning tool. The transfer scenarios are calibrated to require the same decision-making competencies as the training scenarios—resource identification, adaptive problem framing, and stakeholder negotiation—thereby enabling a direct comparison between assisted and unassisted performance.

3. Anticipated Results and Discussion

As this is an ongoing study, the following discusses hypothesised findings and their anticipated implications. The project hypothesises that the expertise-enhanced AI agent (Treatment) will significantly outperform both the Control and Baseline conditions. ANCOVA will compare post-test DQS scores across the three arms, controlling for baseline performance. Key pairwise contrasts will test the utility of general AI (Baseline vs. Control), the value of expert knowledge integration (Treatment vs. Baseline), and overall system effectiveness (Treatment vs. Control).

We anticipate that the cognitive study will support H1 (Expert entrepreneurs prefer adaptive approaches consistent with effectual logic) and H2 (Expert entrepreneurs prefer resource-based approaches over goal-driven planning), aligning with established effectuation theory (Sarasvathy, 2022). The subsequent AI trial is expected to show measurable gains in DQS for the Treatment group. Specifically, we hypothesise that expert experiential knowledge can be successfully codified and integrated into a conversational AI agent, and that AI-assisted learning informed by expert cognitive patterns will significantly enhance student performance beyond what is achieved with a generic AI tool or no tool.

Most critically, we anticipate that performance improvements will be maintained during the unassisted transfer phase. This finding would provide evidence of authentic skill transfer and internalisation of expert reasoning patterns, rather than mere dependence on the system during use.

4. Pedagogical Implications

Should the anticipated results be confirmed, this study will offer significant implications for educational practice and curriculum design within higher education.

Regarding curriculum transformation, the validated AI agent could be integrated into design entrepreneurship courses as a scalable mentorship supplement. Rather than replacing human instruction, the tool would provide students with 24/7 access to guidance informed by expert cognitive patterns, enabling instructors to focus classroom time on higher-order collaborative activities while the AI handles individualised concept reinforcement.

For teaching practice, the methodology demonstrates how tacit expert knowledge can be systematically captured, annotated, and embedded into pedagogical tools. This RAG-based approach offers a replicable blueprint for other disciplines seeking to scale expertise-informed instruction and address the per-

sistent challenge of providing high-quality mentorship in resource-constrained environments.

In assessment design, the transfer phase methodology provides educators with a rigorous approach to evaluating whether AI-assisted learning leads to genuine competency development. This distinction is essential for responsible AI integration, ensuring that educational technologies enhance rather than substitute for authentic cognitive growth.

For the broader adoption of expert-informed GenAI tools in higher education, demonstrated effectiveness would support institutional investment in domain-specific AI development rather than reliance on generic tools. The project's open science commitment—making anonymised datasets, the theoretical framework, and coding manual publicly available through the Design Thinking Research Symposium community—would facilitate international replication and adaptation.

Finally, by equipping Hong Kong's design students with more effective entrepreneurial thinking skills, the project directly supports the innovation economy (Sarasvathy & Venkataraman, 2021), thereby enhancing graduate employability and venture-creation capacity.

Ethical Declaration Statement

The PolyU Institutional Review Board has approved the ethics application related to this project. All procedures will adhere to strict ethical standards, including informed consent, anonymisation, and secure data storage. The transfer phase design prevents AI dependency by measuring unassisted performance; the expert corpus will be audited for bias, and all participants will receive access to the enhanced agent after the study.

Generative AI Policy

The author used Gemini 3 and Claude Opus 4.5 during the preparation of this work to assist in rephrasing sentences for clarity and adherence to formatting guidelines. After utilising the tool, the author thoroughly reviewed and edited the content and assumed full responsibility for the publication's content.

References

Christensen, B. T., Arendt, K. M., McElheron, P., & Ball, L. J. (2023). The design entrepreneur: How adaptive cognition and formal design training create entrepreneurial self-efficacy and entrepreneurial intention. *Design*

Studies, 86, 101180. https://doi.org/10.1016/j.destud.2023.101181

Cross, N. (2023). *Design Thinking: Understanding how Designers Think and Work* (2nd ed.). Bloomsbury.

Dorst, K., & Cross, N. (2001). Creativity in the design process: co-evolution of problem–solution. *Design Studies*, 22(5), 425–437. https://doi.org/10.1016/S0142-694X(01)00009-6

Sarasvathy, S. D. (2022). *Effectuation: Elements of Entrepreneurial Expertise* (2nd ed.). Edward Elgar. https://doi.org/10.4337/9781848440197

Sarasvathy, S. D., & Venkataraman, S. (2021). An Effectual Entrepreneurial Model of Change: Acting on, reacting to, and interacting with markets as artefacts. In M. S. Poole & A. Van de Ven (Eds.), *The Oxford Handbook of Organisational Change and Innovation* (pp. 131–158). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198845973.013.7

From Prompts to Persuasion: Integrating AI in EFL Copywriting for Critical and Multimodal Literacy

Jo Shan FU [1] & Selim BEN-SAID [2]
[1] Department of English and Communication, The Hong Kong Polytechnic University, Hong Kong
[2] Department of Foreign Languages and Literature, National Sun Yat-sen University, Taiwan

Abstract

Artificial intelligence (AI) is rapidly transforming the ways in which language learners produce, revise, and reflect upon writing. Within English as a Foreign Language (EFL) education, copywriting offers a practical and creative context for developing persuasive reasoning, linguistic precision, and multimodal awareness. This study explores how generative and multimodal AI tools—such as ChatGPT, D-ID, and HeyGen—can function as cognitive partners in EFL copywriting tasks. Drawing on sociocultural and critical-thinking frameworks, the research investigates how learners utilise AI to co-construct meaning, regulate their learning processes, and reflect upon issues of authorship and ethics. Data derived from graduate students' reflective portfolios and AI-assisted video projects were analysed using a rigorous, multi-phase thematic analysis incorporating peer debriefing and triangulation. The findings demonstrate that AI-mediated multimodal writing promotes rhetorical precision, self-regulation, and ethical awareness. Consequently, the paper proposes a pedagogical model positioning AI as a creative and reflective collaborator that facilitates the development of critical and multimodal literacies in higher education.

Keywords: artificial intelligence, EFL pedagogy, copywriting, multimodal literacy, critical thinking

1. Introduction

The growing presence of artificial intelligence (AI) in education is fundamentally reshaping how learners conceptualise writing and communication. Within English as a Foreign Language (EFL) contexts, copywriting—defined as writing designed to persuade or inform specific audiences—demands creativity, linguistic accuracy, and sensitivity to context. However, classroom writing practices often prioritise traditional essays, thereby overlooking the multimodal and critical dimensions of persuasion (Kiili et al., 2021).

Generative AI tools, such as ChatGPT, alongside multimodal applications like D-ID and HeyGen, offer novel possibilities for pedagogical innovation. These technologies provide instant linguistic feedback, stylistic models, and opportunities to visualise or vocalise text. When students transform a written advertisement into an AI-generated talking-head video, they learn to adapt tone, rhythm, and vocabulary to suit their target audience. This process transforms writing from a static exercise into an embodied, performative act that seamlessly blends text, image, and sound (Mayer, 2009).

Furthermore, these technologies support higher-order reasoning and reflection. Learners must critically evaluate which AI suggestions to retain, modify, or reject—effectively transforming machine interaction into a process of critical thinking and metacognitive awareness. From a sociocultural perspective, AI serves as a mediational tool that scaffolds performance within learners' zones of proximal development (Thorne & Lantolf, 2006). From a critical-thinking perspective, it strengthens analytical judgment and ethical discernment (Halpern & Dunn, 2022).

While existing studies have demonstrated that AI can enhance coherence and vocabulary in academic writing (Aslan et al., 2025), few have examined its role in genre-specific, persuasive, or multimodal communication. This study addresses this gap by exploring how AI-supported copywriting tasks facilitate the development of rhetorical precision, self-regulation, and ethical awareness among EFL learners.

2. Methodology

2.1 Research Design

This qualitative study employed thematic analysis to investigate how learners experience AI as a cognitive partner in copywriting tasks. The research was guided by two primary questions:

- **RQ1**: How do EFL learners interpret the role of AI tools as cognitive partners within copywriting tasks?
- **RQ2**: How does AI-mediated multimodal composition influence learners' rhetorical, metacognitive, and ethical development?

2.2 Participants

Fifteen graduate students enrolled in an in-service EFL master's programme at a university in central Taiwan participated in the study. Their English proficiency ranged from intermediate to advanced, and they represented diverse professional fields, including education, business, and design—contexts where persuasive writing is professionally relevant. Participation was voluntary,

and all data were handled with strict confidentiality.

2.3 Data Collection

Two primary sources of qualitative data were gathered over the course of a 15-week semester:

1. **Learning Portfolios:**
   Students submitted weekly reflections documenting their utilisation of AI tools, rationales for tool selection, rhetorical choices, and considerations regarding creativity and ethics. These reflections provided insight into their evolving relationship with AI. To guide these reflections, participants responded to specific prompts, such as: *"Which AI tool did you use this week, and why did you select it?"*; *"How did you evaluate or modify the AI's output?"*; *"What did you learn about your own writing process through this interaction?"*; and *"Did any ethical questions arise regarding authorship or originality?"*

2. **Final Multimodal Projects:**
   Each participant created an AI-assisted persuasive video utilising tools such as D-ID, HeyGen, and Canva, accompanied by a 150–300-word reflective commentary. The commentary detailed the design process, creative choices, and stance on AI authorship. Participants were guided by prompts including: *"How did you adapt your writing for a multimodal format (text, image, voice)?"*; *"What rhetorical strategies did you employ to persuade your target audience?"*; *"How did you evaluate and refine AI-generated content?"*; and *"What is your stance on AI authorship, and how did you ensure originality in your final product?"* Collectively, these multimodal artefacts illustrated how students linked linguistic precision to visual and auditory expression.

These two data sources—the weekly learning portfolios and final project commentaries—were analysed using systematic thematic analysis (Braun & Clarke, 2006), following a rigorous multi-phase procedure to ensure reliability and transparency. The analysis was conducted primarily by the author, an experienced qualitative researcher and instructor of the course. To enhance analytical rigour and address potential researcher bias stemming from the dual instructor-researcher role, a second coder—a doctoral student trained in qualitative methods—independently coded a randomly selected subset of the data (approximately 20%). Inter-rater reliability was assessed through collaborative discussion and consensus-building, with any discrepancies resolved via dialogue and refinement of the codebook. This process yielded a high level of coding consistency across the two coders.

Phase 1: Initial Coding employed a hybrid deductive-inductive approach. Deductive codes were derived from the research questions and theoretical framework (e.g., *AI_partnership_dynamics*, *Rhetorical_adaptation*, *Metacognitive_monitoring*, *Ethical_awareness*), while inductive coding captured emergent patterns (e.g., *Tool_integration_complexity*, *Cross_modal_translation*, *Technical_frustration*, *Strategic_prompting*). All codes were organised into a comprehensive codebook containing explicit definitions, inclusion criteria, exclusion criteria, and illustrative exemplars drawn directly from the data. For example:

- AI_partnership_dynamics was defined as «statements describing the perceived role of AI in the creative process,» with inclusion criteria including «references to collaboration, co-creation, or negotiation with AI» and an exemplar: «I treated ChatGPT as a collaborative thinking partner to expand my initial ideas.»
- Cross_modal_translation was defined as «challenges or strategies related to adapting textual content for visual or auditory modes,» with inclusion criteria including «references to aligning script with video, voice, or image" and an exemplar: "I had to shorten my sentences because the AI avatar spoke too slowly with long text."

Phase 2: Theme Development involved the iterative review and refinement of coded data using NVivo 14. The researcher actively generated initial thematic maps, visually grouping and linking codes to explore relationships and higher-order patterns. Related codes were clustered into thematic families through an iterative process of comparison and consolidation. For RQ1 (AI as Cognitive Partner), codes including *AI_collaboration_levels*, *Agency_negotiation*, and *Ethical_positioning* were synthesised into the overarching theme "Dimensions of AI Partnership: From Instrumental Tool to Creative Collaborator." For RQ2 (Competency Development), codes such as *Rhetorical_refinement*, *Metacognitive_evaluation*, and *Ethical_decision_making* coalesced into the central theme "Critical Mediation Cycle: Evaluation, Adaptation, and Refinement."

Phase 3: Robustness Verification employed three safeguards to ensure analytical trustworthiness:

1. **Analytic Memoing:** Maintained a continuous audit trail documenting coding rationales, interpretive decisions, and reflexive considerations regarding the potential influence of teaching relationships on data interpretation.
2. **Structured Peer Debriefing:** Engaged two external colleagues with expertise in qualitative research and educational technology for a critical

review of the coding framework and thematic organisation. This process provided independent validation and challenged analytical assumptions, prompting richer interpretations of emergent patterns.

3. **Cross-Temporal Triangulation:** Compared thematic patterns across different data sources and collection points, verifying that emergent themes manifested consistently in both process-oriented weekly portfolios and outcome-focused final reflections. This process also revealed nuanced developmental trajectories.This rigorous analytical procedure facilitated a comprehensive understanding of how EFL learners conceptualise AI as a cognitive partner (RQ1) while demonstrating how AI-mediated multimodal composition fosters the development of rhetorical precision, metacognitive control, and ethical awareness (RQ2). The systematic approach ensured that findings remained grounded in empirical evidence while accommodating the complex, multifaceted nature of AI-assisted learning experiences.

The analysis process highlighted how AI shaped both the creative process and learners' understanding of authorship and ethics. The alignment between data sources, research questions, and analytical focus is summarised in Table 1.

Rewritten Content

Table 1: *Data Collection Methods, Artifacts, and Corresponding Research Focus*

Data Collection Method	Primary Artifacts	Research Questions Addressed	Analysis Focus
Learning Portfolios	• Weekly written reflections (digital documents)	**RQ1 & RQ2**	Learner perceptions, rationale for tool selection, evolving relationship with AI, and metacognitive and ethical reflections.
Final Multimodal Project	• AI-assisted persuasive video • 150–300-word reflective commentary	**Primarily RQ2**	Analysis of rhetorical and multimodal choices; application of linguistic precision across audio-visual modes; stance on authorship and final creative responsibility.
Thematic Analysis Process	Coded data from both sources	**RQ1 & RQ2**	Identification of patterns relating to AI's role as a partner, and the development of rhetorical, metacognitive, and ethical awareness.

Note. *The triangulation of data from portfolios and projects ensured depth and consistency in interpreting learners' experiences.*

2.4 Ethical Considerations

All participants provided informed consent. Pseudonyms were used

throughout, and participation had no bearing on academic grades. The study fully complied with institutional ethical standards for human research.

3. Results and Discussion

Three key themes emerged from the data.

3.1 AI as a Creative and Cognitive Partner

Students perceived AI not as a replacement for critical thinking but as a collaborator that assisted them in generating and refining ideas. ChatGPT was frequently described as "a sparring partner" that offered linguistic models or rhetorical suggestions. Learners utilised AI outputs as starting points, subsequently editing them for authenticity and tone. This dynamic mirrored the transition from external guidance to self-regulated learning (Lee & Lee, 2024). Rather than diminishing creativity, AI prompted students to articulate their intentions with greater clarity.

3.2 Building Rhetorical and Multimodal Precision

Upon converting text scripts into short films, students became acutely aware of audience perception and emotional resonance. Adjusting sentence rhythm and intonation for speech compelled them to reconsider word choice and coherence. Tools such as HeyGen and D-ID enabled them to visualise their writing, thereby linking linguistic form to gesture and expression. These experiences fostered their multimodal literacy—the capacity to communicate meaning through multiple modes.

3.3 Reflection, Self-Regulation, and Ethics

Engaging with AI required learners to evaluate reliability, detect bias, and ensure originality. Many students described feeling "responsible" for final decisions, even when AI contributed suggestions. They reflected on the nature of authorship, determined when to acknowledge AI assistance, and considered how to balance efficiency with integrity. Through these reflections, learners demonstrated higher-order thinking skills aligned with ethical reasoning frameworks (Zhai et al., 2024).

4. Implications

4.1 Pedagogical Insights

- AI as a Thinking Partner: Framing AI as a cognitive collaborator fosters learner autonomy and critical engagement rather than passive dependence.
- Reflection-Driven Learning: Structured reflection tasks enable students

to assess their creative decisions and ethical stances.
- Multimodal Expression: Integrating text, image, and sound prepares learners for authentic communication across digital platforms.

4.2 Theoretical Contributions

This study supports sociocultural and critical-thinking perspectives on learning as a mediated, reflective process. By extending learners' cognitive space, AI tools facilitate experimentation, evaluation, and meaning-making across multiple modes. These findings align with emerging views of human–AI collaboration as co-learning rather than substitution (Lu et al., 2025).

4.3 Practical Recommendations

Educators can design copywriting tasks that guide students to:

1. Draft persuasive messages with AI support;
2. Transform those messages into multimodal products (e.g., advertisements, pitches, or digital stories); and
3. Critically reflect on linguistic, rhetorical, and ethical dimensions.

This progression facilitates a shift for learners from technical skill development toward creative judgment and self-awareness.

4.4 Contributions to the GaPI Conference Agenda

This research directly addresses GaPI's core mission of exploring how generative AI can augment higher education and prepare students for future challenges. By investigating AI as a cognitive partner in EFL copywriting, the study offers three key contributions to the conference's global discourse:

- **Advancing Pedagogical Innovation:** The study's phased approach to AI integration—progressing from foundational skill development to guided collaboration and finally autonomous synthesis—provides a replicable model for incorporating generative AI into discipline-specific contexts. This responds to GaPI's call for innovative approaches that transcend generic AI literacy in favour of genre-specific applications designed to develop students' professional communication competencies.
- **Fostering Critical and Ethical AI Engagement:** In an era of rapid generative AI adoption, this research demonstrates how structured reflection and multimodal composition can cultivate learners' critical thinking and ethical awareness—competencies identified as essential for responsible AI use across global higher education contexts. The

findings offer evidence-based strategies for educators worldwide who seek to balance technological integration with humanistic values.

- **Addressing Multimodal Literacy in a Digital Age:** As communication increasingly shifts toward multimodal, AI-mediated platforms, this study provides empirical insights into how learners develop rhetorical precision across text, image, and sound. These findings offer significant relevance for international educators grappling with how to prepare students for authentic, digitally mediated communication in diverse cultural and professional settings.

By illuminating how AI-mediated learning can foster both technical proficiency and critical consciousness, this research contributes to GaPI's vision of reframing higher education for an AI-saturated world. The study's focus on EFL learners in Taiwan offers a valuable East Asian perspective, enriching the global conversation on generative AI's role in fostering inclusive, culturally responsive pedagogies. These insights align with the conference's aim to bring together diverse voices to collectively shape the future of AI-augmented higher education.

Acknowledgments

The participating students are gratefully acknowledged for their collaboration and support throughout the study.

Ethics Statement

This research was approved by the Research Ethics Committee of National Yunlin University of Science and Technology, Taiwan.

AI Usage Statement

All analysis, interpretation, and revisions were conducted and verified by the author, who assumes full responsibility for the final content.

References

Aslan, S., Alyuz, N., Li, B., Durham, L. M., Shi, M., Sharma, S., & Nachman, L. (2025). An early investigation of collaborative problem solving in conversational AI-mediated learning environments. *Computers and Education: Artificial Intelligence*, *8*, 100393. https://doi.org/10.1016/j.caeai.2025.100393

Halpern, D. F., & Dunn, D. S. (2022). *Thought and knowledge: An introduction to critical thinking* (6th ed.). Routledge. https://doi.org/10.4324/9781003025412

Kiili, C., Smith, B. E., Räikkönen, E., & Marttunen, M. (2021). Students' interpretations of a persuasive multimodal video about vaccines. *Journal of Literacy Research*, *53*(2), 196–218. https://doi.org/10.1177/1086296x211009296

Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural theory and the genesis of second language development*. Oxford University Press.

Lee, H., & Lee, J. H. (2024). The effects of AI-guided individualized language learning: A meta-analysis. *Language Learning & Technology*, *28*(2), 134–162. https://hdl.handle.net/10125/73575

Lu, J., Yan, Y., Huang, K., Yin, M., & Zhang, F. (2025). Do we learn from each other: Understanding the human-AI co-learning process embedded in human-AI collaboration. *Group Decision and Negotiation*, *34*(2), 235–271. https://doi.org/10.1007/s10726-024-09912-x

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/cbo9780511811678

Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments*, *11*(1), Article 28. https://doi.org/10.1186/s40561-024-00316-7

Reimagining Teaching and Assessment with AI: A Learn by Design Philosophy and Case Studies

Hongyan GENG [1]
[1] Division of Science, Lingnan University, Hong Kong, China

Abstract

This paper reimagines teaching and assessment through the "Learn by Design in the Digital Era" philosophy. It articulates three pedagogical lenses, namely metacognitive reflection, critical interrogation, and empathetic simulation, and implements five pedagogical designs to empirically evaluate the effectiveness of AI within each lens. Findings indicate that generative AI is highly effective in fostering metacognitive reflection and critical interrogation. Preliminary evidence from small-sample studies suggests that AI may also support the development of empathetic simulation, though further research is required. The paper concludes with implications for pedagogical practice, curriculum design, and faculty development, offering guidance on integrating AI to advance deeper and more reflective learning.

1. Introduction

The rapid rise of generative AI has triggered a paradigm shift across the academic landscape. This shift is not only fundamentally altering established models of teaching and learning but also posing significant challenges to traditional assessment. Prevailing responses are often preoccupied with preserving academic integrity through detection and prohibition (Perkins & Joe, 2024), which overlooks a more profound imperative: the need to fundamentally reimagine teaching and learning for a new digital era by harnessing AI to achieve deeper, more meaningful learning.

In response to this challenge, we introduce and elaborate on a teaching philosophy termed "Learn by Design in the Digital Era." Grounded in constructionist pedagogy—a student-centred approach where learners actively construct their own knowledge by connecting new information to prior experiences rather than passively receiving it—this philosophy treats learning as the active creation of tangible artifacts (Papert & Harel, 1991). In AI contexts, students are not passive consumers but active designers, critical interlocutors, and strategic architects of their learning. Guided by backward design (Wiggins & McTighe, 2005), in this philosophy, technology serves as a catalyst for cognitive development rather than merely a novelty.

This paper demonstrates the practical application and efficacy of this philosophy through an analysis of five concrete pedagogical designs implemented across diverse disciplinary and international contexts. First, we delineate the pedagogical lenses under the "Learn by Design" philosophy. Second, we present and discuss the results from five pedagogical designs, analysing how they foster specific competencies. Finally, we outline the broader implications for pedagogical practice, curriculum design, and faculty development in the age of AI.

2. Methodology

2.1 Three Pedagogical Lenses

The "Learn by Design in the Digital Era" philosophy is enacted through three complementary pedagogical lenses: metacognitive reflection, critical interrogation, and empathetic simulation. Figure 1 illustrates the hierarchical and interdependent relationship among these lenses. Metacognitive reflection serves as the foundational layer through which the twin pillars of critical interrogation and empathetic simulation are built (Fig. 1). This structure shows that a student's capacity to critically evaluate AI-generated content or empathetically engage with simulated scenarios is underpinned by their ability to first reflect on their own knowledge, biases, and thinking processes.

Within each lens, thoughtful pedagogical designs (presented in Table 1) were implemented to empirically assess the effectiveness of AI in fostering these specific competencies.

- **Metacognitive Reflection** forms the foundational layer, enabling higher-order skills. It is where learners examine and refine their own thinking. It helps students examine how they think, uncover misconceptions, and identify knowledge gaps. It serves as a prerequisite for the subsequent layers of critical interrogation and empathetic simulation. In this context, AI's role is that of a *tutor*, and the student's role is that of an *active designer*. Four pedagogical designs were developed for metacognitive reflection (Table 1).
- **Critical Interrogation** is designed to strengthen students' critical thinking by positioning AI as an analytical reviewer. Students collaborate with the AI to identify methodological weaknesses, evaluate evidence, and articulate evidence-based critiques. The AI's role is *analyst*, and the student's role is that of a *critical interlocutor*. Three pedagogical designs were developed for critical interrogation (Table 1).
- **Empathetic Simulation** leverages AI to model complex, real-world scenarios and diverse human perspectives. It provides a sound environ-

ment for practising skills and developing empathy prior to real-world application. AI's role is that of a *dynamic simulator*, and the student's role is that of a *strategic architect*. Two pedagogical designs were developed for empathetic simulation (Table 1).

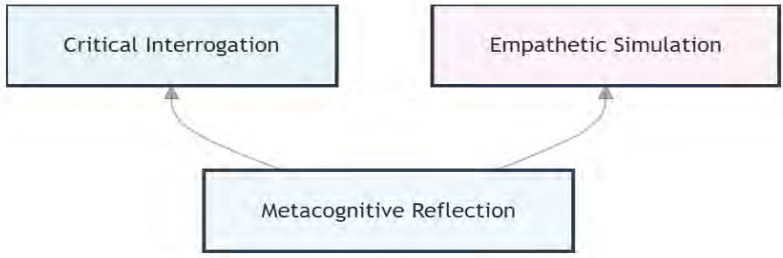


Figure 1. The hierarchical and interdependent structure of the three pedagogical lenses. *(Note: Metacognitive Reflection forms the foundational layer, enabling the higher-order skills of Critical Interrogation and Empathetic Simulation.)*

Table 1. Mapping of Pedagogical Lenses and Pedagogical Designs.

Pedagogical Lenses	**Pedagogical Designs**	**Student Role**	**AI role**
Metacognitive Reflection	- Case 1^: Common Mistake Analytics and Remediation - Case 2^: Short Presentation + Long Q&A - Case 3^: Science News "True or False" (Stage 1) - Case 5^: ESG Simulation Project (Stage 1)	Active Designers	Tutor
Critical Interrogation	- Case 2^: Short Presentation + Long Q&A - Case 3^: Science News "True or False" (stage 2) - Case 5^: ESG Simulation Project (stages 2 and 3)	Critical Interlocutors	Analyst
Empathetic Simulation	- Case 4^: AI-Simulated Client - Case 5^: ESG Simulation Project (stage 3)	Strategic Architects	Dynamic Simulator

- **Case 1^: Common Mistake Analytics and Remediation.** AI is used to identify common mistakes among students' assignments. Based on an analysis of those common mistakes, AI then generates targeted practice exercises and instructional materials.
- **Case 2^: Short Presentation + Long Q&A.** Given the widespread use of AI to generate slide decks, the validity of traditional oral presentation assessments is diminished. To mitigate this, in this activity, students deliver a brief presentation followed by an extended, open-ended Q&A in which peers may pose unlimited questions, enabling more robust evaluation of understanding and reasoning.

- **Case 3^: Science News "True or False".** There are two stages to this activity. In Stage 1, students either curate a verified, evidence-based science news article or, with AI support, compose a plausible but intentionally fabricated science report. In Stage 2, students employ AI to discriminate between authentic articles and fabricated ones, drawing on reputable sources, methodological appraisal, and cross-referencing techniques.
- **Case 4^: AI-Simulated Client.** In service-learning contexts, students often lack an evidence-based understanding of stakeholders' needs and operational constraints. For example, teams planning visits to an elderly care centre may intend to assist the elderly without a clear grasp of the specific support the elderly require. To address this gap, we implement AI-supported pre-service training that simulates daily challenges faced by the elderly, thereby enhancing students' situational awareness and aligning planned activities with actual needs.
- **Case 5^: ESG Simulation Project.** This project includes four courses across four different countries. First, all students engage on a shared digital platform where AI functions as a tutor, guided by a structured prompt library to establish a common foundational ESG knowledge base. Second, students from every two courses use shared themes to conduct thematic case studies with AI serving as a research analyst to synthesise complex data and build disciplinary depth and critical analysis. Third, students from all courses participate in a cross-disciplinary simulation, acting as a global governance body and leveraging AI as a scenario generator to address multi-sector crises by diagnosing problems, anticipating spillover effects, and designing integrated solutions.

Table 1 provides a comprehensive mapping of how specific pedagogical designs operationalise each of the three lenses. The table is structured to clarify the distinct and complementary roles assigned to both the student and the AI in each learning activity. For instance, in the "Science News 'True or False'" activity (Case 3), the student's role evolves from an "Active Designer" in Stage 1 (using AI as a "Tutor" to understand the conventions of science writing) to a "Critical Interlocutor" in Stage 2 (using AI as an "Analyst" to evaluate the veracity of articles). This mapping demonstrates the deliberate design behind each case study, ensuring that the AI's function is precisely aligned with the intended learning outcome.

Questionnaire surveys were administered to assess students' perceptions of the effectiveness of the three pedagogical lenses across the five pedagogical designs.

2.2 AI Roles within the Pedagogical Lenses

This section elaborates on the AI roles within the pedagogical lenses (Table 1).

AI as Tutor: In this role, AI functions as a personalised, on-demand guide to help students externalise and examine their thought processes. Its purpose is not to provide answers, but to prompt student reflection and scaffold self-correction.

- *Example from Case 1:* After students submit an assignment, the AI analyses the aggregate submissions to identify pervasive errors, such as a common misunderstanding of a key concept. It then generates a report for the instructor and creates targeted practice exercises.
- *Example from Case 3 (Stage 1):* When a student asks an AI to write a plausible but fabricated science news story, they usually refine the prompt through several rounds (for example, "make the news sound more authoritative" or "add a credible statistic"). This process works like tutoring: the AI makes the language and structure of science news visible. By watching how the AI builds claims, the student learns the features that make information appear credible.

AI as Analyst: AI serves as a tool for processing and synthesising information, which students then examine critically. The AI supplies raw material or summarised content, but the students are responsible for the higher-order tasks of evaluation, critique, and judgment.

- *Example from Case 3 (Stage 2):* Students ask the AI to analyse both the authentic and fabricated articles obtained in Stage 1. They request it to analyse sources, check internal consistency, or summarise the evidence. In this role, the AI produces a rapid report that flags possible methodological weaknesses and unverified claims. The students then probe the AI's reasoning, verify its points against reputable databases, and determine whether its assessment is sound.
- *Example from Case 5 (Stage 2):* Students from different disciplines use AI to analyse a company's ESG reports. A finance student might ask the AI to assess the financial risks linked to the company's carbon footprint, while a sociology student might ask AI to identify stakeholders and summarise their concerns. Each student then evaluates this analysis through their disciplinary lens, probing the methods and evidence in the company's reporting and identifying potential weaknesses.

AI as Dynamic Simulator: In this role, AI helps create immersive, low-risk settings where students can explore complex situations and encounter dif-

ferent perspectives. The AI is designed to respond dynamically and somewhat unpredictably within clear boundaries, facilitating students to adapt and think strategically.

- *Example from Case 4:* Before visiting an elderly care centre, students interact with an AI chatbot that simulates an older adult with specific needs and limitations (for example, limited mobility, feelings of loneliness, and difficulty using small touchscreens). When students propose activities (such as teaching the elderly to make video calls), the AI responds from the persona's perspective, highlighting practical barriers and preferences. This exchange prompts students to question assumptions, practise active listening, and revise their service plans to be more empathetic and feasible.
- *Example from Case 5 (Stage 3):* Students from all four countries, acting as a global governance body, are presented with a simulated multi-sector crisis generated by the AI (e.g., "A major supplier in your region has been accused of using child labour. The local economy depends on this factory. Social media is calling for an immediate boycott."). As the students deliberate and propose solutions, the AI dynamically introduces new information and spillover effects (e.g., currency devaluation or escalating protests). This dynamic simulation creates a realistic, evolving context that requires students to think strategically, anticipate downstream impacts, and craft integrated, empathetic responses that balance economic, social, and ethical considerations.

3. Results and Discussion

Table 2. Effectiveness of the Pedagogical Lenses.

Pedagogical Lenses	Pedagogical Designs	Effectiveness
Metacognitive Reflection	- Case 1: Common Mistake Analytics and Remediation - Case 2: Short Presentation + Long Q&A - Case 3: Science News "True or False" (Stage 1) - Case 5: ESG Simulation Project (Stage 1)	- 3.8/5 (n=176) - 4.1/5 (n=98) - 4.3/5 (n=184) - 3.4/5 (n=21)
Critical Interrogation	- Case 2: Short Presentation + Long Q&A - Case 3: Science News "True or False" (stage 2) - Case 5: ESG Simulation Project (stages 2 and 3)	- 4.1/5 (n=98) - 4.3/5 (n=184) - 3.4/5 (n=21)
Empathetic Simulation	- Case 4: AI-Simulated Client - Case 5: ESG Simulation Project (stage 3)	- 3.6/5 (n=18) - 3.4/5 (n=21)

Table 2 presents students' perceptions of the effectiveness of AI in supporting each pedagogical lens across the five case studies. These data reveal several patterns with clear practical implications.

First, AI demonstrates strong effectiveness in supporting foundational and analytical skills. When excluding Cases 4 and 5 due to their limited sample sizes (n=18 and n=21 respectively), the average effectiveness scores for Metacognitive Reflection (4.07/5) and Critical Interrogation (4.2/5) are notably high. This suggests that students perceive AI as a valuable tool for both examining their own thinking and critically evaluating information. The practical implication is that course designs positioning AI as a tutor for self-reflection or as an analyst for critique are likely to produce positive learning outcomes. Instructors can therefore integrate AI with confidence into activities that target these core competencies.

Second, the preliminary data on Empathetic Simulation deserve attention. The average score of 3.5/5, drawn from small samples, is lower than the scores for the other two lenses. This could be interpreted in two ways: either AI is genuinely less effective at fostering empathy, or the pedagogical designs for this lens require further refinement. Our qualitative observations from Case 4, where students using the AI-simulated client reported "higher confidence and more nuanced, empathetic project plans," suggest that AI *can* support empathy development, but the effect may be more subtle or difficult to capture through quantitative perception surveys alone. The implication is not to discontinue this approach, but to conduct more rigorous research, using larger samples, mixed methods, and longitudinal designs, to learn how to best use AI for empathetic learning.

Third, the result that students who created false news in Case 3 (Stage 1) became significantly more adept at detecting it in Stage 2 indicates a strong mechanism of active construction, i.e., learning by making, which enhances critical analysis skills more effectively than passive consumption or detection alone. The pedagogical implication is that instruction should prioritise student creation with AI (for example, generating examples, building simulations, or designing outputs) as a pathway to deeper analytical competence.

Fourth, the variation in sample sizes across cases (from n=184 in Case 3 to n=18 in Case 4) highlights practical realities of implementation. Larger, multi-year courses yield more robust data, while newer or smaller-scale innovations (Cases 4 and 5) are still accumulating evidence. The implication for curriculum design and faculty development is that innovation requires sustained support. Promising approaches need time and resources to scale before their full effectiveness can be rigorously evaluated.

4. Implications

The intrusion of AI is extensively shaping the landscape of education (Kasneci et al., 2023; Munaye et al., 2025). This paper discusses its impact from three key perspectives: Pedagogical Practice, Curriculum Design, and Faculty Development.

For Pedagogical Practice, the high effectiveness scores for Metacognitive Reflection (4.07/5) and Critical Interrogation (4.2/5) demonstrate that students value AI not as a source of answers, but as a partner in thinking. This finding supports a necessary evolution in the educator's role. Instructors must move beyond being only content experts to become designers of learning experiences and coaches for critical thinking. The success of the "short presentation + long Q&A" format (Case 2, 4.1/5) exemplifies this shift; because presentations alone can be AI-generated, the focus moves to real-time reasoning, defence of ideas, and process-oriented assessment. In practice, educators should design assessments that evaluate the quality of inquiry, i.e., how students question AI outputs, refine prompts, and explain their reasoning, rather than judging only the final product.

For Curriculum Design, the tiered structure of the ESG project (Case 5) moves students through three stages: foundational literacy with metacognitive reflection (Stage 1), thematic analysis with critical interrogation (Stage 2), and cross-disciplinary, empathetic simulation (Stage 3). This sequence provides a replicable template for programme-level integration. While the effectiveness data for Case 5 are preliminary (3.4/5), the rationale of this progression is supported by the stronger results from the other cases. Specifically, students who created false news became better at detecting it (Case 3) underscores the power of active construction. In practice, curricula should intentionally scaffold AI literacy, guiding students from consumers to creators to critics. Early courses can position AI as a tutor (as in Case 1), intermediate courses can focus on interrogating AI as an analyst (as in Case 3, Stage 2), and capstone experiences can involve navigating AI as a dynamic simulator (as in Case 4). This progression ensures that AI functions as a core, developmental element of students' intellectual growth, rather than as an occasional novelty.

Furthermore, the cross-disciplinary design of the ESG project illustrates AI's value as a catalyst for interdisciplinary learning. By enabling students in different disciplines to work within a shared simulation, AI helps them move beyond disciplinary silos and prepares them for the complex, multifaceted challenges of the contemporary workforce.

For Faculty Development, the variation in effectiveness scores across cases, particularly the lower scores for empathetic simulation (3.5/5), highlights that simply introducing AI tools is insufficient. Effective integration depends on the quality of the pedagogical design that frames the technology. The strong results for Cases 2 and 3 (4.1 and 4.3, Table 2) resulted from carefully structured activities that positioned students as active designers and critical interlocutors. In practice, faculty development should move beyond tool-based training (e.g., "how to use ChatGPT") to cultivate pedagogical design thinking. Workshops should address core questions like: "What learning outcome are we trying to achieve? What role should the AI play? What will the student do? How will the process be assessed?"

The success of the international collaborative project underscores the value of communities of practice, in which instructors from different disciplines co-design, share, and refine AI-enhanced strategies. These communities accelerate innovation, build shared expertise, and help institutions respond adaptively to rapid technological change.

Future Research

The small sample sizes in the empathetic simulation studies (n=18 and n=21) highlight a key institutional implication: meaningful innovation requires sustained investment. Promising pedagogical designs, such as the AI-simulated client (Case 4), need time, resources, and broader implementation to produce strong evidence. Institutions should therefore support iterative cycles of design, implementation, and research, while acknowledging that not every innovation will deliver immediate, large effects.

The most critical step toward harnessing AI's transformative potential is empowering educators with the philosophical framework, practical toolkit, and institutional support to experiment, reflect, and refine.

Acknowledgments

This project was financially supported by the Teaching Development Grants (project codes: 102723 and 102758), Lingnan University.

Ethical Declaration Statement

The project has been approved by the Teaching and Learning Centre of Lingnan University.

Generative AI Policy

The author(s) used ChatGPT during the preparation of this work for English polishing and grammar checking. 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

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (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

Munaye, Y. Y., Admass, W., Belayneh, Y., Molla, A., & Asmare, M. (2025). ChatGPT in Education: A Systematic Review on Opportunities, Challenges, and Future Directions. *Algorithms*, 18(6), 352. https://doi.org/10.3390/a18060352

Papert, S., & Harel, I. (1991). Situating constructionism. *Constructionism*, 36(2), 1-11.

Perkins, M., & Roe, J. (2024). Decoding academic integrity policies: A corpus linguistics investigation of AI and other technological threats. *Higher Education Policy*, 37(3), 633-653. https://doi.org/10.35542/osf.io/z4cru

Wiggins, G. P., & McTighe, J. (2005). *Understanding by design*. ASCD.

Promotion of Technological Social Responsibilities and Academic Integrity through Video-Based Learning in a Leadership Course

Daniel T.L. SHEK [1], Betty MOK [1], Tingyin WONG [1]
[1] Department of Applied Social Sciences, The Hong Kong Polytechnic University, Hong Kong, China

Abstract

With the advance of information technology, there is an urgent need to enhance undergraduates' technological social responsibilities and academic integrity. In the present study, we developed ten videos on technological social responsibilities and academic integrity, which were incorporated into a leadership course at a university in Hong Kong. Adopting a flipped classroom approach, students were required to watch the videos and participate in class activities. A subjective outcome evaluation scale was used to evaluate students' perceptions of the videos (N = 1,748). Results showed that students had positive views of video-based learning, including the perceived benefits on enhancing their technological social responsibilities and academic integrity. The findings suggest that video-based learning is a promising tool to promote students' technological responsibilities and academic integrity.

Keywords: Video-based learning, leadership education, technological social responsibilities, positive youth development, pedagogical innovation

Correspondence: Daniel T.L. Shek, PhD, FHKPS, BBS, SBS, JP, Associate Vice President (Undergraduate Programme), Chair Professor of Applied Social Sciences and Li and Fung Professor in Service Leadership Education, Department of Applied Social Sciences, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, PR China. Email: daniel.shek@polyu.edu.hk

1. Introduction

In Hong Kong, many undergraduates suffer from mental health and developmental issues, but few undergraduate courses have attempted to enhance students' mental health and prevent related problems. Catalano (2004) pointed out that the Positive Youth Development (PYD) approach could help nurture students' capabilities and minimise students' mental health problems. Thus, our leadership course, "Tomorrow's Leaders", was developed based on Catalano's 15 PYD constructs to promote undergraduates' holistic development and

well-being (Shek et al., 2024; Shek & Li, 2025).

Online teaching has been increasingly used to facilitate students' learning, with different interactive pedagogies employed in the higher education sector. For instance, virtual learning and video-based learning were widely used (Shek et al., 2023). However, the virtual learning approach was being challenged as teachers need lots of preparation time and advanced technology to facilitate students' learning (Pope, 2018). Instead, video-based learning was highly accepted because it could help learners learn at their own pace and is easier for educators to prepare teaching materials (Shek et al., 2023).

Besides, as educational technology is continuously evolving, more students use Artificial Intelligence (AI) to complete their assignments in recent years. Students used ChatGPT to summarise or paraphrase their work and seldom examined the policy of using generative AI before submitting their assessments (Yeo, 2023). This becomes a major threat to higher education as students' learning effectiveness and academic integrity are being questioned. Therefore, educating undergraduate students to use AI properly and enhance their technological responsibilities is essential for their holistic development. Using video-based learning could help students learn the guidelines at their own pace, which also provides a channel for students to understand their technological responsibility of using AI.

1.1 Research Gaps

There are few undergraduate courses that adopt video-based learning in promoting technological social responsibilities and academic integrity in students, particularly in the Chinese context. Few evaluation studies have been conducted to understand the effectiveness of using video-based teaching. In this research, we developed 10 videos to promote technological social responsibilities and academic integrity in university students (see Table 1).

Table 1. Ten videos on technological social responsibilities and academic integrity developed in the current project

Video Topic	Video Title	Description
Introduction	Video 1: Service Economy and Technology Enhancement	This introductory video discusses the positive benefits of technology to efficient task completion, its application in the service economy, and its integration into university education and students' personal and professional life.

Video Topic	Video Title	Description
Self-understanding and self-management	Video 2: Accomplishing Lifelong Learning in a Technological World	This video presents a story of a university student who is trying to learn more about the application of AI technology to her future profession by asking for help from a librarian, her classmates, and her professor.
Cognitive competence	Video 3: False Information and Copyright Issues	This video presents a story of a university student using AI tools to search for information for his group project. However, without verifying the information, the student faces the risk of violating academic integrity.
Social-emotional competence	Video 4: Lost in the Virtual World	This video presents a story of a university student feeling very devastated after breaking up with his girlfriend. He resorts to playing online games. But it soon results in addiction and deteriorates physical and mental health. This video aims at raising students' awareness of the risk of being trapped in the "virtual world".
Resilience	Video 5: Physical Activity and Stress-Coping	The video begins with a dialogue between two university students about their academic stress. One of them suggests the use of mobile phone applications, such as "7-minute workout" and mindfulness apps, to improve their physical and mental health. The video also introduces activity trackers that could monitor users' daily activities and physical responses. This video aims at demonstrating how technology can serve as a valuable resource for mental health support and stress coping.
Spirituality	Video 6: Digital Serenity and Navigating the Intersection of AI and Spiritual Well-being	This video begins with a dialogue between two university students about their exam pressure. One of them tries to help the other student by introducing three AI applications—AI-driven meditation and mindfulness apps, AI-powered virtual communities, and AI-guided virtual retreats—that could help them manage stress and explore deeper spiritual experiences. This video shows that proper use of AI applications could enhance the richness of human experiences.
Morality	Video 7: Academic Integrity of PolyU Students When Using GenAI	This is a whiteboard animation video that introduces the university's stance on the use of generative AI and related guidelines for students and provides practical suggestions on using generative AI tools in a responsible, ethical, and effective manner.

Video Topic	Video Title	Description
Cultural competence	Video 8: Benefit of Technology to Mankind	The video begins with three animated characters discussing the benefits of technological advancements in transport and communication, including high-speed trains, 5G technology, and e-commerce. The rapid development of technology has enabled people from various backgrounds to participate in the global marketplace, promoting economic growth and cultural exchange. However, it is equally important for students to understand the drawbacks of being overly dependent on technology, such as the risk related to hacking and excessive Internet use.
Effective communication	Video 9: Case Study on Effective Communication	This video presents a story about a group of university students involved in a cyberbullying incident. The story demonstrates how individuals could use AI technology to create fake pictures and spread rumours, bringing negative consequences to interpersonal relationships and group morale.
Team building	Video 10: Building a Better Tomorrow through Ethics, Teamwork and AI	The Hong Kong Government has proposed three data stewardship values, including "be respectful," "be beneficial," and "be fair," which guide organisations and individuals in the development and use of AI. The video illustrates these values through a case of data misuse in a health technology company.

2. Methodology

In the second semester of the 2024/25 academic year, 1,748 undergraduate students were enrolled in the "Tomorrow's Leaders" course at the Hong Kong Polytechnic University, which is a leadership subject in the General University Requirements (GUR) for all students at the university. 96% of the participants were between 16 and 26 years old. Between 1,005 and 1,309 students completed their evaluation of each of the 10 videos in a voluntary manner (as shown in Table 2). The study has obtained research ethical approval from the institutional review board. All participants have consented to taking part in this study.

Students who registered for the "Tomorrow's Leaders" course were invited to watch one video before or after lectures (from Lecture 1 to 11, except Lecture 8) to gain awareness and knowledge on technological social responsibilities and academic integrity in a voluntary manner. After watching each video, students were invited to complete an evaluation questionnaire to measure their satisfaction and perception towards the video in a voluntary manner.

We have developed a 19-item evaluation scale to evaluate the quality and teaching and learning effectiveness of instructional videos (Noble & Smith, 2015). This scale consists of five subscales: satisfaction with the content of the

video (5 items), implementation or use of the video (4 items), effectiveness of the video (6 items), overall satisfaction with the video (2 items), and overall satisfaction with the lecture (2 items). A six-point Likert scale was applied in each item ("1" represents "strongly disagree" and "6" indicates "strongly agree"). At the end of the semester, the survey data for the 10 videos was compiled and analysed aggregately.

3. Results and Discussion

Results showed that participants generally responded positively to the items in the survey, showing students' positive subjective evaluation of the 10 videos. Items in the survey received positive responses from at least 76.42% of the respondents ("slightly agree" to "strongly agree" to each item; see Table 2). This finding demonstrated that students were generally satisfied with the content, implementation and teaching effectiveness of the videos, and the lectures.

Some examples are as follows:

For the item "the content of the video is stimulating" (Item 5), the average positive response ranged from 91.43% to 95.44% across the 10 videos. The perceived benefits of helping students reflect on the topic (Item 14) ranged from 90.96% to 95.91%. The subscale "overall satisfaction with the video" (Items 16 and 17) got high positive response rates (minimum = 85.46%, maximum = 96.61%, average = 93.74%). In this subscale, the item such as "overall speaking, I am satisfied with the video" had a mean score of 4.97 and an average positive response rate of 95.01% across the 10 videos. The subscale "overall satisfaction with the lecture" (Item 18) received the highest average positive response rates (minimum = 94.12%, maximum = 96.70%, average = 95.35%). For the item "on the whole, I am very satisfied with this lecture" (Item 19), the mean score was 4.98, and the average positive response rate was 95.53% across the 10 videos.

Furthermore, the subscale "content of the video" (Items 1 to 5) also received high positive response rates (minimum = 85.14%, maximum = 97.41%, average = 94.35%). For example, Item 3 "the video is easy to understand" had a mean score of 5.06 and a 95.92% average positive response rate across the 10 videos. Another subscale "effectiveness of the video" (Items 10 to 15) got high positive response rates as well (minimum = 88.21%, maximum = 97.71%, average = 94.47%). For instance, Item 11 "the video has enabled me to understand related concepts" received a mean score of 4.98 and an average positive response rate of 95.64% across the 10 videos. Results from these two subscales show students' positive appraisal of the content and teaching effectiveness of the 10 videos.

Moreover, the subscale "implementation or use of the video" (Items 6 to 9) got positive response rates among the five subscales (minimum = 76.42%, maximum = 96.34%, average = 90.51%). For example, the item "the video-based activity enhanced my interaction with classmates" received a mean score of 4.70 and an average positive response rate of 86.63% across the 10 videos. Additionally, the items related to students' perceived benefits received high positive response rates. For instance, Item 10 "the video-based activity has enabled me to apply the theories in real life cases" received a positive response rate ranging from 88.68% to 96.26%. Also, another perceived benefit, Item 15 "the video can benefit my development", got high positive rates as well. The response rate ranged from 91.27% to 96.26%. The findings indicated that the students derived benefits from watching the 10 videos.

4. Implications

For practical implications, this study showed that video-based teaching could be an effective teaching method to promote positive youth development, as revealed by the subjective outcome evaluation findings. The results suggest that video-based teaching is a promising tool to nurture positive youth development in university students in the higher education context (Shek & Li, 2025). As AI is commonly used by undergraduates, it is essential to adopt a new teaching pedagogy to facilitate students' learning. This study demonstrated students' positive perceptions of using AI after watching the videos. Therefore, this study suggested using video-based learning to promote positive youth development, which could enhance their technological social responsibilities and academic integrity.

In terms of evaluation, this study showed the usefulness of using subjective outcome evaluation to evaluate the teaching and learning effectiveness of instructional videos in higher education. Validated tools are seldom used to evaluate the effectiveness of video-based learning. Subjective outcome evaluation is only commonly used in understanding students' perception towards an intervention (Yu et al., 2022). Thus, this study supported the use of a subjective outcome evaluation scale as a tool to assess the quality of video-based learning.

To conclude, the present study showed that video-based learning is a useful pedagogy to facilitate students' learning of technological social responsibilities and academic integrity. Subjective outcome evaluation was also shown to be an effective methodology to examine students' perceptions towards video-based learning. However, more evidence is essential to understand how video-based learning could help educate students to use AI properly.

Ethical Declaration Statement

The project has been approved by the Institutional Review Board (or its Delegate) at Hong Kong Polytechnic University (Reference Number: HSEARS20230907002).

Generative AI Policy

The authors have not used any generative AI tool during the preparation of this work.

References

Catalano, R. F., Berglund, M. L., Ryan, J. A. M., Lonczak, H. S., & Hawkins, J. D. (2004). Positive youth development in the United States: Research findings on evaluations of positive youth development programs. *The Annals of the American Academy of Political and Social Science, 591*(1), 98–124. https://doi.org/10.1177/0002716203260102

Noble, H., & Smith, J. (2015). Issues of validity and reliability in qualitative research. *Evidence-Based Nursing, 18*(2), 34–35. https://doi.org/10.1136/eb-2015-102054

Pope, H. (2018). Chapter 2. Incorporating Virtual and Augmented Reality in Libraries. *Library Technology Reports, 54*(6), 8.

Shek, D. T. L., & Li, X. (2025). Development, implementation, and evaluation of a leadership subject for university students — "Tomorrow's Leaders". In D. T. L. Shek (Ed.), *Promoting holistic development in university students* (pp. 203–238). Springer. https://doi.org/10.1007/978-981-96-0251-3_9

Shek, D. T. L., Li, X., & Wong, T. (2023). Video teaching materials for leadership education: Subjective outcome evaluation based on university students in Hong Kong. *International Journal of Child Health and Human Development, 16*(1), 73–87.

Shek, D. T. L., Li, X., Wong, T. Y., Chak, L. Y. Y., Cheng, N. S., Chiu, T. Y., Chu, K. M. C., Ho, W. W. L., Lam, W. S., Mok, P. W. B., Shek, K. F., & Xiao, C. (2024). Enhancement of technological social responsibilities in university students: Development of virtual teaching and learning videos. *International Journal of Child Health and Human Development, 17*(3), 159–169.

Shek, D. T. L., Mok, B., & Wong, T. (in press). Promotion of technological social responsibility through video-based learning in a credit-bearing leadership subject. *International Journal of Child Health and Human Development*.

Yeo, M. A. (2023). Academic integrity in the age of Artificial Intelligence (AI) authoring apps. *TESOL Journal, 14*(3), Article e716. https://doi.org/10.1002/tesj.716

Yu, L., Lin, L., Shek, T. L. D., & Chai, W. (2022). Students' Perceived Attributes and Benefits of a Leadership Course: Subjective Outcome Evaluation. *Research on Social Work Practice, 32*(2), 199–214. https://doi.org/10.1177/10497315211042823

Table 2. Positive Response Rates of Post-Lecture Video Evaluation for the 10 Videos.

	Questionnaire Item	L1 N = 1,272	L2 N = 1,222	L3 N = 1,265	L4 N = 1,301	L5 N = 1,309	L6 N = 1,197	L7 N = 1,298	L9 N = 1,149	L10 N = 1,097	L11 N = 1,005
		Positive Response Rate %	Positive Response Rate %	Positive Response Rate %	Positive Response Rate %	Positive Response Rate %	Positive Response Rate %	Positive Response Rate %	Positive Response Rate %	Positive Response Rate %	Positive Response Rate %
1	The video is well designed.	93.55	93.04	95.65	94.39	93.74	94.40	95.53	96.17	95.72	94.53
2	The video is relevant to the lecture topic.	93.63	95.74	96.60	95.77	95.11	95.24	94.99	96.17	96.17	95.62
3	The video is easy to understand.	97.41	96.64	97.08	95.85	94.73	95.15	95.22	96.52	95.44	95.12
4	The video is interesting.	85.14	89.12	91.78	91.85	91.37	91.23	93.37	95.21	94.17	94.03
5	The content of the video is stimulating.	91.43	92.88	93.75	94.54	93.89	93.48	94.14	95.30	95.44	94.43
6	The teacher used the video effectively.	94.42	95.58	96.21	95.47	93.43	94.40	95.22	96.34	95.53	95.52
7	The video-based activity enhanced my interaction with the teacher.	82.31	86.01	87.11	87.86	87.17	89.14	89.29	92.95	92.43	92.24
8	The video-based activity enhanced my interaction with classmates.	76.42	82.41	85.14	86.24	85.33	87.22	88.60	91.47	90.98	92.44
9	I was very engaged in the video-based activity.	86.08	89.36	91.46	90.70	91.37	92.65	91.91	94.34	94.07	93.43
10	The video-based activity has enabled me to apply the theories in real life cases.	88.68	94.35	95.49	94.47	93.20	94.82	95.38	96.26	95.35	95.12

	Questionnaire Item	L1 N = 1,272	L2 N = 1,222	L3 N = 1,265	L4 N = 1,301	L5 N = 1,309	L6 N = 1,197	L7 N = 1,298	L9 N = 1,149	L10 N = 1,097	L11 N = 1,005
		Positive Response Rate %	Positive Response Rate %	Positive Response Rate %	Positive Response Rate %	Positive Response Rate %	Positive Response Rate %	Positive Response Rate %	Positive Response Rate %	Positive Response Rate %	Positive Response Rate %
11	The video has enabled me to understand related concepts.	94.42	95.91	97.71	95.70	94.50	94.49	95.84	96.34	95.99	95.52
12	The video has deepened my understanding of the topic.	92.92	94.93	96.21	95.00	94.50	95.66	95.30	95.82	96.26	95.32
13	The video has enhanced my interests in the topic.	88.21	90.26	91.94	93.08	92.36	92.98	94.07	95.21	94.44	94.93
14	The video has helped me to reflect on the topic.	90.96	93.70	95.10	94.24	92.74	94.15	95.69	95.91	95.17	94.93
15	The video can benefit my development.	91.27	94.44	95.49	94.70	93.35	94.74	95.76	96.26	95.62	94.93
16	I prefer using video in the teaching and learning of this subject.	85.46	90.26	92.57	93.39	92.74	93.48	93.14	95.56	94.17	94.03
17	Overall speaking, I am satisfied with the video.	92.30	94.19	95.57	95.16	94.42	95.57	95.45	96.61	95.99	94.83
18	Overall speaking, the lecture has enriched my development.	95.05	95.01	95.18	95.00	94.35	95.32	95.07	95.91	95.81	95.12
19	On the whole, I am very satisfied with this lecture.	96.70	95.25	95.81	95.47	94.12	94.49	95.45	96.61	96.44	94.93

The Impact of AI on Learners with Basic-Intermediate English Language Proficiency in Virtual Exchange (VE)

Emma ZHANG [1]
[1] Language Centre, Hong Kong Baptist University, Hong Kong, China

Abstract

This study compares two Virtual Exchange (VE) programmes at Hong Kong Baptist University (HKBU): one conducted in 2021 and another in 2025. The research focuses on the impact of Artificial Intelligence (AI) on learners with basic-to-intermediate English proficiency within intercultural communicative contexts. It analyses learners' agency and the quality of cultural exchange using post-programme surveys, focus group interviews, and interaction recordings from HKBU students and their international partners. In 2021, linguistic barriers constrained the agency of weaker learners, causing reliance on peer translation during synchronous communication; this led to frustration and reduced participation. However, exchanges yielded meaningful, authentic cultural insights, with learners appreciating their peers' culturally specific knowledge. By 2025, AI tools partially alleviated linguistic barriers, enabling weaker students to read aloud from AI-generated scripts. However, over-reliance on AI produced generic, detached content, reducing the shared authentic experiences and cultural depth of conversations. Ultimately, while AI democratises linguistic access, it risks undermining cultural exchange and requires VE practitioners to refine their pedagogies if they are to continue assisting with authentic intercultural learning.

Keywords: Virtual exchange, digitally assisted intercultural communication, AI-assisted language learning

1. Introduction

Since the proliferation of telecommunication tools in the 1990s, language teachers in higher education institutions worldwide have increasingly adopted Virtual Exchange (VE) as a scalable method to provide learners with authentic communicative contexts for using their target languages (Turula et al., 2019). VE refers to structured, sustained, peer-to-peer online intercultural learning between groups of students located in different geographical regions, possessing differing cultures and first languages, under the guidance of trained instructors (O'Dowd, 2023, p. 13). Learners are typically motivated to engage in VE

activities to fulfil part of their course credit (Stevens Initiative, 2021).

VE is a value-added practice in English language education in East Asia. Although English is a required subject in both secondary and tertiary education in many countries, the majority of English language classrooms in Asia tend to be teacher-centred and exam-oriented, affording little opportunity for students to use the language for authentic communication. Consequently, many Asian students are unable to speak fluently in English despite years of instruction (Freiermuth, 2025). VE provides a meaningful context for students to use English as a Lingua Franca to communicate with peers who do not share the same first language, thereby preventing learners from reverting to their shared mother tongue when encountering communicative difficulties—a common occurrence during oral practices with peers from the same cultural background (Hagley & Thomson, 2017).

This study compares two VE programmes carried out at HKBU: one in 2021, before AI tools were widely available, and another in 2025, after AI tools had been adopted by language learners and instructors in higher education institutions in Hong Kong. Both programmes involved over 100 participants from different cultural backgrounds and shared the aim of improving students' English language proficiency as well as their intercultural communicative capacity. Both programmes provided opportunities for students to communicate with their international peers synchronously on Zoom and asynchronously through video message exchanges outside class time.

Both programmes served as gateway programmes, aiming to introduce VE practice and prepare students for more advanced online or onsite cultural exchange programmes. They were similar in scale and design, involved HKBU students who were relatively weak in their English language proficiency (at DSE Level 3 or IELTS Band 5.5 or below), and used VE to incentivise students to speak English as a Lingua Franca and communicate with international peers outside class time. Because the key difference between the two programmes was that one was offered in the pre-AI era and the other in the post-AI era, a comparison of their outcomes provides insights into how AI is impacting the quality of VE, especially among students at the basic or intermediate level of English language proficiency.

2. Methodology

2.1 Participants and VE Programme Design

The 2021 programme involved a total of 129 participants: 27 Year 2 students from an Associate Degree programme offered by the College of

International Education (CIE) of HKBU; 3 undergraduate students from the Education University of Hong Kong (EDUHK); 70 sophomores from Georgia State University (GSU); and 29 undergraduate students from Assam Don Bosco University (ADBU) in India. Students were expected to join two Zoom meetings, one at the beginning and another at the end of the eight-week exchange period, and they were required to exchange six asynchronous video messages on Flipgrid. The VE programme was compulsory for students from CIE and GSU, and elective for students of EDUHK and ADBU.

The 2025 programme involved 115 participants: 56 Year 1 students from HKBU; 32 students from Universidad Juárez Autónoma de Tabasco (UJAT), Mexico; 10 students from Kenyatta University (KU) in Kenya; five students from Middle East Technical University (METU); eight students from Holy Child Central Colleges Inc. (HCCCI); and four students from South Cotabato State College (SCSC) in the Philippines. A total of six Zoom discussion meetings involving students from the above institutions were offered at two-week intervals throughout the 13-week semester, and students were expected to attend at least one Zoom discussion of their choice. Students were expected to exchange one video message on Padlet and organise synchronous online meetings with their international peers at least twice during the semester for one-on-one peer discussions. For students of HKBU and UJAT, the programme was compulsory; for students from other institutions, it was elective.

2.2 Data Collection

A post-programme survey comprising 10 questions—including two multiple-choice questions, three Likert-scale questions, and five open-ended questions (all optional)—was circulated among participants who fulfilled a minimum of eight hours of VE tasks, including both synchronous and asynchronous discussions. In 2021, 98 participants were invited and 40 valid responses were received: 22 from Hong Kong students and 18 from international partners. In 2025, 94 participants were invited and 43 valid responses were received: 14 from Hong Kong students and 29 from international partners.

In addition, qualitative data were collected, including students' video message exchanges in both 2021 and 2025, as well as recordings of HKBU students' discussions with their international peers. Three HKBU students in each cohort participated in a post-programme focus group interview. The qualitative data provided insights into the quality and depth of the exchange.

3. Results and Discussion

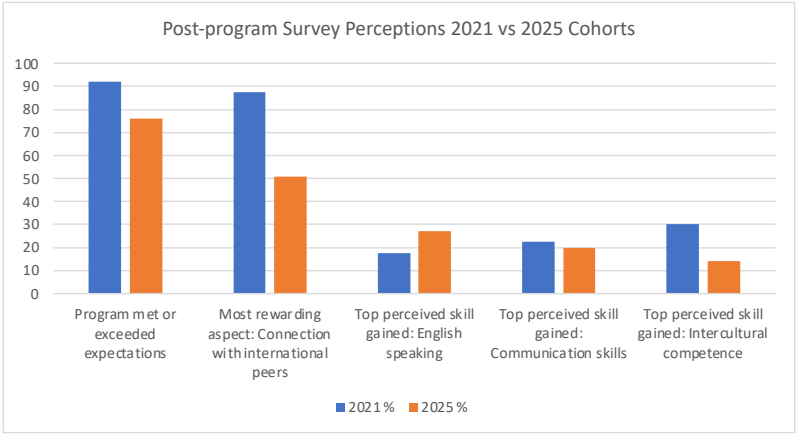
Post-Programme Survey Results

The post-programme survey results indicate that students in both cohorts were generally satisfied with the programme. In 2021, 92% of respondents affirmed that the programme exceeded or met expectations through cultural immersion, enhanced English proficiency, and a global perspective. Some noted the awkwardness of Zoom breakout room discussions due to linguistic and/or cultural barriers. Similarly, many participants in 2025 indicated that the programme met or exceeded expectations due to improved English skills, cross-cultural interactions, and global insights, though a slight decline in programme satisfaction was observed, with only 76% providing confirmation; others desired more in-depth discussions.

Table 1: Post-Programme Survey Characteristics and Perceptions: 2021 vs. 2025 Cohorts

Baseline characteristic	**2021 Cohort**		**2025 Cohort**	
	n	%	n	%
Invited participants	98	-	94	-
Valid responses received	40	40.8	43	45.7
Hong Kong (HKBU) responses	22	55.0	14	32.6
International partners responses	18	45.0	29	67.4
Program met or exceeded expectations	37	92	33	76
Most rewarding aspect: Connection with international peers	35	87.5	22	51
Top perceived skill gained: English speaking	7	17.5	12	27
Top perceived skill gained: Communication skills	9	22.5	9	20
Top perceived skill gained: Intercultural competence	12	30	6	14

Figure 1: Post-programme Survey Perceptions



Focus Group Interview Insights

The focus group interviews reveal that students with weaker language abilities in the 2021 cohort often struggled to speak English or reverted to speaking Chinese, relying on peer translation to keep up with synchronous conversations in Zoom breakout rooms. However, this linguistic barrier was somewhat eased in asynchronous video messaging, as students could rehearse their speech and edit the recording.

For example, one HKBU student's video messages consistently received the highest number of views (over 70) and the highest responses from international peers. Each of her videos received over five responses, whereas the majority of video posts received one or sometimes no responses. Despite her clear popularity in asynchronous discussions, this student did not join either of the two Zoom discussions because of her unwillingness to speak spontaneously. She stated that the biggest challenge for her was: *"to make videos, since I had to make and remake videos to get the best one out of it."*

Short video message exchanges somewhat alleviated the pressure of communicating synchronously and empowered weaker language learners to communicate at a higher level of fluency. Despite this, the limitations of asynchronous communication and the restrictive functions of the Flipgrid platform proved to be one of the major causes of frustration among participants in 2021. Flipgrid made all video posts available to all participants with access to the platform, making it impossible for student partners to exchange friendly private messages unrelated to course content. This design made peer-to-peer exchanges somewhat rigid and formal. Few cross-cultural friendships were formed during the eight-week period of exchange.

Flipgrid allows limited short video posts with a maximum length of 10 minutes, restricting the length and depth of cultural exchange. A student noted: *"Though rewarding, the process of finding things in common is difficult through small video clips."* The slow processing speed of Flipgrid led to a number of participants losing their videos when their internet connection was cut off during uploading. In short, although video messaging somewhat reduced linguistic barriers, it became the biggest challenge to smooth intercultural exchange, as identified by 25% of the participants. Many expressed a desire to communicate more intimately with their partners on an alternative, easy-to-use telecommunication platform.

Despite the linguistic challenges, 87.5% of participants expressed appreciation for the opportunity to connect with and learn from their

international peers, stating that this was the most rewarding aspect of the programme. The top three specific skills gained by joining the VE programme were intercultural competence (30%), English speaking (22.5%), and communication skills (17.5%).

Participants' video exchanges indicate that culturally specific knowledge drawn from students' lived experiences was shared. For example, when discussing environmental protection, students in the US expressed surprise upon learning about the Mainland Chinese policy of limiting car traffic by permitting car owners to drive on alternate days of the week based on their licence numbers. In the discussion about women's empowerment, Chinese students shared stories of their culturally specific heroes—Qiu Jin and Empress Wu—whereas students from India shared their matriarchal tribal traditions.

A Hong Kong student reflected:

"In our topic LGBTQ, I discovered that the discrimination problem regarding LGBTQ rights in America is not really more peaceful than in Hong Kong. In my imagination, America was an open-minded and high-tolerance country. After communicating with my international partner, I realised that Donald Trump isn't really supporting same-sex marriage; he even plans to exclude transgender people. It means that the U.S.A. government won't provide any protection rights for transgender people. Americans seem to treat LGBTQ more tolerantly than Hong Kong because of their citizens' fight and protest over many years to push for improvement in LGBTQ rights. There was nothing perfect at the beginning. Every positive result comes from many protests, strive, and sacrifice."

The survey results indicate that the 2021 VE programme inspired a higher level of awareness of one's own culture as well as a more nuanced understanding of another culture among participants. In contrast, only 51% of students in the 2025 cohort identified connection with international peers as the most rewarding aspect of the VE programme; the remaining students expressed appreciation for gaining global citizenship-related knowledge, personal growth, or poster creation, among others. The top three skills gained through participating in the VE programme were English speaking (27%), communication skills (20%), and intercultural competence (14%).

The Impact of AI Reliance

Participants' video message exchanges and conversation recordings indicate that weaker students from HKBU habitually read aloud AI-generated scripts when communicating with international peers in both synchronous

and asynchronous contexts. Their reliance on AI-generated scripts led to frustration for international peers in some cases. For example, after a Hong Kong participant read her script, her partner from North Cyprus responded: *"You don't have to read from the script. Just tell me what you think. It's okay to make mistakes. I make them too. I'm not a native speaker either. Let's just talk."*

It is worth noting that habitual reliance on AI-generated scripts was more pronounced among HKBU students compared to their international counterparts. Students from other countries tended to acknowledge a lack of preparation or unfamiliarity with answers when confronted with questions they were unprepared to answer. In some cases, HKBU students provided AI-generated scripts to their international partners and asked them to read it aloud to accomplish the task of recording a conversation.

An HKBU student reflected:

"Answering the question provided was a bit challenging. I was afraid to be seen like a fool by reading my answer through notes or AI, but I also cannot provide an answer without valid data which I need to search for first."

This indicates that weaker students tended to rely heavily on AI tools not only because of their lack of fluency in English but also because of a lack of content knowledge and insufficient motivation to actively seek the information they needed. While AI tools may have enabled weaker students to express ideas in English without having to revert to their mother tongues—helping to explain why more students in the 2025 cohort emphasised the English-speaking skills gained through the VE programme—the content generated by AI was often devoid of culturally specific knowledge, diminishing the intercultural learning potential of these exchanges.

For example, HKBU students discussing racial equality simply read aloud content on race-related social movements that took place in the USA, such as "Black Lives Matter," without discussing any of the problems of racial inequality prevalent in Hong Kong, even though such local examples were provided in the relevant module of their textbooks. Although students' self-assessment was that they improved their English-speaking skills, the validity of their assessment is questionable, likely derived in part from their failure to distinguish between the skills of reading aloud and speaking.

The post-programme survey indicates that students gained confidence because they realised that their international peers could understand their reading, despite their sometimes heavily accented English. However, their

ability to speak their minds and share their experiences in English remains limited, as a number of international students noted the persistent silence of HKBU students in breakout rooms during Zoom discussions.

4. Limitations and Further Research

There are several limitations to this study. First, the sample sizes for post-programme surveys (40 in 2021; 43 in 2025) are relatively small, limiting generalisability. Second, reliance on self-reported data introduces potential bias, particularly regarding perceived gains in English-speaking skills, whose validity is questionable since many participants conflated reading aloud AI-generated scripts with genuine spontaneous speaking. Third, the comparison is quasi-experimental, with differences in programme duration, platforms, task requirements, and partner demographics, which may confound the isolated effects of AI. Future research should control programme variables to enhance comparability. Objective measures, such as pre- and post-programme oral proficiency tests, should be incorporated to address the validity issues of self-assessment.

Implications

This study indicates that the easy availability of AI tools can help basic-to-intermediate language learners gain a sense of agency as well as a degree of confidence in their language use. However, this comes at a considerable cost: it reduces students' motivation to seek information independently or share their lived experiences, thus diminishing the quality of peer-to-peer intercultural learning in a VE programme. Furthermore, students' perceived gain in English-speaking skills may be an illusion, as students were reading content they often did not understand, disconnecting language use from authentic personal expression.

The implication is that apart from making AI tools accessible to students, educators need to equip students with adequate AI literacy and alert them to the consequences of over-dependence. VE instructors may consider guiding students in their AI tool usage towards using AI tools for automated translation rather than text generation, ensuring students are sharing their original thoughts and lived experiences in a VE programme to ensure the richness and depth of intercultural learning.

Acknowledgements

Appreciation is extended to all colleagues and students from partner institutions who participated in the virtual exchange programmes with HKBU in 2021 and 2025.

Ethical Declaration Statement

The 2021 project was supported by the TDG Seed Fund provided by the Research Committee of the School of Continuing Education (SCE); and the 2025 project was supported by the TDLEG fund provided by the Centre of Holistic Education of The Hong Kong Baptist University.

Generative AI Policy

The author used Grok in the preparation of this work to improve phrasing and format references. After utilising the tool, the author thoroughly reviewed and edited the content as necessary and assumes full responsibility for the publication's content.

References

Freiermuth, M.R. (2025). *The status of 'English as a Foreign Language' in East Asian Countries: A second language teacher's perspective*. Learning and Teaching English in Multilingual Educational Environments, 45. DOI: 10.61736/alizes,465

Hagley, E. & Thomson, H. (2017). *Virtual exchange: Providing international communication opportunities for learners of English as a foreign language*. 北海道言語文化研究 , 15, 1-10.

O'Dowd, R. (2023). *Internationalising higher education and the role of virtual exchange*. Routledge. https://doi.org/10.4324/9781315393704

Stevens Initiative. (2021). *Virtual exchange typology*. The Aspen Institute. https://www.stevensinitiative.org/resource/virtual-exchange-typology/

Turula, A., Kurek, M., & Lewis, T. (2019). *Telecollaboration and virtual exchange across disciplines: in service of social inclusion and global citizenship*. Research-publishing.net.

Leveraging GenAI in Creating Audio and Video-Based Flipped Classroom Materials

Eric FRIGINAL [1] & Jenifer HO [1]
[1] Department of English and Communication, The Hong Kong Polytechnic University, Hong Kong, China

Abstract

The move from traditional teacher-centred classrooms to blended and flipped classrooms that are student-centred has called for educators to experiment with new formats of teaching (Friginal & Ho, 2025). Nevertheless, this shift requires educators to develop both pedagogical and technical expertise to design new materials. To better equip educators for this shift, this workshop aims to introduce participants to creating audio and video-based flipped classroom materials using GenAI-powered platforms: ElevenLabs (for audio) and HeyGen (for video). Through interactive lectures, discussions, and hands-on activities, participants will leverage the power of ElevenLabs and HeyGen to create flipped classroom materials that can promote personalised learning and are relevant to students with different needs.

Outcomes-Based Learning Objectives

By the end of the workshop, participants will be able to:

- Identify and select appropriate GenAI tools for creating engaging, interactive, and personalised audio and video-based materials for flipped classrooms.
- Design (interactive) AI avatars for disciplinary-specific learning.
- Develop best practices for integrating GenAI into designing audio and video-based flipped classroom materials.

Target Audience

The expected background of participants is instructors teaching in higher education contexts, including but not limited to language-related disciplines. Educators from other disciplines, such as health, social sciences, aviation, and hotel management, will also be able to participate and adapt the pedagogical approaches introduced in this workshop in their own areas of teaching. In general, this workshop will be useful to educators whose disciplines require the simulation of communicative scenarios.

Topic Relevance

The workshop is closely connected to the sub-themes 'AI tools and pedagogy', 'ethics and integrity', and 'incorporating GenAI in learning, teaching and assessment'. The workshop addresses a university-wide initiative to move to a flipped classroom format, as well as aligning with the university's call for a higher level of integration of GenAI in learning and teaching activities. The workshop also facilitates discussions on the ethical issues of using GenAI tools.

Workshop Design

1. Workshop Format and Activities: Introduction

Multimodal materials, such as audio and video, have become indispensable resources for English Language Teaching (ELT). The trend of using audio and video in ELT coincides with the latest developments in Generative Artificial Intelligence (GenAI) technology. Not only do GenAI chatbots enable students and teachers to interact with tools such as ChatGPT and DeepSeek to generate text-based responses, but the technology has also developed the capability to produce multimodal outputs such as voices and videos.

GenAI technology has also made great strides in enhancing its interactions with users mediated by virtual avatars that sound and look like human teachers. Not only can these virtual instructors present content in video lectures, but with sufficient training, they can also provide answers and engage in conversations in multiple languages.

This workshop focuses on GenAI's ability to transform the creation of audio and video-based materials in language education, with a focus on interactive avatars and their implications for English for Specific Purposes (ESP). This includes simulating communication scenarios in various disciplines, such as delivering announcements in a flight cabin or managing hotel service encounters.

We start by reviewing literature on the different ways audio and video have been used to support language learning and teaching, such as providing authentic L2 input to improve pronunciation (Scheffler & Baranowska, 2023), speaking and listening skills (Verdugo & Belmonte, 2007; Zhussupova & Shadiev, 2023), followed by a discussion on one specific way of integrating videos in language education – the use of virtual avatars. The workshop concludes by discussing the potential opportunities and challenges, including ethical issues, of using these GenAI tools in language education and beyond.

To establish rapport among participants, the workshop will begin with a quick round of self-introductions and invite participants to share how they have been using GenAI in their own teaching contexts.

2. Main Content Delivery

The workshop discusses the potential of using interactive AI avatars as knowledge partners to provide a personalised learning experience for students. The use of virtual instructors that resemble the appearance of human teachers can streamline the material creation process and ensure consistency in the delivery of materials, especially for courses that last over several months.

Virtual instructors provide multilingual support by translating from the content provided by the human teacher, thus catering to individual learners' needs. This is especially helpful in multilingual classrooms with students of different English proficiency levels. Teachers can also develop materials on the same topic with varying levels of difficulty by adjusting the vocabulary used and the pace of speech. They can provide a script for the avatars to deliver using different accents and a range of voices to match the content.

The recent development of interactive avatars allows them to be trained to converse with users on various topics, taking on different roles such as customer support, therapist, doctor's assistant, fitness coach, and more. This potential broadens the scope of their use, demonstrating that they are valuable not only for language teachers but also for educators in other fields with distinct communicative demands.

Besides making use of the available roles, in some platforms, teachers can customise the avatar by training it with the 'knowledge base' necessary for them to perform their role. The prompt is easily customisable, thus allowing teachers to tailor the 'knowledge base' of the avatar with discipline-specific vocabulary and disciplinary norms of interaction to maintain relevance to learners. The avatar can also provide immediate feedback and tailor its language use to be appropriate to the learners. Teachers can easily update the 'knowledge base' of the avatar based on students' needs.

The workshop will first introduce some background studies on audio and video-based material design and GenAI-assisted material design (Friginal & Ho, 2025) using the mode of interactive lecture. Then, participants will engage in hands-on interactive activities to try creating avatars by themselves. Lastly, the workshop will end with a discussion on the potential opportunities and limitations that interactive AI avatars will bring to ELT and ESP. It is our

intention that the workshop will strike a good balance between drawing on research findings and engaging participants in hands-on activities.

3. Interactive Activities

Activity 1: Creating Audio and Video-Based Flipped Classroom Materials

Participants will learn how to create multimodal flipped classroom materials using ElevenLabs (https://elevenlabs.io/) and HeyGen (https://www.heygen.com/). To increase engagement, the workshop will also demonstrate how to insert human-like AI avatars into video templates or PowerPoint slides that instructors may already have to make the content more engaging for students.

Activity 2: Designing an Interactive AI Avatar for Specific Purposes

Participants will engage in designing the 'knowledge base' required for AI avatars that can perform different roles, such as customer support, therapist, doctor's assistant, fitness coach, and more. Participants will be shown how to update the 'knowledge base' of the avatar based on students' needs and disciplinary requirements.

Participants will be guided to design the 'knowledge structure' of the avatar according to the following perspectives as suggested by HeyGen:

1. Persona of the avatar
2. Knowledge base
3. Language practice activities
4. Feedback and encouragement, among others.

4. Discussion and Q&A

- How do we, as educators, ensure that these tools augment rather than replace human interaction?
- How do we guard against the misuse of cloning technology, such as creating deceptive or harmful content?

5. Closing Remarks

The rise of voice and video cloning technologies, once reserved for experts, is now within reach for language teachers at a relatively low cost. Virtual instructors open doors to personalised, immersive, and flexible language learning experiences like never before (Friginal & Ho, 2025). But we must also confront the dangers that come with this transformation. These increasingly 'human-like' virtual instructors blur the line between what is real and artificial, raising critical questions about authenticity, ethics, and the potential over-

reliance on technology in the classroom. As these tools grow more sophisticated and powerful, the urgency to define their role in education—and to safeguard their use—cannot be overstated.

After the workshop, participants are encouraged to start using the (interactive) AI avatar function to create flipped classroom materials. Once they are comfortable with the technology, they can proceed to more advanced uses offered by the platform, such as exploring AI cloning to create a digital clone of themselves, or a customised interactive AI avatar that is relevant to the learning outcomes of their courses.

Acknowledgments

Our work in this workshop was supported by the Financial Support for Book Writing 2023/24 of The Hong Kong Polytechnic University.

Ethical Declaration Statement

N.A.

Generative AI Policy

The author(s) used ChatGPT-4.1 during the preparation of this work to refine their writing. They also studied the functions offered by HeyGen and ElevenLabs to prepare the materials for this workshop. 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

Friginal, E., & Ho, W.Y.J. (2025). Technology and language learning: Effective podcasting and video-based language teaching practices in the age of generative AI. Routledge.

Scheffler, P., & Baranowska, K. (2023). Learning pronunciation through television series. Language Learning & Technology, 27(1), 1–16. https://doi.org/10.64152/10125/73520

Verdugo, D. R., & Belmonte, I. (2007). Using digital stories to improve listening comprehension with Spanish young learners of English. Language Learning & Technology, 11(1), 87–101. https://doi.org/10.64152/10125/44090

Zhussupova, R., & Shadiev, R. (2023). Digital storytelling to facilitate academic public speaking skills: Case study in culturally diverse multilingual classroom. Journal of Computers in Education, 10(3), 499–526. https://doi.org/10.1007/s40692-023-00259-x

Designing Workflow in Teaching and Learning for Automation through a Pedagogical Partnership Approach

Peter LAU [1], Dimple THADANI [2], Louis CAI [3]
[1] Teaching and Learning Innovation Centre, The University of Hong Kong, Hong Kong, China
[2] Department of Management, Marketing and Information Systems, Hong Kong Baptist University, Hong Kong, China
[3] Centre for Holistic Teaching and Learning, Hong Kong Baptist University, Hong Kong, China

Abstract

Integrating Artificial Intelligence (AI) into education offers both remarkable opportunities and serious challenges (Bittle & El-Gayar, 2025). Educators remain cautious, concerned about unintended consequences such as unproductive cognitive offloading (Ahmedtelba et al., 2025) and increased risks of academic dishonesty (Bittle & El-Gayar, 2025). As an entry point, we propose beginning with low-risk teaching routines, for example, responding to student enquiries, facilitating active learning, and providing formative feedback. This workshop will guide participants to identify a routine workflow in teaching and learning, then redesign it for AI-enabled automation. Using the n8n platform, examples of workflow automation will be demonstrated, followed by hands-on practice. Through a pedagogical partnership approach (Cook-Sather, 2014), participants will explore the thoughtful integration of AI tools. The Trio model—a Student–Teacher–AI partnership framework inspired by Williams & Ingle's (2025) case study—will be introduced as a reflective checklist to support redesign and ensure balanced adoption.

Learning Objectives:

1. Participants will be able to identify the principles of the Trio pedagogical model.
2. Examine a routine teaching and learning workflow and redesign an AI-supported automation plan for it.
3. Create and test an AI tool using the n8n platform.

Takeaways:

1. A Trio model checklist for Student–Teacher–AI decision-making.
2. A draft AI-supported workflow map.

Significance:

This workshop moves beyond general conversations about AI by guiding participants to analyse and redesign a teaching workflow (e.g., feedback for formative assessment), trial it in practice, and reflect on the process through the lens of the Trio model (Student–Teacher–AI partnership).

Target Audience:

This workshop is specifically designed for Higher Education (HE) educators with a focus on the following roles:

- **Faculty:** All levels (including teaching assistants) across all disciplines.
- **Instructional Designers and Curriculum Development specialists.** *(Note: Programming experience is not necessary.)*

Topic Relevance:

This workshop fits well within the following sub-themes:

Sub-theme 1: AI Tools and Pedagogy

The workshop will focus on the strategic use of AI tools to enhance pedagogical practices. Facilitators will introduce the **n8n** platform, which can automate teaching routines such as content/question creation, feedback generation, etc. Participants will be guided to examine what teaching routines or instructions are suitable for automation by n8n and to reflect on the process through the lens of the Trio model (Student–Teacher–AI partnership).

Sub-theme 5: Incorporating GenAI in Learning, Teaching and Assessment

To ensure that technology serves pedagogical goals rather than dictating them, Trio model principles ("pedagogy-first", reciprocal process, shared decision-making) will be introduced to guide the redesign. Participants will be engaged in hands-on redesign of a teaching routine by using the n8n platform.

Sub-theme 6: Teacher Professional Development

Participants will be encouraged to further develop and implement small-scale experiments and build capacity through experiential learning and reflective dialogue. Peer mentoring will be arranged for those who require additional support. Participants will be invited to join the Community of Practice for Student Partnership (CoP-SP) to share good practices across local institutions in Hong Kong.

Workshop Design

Workshop Format and Activities:

1. Introduction (5 min)

The workshop begins with a brief overview of the latest trends on AI in education, highlighting the urgency of moving beyond hype toward strategic implementation. Learning outcomes and the agenda are presented.

2. Main Content Delivery (15 min)

Facilitators will introduce how **n8n** (an AI workflow automation platform) can support teaching routines such as student enquiries, feedback generation (Williams & Ingle, 2025), etc. Examples will be explained to demonstrate how AI tools can augment, rather than replace, educator expertise. See an example in the Appendix. To ensure that technology serves pedagogical goals rather than dictating them, the workshop will introduce pedagogical partnership principles to guide the co-design of AI-enhanced learning experiences.

3. Interactive Activities (20 min)

The workshop will guide participants in exploring any teaching routines in their own specific contexts that could be automated. Participants will work in groups to redesign a workflow that integrates AI tools into the process. Interactive activities include:

- **Discussion:** Identifying suitable teaching routines for redesign (5 min).
- **Co-creation:** Redesigning the workflow (5 min).
- **Hands-on engagement:** Building and testing an AI-enabled protocol using the n8n platform (10 min).

4. Sharing and Discussion (10 min)

Participants will be invited to present their ideas, share their redesigned workflow, and discuss their test results. If time allows, they will also be encouraged to explore possible future development actions or collaboration.

5. Closing Remarks (3 min)

Participants will be encouraged to further develop and test their protocols on the n8n platform. Peer mentoring will be arranged for those who require additional support. Participants will be invited to join the Community of Practice for Student Partnership (CoP-SP) to share good practices across local institutions in Hong Kong.

6. Workshop Evaluation (2 min)

Participants will be invited to complete a post-workshop evaluation to share their comments on the workshop contents, arrangements, and suggestions for improvement.

Acknowledgments

This workshop is supported by a TDLEG-IICA project (2025–2028 Triennium), titled *Exploring GenAI Use in Teaching and Learning Enhancement through a Pedagogical Partnership Approach and Collaborative Showcase*, funded by the University Grants Committee (UGC).

Ethical Declaration Statement

This project has been approved by the Human Research Ethics Committee at The University of Hong Kong (HREC Reference Number: EA260055).

Generative AI Policy

The author(s) used CoPilot during the preparation of this work for grammar checking. 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

Ahmedtelba, A., Elycheikh, M., Mitereva, S., & Ponce, M. (2025). Critical integration of generative AI in higher education: Cognitive, pedagogical, and ethical perspectives. *London Journal of Research in Humanities and Social Sciences*, 25(13), 45–62.

Bittle, K., & El-Gayar, O. (2025). Generative AI and academic integrity in higher education: A systematic review and research agenda. *Information*, 16(4), 296. https://doi.org/10.3390/info16040296

Cook-Sather, A. (2014). Student-faculty partnership in explorations of pedagogical practice: A threshold concept in academic development. *International Journal for Academic Development*, 19(3), 186–198. https://doi.org/10.1080/1360144X.2013.805694

Winstone, N. E., & Carless, D. (2020). *Designing effective feedback processes in higher education: A learning-focused approach*. Routledge. https://doi.org/10.4324/9781351115940

Williams, A., & Ingle, E. (2025). Assessment design through co-creation: Student-staff partnership in evaluating the impact of artificial intelligence. *International Journal for Students as Partners*, 9(1). https://doi.org/10.15173/ijsap.v9i1.5802

Appendix: Example – Designing AI Automation for a Formative Classroom Assessment in a Large Lecture

A teacher used Moodle multiple-choice questions as a formative classroom assessment to check students' understanding of key concepts in a large lecture (200+ students). However, students reported that the pre-set feedback in the system was too general and not very helpful. The teacher considered using a two-minute paper as an alternative assessment task but hesitated due to the heavy workload involved in reading responses and providing feedback. This creates tension between the teacher (workload) and the students (timely and personalised feedback).

According to the Trio model, the AI partner occupies a vantage position that can offer solutions, such as automating parts of the formative assessment and feedback process. However, simply asking students to use a Chatbot to check their understanding does not guarantee an effective learning experience. Following a "pedagogy-first" principle of the Trio model, the teacher, as a content and pedagogy expert, should take the lead in redesigning an effective assessment workflow suitable for AI-supported automation.

A possible redesigned workflow could be:

1. Students spend two minutes responding to a short-answer question and indicate whether they prefer:
   Option A: See the model answer and explanation immediately.
   Option B: Receive formative feedback from the system and make a second attempt.
2. The system should comment on student responses and prompt them to consider what key points might be missing. It should avoid revealing the model answer or giving keywords.
3. A constructive and non-judgmental tone should be used.
4. After the second attempt, students will see the model answer and explanation.
5. The system generates a brief report summarising student performance and responses, including any confused or unclear concepts that may require further revision. It may also refer students to relevant chapters or readings when necessary.
6. Students provide comments on (or rate) their experience before exiting.
7. The system stores all the responses and AI-generated contents in an Excel file.
8. The teacher prompts the system to generate a report summarising student performance and experience accordingly, and prepares group feedback for the next class.

9. The teacher may revise the workflow or system prompt as needed.

Notes:

For other assessment tasks, if the teacher wishes to promote dialogic feedback, an optional input field can be added for students. As suggested by Winstone and Carless (2020):

- A **feedback request form** allows students to explain what they struggled with and which areas they want feedback on.
- A **feedback cover page** allows students to explain how they have used previous feedback.

GenAI and AI Agents: Your Teaching Sidekicks (No Coffee Runs Required)

David SANTANDREU CALONG [1] , Theresa KWONG [2] & Louis CAI [2]

[1] Center for Teaching and Learning, Mohamed bin Zayed University of Artificial Intelligence, United Arab Emirates

[2] Centre for Holistic Teaching and Learning, Hong Kong Baptist University, Hong Kong, China

Abstract

This high-energy workshop explores the strategic integration of the latest and most relevant Generative AI (GenAI) tools and AI agents for Teaching and Learning (T&L). We will cut through the hype to provide a comprehensive, practical understanding of GenAI tools and agents' capabilities (and, yes, their limitations), supported by numerous examples, use cases, and activities across various disciplines. You will leave with hands-on skills to design and implement innovative teaching and learning activities that can dramatically improve student engagement, motivation, and learning outcomes, while drastically reducing your workload.

Outcomes-based Learning Objectives

By the end of this workshop, participants will be able to:

1. **Identify** how GenAI tools can be utilised to augment the efficiency of Higher Education (HE) teaching and learning tasks, comparing and contrasting their pedagogical implications.
2. **Apply** specific Generative AI tools and AI agents to augment and improve at least one learning, teaching, or assessment task within their disciplinary context, explaining the rationale for their choice of tools and the expected pedagogical benefit.

Target Audience

This workshop is specifically designed for Higher Education (HE) educators with a focus on the following roles:

- **Faculty:** All levels (including teaching assistants) across all disciplines.
- **Instructional Designers** and Curriculum Development specialists.
- **Educational Technologists.**

Topic Relevance

This workshop is an ideal fit for the following sub-themes:

- **Sub-theme 1: AI Tools and Pedagogy**
  The workshop's central promise is the strategic integration of cutting-edge Generative AI tools and AI agents. This directly addresses the sub-theme by (1) providing hands-on instruction on the latest technologies, and (2) focusing on designing genuinely innovative educational experiences augmented by AI. Whilst Hultberg et al. (2024) argued that generative language models can help students learn and perform any assessment task if they have access to a computer (or smartphone), Santandreu Calonge et al. (2025) highlighted that "educational institutions must train their staff and students to use technology and AI ethically and effectively to ensure work-ready graduates" (p. 21).

- **Sub-theme 5: Incorporating GenAI in Learning, Teaching and Assessment**
  This sub-theme precisely describes the workshop's main scope: to equip participants with hands-on skills to design and implement AI-powered innovative teaching and learning activities and assessment tasks. Kamalov et al. (2025) indicated that "AI planning can dynamically adjust these learning pathways in real-time based on ongoing assessments, ensuring the educational experience remains relevant and effective as students progress (hyper-personalisation)" (p. 20).

- **Sub-theme 6: Teacher Professional Development**
  The workshop is fundamentally a professional development activity for Higher Education (HE) educators, para-academics, and personnel with teaching responsibilities. The focus on providing a comprehensive, practical understanding of GenAI's capabilities and limitations, supported by numerous examples, use cases, and activities across various disciplines... to improve student engagement, motivation, and learning, in short, developing new skills and improving efficiency, is the primary purpose of professional development. According to Kamalov et al. (2023), Large Language Models (LLMs) could help faculty grow as teachers.

Workshop Format and Activities

1. Introduction

Latest research and news on the topic will be briefly introduced to contextualise the sub-themes. The necessity of moving beyond doubt, reluctance, or panic towards the strategic implementation of Generative AI and AI agents to improve both teaching efficacy and student learning will be highlighted. The agenda for the workshop will then be presented.

2. Main Content Delivery

- **Part 1: Video Generation for Teaching:** This section will illustrate immediate content augmentation, summarisation, application, and time-saving benefits.
- **Part 2: AI Apps and Agents for HE Faculty:** This section will focus on the latest AI agents applied to teaching and learning.

3. Interactive Activities

The entire workshop focuses on showcasing practical examples of GenAI automating, improving, or augmenting common HE tasks to enhance professional efficiency (addressing Sub-themes 1, 5, and 6). Participants will break into pairs or teams.

(1) Each pair/team must identify one (challenging or routine) HE teaching and/or learning task in their field/discipline(s) and select the appropriate AI tool or agent to augment/simplify that task. They must justify their choice and application before showcasing it.

4. Discussion and Q&A

- **Question 1:** What is one administrative or teaching preparation task that you can immediately offload to an AI tool/agent to reclaim at least 30 minutes of your time this week?
- **Question 2:** How does this (ethical) application of AI in the classroom better prepare our students for an AI-driven professional workforce?

5. Closing Remarks

Your assignment is to apply one of the following (Video, tool, or Agent) to your course within the next 7 days. Be confident; start small. Start now.

Acknowledgments *(If applicable)*

Ethical Declaration Statement

NIL

References

Hultberg, P. T., Santandreu Calonge, D., Kamalov, F., & Smail, L. (2024). Comparing and assessing four AI chatbots' competence in economics. *PLOS ONE*, 19(5), e0297804. https://doi.org/10.1371/journal.pone.0297804

Kamalov, F., Calonge, D. S., Smail, L., Azizov, D., Thadani, D. R., Kwong, T., & Atif, A. (2025). Evolution of AI in education: Agentic workflows. *arXiv preprint arXiv:2504.20082*.

Kamalov, F., Santandreu Calonge, D., & Gurrib, I. (2023). New era of artificial intelligence in education: Towards a sustainable multifaceted revolution. *Sustainability*, 15(16), 12451. https://doi.org/10.3390/su151612451

Santandreu Calonge, D., Kamalov, F., Medina Aguerrebere, P., Hassock, L., Smail, L., Yousef, D., ... & Abdulla, N. (2025). Upskilling and reskilling in the United Arab Emirates: Future-proofing careers with AI skills. *Journal of Adult and Continuing Education*, 14779714251315288. https://doi.org/10.1177/14779714251315288

An AI Simulator for Enhancing Negotiation Practices in Business Education

Beatrice CHU [1], Esme ANDERSON [1], Kubert WANG [1], Yonathan Akililu KIDANEMARIAM [2]

[1] Centre for Education Innovation, The Hong Kong University of Science and Technology, Hong Kong, China

[2] School of Business and Management, The Hong Kong University of Science and Technology, Hong Kong, China

Abstract

This poster presentation introduces a novel, student-developed AI negotiation simulator powered by a Retrieval-Augmented Generation (RAG) agentic platform. Aimed at enhancing business education, the system creates a dynamic, responsive learning environment to prepare students for real-world business negotiation practices. The development of the platform involved a meticulous knowledge ingestion process into a scalable cloud-based vector database, facilitating RAG-driven interaction. Initial testing demonstrates the potential of RAG-based solutions in supporting the development of students' negotiation skills linked to business pedagogy. This study highlights the potential of emerging technologies, offering practical implications for educators regarding deployment.

Keywords: Business pedagogy; negotiation; case-study student-driven practice

1. Introduction

Recent advancements in generative AI are reshaping educational practice, enabling increasingly adaptive and personalised learning experiences (Alfredo et al., 2024; Li et al., 2025). This case study examines the design and deployment of a RAG-based business AI platform aimed at supporting negotiation practices for business students. Currently, there is limited empirical evidence on how these tools can be intentionally designed to balance cognitive understanding with context-specific, practice-oriented skills in negotiation pedagogy. Addressing these gaps, this case study analyses a RAG-based negotiation platform co-developed by an undergraduate student with faculty support, offering pedagogical insights for AI-driven negotiation practices in business education.

2. Methodology

Technical Architecture

The system utilises a RAG framework optimised for academic contexts (Afzal et al., 2024). To address the complexity of negotiation concepts, semantic chunking was employed to maintain contextual integrity. The student-developer partnered with faculty in providing discipline-accurate content. Retrieval was enhanced by generating potential user questions for each text prior to indexing (Guinet et al., 2024). This enabled a "question-to-question" semantic search strategy, ensuring higher retrieval precision than standard keyword matching (Edwards, 2024).

Pedagogical Implementation

The platform has been designed for business negotiation exercises to be used across MGMT courses. To prepare the platform for deployment, 1,660 external tester participants (MTurk) with self-reported backgrounds in business used and provided feedback on the platform. Faculty provided an authentic business case—the 'Prisoner's Dilemma'—for the chatbot. Participants subsequently practised the exercise against challenging and aggressive personas, which can be difficult to replicate in real life. With 24/7 access, participants could continuously practice and better enhance their negotiation performance.

The platform also captured interaction transcripts; users were invited to reflect on their performance, and the platform provided an opportunity for assessors to analyse users' decision-making and application of course concepts (e.g., cross-cultural negotiation, Nash equilibrium). System accuracy was validated through this large-scale external testing to mitigate hallucinations and ensure reliability; there is currently a relatively low error rate of 5% in maintaining character/factual adherence. This platform was awarded a Teaching and Learning Innovative Project fund and will be fully deployed in three UG and PG MGMT courses for negotiation exercises during Spring 2026.

3. Results and Discussion

Initial testing with these external participants yielded the following insights, demonstrating promise for course deployment:

- **High Engagement Volume:** Testers exchanged over 12,000 messages across 1,696 practice sessions, demonstrating sustained interest and voluntary participation.
- **Meaningful Interaction:** On average, testers spent 12 minutes per

session practising negotiation strategies, indicating sustained engagement.
- **Quality of Engagement:** User input analysis revealed an average sentence length of 12 words, demonstrating that students went beyond simple "yes/no" responses.
- **Total Learning Time:** The platform facilitated over 550 hours of active negotiation practice.

These insights highlight the potential of the platform to foster active learning experiences. Over the next eight weeks, comprehensive data will be collected on the 'Prisoner's Dilemma' negotiation exercise performance of 311 undergraduate and postgraduate students using the platform. Next steps include faculty analysis of current student platform interactions and evaluating whether the platform has enhanced student performance in the exercise.

4. Implications

This case study highlights key implications for pedagogy and learning design. RAG-based platforms have the potential to serve as structured, scaffolded environments for students to iteratively rehearse negotiation, receive targeted feedback, and connect conceptual understanding with contextual practice difficult to replicate in classrooms (e.g., aggressive counterparts).

The study also underscores the need for institutional support in guiding faculty on GenAI integration, whilst highlighting the critical role of faculty in ensuring discipline-specific content and skills are effectively addressed. Finally, this project demonstrates the untapped potential of student partnership, as the platform was developed by an undergraduate student.

Acknowledgements

Professor T. Bradford Bitterly (SBM, HKUST), The Centre for Education Innovation (HKUST), and the CEI-run project 'Fostering Digital Competency in the GenAI-GenZ Era: A Student-Driven Community of Practice'.

Ethical Declaration Statement

Approved by the Ethics Committee of The Hong Kong University of Science and Technology (Reference Number: HREP-2024-0471).

Declaration of Use of AI in the Writing Process

The author(s) used Jenni to identify RAG-related research. After utilising the tool/service, the author(s) thoroughly reviewed and edited the content and assumed full responsibility for the publication's content.

References

Afzal, A., Vladika, J., Fazlija, G., Staradubets, A., & Matthes, F. (2024). Towards Optimizing a Retrieval Augmented Generation using Large Language Model on Academic Data. In: *Proceedings of the 2024 8th International Language Processing and Information Retrieval*, p. 250. https://doi.org/10.1145/3711542.3711575

Alfredo, R., Echeverría, V., Jin, Y., Yan, L., Swiecki, Z., Gašević, D., & Martínez-Maldonado, R. (2024). Human-centred learning analytics and AI in education: A systematic literature review. *Computers and Education Artificial Intelligence*, 6, 100215. https://doi.org/10.1016/j.caeai.2024.100215

Edwards, C. (2024). Hybrid Context Retrieval Augmented Generation Pipeline: LLM-Augmented Knowledge Graphs and Vector Database for Accreditation Reporting Assistance. *ArXiv*, abs/2405.15436.

Guinet, G., Omidvar-Tehrani, B., Deoras, A., & Callot, L. (2024). Automated Evaluation of Retrieval-Augmented Language Models with Task-Specific Exam Generation. In *Proceedings of the 41st International Conference on Machine Learning*. https://doi.org/10.48550/ARXIV.2405.13622

Li, Z., Wang, Z., Wang, W., Hung, K., Xie, H., & Wang, F. L. (2025). Retrieval-Augmented Generation for Educational Application: A Systematic Survey. *Computers and Education Artificial Intelligence*, 100417. https://doi.org/10.1016/j.caeai.2025.100417

3 Learning Analytics

Fluent but Misaligned: An NLP Approach to Measuring Uptake in Classroom AI-Avatar Dialogue

Paolo MENGONI [1], Daniel SHEN Jiandong [2], Aru NURGISSAYEVA [2], Kristen LI Yuanxi [3], Renia LOPEZ-OZIEBLO [4], and Paulina WONG Pui Yun [5]

[1] Department of Interactive Media, Hong Kong Baptist University, Hong Kong SAR, China

[2] Soqqle Hong Kong Limited, Hong Kong SAR, China

[3] Department of Computer Science, Hong Kong Baptist University, Hong Kong SAR, China

[4] Department of English and Communication, The Hong Kong Polytechnic University, Hong Kong SAR, China

[5] Division of Science, Lingnan University, Hong Kong SAR, China

Abstract

Generative AI can sound fluent yet still miss what students mean, especially during brief classroom brainstorming. We analysed chat logs from a mathematics-for-game-development lesson in which 25 Hong Kong undergraduates worked in small groups and interacted with a generative AI avatar. We treated each student turn and the following avatar reply as a dyad and used Natural Language Processing (NLP) to measure three classroom-relevant signals: meaning continuity (Sentence-BERT cosine similarity), keyword carryover (Jaccard overlap of content-word lemmas), and interaction mode (Language Style Matching). Results showed an opportunity-structure effect: when students wrote longer, more content-rich turns, the avatar's replies were more consistently on-topic and carried forward more task terms. We also observed two productive patterns (on-topic support with shared terminology or with paraphrase) and a risk pattern, *keyword echo*, where term reuse looked responsive but meaning continuity was weak. These measures offer educators a practical way to interpret classroom GenAI dialogue beyond surface fluency.

Keywords: AI in education (AIEd), learning analytics, classroom discourse, dialogue analysis, semantic similarity

1. Introduction

Generative AI is increasingly used in classrooms for brainstorming, but its value depends on how it handles short or unclear input. The AI can sound fluent

yet still miss the student's intent. Teachers need a way to check if the AI responds to the student's idea, not just plausible text (Chiu, 2021; Giannakos, 2024).

To judge response quality turn by turn, we examine three alignment signals: meaning (idea continuity), wording (key-term carryover), and style (function-word mode). We separate these signals because good brainstorming support often paraphrases or expands the idea, so the meaning can match even when the wording changes. By contrast, the AI can repeat keywords but still misunderstand the student's point.

Evidence from classroom brainstorming logs is limited, and evaluations often mix meaning continuity with surface repetition, which can hide *keyword echo*, where word repetition appears responsive despite weak meaning continuity. Without separating these signals, conclusions about classroom support can be misleading. Short turns can depress overlap scores even when a reply is helpful, while keyword repetition can inflate apparent responsiveness even when intent is missed.

To study this in a real setting, we analyse classroom chat logs from a brainstorming lesson where students interacted with a generative AI avatar. Each student turn and the immediately following avatar reply is treated as an adjacent dyad. We measure meaning alignment using Sentence-BERT similarity between the student turn and the AI reply (Reimers & Gurevych, 2019). We also measure lexical overlap and Language Style Matching (LSM) to describe interaction mode. We report turn length to support interpretation of these scores.

Our research question is: In classroom brainstorming dialogue, to what extent does the AI avatar follow students' immediately prior ideas, and how do lexical grounding and interaction style relate to this continuity under short or noisy student turns?

2. Methodology

We analysed a brainstorming activity where 25 Hong Kong undergraduates interacted with a generative AI avatar in a mathematics-for-game-development lesson (e.g., 3D rotation and animation). Students worked in small groups. They discussed ideas together, and one student typed prompts to the avatar for the group. The task involved open-ended brainstorming and refinement of technical ideas, such as explaining concepts, suggesting uses, or clarifying steps.

We chose brainstorming because classroom prompts are often brief or unclear. In this setting, we focus on whether the avatar maintains alignment with

the student's immediately prior idea across turns. Avatar/interface configuration was not treated as an analytic factor in this paper; all adjacent student–AI dyads were analysed together at the turn level.

Data Source and Unit of Analysis

Data were drawn from platform chat logs recorded in Soqqle's Classlet system, which stores turn-by-turn student messages and avatar replies during classroom activities. We analysed interaction at the level of adjacent *student → AI* dyads: each student turn was paired with the immediately following avatar response. This dyad unit captures the most direct test of whether the AI follows the student's immediately prior contribution.

Measures and Analysis

We computed three dyad-level indicators to describe classroom interaction quality in a way that is interpretable for teaching practice: whether the avatar takes up a student's immediately prior contribution (meaning continuity), whether it carries forward key task terms (light lexical grounding), and whether it mirrors the interaction mode of the student turn (brief/social versus task-focused). These indicators are not treated as learning outcomes; they are used as behavioural traces of contingent responding during brainstorming in authentic classroom conditions (e.g., short, messy prompts and mixed communicative purposes).

Semantic Alignment (Meaning Continuity)

In classroom brainstorming, a helpful AI response continues the student's idea, even when the prompt is brief or informal. To measure this uptake, we computed cosine similarity between Sentence-BERT embeddings of each student turn and the next avatar reply (Reimers & Gurevych, 2019). Higher similarity indicates stronger meaning continuity, even when wording differs.

Lexical Alignment (Keyword Reuse / Grounding)

In learning dialogue, shared terminology can support mutual understanding, especially in technical topics where key terms carry the main idea. Lexical overlap is constrained by the amount of topical content in the student turn. We lemmatised content words and removed non-content tokens, then computed Jaccard similarity between the student and avatar lemma sets. If a turn contained no content lemmas, overlap was undefined and retained as an indicator of minimal topical content.

Language Style Matching (Interaction Mode)

In classrooms, not all turns are task-focused. Students may greet the system, test it, or type partial prompts before asking a fuller question. To describe interaction mode in each dyad, we computed Language Style Matching (LSM) from similarity in function-word categories between the student turn and avatar reply (Ireland & Pennebaker, 2010). We use LSM to describe interaction mode and contextualise alignment scores, not as evidence of rapport or learning.

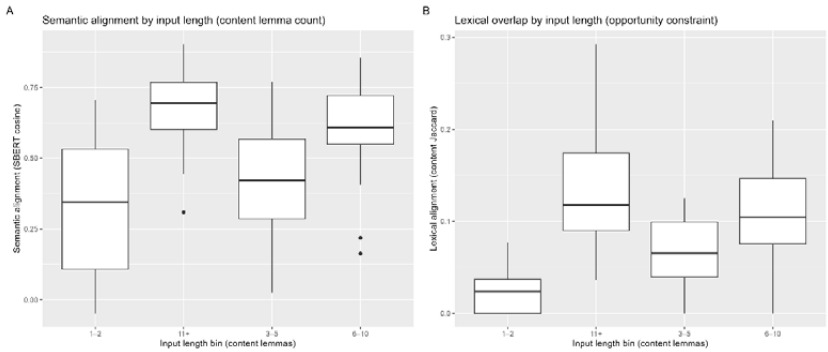
3. Results and Discussion

Corpus Overview

The primary dataset contained 175 dyads for measures available on all turns (semantic alignment and style matching). For lexical overlap, 160 dyads contained usable content after filtering; remaining cases had no content lemmas and were retained as informative "thin/non-content" turns. Dyads were indexed by turn order (sequence), ranging from 3 to 41.

Result 1: Opportunity Structure (Student Input Length)

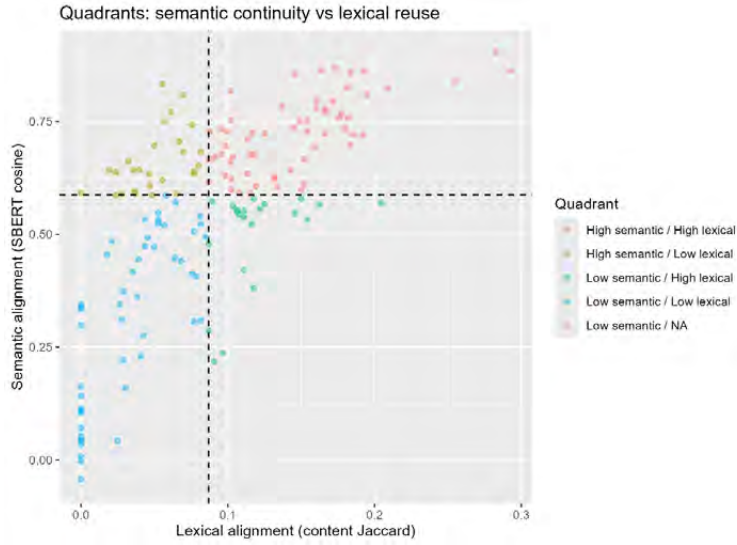
Figure 1 shows a clear "opportunity structure" effect: when students contribute very little usable content, alignment measures have limited material to work with and the avatar's continuity appears less stable. As student prompts become more content-rich, the interaction becomes more consistently on-topic and shared task vocabulary becomes easier to sustain. This suggests that apparent misalignment in classroom logs can reflect brief or fragmentary student input rather than purely the quality of the AI response.



(Figure 1: Opportunity structure by student input length. Semantic alignment (cosine similarity) and lexical overlap (Jaccard) by student content-lemma bins. Meaning continuity increases with richer prompts (median ≈ 0.35 for 1–2 lemmas vs ≈ 0.70 for 11+ lemmas), and lexical overlap rises in parallel (median ≈ 0.02 vs ≈ 0.10–0.12), showing that short classroom turns provide limited opportunity for measurable alignment.)

Result 2: Meaning Continuity Overall, Plus "Keyword Echo" Risk

Figure 2 shows that "good" brainstorming support can occur through more than one productive response mode. The avatar can stay on the student's idea while reusing the student's terminology, or it can stay on the same idea while re-expressing it in different words (paraphrase/elaboration). Importantly, the map also reveals a classroom-relevant risk pattern: responses can echo keywords without reliably tracking the student's intended meaning, which may look responsive on the surface but can undermine trust if students treat keyword reuse as evidence of understanding.



(Figure 2: Quadrant map of meaning continuity (semantic alignment, y-axis) versus keyword reuse (lexical overlap, x-axis). Two productive modes dominate—high semantic + high lexical (34.9%) and high semantic + low lexical (15.4%; paraphrase)—while low semantic + high lexical (12.6%) flags "keyword echo" (term reuse without strong meaning continuity).)

Finding 1: Classroom Prompting Conditions Shape What Responsiveness Can Look Like

Students often contribute short fragments, partial ideas, or social turns, which limits how much alignment can appear in the transcript. In Figure 1, meaning continuity and term carryover increase when students provide more usable content (semantic median ≈ 0.35 at 1–2 lemmas vs ≈ 0.70 at 11+; lexical median ≈ 0.02 vs ≈ 0.10–0.12). Educationally, low alignment in classroom logs may reflect brief student input rather than AI failure. This is relevant because perceived responsiveness may influence engagement and willingness to participate (Chiu, 2021; Giannakos, 2024).

Finding 2: Educationally Useful Uptake Can Occur Through Shared Terminology or Paraphrase

In brainstorming, the reply does not need to repeat student wording; it needs to continue the student's idea in a way that supports ideation. Figure 2 shows two productive patterns: high semantic + high lexical alignment (34.9%), where the avatar continues the idea while carrying forward key terms, and high semantic + low lexical alignment (15.4%), where the avatar continues the idea while rewording it, consistent with paraphrase or elaboration. This suggests that useful AI uptake does not depend on exact keyword repetition; on-topic support may also appear through paraphrase or elaboration (Chiu, 2021). We capture meaning continuity with embedding-based semantic similarity rather than surface matching alone (Reimers & Gurevych, 2019).

Finding 3: Surface Cues of Understanding Are Not Always Reliable in Learning Dialogue

In classroom talk, visible term reuse can be taken as a cue that the system has understood, which may affect whether students and teachers trust the reply. Figure 2 highlights a counter-pattern: *keyword echo*, where lexical overlap is high but meaning continuity is low (12.6%). This pattern shows that word repetition can look responsive even when the reply does not track the student's intended idea. For classroom use, semantic continuity should be interpreted separately from word reuse to avoid overestimating uptake.

4. Implications

Teachers should not judge responsiveness by surface fluency or keyword reuse alone. Interpreting semantic continuity alongside lexical overlap helps distinguish productive paraphrase from *keyword echo*, enabling more reliable review and timely intervention. The opportunity-structure effect suggests that short student prompts limit measurable uptake. Low alignment scores should be interpreted alongside student input length and turn type, not seen as poor AI performance.

Acknowledgments

This study was funded by Lingnan University, Hong Kong, through the Fund for Innovative Technology-in-Education (FITE) of the University Grants Committee (Project Code: 120042) and "UGC FITE IICA #10 Beyond Reality: Unleashing Generative Metaverse Avatars in Education" fund, University Grant Committee, Hong Kong Baptist University, Hong Kong, China.

Ethical Declaration Statement

The research was ethically approved by the Sub-Committee on Research Ethics and Safety of the Research Committee at Lingnan University, Hong Kong (Ref. no. EC176-2324).

Declaration of Use of AI in the Writing Process

The author(s) used ChatGPT during the preparation of this work to improve clarity and concision of the writing. 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

Chiu, T. K. F. (2021). Digital support for student engagement in blended learning based on self-determination theory. *Computers in Human Behavior*, 124, 106909. https://doi.org/10.1016/j.chb.2021.106909

Giannakos, M., Azevedo, R., Brusilovsky, P., Cukurova, M., Dimitriadis, Y., Hernandez-Leo, D., Järvelä, S., Mavrikis, M., & Rienties, B. (2024). The promise and challenges of generative AI in education. *Behaviour & Information Technology*, 1–27. https://doi.org/10.1080/0144929X.2024.2394886

Ireland, M. E., & Pennebaker, J. W. (2010). Language style matching in writing: Synchrony in essays, correspondence, and poetry. *Journal of Personality and Social Psychology*, 99(3), 549–571. https://doi.org/10.1037/a0020386

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. *arXiv*. https://doi.org/10.48550/arXiv.1908.10084

4 Incorporating GenAI in Learning, Teaching and Assessment

ESL Learners' Acceptance of Generative AI Chatbot for IELTS Speaking Practice: An Extended TAM Approach

Ka Yee CHAN [1] & Lin Qian DING [2]
[1] Centre for Language in Education, The Education University of Hong Kong
[2] Department of English Language Education, The Education University of Hong Kong

Abstract

The rise of generative artificial intelligence (GAI) tools has transformed speaking practice by enabling real-time interactions and personalised feedback. However, learners' acceptance of these tools remains under-explored. This study created a GAI chatbot and investigated whether GAI-supported speaking practice influences students' learning and attitudes. An extended Technology Acceptance Model (TAM) was employed, incorporating perceived trust, technology anxiety, and computer self-efficacy alongside traditional constructs. Results showed significant improvements in perceived usefulness, ease of use, trust, self-efficacy, attitudes, and behavioural intention after the GAI intervention, although anxiety levels remained unchanged. Structural equation modelling showed that perceived ease of use strongly influenced both perceived usefulness and attitude towards use, which in turn significantly predicted students' intention to continue using the chatbot. Computer self-efficacy and perceived trust also played important roles in shaping learners' perceptions. The study offers practical insights into integrating GAI into formal classroom language learning through human-AI collaboration.

Keywords: Generative artificial intelligence, chatbot, second language speaking

1. Introduction

Generative artificial intelligence (GAI), particularly large language model (LLM)–based chatbots, is increasingly used to support language learning by enabling real-time, context-aware conversations that resemble authentic interaction (Wan & Moorhouse, 2024). In speaking practice, GAI chatbots help learners identify their weaknesses and improve over time. They also reduce learners' anxiety about making mistakes and encourage more active participation in speaking tasks. However, these technologies are still relatively new in educational contexts, and the potential of GAI chatbots in language learning still needs more empirical evidence (Liu & Ma, 2024).

To evaluate their effectiveness, it is essential to examine how these GAI chatbots are perceived and accepted by English language learners, as understanding learners' attitudes is the first step towards helping make decisions about incorporating these technologies into pedagogical settings (Liu & Ma, 2024). To address this gap, this study applies an extended version of the Technology Acceptance Model (TAM) (Davis, 1989) to examine students' attitudes towards using GAI chatbots for English speaking practice. In addition to core TAM constructs (perceived usefulness and perceived ease of use), the model includes three additional factors: perceived trust, technology anxiety, and computer self-efficacy. The aim is to understand how these variables influence students' acceptance and use of GAI chatbots, and to provide practical insights for educators and developers working to integrate GAI into language education.

2. Literature Review

2.1 Technology Acceptance Model (TAM) and Language Learning

The TAM, developed by Davis (1989), is one of the most frequently cited frameworks for explaining individuals' attitudes and behaviour towards new technologies. It is constructed with four core components: Perceived Usefulness (PU), Perceived Ease of Use (PEOU), Attitude Towards Use (ATU), and Behavioural Intention (BI) (Davis, 1989). PU is the belief that a system improves performance, while PEOU refers to how easy users think the system is to use (Davis, 1989). In TAM, PEOU positively influences PU, and both PU and PEOU have a direct positive influence on users' attitudes towards technology use, which in turn predicts BI (Davis, 1989; Venkatesh & Bala, 2008).

In language learning contexts, PU is often especially salient because learners tend to value technologies that deliver clear learning gains. In our research context, actionable and high-quality feedback tends to make the perceived usefulness of GAI feedback a central driver of acceptance. Therefore, based on the TAM, this study proposes the following hypotheses:

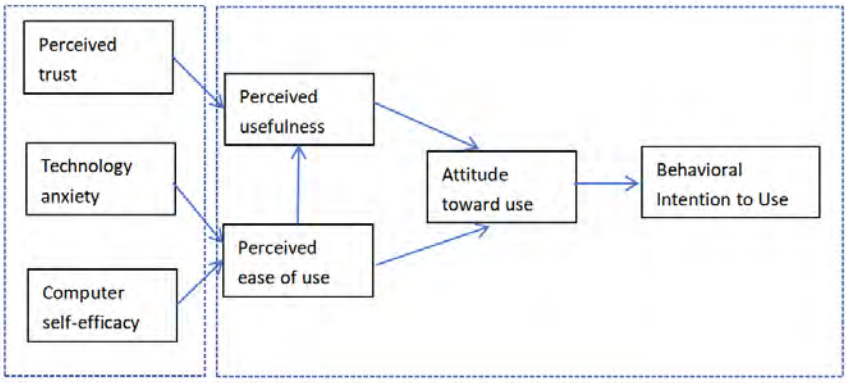
- **H1:** PEOU positively influences PU.
- **H2:** PU positively influences ATU.
- **H3:** PEOU positively influences ATU.
- **H4:** ATU positively influences BI.

2.2 Extended TAM and GAI Execution

The Extended TAM is an expansion of the original TAM. It integrates additional variables that influence users' attitudes and behavioural intentions either directly or indirectly, allowing researchers to better understand user behaviour in more complex environments (Venkatesh & Bala, 2008). This

study adopts the Extended TAM framework to investigate learners' acceptance and usage of GAI for speaking practice. It explores how perceived trust (PT), technology anxiety (ANX), and computer self-efficacy (CSE) influence users' acceptance and behavioural intentions of using GAI for speaking practice (see Figure 1).

Figure 1: The Extended TAM Framework



In this model, PT refers to the belief that a system is reliable, secure, and acts in the users' best interests. ANX refers to users' discomfort or apprehension when interacting with technological systems. CSE refers to an individual's belief in their capability to effectively use computer technologies. Specifically, the study posits that PT positively influences PU, as learners must trust the effectiveness of GAI's feedback to aid their understanding and learning process. Furthermore, PEOU is determined by the interplay of CSE and ANX. While high self-efficacy promotes ease, anxiety acts as a barrier. Crucially, this research highlights that mitigating technology anxiety by enhancing students' AI literacy is essential, as a less anxious, more literate approach ensures that the integration of GAI tools leads to successful and sustainable language learning experiences.

Therefore, this study proposes the following hypotheses:

- **H5:** Perceived trust (PT) positively influences perceived usefulness (PU).
- **H6:** ANX negatively influences PEOU.
- **H7:** CSE positively influences PEOU.

3. Methodology

3.1 Participants and Procedure

The study involved 90 Year 2 to 3 students from a university in Hong Kong. Most were second-year university students aged 18–20 enrolled on a compulsory IELTS Speaking course. All participants were English as a Second Language (ESL) learners, with Cantonese or Mandarin as their first language. They followed the course schedule and completed GAI tasks in Part 2 of the course (see Table 1).

Table 1: Course Timeline, Teaching Content, and Data Collection Schedule

Sessions	Teaching Content	Data Collection
Session 1–6	Part 1: Learning Writing in IELTS	Pre-questionnaire
Session 7–9	Part 2: Learning Speaking in IELTS	Before GAI intervention
Session 10	GAI feedback on speaking practice as a self-learning task; GAI Chatbot task with reflection writing on learning experiences	During/After Session 10
Session 10–11	Teacher-student consultation on their speaking performance	N/A
Session 12–13	Mock Test on IELTS Speaking	Post-questionnaire
After Session 13	Official IELTS Test	N/A

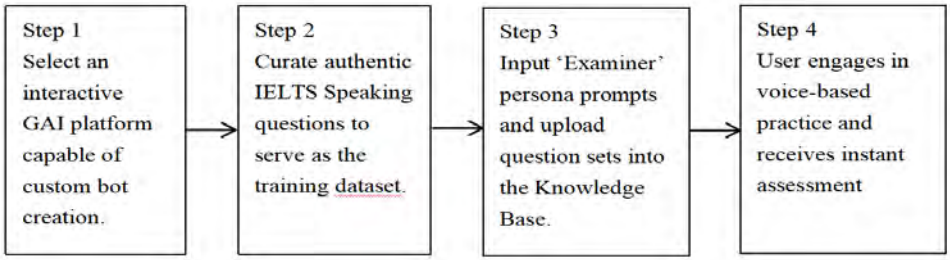
3.2 Design of GAI Chatbot

In this study, we developed a custom GAI chatbot on the Monica platform to provide structured IELTS Speaking practice with automated feedback. Monica was selected because it allowed the creation of a customised chatbot with stable role instructions and an interface that students could access conveniently outside class. The procedure is shown in Figure 2, and the prompts for the GAI chatbot setup are suggested below.

3.2.1 Procedure to Set Up GAI Chatbot

The chatbot was designed to simulate the role of an IELTS Speaking examiner. To achieve this, we created a custom instruction set that defined the chatbot's pedagogical role, interaction flow, and feedback principles. First, the chatbot was instructed to conduct the speaking task according to the three-part structure of the IELTS Speaking test. Second, it was guided to interact in an examiner-like manner by asking questions sequentially, maintaining the test format, and avoiding excessive prompting during students' responses. Third, after the speaking task, it was instructed to generate feedback aligned with the IELTS Speaking Band Descriptors rather than giving only generic language-learning advice. In particular, the feedback focused on criteria such as fluency and coherence, lexical resource, grammatical range and accuracy, and pronunciation.

Figure 2: Design and usage workflow of the custom GAI chatbot



3.2.2 Prompts Provided to GAI Chatbot

The core prompt given to the chatbot defined both its examiner role and its feedback function. In summary, the chatbot was instructed to:

(a) act as an IELTS Speaking examiner;
(b) guide students through the speaking test in an ordered sequence;
(c) ask one question at a time;
(d) avoid over-helping during the interaction;
(e) provide feedback after the task based on the IELTS Speaking Band Descriptors; and
(f) offer concise, actionable suggestions for improvement.

The prompt also reminded the chatbot to frame its comments as formative feedback rather than as a definitive official score. During the intervention, all participants followed the same workflow. Students were asked to complete the chatbot task independently. They used the custom chatbot to simulate an IELTS Speaking test, responded to the prompts, reviewed the automated feedback, and wrote a short reflection on their learning experience. Question prompts were provided to guide students to reflect more critically and in depth on their learning journey through the GAI chatbot. The GAI chatbot task guideline and the reflection guided questions for students are available in Appendix 1.

3.3 Data Collection and Data Analysis

A quantitative approach was utilised to collect students' perceptions of using GAI for speaking practice. The data was collected through pre- and post-questionnaires based on our extended TAM framework. Two questionnaires were identical in content. They contained 26 items across seven constructs: PT, ANX, CSE, PU, PEOU, ATU, and BI. All items were measured using a five-point Likert scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree). All items were adapted from previously validated instruments (e.g., Zou et al., 2025).

First, we conducted paired-sample t-tests to compare participants' pre- and post-questionnaire results across the seven constructs in the extended TAM model. This analysis allowed us to observe any significant shifts in students' perceptions after engaging with the GAI chatbot. With regard to the test of hypotheses 1 to 7, a two-step structural equation modelling (SEM) approach was applied. In the first step, confirmatory factor analysis (CFA) assessed the measurement model. This step examined the reliability of each construct through Cronbach's alpha and composite reliability, and evaluated convergent validity. In the second step, the structural model tested the hypothesised relationships among the variables. The analysis focused on path coefficients and overall model fit.

4. Results and Discussion

4.1 Changes in Students' Perceptions Before and After Using GAI for English-Speaking Practice

Table 2 presents the results of paired-sample t-tests. Statistically significant improvements were found in six of the seven constructs after the GAI chatbot intervention. There were statistically significant improvements in PT, CSE, PU, PEOU, ATU, and BI. These results indicate that participants' perceptions in these areas significantly increased after the GAI intervention. However, the difference in ANX scores between pre-questionnaire and post-questionnaire was not statistically different. This suggests that the intervention had no significant effect on participants' anxiety levels.

Table 2: Changes in students' perceptions before and after AI-supported practice

Construct	Mean Difference	SD Difference	t	df	p (two-tailed)
PT	-0.41	0.86	-3.80	63	< .001
ANX	-0.08	1.29	-0.49	63	.629
CSE	-0.39	0.83	-3.73	63	< .001
PU	-0.46	0.86	-4.35	63	< .001
PEOU	-0.82	0.72	-9.03	63	< .001
ATU	-0.52	0.84	-4.94	63	< .001
BI	-0.58	0.77	-6.01	63	< .001

Overall, the findings suggest that after using the GAI chatbot for speaking practice, students developed a more favourable attitude towards its use. Table 2 shows changes in students' perceptions before and after AI-supported practice. Moreover, it suggests that the GAI chatbot successfully empowered students by providing an accessible, low-pressure environment for speaking practice, which enhanced their self-efficacy and trust in the technology. However, the

persistence of anxiety indicates that the stress of high-stakes exam simulations remains a challenge that may still require human emotional support or scaffolded interventions.

To conclude, the high perceived usefulness and ease of use confirm that such tools are viable, student-friendly supplements for fostering autonomous language learning. Our qualitative data from students' reflections further illuminate how the chatbot intervention was experienced in practice. The most frequently mentioned strengths were flexibility, accessibility, and immediacy of feedback. Many students reported that the chatbot enabled them to practice "anytime and anywhere without the need for a human instructor." This suggests that the tool supports autonomous learning beyond scheduled class time.

Students also described the feedback as "immediate," "holistic," and "personalised." One student wrote, "I think the feedback from ChatGPT is very helpful and detailed. It gives me a lot of comments on my performance at each part and states the strengths and improvements for me to focus more." These responses help explain the significant post-intervention increases in perceived usefulness and perceived trust.

Moreover, several participants noted that they felt more relaxed and less self-conscious when speaking to a machine than to a human interlocutor. This reduced sense of being judged may have created a psychologically safer environment for repeated practice, which in turn contributed to more favourable attitudes towards the tool. At the same time, the reflections also reveal important limitations. Some students felt that the chatbot interaction was less realistic than a face-to-face speaking test because it lacked eye contact, facial expressions, and authentic human interaction. As one student commented, "I worry that I may not get used to the real IELTS speaking test." This concern may help explain why anxiety did not significantly decrease after the intervention. Although the chatbot reduced interpersonal pressure, it may not have fully addressed the broader anxiety associated with high-stakes oral examinations.

In addition, many students reported increased confidence in their ability to use AI for learning, which is consistent with the observed improvement in computer self-efficacy. Students generally found the interface manageable and the activity easy to complete independently. However, some also mentioned technical limitations, including inaccurate voice-to-text transcription, the lack of audio responses, and the absence of an integrated timer for IELTS Part 2 preparation. These issues may have constrained perceptions of ease of use for some learners.

Regarding future use, most participants expressed a willingness to continue using the chatbot for its convenience and feedback value, although some still preferred human teachers, whom they considered more authentic and better able to provide nuanced feedback based on learners' actual proficiency and needs.

4.2 Validation of the Research Model and Hypothesis Testing

To ensure the robustness of our findings, we first confirmed the measurement model's reliability and validity. We observed that all constructs exceeded recommended thresholds for Cronbach's alpha (>0.70), composite reliability (>0.70), and average variance extracted (>0.50), with discriminant validity confirmed via the Fornell-Larcker criterion. Following this validation, we evaluated the structural model using the coefficient of determination (R^2). The R^2 values for PU, ATU, and BI were 0.479, 0.604, and 0.602 respectively. This indicates moderate to substantial explanatory power. However, PEOU's R^2 value was relatively low (0.184), which indicates that the current predictors explained only a small portion of its variance. This weak explanatory power suggests that additional factors may be necessary to better account for learners' perceptions of ease of use when interacting with GAI tools.

Table 3: Hypothesis Test Results

Hypothesis	Path	β	t-value	p-value	Supported
H1	PEOU → PU	0.463	4.05	<0.001	Yes
H2	PU → ATU	0.213	1.27	0.203	No
H3	PEOU → ATU	0.625	4.72	<0.001	Yes
H4	ATU → BI	0.776	12.16	<0.001	Yes
H5	PT → PU	0.332	3.44	<0.001	Yes
H6	ANX → PEOU	0.080	0.65	0.519	No
H7	CSE → PEOU	0.400	2.95	0.003	Yes

Note: Bootstrapping with 5,000 subsamples was used to test significance. Significance threshold: $p < 0.05$.

Next, the path coefficients and their statistical significance were evaluated to examine the structural relationships within the model. As shown in Table 3, most hypothesised relationships were supported. ATU had the strongest direct influence on BI, while PEOU significantly influenced both PU and ATU. In addition, CSE significantly affected PEOU, and PT positively predicted PU. However, not all relationships were significant. The path from PU to ATU did

not reach statistical significance. Similarly, ANX did not significantly influence PEOU, which suggests that technology-related anxiety may have a limited role in shaping perceived ease of use.

The strong influence of H3 compared to the non-significant path of H2 suggests that for students using GAI tools, a user-friendly interface is a more critical driver of positive attitudes than the tool's perceived utility alone. Pedagogically, this implies that while the GAI feedback quality is important, students are more likely to embrace the technology if they feel confident in their ability to navigate it rather than simply being told it is useful. Furthermore, the lack of impact of H6 indicates that even anxious learners found the tool accessible, which reinforces the idea that well-designed AI interfaces can lower the affective barriers typically associated with language learning technology.

5. Implications

Our study utilised the extended TAM to explore how university-level ESL learners perceived and adopted a GAI chatbot for IELTS speaking practice. The findings show that the intervention significantly enhanced students' trust, self-efficacy, and behavioural intention, with perceived ease of use and attitude emerging as the critical drivers of adoption, even though anxiety levels remained unchanged.

Aligned with Wang et al. (2025), the findings suggest that the user-friendly interface reduced students' cognitive load during speaking practice and allowed them to concentrate on functionality. Therefore, educators should select GAI tools with intuitive designs to lower technical barriers to ensure sustained student engagement. Furthermore, the significant roles of PT and CSE mirror the conclusions of Shahzad et al. (2025), who argue that trust in a system's reliability and security is a prerequisite for its use. As a result, institutions should implement training programmes that not only teach technical skills but also build trust in the AI's feedback accuracy (Shahzad et al., 2025).

Finally, the persistence of technology anxiety in this study suggests that GAI cannot fully replace human emotional support. A blended learning approach is therefore recommended: GAI should act as a "native-speaker" simulator for routine practice (Wang et al., 2025), while instructors focus on scaffolding and addressing the affective dimensions of language anxiety that AI cannot alleviate.

Our study offers practical implications not only for technology acceptance research but also for the design of GAI-supported speaking interventions.

The chatbot created in the study was not used as a general conversational AI tool; rather, it was intentionally customised for IELTS Speaking practice by embedding the test structure, examiner role, and IELTS Speaking Band Descripto rs into the chatbot's instructions. This suggests that the pedagogical value of GAI in language learning may depend heavily on task-specific design.

For teachers and institutions, an important implication is that AI tools should be configured around clearly defined assessment criteria to provide feedback that students perceive as relevant, credible, and actionable. The intervention also demonstrates how GAI can be incorporated into a blended learning sequence. In our course, the chatbot served as a formative feedback channel before teacher consultation and mock testing. This allowed students to practice independently, receive immediate AI-generated feedback, and then revisit their performance with human guidance. Such a design may be particularly useful in large classes or test-preparation courses where opportunities for individualised speaking feedback are limited.

In addition, the activity helped students develop a clearer understanding of how speaking performance is evaluated, thereby enhancing their AI-criticality. With understanding, students can make judgments about the quality of feedback from the GAI, which allows them to become more critical self-markers across a variety of formative feedback from peers, the GAI, and teachers. This may be transferable to other public examination contexts in which students need repeated practice together with criterion-based formative feedback.

At the same time, the findings highlight the limits of AI-only support. Although students generally valued the chatbot's flexibility, immediacy, and low-pressure environment, some still preferred human feedback, and some remained concerned about the lack of realistic face-to-face interaction. This suggests that GAI is best used to supplement rather than replace teachers. For speaking instruction, a balanced model may be especially effective: AI can support repeated independent practice and rapid formative feedback, while teachers provide emotional support, contextualised feedback, and preparation for the interpersonal demands of high-stakes oral assessment.

Ethical Declaration Statement

The project has been approved by the Faculty of Humanities of The Education University of Hong Kong (Reference Number: T0293).

Generative AI Policy

The authors used POE during the preparation of this work to check

for grammatical errors and improve language expression. After using the tool/service, the authors thoroughly reviewed and edited the content as needed and assume full responsibility for the final version of the manuscript.

References

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340. https://doi.org/10.2307/249008

Liu, G., & Ma, C. (2024). Measuring EFL learners' use of ChatGPT in informal digital learning of English based on the technology acceptance model. *Innovation in Language Learning and Teaching*, 18(2), 125–138.

Shahzad, M. F., Xu, S., & Asif, M. (2025). Factors affecting generative artificial intelligence, such as ChatGPT, use in higher education: An application of technology acceptance model. *British Educational Research Journal*, 51(2), 489-513.

Venkatesh, V., & Bala, H. (2008). Technology acceptance model 3 and a research agenda on interventions. *Decision Sciences*, 39(2), 273–315. https://doi.org/10.1111/j.1540-5915.2008.00192.x

Wan, Y., & Moorhouse, B. L. (2024). Using Call Annie as a generative artificial intelligence speaking partner for language learners. *RELC Journal*. https://doi.org/10.1177/00336882231224813

Wang, C., Li, X., & Zou, B. (2025). Revisiting Integrated Model of Technology Acceptance Among the Generative AI-Powered Foreign Language Speaking Practice: Through the Lens of Positive Psychology and Intrinsic Motivation. *European Journal of Education*, 60(1), e70054.

Zou, B., Wang, C., Yan, Y., Du, X., & Ji, Y. (2025). Exploring English as a foreign language learners' adoption and utilisation of ChatGPT for Speaking practice through an extended technology acceptance model. *International Journal of Applied Linguistics*, 35(2), 689-704. https://doi.org/10.1111/ijal.12658

Appendix 1: GAI Chatbot Task Guidelines for Students

Blended Learning Task 2 Guidelines
Activity 2a: Attempting a speaking task (5%)

Objectives

Blended learning aims to extend learning beyond classroom hours. It also enhances your skills and awareness specified in the University's Generic Intended Learning Outcomes (GILOs). Although blended learning is completed out of class, what you work on is course-related and requires a similar amount of time and effort.

Submission Deadline:

Activity 2a: Attempting a speaking task on ChatGPT and writing a reflection.

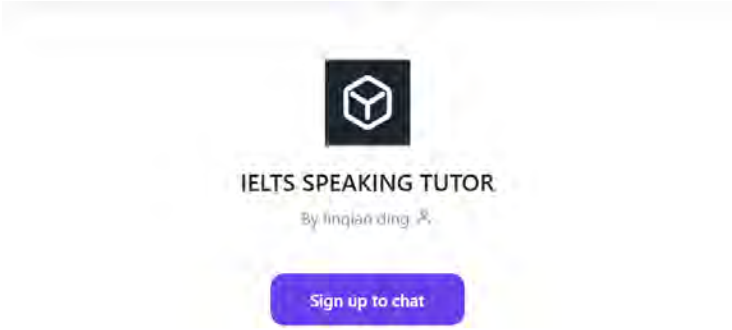
- By the end of class time in Session 10.
- On ZOOM – Weeks 10 & 11.

Instructions:

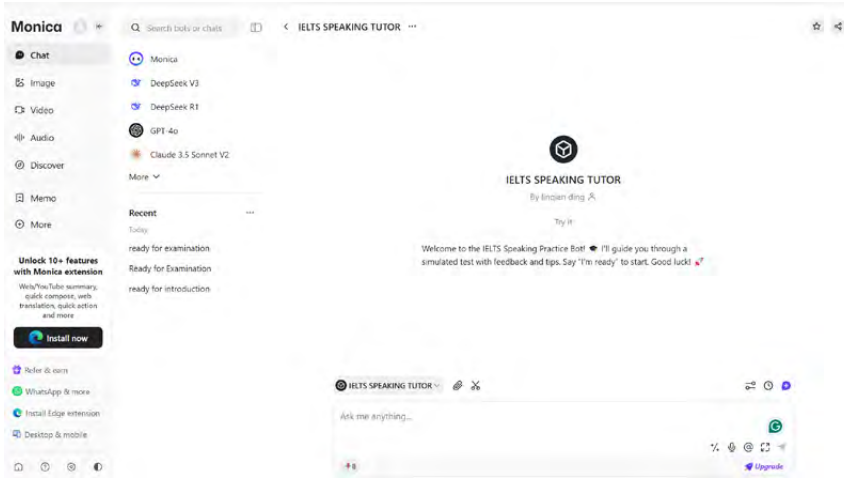
In order to complete this activity, you need to complete speaking practices on a GenAI platform and write a reflection on your learning experiences by using the reflection form below, and submit it to Moodle Turnitin by the end of the lesson time in Session 10. Afterwards, you will need to attend a speaking session with your teacher to practice your speaking skills. Please use the guidelines below to complete the task.

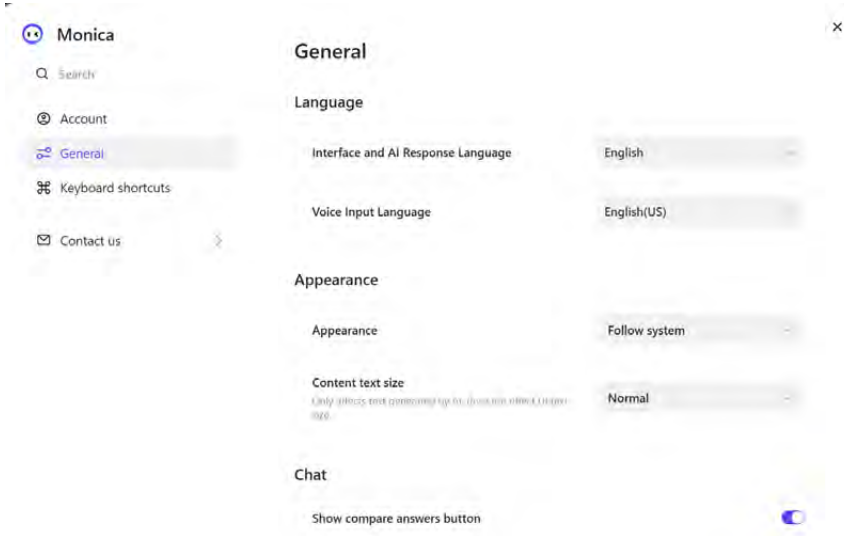
Guidelines for Using the IELTS Speaking Practice Bot:

- **Task 1:** Click on the link: `https://monica.im/share/bot?botId=avmxkyYR`. Use your school email to register and start chatting.

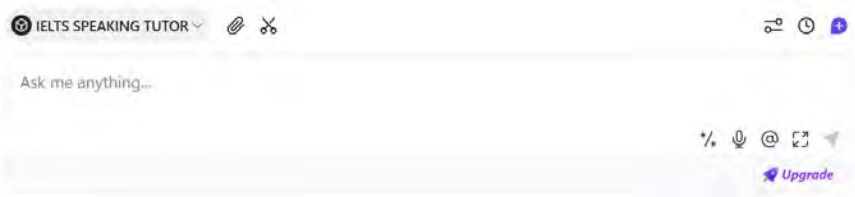


- **Task 2:** You will enter a realistic IELTS Speaking test environment designed to help you practice and improve your speaking skills. Before you start, click the **Settings** button at the bottom-left corner of the page. Set both the interface and AI response language, as well as the voice input language, to English.





- **Task 3:** Click the microphone icon in the bottom-right corner of the chat box and say "I'm ready" to start the speaking test. Please note that during the test, you must use your microphone to answer questions—do not type your answers.



- **Task 4:** The bot will guide you through the three parts of the IELTS Speaking test:
  * **Part 1:** Introduction and Interview (you will answer basic questions about yourself).
  * **Part 2:** Long Turn (you will have 1 minute to prepare and 1–2 minutes to speak on a given topic).
  * **Part 3:** Discussion (you will answer more in-depth and abstract questions related to Part 2). Read carefully to the bot's instructions and respond to each question as if you were in a real IELTS Speaking test.
- **Task 5:** After completing each part, the bot will provide detailed feedback on your performance, including your strengths and areas for improvement. Use this feedback to enhance your speaking skills.
- **Task 6:** When the test is complete, the bot will provide an overall evaluation of your speaking ability. Screenshot the overall evaluation and paste it on the reflection form. Write up a 150-word reflection by following the reflection guided questions. You can also take note of the feedback and work on the recommendations to improve for your next practice session.

Blended Learning Component 2a: Reflection Record Sheet

Class group: xxE

Name: _____ (**Student no.:** _____)

Speaking Test ChatGPT Feedback Screenshot	**150-word Reflection on the Speaking Practices**
	Please use the guided questions to help you write the speaking practice process reflection: 1. Is it the first time you worked on ChatGPT to practice speaking? 2. Do you find this useful to help you practice alone? 3. Which part(s) did you find the most challenging? Why? 4. What are the other difficulties? 5. Based on the feedback from ChatGPT in this practice experience, what can you do to improve your performance in the next test? 6. Will you plan to use ChatGPT to practice more? You can browse the following website to explore some tips for handling different question types: Listening tips: https://ieltsliz.com/ielts-speaking-free-lessons-essential-tips/ Suggested platform (Monica): https://monica.im/?ref=yzq2zjj&ref_aff=yzq2zjj&gad_source=1&gclid=EAIaIQobChMIyNHwnPCjiwMVT_pMAh3dXg4hEAAYASAAEgI0T_D_BwE/ AI Speaking at ASLLC

Student-Staff Co-Developed Immersive Virtual Reality and Engineered AI Prompts to Enhance Therapeutic Communication

Debbie WONG [1], Henry CHAN [1], Ivan CHAN [2], Kitty CHAN [2], Maggie CHU [3], Mephisto WU [2], Vivian HUI [2]

[1] Industrial Centre, The Hong Kong Polytechnic University

[2] School of Nursing, The Hong Kong Polytechnic University

[3] Department of Electronic and Electrical Engineering, The Hong Kong Polytechnic University

Abstract

Although students frequently use generative AI for various purposes, their limited understanding of effective usage and critical evaluation often reduces its educational value. This study examines the impact of AI-assisted prompt engineering and explicit instruction in therapeutic communication on nursing students' clinical communication skills during immersive virtual reality (IVR) simulations. Analysis of thirty-seven student scripts across three stages—initial, AI-refined, and therapeutically enhanced—revealed that initial responses were factual and impersonal, reflecting limited AI literacy. Without explicit guidance, students struggled to optimise AI outputs for therapeutic communication. However, prompt engineering led to more structured, empathetic, and patient-centred responses. Thematic analysis of student reflections indicated increased confidence and self-awareness. These findings underscore the essential role of prompt engineering in empowering students to maximise AI's educational benefits and develop advanced communication skills, highlighting the need for targeted instruction to fully realise AI's potential in nursing education.

Keywords: Nursing Education; Virtual Reality; Communication; Prompt Engineering; Generative AI

1. Introduction

Artificial intelligence (AI) is rapidly transforming higher education through applications such as adaptive learning and personalised feedback (Sallam, 2023). However, despite students' frequent use of generative AI, many lack the skills in effective prompt engineering and critical evaluation, which often limits its educational value. While AI-driven simulations and immersive virtual reality (IVR) have shown promise in enhancing experiential learning and communication skills among healthcare students (Elendu et al., 2024), evidence

on using AI and prompt engineering to improve therapeutic communication in scenario-based activities remains scarce (Patil et al., 2024).

This study addresses this gap by evaluating how an IVR game—developed through a student-staff partnership and incorporating AI-driven prompt engineering—affects undergraduate nursing students' therapeutic communication skills and self-reflection in simulated clinical scenarios. By combining immersive technology with structured prompt engineering, the study explores a practical approach to maximising AI's potential in nursing education.

2. Methodology

Study Design

This study employed a retrospective, qualitative descriptive design to explore undergraduate nursing students' experiences and perceived learning outcomes following participation in an immersive VR-based communication skills tutorial. The intervention was delivered as a three-hour tutorial session within the regular curriculum for Year 1 undergraduate nursing students.

Intervention Development

The IVR game and three-hour tutorial were co-developed by two nursing students and one engineering student with faculty guidance from the School of Nursing and the Industrial Centre. During the tutorial, students worked in pairs, participated in two scenarios, and alternated as communicator and observer. The learning process followed three structured steps:

1. **Interactive Gameplay and Peer Critique:** Students interacted with the virtual patient avatar, recorded their spoken responses, and generated speech-to-text transcripts for peer feedback.
2. **Prompt Engineering for Dialogue Refinement:** Using prompt engineering techniques, students iteratively refined their dialogues using various AI tools according to their preference and selection.
3. **Application of Therapeutic Communication Principles:** Students further enhanced their conversations by applying therapeutic communication principles.

Students were also involved in the game's development and reflected on their learning experiences.

Data Collection

Data were collected from undergraduate nursing students during the IVR

simulation tutorial session using an in-class worksheet to document their communication scripts and game reflections.

Outcome Measures

Stepwise Communication Competency

To evaluate the impact of prompt specificity on the quality of therapeutic communication in a stepwise process, the 5-item DPCC-MS (Doctor-Patient Communication Coding – Medical Students) framework was used as the primary assessment tool (Renaud & Côté, 2017) to assess healthcare students' performance with a 4-point Likert scale (Superior = 4, Expected = 3, Borderline = 2, Insufficient = 1), plus the 'not applicable' option (Cronbach's alpha = 0.90).

Students interacted with virtual patients and recorded their initial spoken responses using speech-to-text technology. All student responses were categorised into three stages to examine their progression in communication competence, benchmarking against the DPCC-MS:

1. Initial Student Response
2. AI-refined Response
3. AI-refined and Therapeutically Enhanced Responses

Game Reflection: Perceived Impact on Communication Skills and Self-Reflection

Students were invited to submit a written reflection as part of their in-class activity. The guiding question for reflection was: "Share your experience with the game briefly."

Data Analysis

Content analysis of the student-generated nurse-patient communication scripts from the educational activities was conducted. The initial analysis was performed using an AI-based assessment tool, which identified the presence and quality of these communication elements. For the reflections, key themes were extracted from the written reflections, focusing on students' experiences, perceived improvements, challenges, and insights gained throughout the intervention. The analysis was subsequently reviewed and refined by two independent researchers to ensure credibility and depth of interpretation.

3. Results and Discussion

Forty Year 1 students enrolled in the undergraduate nursing programme participated in the VR-based prompt engineering learning activity as part of their regular tutorial class. Of these, 37 students (female: n = 23) submitted

written reflections on their experiences during the session. All submitted responses were included in the analysis, with implied consent for use in publication. The qualitative feedback was systematically reviewed to identify key themes in students' perceptions of the activity.

Patterns of Communication Progression, Prompt Utilisation, and Benchmark Achievement

Analysis of stepwise communication competency showed clear improvement across three stages. Initially, students' responses were factual and lacked emotional engagement. After AI-assisted self-refinement, responses became more structured and empathetic. Explicit prompts consistently led to superior or expected DPCC-MS ratings, while vague or generic prompts resulted in only minimal or superficial changes, often remaining at borderline or insufficient levels. Moderately explicit prompts produced some improvement but did not consistently reach the superior level.

Four distinct patterns of progression were identified (Table 1).

Key Themes from Student Reflections on VR-Based Communication Learning

Analysis of students' reflections after the VR exercise revealed key themes regarding how AI-assisted learning improved their communication skills, confidence, self-awareness, and approach to patient interactions:

1. **Immersive Realism and Engagement**
   "The overall design for the game is smooth; patients have realistic responses just like what a patient may ask in a hospital setting. It really inspires me when I am faced with such situations." (Student 1)

2. **Communication Skill Development**
   "In this activity, after I give out my words, the system will introduce an appropriate response which helps me to understand how I can improve my communication skills." (Student 34)

3. **Emotional Awareness and Patient-Centred Care**
   "The conversation simulation helped me recognise what I am going to face while communicating with clients and building a sense of sensitivity to clients' emotions, fostering a trust-caring relationship." (Student 5)

4. **Reflection and Self-Improvement**

"This VR game allows me to reflect on my immediate response to the

patient's concerns and questions, and thus I can practice my communication skills based on the weakness shown during the conversation with Roy." (Student 19)

5. **Creativity and Innovation in Learning**
"The pills machine is quite a creative idea for us as new users in VR and attracts us to keep playing it." (Student 18)

Discussion

The integration of AI-assisted prompt engineering with explicit instruction in therapeutic communication represents a significant advancement in nursing education, enhancing students' clinical communication skills. Blended approaches facilitate skill transfer (Ching & Ho, 2025), with VR simulations promoting critical thinking, practice, feedback, and empathy. Furthermore, higher levels of AI literacy are associated with improved communication competency.

Consistent with recent studies, the combination of AI technologies, particularly in IVR environments, with structured teaching—as scoping reviews highlight the benefits of AI-driven personalisation, iterative feedback, and VR's immersive learning experiences (Cucci et al., 2025)—is effective. Generative AI tools enhance confidence and learning outcomes when supported by deliberate pedagogical scaffolding (Gonzalez-Garcia et al., 2025).

Our findings reveal that students' initial responses were predominantly factual and impersonal, lacking emotional validation, indicating limited AI literacy and an inability to effectively leverage generative AI for self-learning. Without explicit guidance, students struggled to optimise AI outputs for therapeutic communication. The introduction of explicit prompt engineering markedly transformed response quality, resulting in more structured, empathetic, and patient-centred communication, with the highest competency levels observed in the final stage.

Despite these promising results, the study has several limitations, including a single-cohort design, lack of a control group, scripted simulations, and reliance on self-reported reflections. Future studies should employ objective measures and assess long-term skill transfer.

4. Implications

This study concludes that integrating AI-driven prompt engineering with IVR training effectively enhances nursing students' empathetic, patient-centred communication, self-reflection, self-learning, and advanced skills. Structured

prompts yielded the strongest improvements, while reflections revealed greater engagement and self-awareness. These findings underscore the pivotal role of prompt engineering in realising the full educational potential of AI. On a global scale, they signal AI-integrated education as a promising future trajectory in healthcare pedagogy. Combining structured AI prompts with immersive simulation provides a scalable and effective model for therapeutic communication training. Continued guidance and feedback are recommended, alongside future multi-site, controlled studies to evaluate long-term retention and clinical transferability.

Acknowledgements

This research was supported by the Makubator Fund, Industrial Centre, The Hong Kong Polytechnic University. The authors gratefully acknowledge this funding support.

Ethical Declaration Statement

The project has been approved by the Departmental Research Committee of The Hong Kong Polytechnic University (Reference Number: HSEARS20251004007).

Generative AI Policy

The authors used GPT-4.1 to extract preliminary themes and elements from communication response and reflection scripts, as described in the data analysis section, and for grammar editing. Scite was utilised during the literature review process. 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

Ching, Y.-C., & Ho, Y.-C. (2025). AI virtual human-augmented game-based teaching to enhance emotional intelligence in nursing students: Protocol for a single-group pretest-posttest action research study. *JMIR Research Protocols*, 14, e80290. https://doi.org/10.2196/80290

Cucci, F., Marasciulo, D., Romani, M., Soldano, G., Cascio, D., De Nunzio, G., Caldararo, C., Rubbi, I., Vitale, E., Lupo, R., & Conte, L. (2025). The contribution of artificial intelligence in nursing education: A scoping review of the literature. *Nursing Reports*, 15(8), 283. https://doi.org/10.3390/nursrep15080283

Elendu, C., Amaechi, D. C., Okatta, A. U., Amaechi, E. C., Elendu, T. C., Ezeh, C. P., & Elendu, I. D. (2024). The impact of simulation-based training

in medical education: A review. *Medicine (Baltimore)*, 103(27), e38813. https://doi.org/10.1097/MD.0000000000038813

Gonzalez-Garcia, A., Bermejo-Martinez, D., Lopez-Alonso, A. I., Trevisson-Redondo, B., Martín-Vázquez, C., & Perez-Gonzalez, S. (2025). Impact of ChatGPT usage on nursing students education: A cross-sectional study. *Heliyon*, 11(1), e41559. https://doi.org/10.1016/j.heliyon.2024.e41559

Patil, R., Heston, T. F., & Bhuse, V. (2024). Prompt engineering in healthcare. *Electronics (Basel)*, 13(15), 2961. https://doi.org/10.3390/electronics13152961

Renaud, J.-S., & Côté, L. (2017). Validation of the 5-item doctor-patient communication competency instrument for medical students (DPCC-MS) using two years of assessment data. *BMC Medical Education*, 17(1). https://doi.org/10.1186/s12909-017-1026-9

Sallam, M. (2023). ChatGPT utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns. *Healthcare*, 11(6), 887. https://doi.org/10.3390/healthcare11060887

Table 1. Communication Progression Patterns, Example Prompts, and DPCC-MS Benchmark Ratings (n = 34)

Communication Pattern	Scripts (n)	Example of Prompt Used (Student code number)	DPCC-MS Benchmark at Final Stage	Description	Example Response	
					Initial	Final
Most Effective Initial Conversations	3	No prompt needed or simple prompt: Please check for clarity." (S03)	Superior (4) or Expected (3)	Consistently strong patient-centered communication from the outset, maintained or enhanced through each stage.	"I know you are negative now. You can have your autonomy to choose your things to do before you die. But our nursing team will help you and try our best to help you recover."	
Least Effective Initial but Best Progressions	14	Explicit prompt: "Refine this response using therapeutic communication techniques such as empathy, validation, normalization, and open-ended questions." (S11)	Superior (4) or Expected (3)	Marked improvement from borderline/insufficient to superior/expected after explicit prompting.	"It is just for the health screening and don't be worried about this."	"It sounds like you're wondering why this is necessary, which is a very common question, especially for someone your age. Can you tell me what specific part of the procedure is on your mind right now?"
Moderate/Mixed Progression	10	Moderately explicit prompt: "Make this response more supportive and clear." (S22)	Expected (3) or Borderline (2)	Some improvement, often reaching expected but not superior; communication is supportive but not fully patient-centred.	"You should not worry."	"It's understandable to be concerned. Let's talk about your worries."
Least Effective Performance Without Significant Progression	7	Vague/generic prompt: "Can you improve this response?" (S29)	Borderline (2) or Insufficient (1)	Minimal change, typically due to vague prompts; communication remains factual, directive, or lacking empathy.	"I understand you are afraid of it but the information from Google must not be 100% correct…"	"I understand you are afraid of it but the information from Google must not be 100% correct…"

Note:

- **DPCC-MS Benchmark at Final Stage:** Indicates the typical rating achieved in the final response for scripts in this pattern, based on the DPCC-MS scale (Superior = 4, Expected = 3, Borderline = 2, Insufficient = 1).
- **Example of Prompt Used:** Illustrates the type of AI prompt that led to the observed communication pattern.

The Impact of AI Writing Tools on University Students' Academic Writing in English

Hsuan-Yu CHEN [1] & Yi-Chun KUO [1]
[1]Department of Foreign Languages, National Chiayi University, Taiwan

Abstract

Academic English writing presents significant, universal challenges even for native speakers, let alone English as a Foreign Language (EFL) learners. This difficulty is particularly pronounced among English majors, who face the most rigorous demands in English composition. The advent of Generative AI may offer valuable assistance to address these hurdles. This mixed-methods study investigated the effects of AI writing tools on the academic English writing performance and perceptions of 18 second-year English majors in a Taiwanese EFL context, alongside a control group of 21 sophomore English majors from a closely comparable cohort. The AI-supported group improved from Draft 1 to Revision (M = 77.67 to 82.83), yielding an average gain of 5.17 points. However, the control group demonstrated a larger descriptive gain (M = 72.81 to 82.29; gain = 9.48), suggesting that AI-supported revision did not produce greater short-term score improvement in this specific dataset. Qualitative reflections nevertheless indicated that students used AI strategically to reorganise structure, clarify topic sentences, improve transitions, and strengthen evidence. Simultaneously, they emphasised human oversight and responsible use, indicating the development of Critical AI Literacy (CAIL). The findings suggest that AI may function more effectively as a revision scaffold than as a guaranteed route to higher scores. Pedagogically, AI integration should therefore be paired with explicit revision instruction and CAIL training.

Keywords: Generative AI, Academic English Writing, Writing Performance, Critical AI Literacy

1. Introduction

Academic English writing presents significant challenges for English as a Foreign Language (EFL) learners at the tertiary level. These students frequently struggle with linguistic accuracy, complex sentence structures, and the logical organisation required for scholarly communication (Aldabbus & Almansouri, 2022). This difficulty is particularly pronounced for English majors, who face intensive writing demands across their curriculum, underscoring the critical need for effective writing support.

The emergence of generative AI, such as ChatGPT and Gemini, has introduced AI-powered tools that offer diverse functions to address these challenges, ranging from correcting surface-level errors to providing generative assistance for brainstorming, outlining, and content refinement (Durrani et al., 2025). Previous research suggests that AI can enhance writing proficiency, boost self-assurance, and reduce writing anxiety among EFL learners (Eragamreddy, 2024; Meniado et al., 2024). However, the pedagogical application of these tools is accompanied by ethical risks, notably concerns about overreliance on automation and threats to academic integrity, which may compromise the development of critical thinking and originality (Octaberlina et al., 2024).

Empirical studies on AI use in the Taiwanese EFL context, particularly with English majors, remain scarce. This research is therefore timely and aims to address this gap by systematically examining the impact of AI tools on a discipline-specific cohort. The study is conceptually grounded in four frameworks:

1. **The Technology Acceptance Model (TAM)** to interpret students' usage intentions;
2. **The Writing Process Approach** to situate AI within recursive composition stages;
3. **Cognitive Load Theory (CLT)** to understand how AI manages mental effort for higher-order thinking; and
4. **Critical AI Literacy (CAIL)** to ensure ethical and reflective use.

By integrating these perspectives, the study seeks to answer the following research questions:

RQ1: What are the effects of AI tools on students' academic English writing performance, and how did that change compare with a control group?

*(Hypothesis: A larger proportion of students will show improvement in their academic English writing performance after using AI writing tools.)**

RQ2: What are students' perceptions and attitudes toward the use of AI tools in academic writing?

*(Hypothesis: Although students will report predominantly positive attitudes toward AI writing tools, they will simultaneously raise critical concerns regarding overreliance, academic integrity, and the authenticity of their work.)**

2. Methodology

This study ad opted a mixed-methods, two-group pre-test/post-test design to investigate the effects of AI tools on academic writing performance and student perceptions. The participants were 18 sophomore English majors in

the experimental group and 21 in the control group at a university in southern Taiwan. All students were enrolled in the fourth required English Composition course, which focused on academic writing. All participants provided informed consent.

Data collection relied on two primary instruments:

1. Academic Writing Performance (Pre- and Post-tests): This involved collecting and scoring students' essays with the same instructor following the same syllabus, except for the use of AI. The topic for the experimental group was *"AI: Friend or Foe for English majors"*; the topic for the control group was *"Bilingual education: Friend or Foe for English majors"*. These topics were chosen in 2024 when AI was not yet ubiquitous. Draft 1 served as the pre-test, and the revision served as the post-test.

2. Student Learning Feedback: An open-ended written form was administered after the intervention to capture experimental group students' reflections on perceived benefits and challenges.

The study was conducted over an 18-week semester, integrated into the course syllabus:

Phase 1 (Weeks 1–10): Established baseline performance by collecting Draft 1 after instruction on fundamental writing skills.

Phase 2 (Weeks 11–15): Introduced the AI intervention in the experimental group, where students were explicitly guided to use ChatGPT for language refinement and idea elaboration, and Consensus for literature searching and synthesis.

Phase 3 (Weeks 16–18): Involved collecting the final Revision essays and the open-ended feedback forms (for the experimental group only).

Because the scores for the AI-supported group were paired and the sample size was small ($n = 18$), a Wilcoxon signed-rank test was conducted to examine whether Revision scores differed significantly from Draft 1 scores. Descriptive statistics (means, standard deviations, and gain scores) were also calculated to summarise the direction and magnitude of score changes. Given the historical and non-equivalent nature of the control cohort, between-group differences were interpreted cautiously. For RQ2, student reflections were analysed thematically following Braun and Clarke (2006).

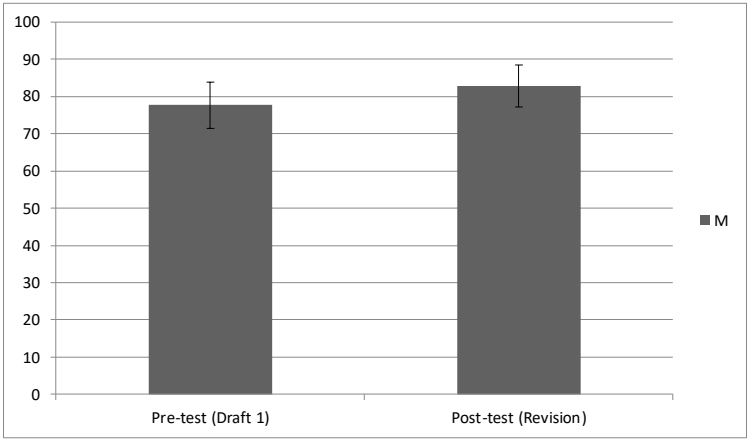
3. Results and Discussion

RQ1: What are the effects of AI tools on students' academic English writing performance, and how did that change compare with a control group?

For the AI-supported group, writing scores increased from Draft 1 (M = 77.67, SD = 6.25) to Revision (M = 82.83, SD = 5.64), yielding an average gain of 5.17 points. Fourteen of the 18 students (77.8%) received a higher score on the Revision than on Draft 1, while four students (22.2%) scored lower. A Wilcoxon signed-rank test indicated that Revision scores were higher than Draft 1 scores ($W = 32.0$, $Z = -2.33$, $p = .020$, $r = .55$). This result suggests a statistically reliable improvement in the AI-supported group, although the finding should still be interpreted in light of the study's small sample.

The control group also improved from Draft 1 to Revision. Its mean score rose from 72.81 to 82.29, yielding an average gain of 9.48 points; 15 of 21 students improved, one remained unchanged, and five scored lower. Descriptively, the control group's gain was larger than that of the AI-supported group. Because the groups were not randomly assigned, the prompts were not identical, and the data came from two different cohorts, this pattern should not be read as a causal refutation of AI. It does, however, show that AI-supported revision did not automatically produce greater short-term score gains than conventional revision in the present dataset.

(Figure 1: Mean Pre- and Post-test Academic Writing Scores for the AI-supported Group (n = 18). Note: Error bars represent standard deviations.)



RQ2: What are students' perceptions and attitudes toward the use of AI tools in academic writing?

The thematic analysis of student reflections identified two major themes concerning the use of AI tools: **(1) Facilitation of High-Order Writing and Enhanced Criticality** and **(2) Awareness of Human Agency and Academic Rigor**.

Under the first theme, students primarily reported that AI assisted them in structural and conceptual refinement, moving beyond mere surface correction. One student noted: *"I not only reorganised my structure but also added clearer topic sentences and smoother transitions. More importantly, I included research to back up my claims."* This finding supports the Writing Process Approach, demonstrating that students strategically integrated AI into the revision phase to improve coherence and scholarly backing. Furthermore, the perceived usefulness of AI in making ideas *"easier to follow"* and the writing *"more professional"* aligns strongly with the Technology Acceptance Model (TAM), indicating a high level of perceived usefulness.

The second theme revealed students' growing Critical AI Literacy (CAIL). Students explicitly acknowledged the necessity of human oversight, with one participant stating: *"All in all, even though AI helped a bit, it was the human effort that made the real difference."* Critically, the process led students to reflect on the nature of academic writing itself, realising it *"focuses on critical thinking, exploring different aspects, and supporting ideas with proper evidence,"* rather than just personal feelings. This theme confirms the hypothesis, showing positive acceptance alongside critical self-reflection on agency and ethical authorship. Students did, however, note lingering struggles with complex conventions like APA format and task interpretation, areas where AI needs to be guided more effectively by human critical input.

4. Implications

Pedagogically, AI should be integrated into writing courses as a guided revision aid rather than as a guaranteed route to better scores. Teachers may gain more by using AI to support brainstorming, organisation, and language polishing while requiring students to justify, revise, and evaluate AI suggestions. This approach aligns with CAIL because it keeps human agency central.

Second, AI effectively addresses the challenges highlighted by CLT. By managing the lower-level burden of linguistic accuracy, AI allows EFL students to concentrate on advanced rhetorical skills (argument structure, evidence integration). This finding carries global relevance for non-native English speaking (NNS) students and institutions in English as a Medium of Instruction (EMI) programmes. It suggests a transformative pedagogical shift: writing instruction should leverage AI to automate low-level editing, freeing class time for critical thinking, source evaluation, and deep disciplinary argumentation, reinforcing the Writing Process Approach and promoting equity for NNS students in high-stakes academic environments.

Several limitations should be acknowledged. First, the control group was historical rather than concurrent. Second, the two cohorts wrote on different prompts, and the AI-supported group revised the same topic across drafts; thus, repeated practice and task familiarity may also explain part of the score change. Third, the study relied on course-based scoring and did not report formal inter-rater reliability. Future research should therefore use matched prompts, concurrent control groups, and paired non-parametric tests based on complete raw scores.

Ethical Declaration Statement

The study followed the ethical guidelines for educational research at National Chiayi University. Participation was voluntary, informed consent was obtained from all student participants, and course grades were not affected by participation or non-participation. All data were anonymised prior to analysis, and no identifying information is disclosed in this manuscript.

Generative AI Policy Statement

The authors used ChatGPT, Gemini, and Consensus during the preparation of this work to assist with grammar correction, sentence refinement, initial literature review search, and organisation of ideas. 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

Aldabbus, S., & Almansouri, E. (2022). Academic writing difficulties encountered by university EFL learners. *British Journal of English Language Linguistics, 10*(3), 1-11. https://doi.org/10.37745/bjel.2013/vol10n3111

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Durrani, N., Imran, M., & Saeed, R. (2025). Evolution of artificial intelligence writing tools in domain of scientific writing: A threat to human writing and its ethical considerations. *Journal of Family Medicine and Primary Care, 14*(5), 1580-1583. https://doi.org/10.4103/jfmpc.jfmpc_1615_24

Eragamreddy, N. (2024). The Impact of AI in Creating Writing Skills in English Language Learners. *International Journal of Social Science Humanity & Management Research, 3*(12), 1573-1586. https://doi.org/10.58806/ijsshmr.2024.v3i12n05

Meniado, J. C., Huyen, D. T. T., Panyadilokpong, N., & Lertkomolwit, P. (2024). Using ChatGPT for second language writing: Experiences and perceptions of EFL learners in Thailand and Vietnam. *Computers and Education: Arti-*

ficial Intelligence, 7, 100313. https://doi.org/10.1016/j.caeai.2024.100313
Octaberlina, L. R., Muslimin, A. I., Chamidah, D., Surur, M., & Mustikawan, A. (2024). Exploring the impact of AI threats on originality and critical thinking in academic writing. *Edelweiss Applied Science and Technology, 8*(6), 8805-8814. https://doi.org/10.55214/25768484.v8i6.3878

Sequencing Peer and AI Collaboration in EFL Writing: Student Performance and Perceptions

Jiale YAO [1], Jiaqi QIN [1], Nuo CHEN [1], Ruiya GAO [1], Xuchuan (Winnie) CHEN [2]
[1] College of Science, National University of Defense Technology, China
[2] Department of Foreign Languages, Basic Education College, National University of Defense Technology, China

Abstract

This study investigates how different sequences of peer and AI-supported collaboration relate to EFL students' graph description performance and perceptions. Using a counterbalanced design, 116 undergraduate students completed two writing tasks across two modes: peer collaboration and AI-assisted collaboration. Writing performance was assessed via an analytic rubric, and perceptions were collected through open-ended surveys. Results from paired-samples t-tests revealed that peer collaboration yielded significantly higher scores in content and organisation than AI collaboration. Furthermore, a significant sequence effect emerged: the "peer-first" group significantly outperformed the "AI-first" group in the second task, suggesting that human negotiation provides a conceptual "warm-up" for subsequent AI use. Qualitative analysis revealed a divide between students prioritising AI's instrumental efficiency and those valuing the communicative depth of peer interaction. These findings suggest a "human-first" hybrid model for integrating GenAI into L2 writing curricula to ensure conceptual rigour and ethical engagement.

Keywords: Peer collaboration, AI collaboration, collaborative writing, task sequence, EFL writing

1. Introduction

Collaborative writing has long been shown to benefit L2 learners by fostering negotiation of meaning and co-construction of new knowledge (Storch, 2013). Recent advances in generative artificial intelligence (GenAI) have introduced a disruptive collaborative partner for L2 writers (Wiboolyasarin et al., 2024). Large language models (LLMs) can suggest ideas, offer alternative phrasings, and provide immediate feedback on language form and organisation (Salman & Chaya, 2024). As these tools enter higher education, established practices such as peer interaction are being re-examined: rather than relying solely on human partners, students can now consult GenAI during the writing process as an interactive, context-sensitive assistant (Zheng et al., 2024).

However, it remains unclear whether collaborating with an AI tool under guided conditions affords learning opportunities comparable to those provided by a human peer. In addition, students may move between human-human and human-AI collaboration across tasks. Yet, little is known about whether starting with peer collaboration and then working with AI, or vice versa, is differentially associated with subsequent writing performance. Understanding such sequence-related patterns is crucial for designing technology-enhanced curricula that integrate GenAI in a principled way.

In this study, peer collaboration is defined as two students jointly planning and producing a single text through face-to-face interaction, and AI collaboration as an individual student composing a text while interactively consulting a GenAI tool via text prompts. To ensure a meaningful comparison and address ethical concerns, AI was framed as a supportive partner for feedback rather than a tool for outsourcing text production. Crucially, this study utilises graph description tasks to further mitigate non-ethical AI use. Graph descriptions require strict adherence to task-specific visual data; this constraint necessitates that students critically evaluate and reconcile AI suggestions with factual evidence, thereby discouraging uncritical copying.

Despite the surge in GenAI research, relatively few studies have compared writing performance under peer versus AI collaboration, or examined how the sequence of these experiences relates to later performance. Less is known about how students evaluate the strengths and limitations of each mode after experiencing both. Against this background, the present study investigates EFL university students' graph description writing through a counterbalanced design to address the following research questions:

- **RQ1:** When the two collaboration modes are counterbalanced across two tasks, how do students' writing scores differ between peer collaboration and AI collaboration?
- **RQ2:** How is the sequence of experience (peer-first vs. AI-first) associated with performance in the second graph description task?
- **RQ3:** After experiencing both modes, how do students perceive the strengths and limitations of peer and AI collaboration in graph description writing?

By addressing these questions, the study aims to offer practical insights for integrating peer and AI collaboration into L2 writing courses and to clarify under what conditions GenAI can function as a pedagogically meaningful collaborative partner.

2. Methodology

2.1 Participants and Design

A total of 116 undergraduate students participated in the study and were randomly assigned to two groups. All participants had passed the CET-4, and there was no significant difference in initial English proficiency between the two groups. A counterbalanced experimental design was implemented over two weeks to investigate the effects of collaboration modes (Peer vs. AI) and their sequencing.

2.2 Procedures

The study followed a structured four-stage pedagogical intervention:

- **Week 1 (Instruction & Task 1):** Following a lecture on graph description and AI ethics, students completed Task 1. Group 1 (n=58) engaged in *Peer Collaboration* (pairs), while Group 2 (n=58) utilised *AI Collaboration* (GenAI as a consultant).
- **Week 2 (Feedback & Task 2):** After a collective feedback session addressing common errors and AI-sharing experiences, the groups switched modes. Group 1 performed *AI Collaboration*, and Group 2 performed *Peer Collaboration*.

2.3 Instruments and Ethical Controls

Two graph description tasks of equivalent difficulty were used. To ensure academic integrity, the teacher emphasised AI ethics and plagiarism policies. The graph description genre was specifically chosen as it requires data-to-text transformation based on specific visual input, thereby preventing unreflective verbatim copying from GenAI.

2.4 Data Analysis

Two experienced raters blind-graded all 232 scripts using a standardised rubric. The inter-rater reliability reached an acceptable level (Cronbach's $\alpha = .82$), and any significant discrepancies were resolved through discussion. The mean of the two raters' scores was adopted as the final grade for analysis. Paired-samples t-tests were conducted to compare performance between Peer and AI modes across the entire sample (RQ1). Independent-samples t-tests were used to analyse Task 2 scores to identify sequence effects (RQ2). A post-task survey collected data on students' perceptions of the strengths and limitations of each mode. These responses were analysed via thematic analysis.

3. Results and Discussion

3.1 Comparison of Collaborative Modes (RQ1)

A paired-samples t-test revealed that peer collaboration yielded significantly higher overall writing scores ($M = 14.28$, $SD = 1.82$) than AI collaboration ($M = 13.52$, $SD = 2.54$), $t(115) = -3.12$, $p = .002$. Dimensional analysis suggested that, although AI-assisted texts were comparable in lexical resource and grammatical accuracy, they were significantly weaker in content coverage and organisation than peer-negotiated texts. The greater variance in the AI condition indicates that students' ability to leverage generative tools remains inconsistent.

3.2 The Impact of Sequencing (RQ2)

An independent-samples t-test demonstrated a significant sequence effect. Students in the Peer-first group outperformed those in the AI-first group on Task 2, $t(114) = 2.45$, $p = .016$. Specifically, participants who engaged in peer negotiation before using AI achieved the highest performance (Peer-first Task 2: $M = 14.86$), suggesting that human collaboration provides a conceptual "warm-up" that enhances subsequent human–AI interaction. This pattern indicates that pedagogical sequencing should prioritise human-centred brainstorming before introducing algorithmic support.

3.3 Students' Perceptions (RQ3)

Among the 116 respondents, preferences were relatively balanced: 66 students (56.90%) indicated a preference for AI collaboration, while 50 (43.10%) favoured peer collaboration. Qualitative analysis of open-ended responses revealed that, although the proportions were similar, the underlying rationales differed. Students who preferred AI collaboration were primarily driven by instrumental considerations, emphasising efficiency, convenience, and reliability. They described AI as a fast, always-available "productivity tool" that can reduce effort, provide stable language support, and avoid interpersonal disagreement.

In contrast, students who favoured peer collaboration highlighted interactional and developmental values. They stressed idea exchange, co-construction of meaning, and rich dialogue, viewing human partners as irreplaceable for generating creative ideas, negotiating different perspectives, and experiencing emotional support and a sense of shared achievement.

The findings show that peer collaboration continues to offer advantages for graph description writing, particularly in content and organisation, even when students have access to powerful GenAI tools. The mode comparison

(RQ1) suggests that AI support alone does not automatically translate into better holistic performance, and that students differ considerably in their ability to use AI productively. The sequencing effect (RQ2) further indicates that beginning with peer collaboration can prime learners conceptually and rhetorically, leading to stronger subsequent performance than starting with AI. Students' perceptions (RQ3) reveal a tension between efficiency and learning-oriented interaction. While many value AI for its speed and linguistic support, others prioritise the deeper cognitive and affective benefits of human-to-human collaboration. For pedagogy, this points to the need for deliberate task design that integrates, rather than replaces, peer work with AI tools—for example, using peer discussion for idea generation and structuring, followed by guided AI use for language refinement under clear ethical guidelines.

4. Implications

This study demonstrates that while GenAI is a powerful tool for linguistic refinement, its effectiveness in L2 writing is significantly enhanced when preceded by human-centric peer collaboration. The primary conclusion is that pedagogical sequencing matters: a *human-then-AI* workflow fosters deeper conceptual engagement and superior writing performance compared to an AI-first or AI-only approach.

Theoretically, these findings extend sociocultural theory into the digital age by positioning GenAI not as a substitute for human interaction, but as a specialised scaffold that operates most effectively after a foundation of social negotiation has been established. Practically, the results offer a scalable framework for educators globally to integrate GenAI ethically and effectively. Instead of banning AI, institutions should design curricula that prioritise human-led brainstorming and critical thinking as a prerequisite for algorithmic assistance. In a global context where GenAI is rapidly reshaping higher education, this "human-first" hybrid model ensures that students remain active agents in the writing process, utilising technology to refine—rather than replace—their cognitive and communicative development.

Ethical Declaration Statement

The project has been approved by the Ethics Committee of The National University of Defense Technology.

Declaration of Use of AI in the Writing Process

The author used Gemini and ChatGPT during the preparation of this work to refine the linguistic clarity and academic tone of the manuscript. After utilis-

ing the tools, the author thoroughly reviewed and edited the content as necessary and assumed full responsibility for the publication's content.

References

Salman, S., & Chaya, R. (2024). The influence of AI-powered learning platforms on student engagement and performance: Emerging technologies in education. *International Journal of Research Publication and Reviews, 5*(7), 1816–1824. https://doi.org/10.55248/gengpi.5.0724.1750

Storch, N. (2013). *Collaborative writing in L2 classrooms*. Multilingual Matters.

Wiboolyasarin, W., Wiboolyasarin, K., Suwanwihok, K., Jinowat, N., & Muenjanchoey, R. (2024). Synergizing collaborative writing and AI feedback: An investigation into enhancing L2 writing proficiency in wiki-based environments. *Computers and Education: Artificial Intelligence, 6*, 100228. https://doi.org/10.1016/j.caeai.2024.100228

Zheng, L., Fan, Y., Chen, B., Huang, Z., LeiGao, & Long, M. (2024). An AI-enabled feedback-feedforward approach to promoting online collaborative learning. *Education and Information Technologies, 29*(9), 11385–11406. https://doi.org/10.1007/s10639-023-12292-5

Generative Artificial Intelligence (GenAI) in Higher Education: Human-AI Collaborative Learning in Hong Kong

CHENG Shuk Ling [1]
[1] Faculty of Humanities, The Hong Kong Polytechnic University, Hong Kong

Abstract

This paper reports on a university-funded teaching development project conducted to integrate collaborative learning among learners, teachers, and Generative Artificial Intelligence (GenAI) to enhance the effectiveness of student learning in higher education in Hong Kong. Students first acquired theoretical knowledge of business leadership in traditional lecture settings; they then applied these concepts by proposing promotional projects for local companies, developing role-play scripts, and consulting GenAI for formative feedback. The process involved learner-AI collaboration for feedback-giving and idea improvement through iterative cycles. Findings highlight the effectiveness of human-AI collaboration in learning, deepening understanding, and improving leadership behaviours in scenario-based contexts. The study demonstrates best practices for leveraging GenAI in collaborative learning, offering transferrable insights for higher education across disciplines. It promotes enhanced student engagement and personalised learning through human-AI interaction.

Keywords: Generative Artificial Intelligence (GenAI) in higher education; collaborative learning; human-AI interaction; business management education

1. Introduction

Recent advances in Generative Artificial Intelligence (GenAI) since 2022, particularly generative models such as ChatGPT, have presented new opportunities to enhance collaborative learning and collective idea improvement in higher education. Academic discussions within the educational sphere concerning the usage of ChatGPT in higher education range from a focus on student-ChatGPT interaction in collaborative learning (Kim et al., 2024) and student identity (Darvin, 2025), to concerns over the ethical use of GenAI and plagiarism (Powers et al., 2025), as well as teaching and learning practices in Science, Technology, Engineering, and Management (STEM) education (Kumar et al., 2024). Whilst the application of human-AI interaction in educational settings is still emerging, scholars have begun to recognise its significant potential to redesign teaching and learning processes (Darvin, 2025). Nam and Bai

(2023) revealed that GenAI can act as a support for learners, providing valuable insights in academic settings. In particular, Kumar et al. (2024) investigated the transformative potential of ChatGPT in enhancing learning experiences and contributing value to business management higher education.

In the contemporary landscape of business management higher education, the cultivation of leadership skills has become a critical graduate attribute for business schools, highly valued by employers across industries. Leadership is not merely the acquisition of theoretical knowledge; it is the demonstration of effective behaviours and decision-making in real-world contexts. Despite its importance, bridging the gap between theory and practice remains a persistent challenge for both educators and learners. Traditional pedagogical approaches, such as lectures and case studies, often fall short in fostering the active application and exhibition of leadership skills, particularly in dynamic business environments. The emerging trend of using GenAI in higher education contexts serves as a catalyst for pedagogical researchers to rethink and refine current pedagogies and to redesign teaching practices strategically.

In light of this, a teaching development project titled *"Learning Experience Enhancement: Teaching Millennials Ethical Leadership through Role-Play Simulation with Creative Media Techniques"* was proposed and supported by the Teaching Development Grants (TDGs) funding at City University of Hong Kong (Project Number: 6000900). This project was implemented in a major elective course for Bachelor of Business Administration (BBA) students at City University of Hong Kong. Teaching and learning activities (TLAs), assessment tasks (ATs), and course intended learning outcomes (CILOs) were strategically aligned to foster learner-AI collaboration, teacher-AI collaboration, and formative feedback cycles. Students were coached in teams of four to enact managerial perspectives, propose promotional projects for local companies, and apply contingent leadership styles in realistic business scenarios. They were tasked with writing a role-play script and producing a video output based on the script.

In terms of GenAI, ChatGPT (powered by CityU) was adopted to function as a coach offering feedback to students. This paper aims to share best practices under the principle of collaborative learning in business management education, emphasising collective idea improvement among three key parties: learners, teachers, and ChatGPT. The teaching development project can be perceived as a pilot study demonstrating how human-AI collaboration can be leveraged to facilitate collective idea improvement in business management education. It identifies best practices for aligning CILOs through learner-AI and teacher-AI collaboration and explores the implications of this approach for enhancing stu-

dent agency, engagement, and the practical application of leadership skills. By disseminating findings and observations from this action research project, the paper also provides transferrable insights for instructors and course designers across different majors and institutions.

2. Methodology

This study adopted an action research methodology, integrating both qualitative and quantitative approaches to evaluate the effectiveness of AI-enabled collaborative learning in leadership education. The intervention was designed to align TLAs, ATs, and CILOs, with a focus on collective idea improvement through iterative feedback and collaboration among learners, teachers, and GenAI.

The participants were approximately seventy undergraduate students from different years of study and majors (e.g., business, social sciences, engineering, science, law, arts, etc.) enrolled in a "Leadership" major elective course. Students were organised into small groups, each tasked with developing a business promotional proposal for a local company in Hong Kong. The diversity of the cohort in terms of academic backgrounds and year of study provided a rich context for collaborative learning.

Teaching and learning activities were structured to promote active engagement, scenario-based learning, and iterative improvement. Students first attended lectures and seminars on leadership theories, including transformational, situational, and servant leadership, with emphasis placed on understanding how these theories apply in business contexts. Each group then selected a local company and developed a scenario for a promotional project, considering relevant business operational and environmental factors from a managerial perspective. Groups wrote role-play scripts that explicitly applied leadership concepts to their chosen scenario; the scripts were designed to persuade a board of directors to support and finance the promotional project.

Learner-AI collaboration was a central feature of the intervention. Students engaged with ChatGPT to propose ideas, receive formative feedback, and refine their scripts. ChatGPT acted as a coach, prompting students to reflect on their leadership choices and strategies. Teacher-AI collaboration was also integral, as teachers used ChatGPT to summarise and assess students' ideas, providing additional formative feedback and ensuring alignment with CILOs.

Assessment tasks were designed to evaluate both the process and the outcomes of collaborative learning. Each group submitted a role-play script

demonstrating the application of leadership theories. Students received iterative feedback from ChatGPT and teachers, with opportunities to revise and improve their work. The intervention was aligned with two key CILOs: demonstrating understanding of leadership theories and their application in business contexts, and exhibiting leadership behaviours and decision-making skills in realistic scenarios. Lastly, students were required to produce a video based on the scripts they had written, either in the form of animation or by video recording.

3. Results and Discussion

The integration of ChatGPT as a coach was found to significantly enhance collective idea improvement among students. A large majority of students reported in their individual self-reflection reports that engaging with ChatGPT helped them rethink and refine their assignments. The iterative feedback process encouraged students to refine their scripts, resulting in more persuasive and well-structured role-play outputs. Formative feedback cycles, enabled by both learner-AI and teacher-AI collaboration, were instrumental in driving collective idea improvement. Irons and Elkington (2022) noted that formative feedback "plays an integral role in higher education" and fosters students' continual improvement. Students received feedback before refining their scripts; this process fostered a culture of collaboration, with students actively seeking input from teachers and GenAI.

Analysis of the role-play scripts revealed that the majority of groups explicitly applied leadership theories in their scenarios, demonstrating a thorough understanding of concepts such as transformational leadership. The intervention incentivised student agency by encouraging learners to take ownership of their projects, experiment with different leadership styles, and articulate their reasoning. The integration of GenAI feedback with teacher evaluation ensured that proposals were both theoretically sound and practically relevant.

Students were required to submit individual written reports reflecting on the overall learning experience, including the briefing process conducted by the instructor, the script-writing process, consultation with AI, and the video production process. Students provided positive feedback regarding the effectiveness of adopting script-writing and role-play ideas for learning leadership in business:

> *"We had sessions on leadership and communication skills, but* ***the real magic happened when we started collaborating as a team****. We improve a lot as a team and individually because of constructive feedback."*

"By writing the script and producing a role-play animated video of these two scenarios, ***I was able to transform ambiguous ideas into clear ways to understand leadership in business easily.****"*

"My experience with ***this project motivates me to become a transformational leader*** *as well as build constructive workplace environments regardless of my future roles."*

"Through the real-world application of theories, ***I gained practical knowledge that will shape my leadership skills in my upcoming career.****"*

*"****My professional and academic path*** *forward shows greater assurance when dealing with ethical challenges because I understand that my decisions will affect my team members as well as myself."*

"This project not only educated me about leadership, but also helped me realise their ***practical implication*** *with empathy, fairness, and accountability."*

On the other hand, students reported difficulties encountered during this learning experience, including steep learning curves due to a lack of prior experience in using animation software for video production, and the difficulty of writing an effective script that captures the application of theories within a limited video timeframe. Based on student feedback, the Principal Investigator (PI) and Co-Investigator (Co-I) would revise the design of the learning experience, ensuring students receive sufficient training on the use of animation software and acquire intermediate script-writing skills prior to the learning experience.

"Our team ***members faced steep learning curves because animation software tools were new to us****. The obstacles we encountered forced us to collaborate as a group, which enriched our understanding of leadership while building our technical capabilities to improve communication effectiveness."*

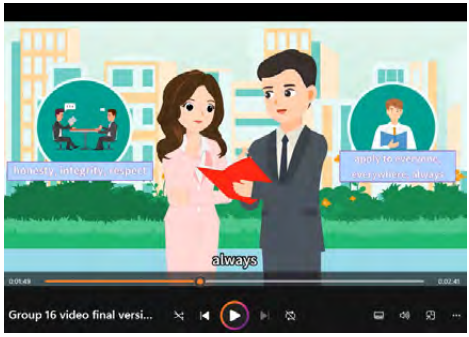
"At first, we found it ***difficult to precisely apply leadership concepts into our scripts within the time limit of the video****. Then we worked on searching lecture notes and online resources to find ideas for the scripts."*

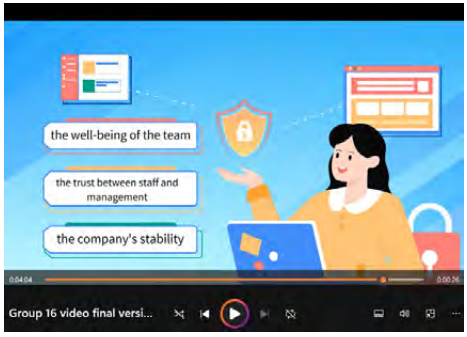
Overall, the study demonstrates that AI-enabled collaborative learning can effectively bridge the gap between leadership theory and practice. By fostering collective idea improvement among learners, teachers, and GenAI, the intervention promoted deeper understanding, active application, and the development of practical leadership skills.

4. Pedagogical Implications

The integration of GenAI into collaborative learning environments facilitated collective idea improvement, enhanced student agency, and promoted the practical application of leadership skills. The findings contribute to the theory of collaborative learning by illustrating how human-AI partnerships can enrich the learning process. The study supports the view that learning is a social and iterative endeavour, enhanced by diverse sources of feedback and collective idea generation.

The role of GenAI as a coach aligns with constructionist principles, in particular a participatory approach that enabled students to actively construct knowledge through dialogue, reflection, and experimentation. The Situated Learning Model (SLM), also known as participatory learning theory, suggests that the nature of knowledge is context-dependent and situational, and the need to learn is driven by the context or environment. Knowledge could be acquired through a collaborative approach. This echoes the design of the role-play script project, where management students collaborated under a given context (i.e., developing a scenario for a promotional project to persuade a board of directors to support and finance the promotional project).





On an institutional level, this experimental study offers practical recommendations for implementing AI-enabled collaborative learning. It is suggested that Higher Education Institutions (HEIs), and in particular course designers, should begin to equip students with skills to communicate effectively with GenAI (i.e., AI literacy) and to conduct learning activities that incorporate iterative

feedback from both GenAI and human sources.

This AI-assisted video production learning experience can be perceived as a blueprint that can be modified and adapted to other relevant academic disciplines, such as public management and social work, for learning enhancement purposes. This teaching development project highlights the effectiveness of human-AI collaboration in learning, deepening understanding, and improving leadership behaviours in scenario-based contexts. It demonstrates best practices for leveraging GenAI in collaborative learning, hence offering transferrable insights for higher education across disciplines, promoting enhanced student engagement and personalised learning through human-AI interaction.

5. Conclusion

In conclusion, AI-enabled collaborative learning represents a promising approach to enhancing leadership education in higher education. By fostering collective idea improvement among learners, teachers, and GenAI, the intervention bridged the gap between theory and practice, promoted student agency, and facilitated the development of practical leadership skills. The principles and practices identified in this study can be generalised and are transferrable to other academic disciplines. Future research could explore the use of GenAI in various academic disciplines such as STEM education, and the feasibility of combining GenAI in higher education with Task-Based Learning (TBL). As GenAI technologies continue to evolve, their integration into collaborative learning environments will be essential for preparing graduates to lead and innovate in complex settings.

Ethical Declaration Statement

The project *"Learning Experience Enhancement: Teaching Millennials Ethical Leadership through Role-Play Simulation with Creative Media Techniques"* was supported by the Teaching Development Grants (TDGs) funding at City University of Hong Kong (Project Number: 6000900). This project has received ethical approval by the Human and Artefacts Ethics Sub-Committee, Research Grants and Contracts Office, City University of Hong Kong (Reference Number: HU-STA-00001260).

Generative AI Policy

The author used PolyU GenAI during the preparation of this work to improve readability and language. After utilising the tool, the author thoroughly reviewed and edited the content as necessary and assumed full responsibility for the publication's content.

References

Darvin, R. (2025). Identity and investment in the age of Generative AI. *Annual Review of Applied Linguistics, 45*, 10–27. https://doi.org/10.1017/s0267190525100135

Irons, A., and Elkington, S. (2022). *Enhancing learning through formative assessment and feedback*. Routledge.

Kim, H. K., Roknaldin, A., Nayak, S., Zhang, X., Twyman, M., Hwang, A. H.-C., and Lu, S. (n.d.). Empowering computer-supported collaborative learning with chatgpt: Investigating effects on student interactions. *2024 ASEE PSW Conference Proceedings*. https://doi.org/10.18260/1-2--46032

Kumar, S., Rao, P., Singhania, S., Verma, S., and Kheterpal, M. (2024). Will artificial intelligence drive the advancements in higher education? A tri-phased exploration. *Technological Forecasting and Social Change, 201*, 123258. https://doi.org/10.1016/j.techfore.2024.123258

Nam, B. H., and Bai, Q. (2023). CHATGPT and its ethical implications for STEM Research and Higher Education: A media discourse analysis. *International Journal of STEM Education, 10*(1). https://doi.org/10.1186/s40594-023-00452-5

Powers, E., Arif, R., Fedner, M., Conahan, E., and Ezechukwu, A. (2025). "A starting point and a crutch": Student and faculty perceptions of generative AI. *Teaching Journalism & Mass Communication, 15*(1), 2-12. Retrieved from https://lbapp01.lib.cityu.edu.hk/ezlogin/index.aspx?url=https://www.proquest.com/scholarly-journals/starting-point-crutch-student-faculty-perceptions/docview/3251471494/se-2

Integration of GenAI-Simulated Patient Pedagogy in Tutorials to Facilitate First-Year Occupational Therapy Students' Learning

Chi-Wen CHIEN [1]
[1] Department of Rehabilitation Sciences, The Hong Kong Polytechnic University, Hong Kong, China

Abstract

Generative artificial intelligence (GenAI), utilised as simulated patients, represents an emerging pedagogy for enhancing clinical skills in health professions education. This approach was integrated into the author's tutorials to facilitate learning among first-year occupational therapy students and to explore their engagement and satisfaction. A stroke case with minimal background information and three treatment notes was created and input into ChatGPT-4o. Students interacted with the GenAI-simulated patient to gather information, conduct assessments, and explain interventions sequentially across three tutorials. Qualitative analysis of the chat records demonstrated signs of student learning through these interactions. Learning gains included real-time adaptations of communication strategies and occupation-based reasoning. Of the 94 students, 30 indicated that the tutorials strongly facilitated their active learning (mean = 4.6). These findings suggest that integrating a GenAI-simulated case into tutorials is a promoting strategy for rehabilitation education, even for students with limited clinical experience.

Keywords: Occupational Therapy, Generative Artificial Intelligence, Simulated Patients

1. Introduction

Generative artificial intelligence (GenAI) is an emerging technology capable of creating new content and has recently been used as simulated patients to support students' learning in history-taking, interviewing, and clinical reasoning within health professions education (Pham et al., 2025). However, most such applications to date are concentrated in medical and nursing fields, highlighting the need to explore their relevance to other health professions, such as occupational therapy (OT).

OT is a rehabilitation profession grounded in client-centred practice, using everyday activities therapeutically to enable individuals to participate in mean-

ingful occupations. To design tailored interventions, occupational therapists must interview clients to gather their occupational history, values, patterns, and interests, as well as conduct assessments and explain treatment options. These tasks demand strong communication skills and clinical reasoning, typically learnt through small-group tutorials and clinical placements during training.

Teaching these communication and clinical skills is challenging, particularly for first-year OT students who have limited clinical exposure and professional knowledge. Existing research has demonstrated the potential of using GenAI in OT education. For example, Mansour and Wong (2024) found that GenAI could facilitate last-year students' generation of creative intervention ideas in case studies during fieldwork seminars. While their findings are promising, the usage of GenAI can be extended beyond paper-based case brainstorming to simulated patient interactions that enable students (especially novices) to practise interviewing, assessment reasoning, and intervention planning in a safe environment. Therefore, the author explored how GenAI-simulated patient pedagogy could be integrated to facilitate first-year OT students' learning of the fundamental therapy process during small-group tutorials. The two objectives were: (1) to explore students' learning from interacting with a GenAI-simulated patient; and (2) to investigate students' satisfaction with this approach.

2. Methodology

2.1 Implementation Design

The author adopted a three-tutorial learning design to implement the use of a GenAI-simulated patient in a fundamental subject on OT theory and process at a university in Hong Kong. These tutorials were regular learning activities embedded within the subject curriculum.

2.2 Participants and Educational Context

Participants were 94 first-year Bachelor of Science in Occupational Therapy students enrolled in the second semester of the 2024–2025 academic year. At this stage of their study, students had minimal to no clinical experience, and most had never interviewed a patient or conducted standardised assessments.

These students were divided into four tutorial groups of 24–25 students each. Every group attended six tutorials in total; three of which incorporated the GenAI-simulated patient pedagogy. These tutorials were aligned with corresponding lectures on the concepts of occupation, assessment, and intervention, delivered at Week 2, Week 9, and Week 11, respectively. They were intended to strengthen students' ability to apply knowledge and skills introduced in lectures.

2.3 Development of the GenAI-Simulated Patient

A single stroke case, "Mrs Felicia Scott," was developed to serve as the basis for all three tutorials. This case described a 43-year-old woman who suffered a right-sided cerebrovascular accident two weeks earlier and was referred to rehabilitation services due to impaired motor control on the left side of her body. The author created a comprehensive yet concise case package, including:

(a) a 1-page sheet introducing the case background;
(b) a 2-page time-use profile describing her pre-stroke daily routines;
(c) a 2-page initial note summarising her neuromuscular, cognitive-perceptual-motor, psychological, and self-care status;
(d) summaries of her work history (98 words), hobbies and interests (52 words), and current performance across different services (253 words);
(e) a 3-page document outlining what the client can and cannot do midway through rehabilitation; and
(f) a 1-page 2-week progress note.

These materials were input into the ChatGPT-4o model using carefully crafted prompts for different tutorials (i.e., the first four materials for Tutorial 1 and the fifth and sixth materials for Tutorials 2 and 3, respectively). The prompts instructed the model to respond solely based on provided case information and, when asked questions beyond its scope, to either generate reasonable make-up responses or indicate that no ideas were available. This GenAI-simulated patient was designed to provide a realistic but safe learning environment where the AI acted in role as the client and responded dynamically to students.

2.4 Tutorial Contents

Table 1 shows the details of each tutorial. In Tutorial 1, students in each group first brainstormed potential interview questions. Three students were randomly selected by lucky draw to conduct the interview and pose these questions to the GenAI-simulated patient to obtain her responses. In Tutorial 2, another set of three randomly selected students verbally administered an assessment to the client using a tool they had identified. The GenAI-simulated patient described her performance (or inability to perform) according to the documented impairments. Subsequently, students communicated a proposed treatment goal to the client. If an unrealistic goal was proposed, the GenAI-simulated patient was prompted to gently decline and redirect students toward a more attainable goal.

In Tutorial 3, three randomly selected students articulated their selected treatment activities and explained the intervention processes, allowing the Ge-

nAI-simulated patient to respond with cooperation or questions. In the latter two tutorials, students interacted with the client using the same chat channel as previous sessions, with one group having its own dedicated channel. Alongside the GenAI-simulated patient, each tutorial included additional in-class activities that introduced further skills. Homework in the first two tutorials required students to identify assessments and interventions for reporting in later sessions.

Table 1: Structure and Details of the Three Tutorials

	Session Purpose	Structure (Time Spent)
Tutorial 1	Interviewing the client	• Guided discussion of trigger information (20 minutes) • Completion of a time-use profile worksheet (30 minutes) • Group-based brainstorming of interview questions (20 minutes) • Live interview with the GenAI-simulated patient (30 minutes) • Homework explanation (10 minutes)
Tutorial 2	Assessment and goal setting	• Teacher's reporting of one assessment and demonstration of assessing the GenAI-simulated patient (20 minutes) • Students' reporting of one assessment and one proposed goal, followed by conducting the assessment with the GenAI-simulated patient (30 minutes) • Participation in instructor-led debriefing activities on client progress (50 minutes)
Tutorial 3	Intervention planning	• Homework explanation (10 minutes) • Students' reporting of treatment activities, followed by explaining these activities to the GenAI-simulated patient (40 minutes) • Teacher's debriefing and demonstration of possible treatment activities (10 minutes) • Introduction of a self-management programme, followed by in-class activities (60 minutes)

2.5 Data Collection

All group interactions with the GenAI-simulated patient across three tutorials were automatically recorded in the author's ChatGPT account. Additionally, one item was added to the subject's electronic Student Feedback Questionnaire (eSFQ), asking students to rate how much the GenAI-simulated patient activity facilitated their active learning on a 5-point scale. Students also

provided written comments about which learning activity within this subject they liked most. The response rate of the eSFQ was 31.9% (30/94).

2.6 Data Analysis

Content analysis of the chat transcripts from all tutorial groups was conducted to identify themes related to students' learning. Descriptive analysis was used to summarise attendance rates and eSFQ rating scores, providing an overview of students' engagement and satisfaction with the GenAI-simulated patient pedagogy.

3. Results and Discussion

3.1 Themes and Examples Identified from Chat Interaction

One theme, *Learning from the GenAI's responses*, was identified from the chat transcripts of the four tutorial groups. For example, students demonstrated a learning pattern when interacting with the GenAI-simulated patient in Tutorial 1. They began by asking emotionally-oriented questions (e.g., "Are there any emotions or thoughts that stand out to you?") and gradually progressed to explore the client's occupational identity and environments (e.g., "What do you like to do in your leisure time?").

In Tutorial 3, one student provided a step-by-step explanation of constraint-induced movement therapy (i.e., restricting use of the unaffected hand to promote function of the affected hand). In response, the GenAI-simulated patient expressed concern about her current difficulty using the affected hand and suggested that the treatment might be too challenging at this stage. This feedback prompted the student to reconsider the intervention and explore more graded, achievable treatment activities. The examples indicate that students began to adopt more structured communication strategies and occupation-based reasoning. Although these transformative skills were still emerging among first-year students, developing them aligns with core OT competencies outlined in documents from major professional associations (American Occupational Therapy Association, 2015).

3.2 Students' Satisfaction

Among the 30 students who responded to the eSFQ, the mean perceived active-learning rating for the GenAI-simulated patient activity was 4.6 out of 5.0 (standard deviation = 0.5). More than half of these respondents (n=16) identified the interactive tutorials as the activity they liked most among all activities, excluding clinical visits. For example, one student commented: "I like how we can talk to Mrs Scott 'in-person' using ChatGPT. It was very creative and I

love it very much." These responses indicate that students valued the safe, low-stakes practice opportunity offered by this innovative activity, highlighting the motivational benefits of learning.

3.3 General Discussion of the Findings

Overall, integrating the GenAI-simulated patient across multiple tutorials facilitated students' reflective and self-directed learning. This was evident from students' real-time adaptations to the AI's feedback and completion of homework tasks. The GenAI also provided non-threatening responses when students showed communication insensitivity or proposed unrealistic intervention suggestions, providing a safe environment to refine their skills without fear of harming a real client or feeling embarrassed in front of peers. The finding aligns with simulation literature advocating psychological safety as a cornerstone of experiential learning (Madsgaard & Svellingen, 2025). Additionally, once developed, the GenAI-simulated patient could be reused across all tutorial groups, enhancing the sustainability and scalability of the approach in OT education.

3.4 Limitations

No formal evaluation surveys were conducted to examine students' learning gains. The implementation of this approach also required substantial logistical preparation. In addition, students interacted with the GenAI-simulated patients only verbally and were unable to observe nonverbal communication cues or conduct physical assessments, which may have restricted opportunities for full skill practice.

4. Implications

The findings have important implications for OT education. GenAI-simulated patients may provide first-year students with a structured and low-risk opportunity to practise interviewing, explaining assessments, and developing early occupation-based clinical reasoning before real clinical exposure. This approach may help bridge the gap between classroom learning and practice, particularly for students with limited clinical experience. The case design and prompts developed in this study may also be adapted for use across subjects and in senior-year levels to support progressive development of communication and reasoning skills. In a global context, the GenAI-simulated patient approach may be especially relevant to other health professions education programmes seeking scalable, flexible, and cost-effective teaching strategies. As the materials can be tailored to different cultural, linguistic, and professional contexts, the teaching approach may be transferable across countries. Future work may explore the incorporation of physical agents and evaluate the impact on students' learning outcomes.

Acknowledgements

None.

Ethical Declaration Statement

This study did not require formal ethical review because it analysed de-identified data derived from normal classroom activities and routine course evaluations (including engagement and eSFQ). No identifiable personal information was collected.

Generative AI Policy

The author(s) used ChatGPT-5.1 during the preparation of this work to improve the readability and language of the manuscript. 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

American Occupational Therapy Association. (2015). *Standards of practice for occupational therapy*. American Journal of Occupational Therapy, 69(Suppl. 3), 6913410057. https://doi.org/10.5014/ajot.2015.696S06

Madsgaard, A., & Svellingen, A. (2025). The benefits and boundaries of psychological safety in simulation-based education: An integrative review. *BMC Nursing*, 24(1), 922. https://doi.org/10.1186/s12912-025-03575-y

Mansour, T., & Wong, J. (2024). Enhancing fieldwork readiness in occupational therapy students with generative AI. *Frontiers in Medicine*, 11, Article 1485325. https://doi.org/10.3389/fmed.2024.1485325

Pham, T. D., Karunaratne, N., Exintaris, B., Liu, D., Lay, T., Yuriev, E., & Lam, A. (2025). The impact of generative AI on health professional education: A systematic review in the context of student learning. *Medical Education*, 59(12), 1280–1289. https://doi.org/10.1111/medu.15746

Promoting Generative AI in Learning, Teaching, and Assessment: A Case Study in Human Anatomy Education

Chileka CHIYANIKA[1]
[1] Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

Abstract

This case study examined the impact of integrating Generative Artificial Intelligence (GenAI), specifically ChatGPT, into human anatomy education, comparing the academic performance of two cohorts from different academic years. Cohort 1 () studied using traditional methods, while all students in Cohort 2 () engaged with GenAI tools during collaborative learning sessions. Survey responses from Cohort 2 () indicated that 69.8% rated their GenAI experience positively, and 60.6% reported an enhanced understanding of anatomical concepts. Academic performance analysis revealed that Cohort 2, which utilised GenAI universally, achieved a higher mean score (78.73% vs 77.31%) and a greater proportion of top grades (27.6% vs 16.5%) compared to Cohort 1. While students valued GenAI for its ease of use and interactivity, they expressed concerns regarding information accuracy and highlighted the need for further improvements. The study underscores GenAI's potential as a valuable adjunct to traditional teaching, supporting digital literacy and pedagogical innovation in anatomy education.

Keywords: Generative Artificial Intelligence, Human Anatomy Education, ChatGPT, Collaborative learning, Digital literacy

Introduction

Generative Artificial Intelligence (GenAI) tools, such as ChatGPT, are reshaping higher education by granting students unprecedented access to information and enabling more interactive, personalised, and self-directed learning experiences. In response to this technological evolution, universities are increasingly embracing GenAI, prompting academic departments to rethink and update their approaches to Learning, Teaching, and Assessment (LTA).

Human anatomy is a cornerstone of medical and health sciences education, traditionally taught through cadaveric dissection, textbooks, and physical models. While effective, these methods are often limited by issues such as ac-

cessibility, cost, and the static nature of learning materials (Fahrni & Sabatasso, 2025). GenAI offers new opportunities to enhance anatomy education through interactive, adaptive, and scalable digital platforms, potentially overcoming many limitations of conventional approaches.

In line with our university's digital innovation strategy, ChatGPT was piloted to supplement traditional teaching and promote collaborative, critical engagement with technology. Given the limited evidence on GenAI's effectiveness in anatomy education, this study evaluated ChatGPT integration, assessed student experiences and perceptions, and compared academic performance between a traditional cohort and one with GenAI integration.

Methodology

Study Design and Setting

This study was conducted during the autumn semester of the 2024/25 academic year with ethics approval (HSEARS20250409003), targeting all 121 first-year undergraduate radiography students at our university.

Participants

Two cohorts were compared: Cohort 1 (2023/24) utilised traditional methods, while Cohort 2 (2024/25) participated in collaborative learning sessions enhanced with Generative AI (GenAI), specifically ChatGPT. All students were included in the academic performance analysis. Survey participation was voluntary and anonymous; responses were received from 43 students within Cohort 2.

Traditional Learning Methods

Cohort 1 followed traditional learning approaches, including structured lectures and seven collaborative sessions over 13 weeks, focused on revising key content. Activities such as real-life case scenarios, mini-presentations, and manikin practice reinforced theory and application. The remaining six weeks were reserved for optional individual consultations with the tutor, either in person or online.

Integrating ChatGPT into Collaborative Anatomy Learning

In Cohort 2, selected collaborative learning sessions (four out of seven in total) were supplemented with clinically relevant case scenarios designed by teaching staff to reflect real-world challenges. Students worked in small groups to analyse these scenarios, utilising ChatGPT to research and generate responses. Following AI-assisted exploration, students engaged in group discussions to critically evaluate the accuracy and relevance of AI-generated content, cross-referencing it with textbooks and trusted sources. Tutors facilitated these

sessions, providing expert feedback and clarifying misconceptions to reinforce theoretical understanding and promote digital literacy.

Survey Instrument

At the end of the semester, Cohort 2 students completed a structured questionnaire assessing their GenAI experience in anatomy sessions. The ten-item instrument used a 1–10 Likert scale to evaluate overall experience, ease of use, perceived enhancement of understanding, helpfulness, challenges, frequency of use, accuracy, comprehensiveness, likelihood of recommending GenAI, comparison to traditional methods, and suggestions for improvement.

Student Performance Assessment

Academic outcomes were compared between cohorts using final grades for the subject. Grade distributions and mean scores were extracted from institutional records, reflecting either traditional teaching (Cohort 1) or GenAI-supported learning (Cohort 2).

Data Management and Confidentiality

All data were managed in accordance with institutional ethics and data protection guidelines. Survey responses were anonymous, securely stored, and stripped of personal identifiers.

Data Analysis

Descriptive statistics summarised survey and academic data. Grade distributions and mean scores were compared using independent samples -tests or Mann-Whitney tests, as appropriate. All tests were two-sided, with considered significant (SPSS version 28).

Results

A total of 43 students completed the survey. GenAI (ChatGPT) received positive feedback, with 69.8% rating it 7 or higher (the most frequent rating was 8 at 30.2%). Ease of navigation was also rated highly, with 69% giving a score of 7 or above. Regarding learning impact, 60.6% felt GenAI enhanced their understanding of human anatomy, and 60.5% found its features helpful. Few students reported significant challenges; only 30.2% noted minimal difficulties. Usage was moderate (67.5% rated it 5 or higher), and 48.8% rated its accuracy and comprehensiveness at 7 or higher. The likelihood of recommending GenAI was moderate (58.2% rated it 7 or above), and 55.8% considered it better than traditional methods. However, 78.7% expressed a desire for further improvements.

Baseline academic abilities were similar between cohorts, as shown by comparable university admission scores (Cohort 1: ; Cohort 2: ;). At the end of the semester, Cohort 2 (GenAI users) had a slightly higher mean anatomy score () than Cohort 1 (), though this difference was not statistically significant (). Notably, Cohort 2 achieved significantly more top grades (A+ and A: 27.6% vs 16.5%;). With similar baseline abilities, these results suggest that GenAI may have helped improve top-level performance in human anatomy. See Tables 1 and 2 for details.

Table 1. Survey results regarding student perceptions of Generative AI (ChatGPT) use in human anatomy education.

Survey Question	Likert scale responses																			
	1		2		3		4		5		6		7		8		9		10	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
Overall Experience	0	0	0	0	1	2.3	0	0	7	16.3	5	11.6	11	25.6	13	30.2	4	9.3	2	4.7
Ease of Use	0	0	0	0	0	0	1	2.4	9	21.4	3	7.1	8	19.0	9	21.4	5	11.9	7	16.7
Enhanced Understanding	0	0	1	2.3	0	0	2	4.7	8	18.6	6	14.0	6	14.0	11	25.6	6	14.0	3	7.0
Helpful Features	0	0	0	0	2	4.7	3	7.0	6	14.0	6	14.0	8	18.6	10	23.3	5	11.6	3	7.0
Challenges Encountered	0	0	4	9.3	8	18.6	1	2.3	14	32.6	4	9.3	5	11.6	4	9.3	2	4.7	1	2.3
Frequency of Use	0	0	2	4.7	7	16.3	5	11.6	6	14.0	9	20.9	7	16.3	4	9.3	1	2.3	2	4.7
Accuracy/ Comprehensiveness	0	0	0	0	4	9.3	2	4.7	8	18.6	8	18.6	10	23.3	9	20.9	1	2.3	1	2.3
Recommend to Others	1	2.3	0	0	3	7.0	2	4.7	7	16.3	5	11.6	7	16.3	10	23.3	3	7.0	5	11.6
Use of GenAI compared to Traditional Methods	0	0	0	0	2	4.7	2	4.7	9	20.9	6	14.0	8	18.6	8	18.6	5	11.6	3	7.0
Want Improvements	0	0	1	2.4	2	4.8	6	14.3	7	16.7	7	16.7	9	21.4	7	16.7	2	4.8	1	2.4

Note. for each item. Scores are based on a 1–10 Likert scale, where 1 indicates a strongly negative response and 10 indicates a strongly positive response.

Table 2. Comparison of human anatomy assessment results between student cohorts with and without Generative AI (GenAI) tool use in collaborative learning sessions.

Grade	Cohort 1 (without GenAI tools) n=121		Cohort 2 (with GenAI tools) n=123		P-Value
	n	%	*n*	%	
A+	7	5.8	18	14.6	**0.023**
A	13	10.7	16	13	0.586
A-	32	26.4	33	26.8	0.946
B+	27	22.3	18	14.6	0.123
B	24	19.8	15	12.2	0.104
B-	15	12.4	15	12.2	0.962
C+	2	1.7	6	4.9	0.158
C	1	0.8	1	0.8	0.991
C-	0	0	1	0.8	0.321
D+	0	0	0	0	-
D	0	0	0	0	-
F	0	0	0	0	-

Note. The Mann-Whitney test was used for statistical comparison. Grade classifications are defined as follows: A+ (90–100), A (85–89), A− (80–84), B+ (75–79), B (70–74), B− (65–69), C+ (60–64), C (55–59), C− (50–54), D+ (45–49), D (40–44), and F (0–39).

Discussion

This study highlights a broadly positive response to Generative AI (GenAI) tools, particularly ChatGPT, among undergraduate radiography students studying human anatomy. Students reported that GenAI enhanced their understanding of anatomical concepts, was easy to navigate, and was a tool they would recommend to their peers. These findings reflect the growing value placed on GenAI as a supplementary educational resource, aligning with trends in higher education where digital tools are increasingly integrated into learning, teaching, and assessment (LTA) (Khlaif et al., 2024).

GenAI usage was moderate, with 67.5% of students reporting usage levels between 5 and 10 on the Likert scale. This suggests that these tools are becoming a regular part of students' study routines. This shift indicates changing preferences, with students turning to digital support alongside traditional methods. The convenience and immediacy of GenAI—such as on-demand access to information and clarification of concepts outside scheduled teaching sessions—are clear advantages. This aligns with the broader move towards technology-enhanced learning across universities, where intuitive and accessible platforms drive adoption (Zenios, 2020).

Ease of use is a significant factor in GenAI uptake; 69% of students rated navigation highly. When educational technology is user-friendly, students are more likely to incorporate it into their daily study habits (Holmes et al., 2019). The perceived improvement in understanding, reported by 60.6% of students, further demonstrates GenAI's effectiveness in supporting learning. Human anatomy results reinforce this: the GenAI cohort performed better overall and achieved a higher proportion of top grades compared to the non-GenAI cohort. AI-powered platforms offer personalised explanations and instant feedback, which are particularly valuable in complex subjects like anatomy (Holmes et al., 2019). These features facilitate deeper learning and encourage active engagement (Lam, 2025).

Despite these advances, there remain areas for improvement. Many students expressed a desire for improved features and greater accuracy, reflecting ongoing concerns about current AI's ability to contextualise complex anatomical variations and ensure valid content (Davis et al., 2025). Only 48.8% rated GenAI's accuracy and comprehensiveness highly, highlighting the need for caution. Frequent reliance on these tools raises questions about their educational role and underscores the importance of educators guiding students in critically evaluating AI-generated content. GenAI should complement, not replace, expert instruction and validated resources. Issues such as group dynamics, fairness in

assessment, and academic integrity in AI-assisted assignments require attention to ensure effective learning and ethical standards.

This study has several limitations. The sample size was small, with a low survey response rate, introducing potential non-response and selection bias. The non-randomised, cross-sectional design compared two cohorts from different academic years without controlling for confounding factors such as demographic differences and changes in teaching staff. Furthermore, the lack of parallel survey data from Cohort 1 makes it difficult to determine whether the experiences reported by the GenAI-supported cohort reflect actual improvements or simply perceived differences, limiting the strength of comparative claims. Reliance on self-reported perceptions introduces response bias and may not reflect objective learning outcomes or long-term retention. Conducted at a single institution and focused on undergraduate radiography students, the findings may not generalise to other disciplines or settings. While Cohort 2 showed a modest improvement in mean scores and more top grades, these outcomes should be interpreted cautiously. Future research should use larger, more diverse samples, objective measures, and longitudinal or randomised controlled designs to better assess GenAI's educational impact.

Implications

Integrating GenAI tools like ChatGPT into anatomy education can enhance student engagement, understanding, and academic performance. However, concerns regarding information accuracy and the need for critical evaluation indicate that GenAI should supplement, not replace, expert instruction. Ethical considerations, including fairness and academic integrity, must also be addressed to ensure responsible use.

These findings suggest that GenAI could benefit diverse disciplines and contexts by promoting digital literacy and critical thinking. It can guide ethical integration and standards while helping students adapt to evolving technology, thereby supporting global educational priorities in a rapidly changing landscape.

Acknowledgments

This study was supported by a grant from The Hong Kong Polytechnic University on promoting the use of GenAI in LTA. The author gives special thanks to all the students for their participation.

Ethical Declaration Statement

The project has been approved by the Human Subjects Ethics Application Review Committee of The Hong Kong Polytechnic University (Reference No: HSEARS20250409003).

Generative AI Policy

The author used GPT-4.1 during the preparation of this work to correct grammatical errors and improve readability. After utilising the tool, the author thoroughly reviewed and edited the content as necessary and assumes full responsibility for the publication's content.

References

Chan, K. S., & Zary, N. (2019). Applications and challenges of implementing artificial intelligence in medical education: Integrative review. *JMIR Medical Education*, *5*(1), e13930. https://doi.org/10.2196/13930

Davis, P., Napole, A., Reddy, C., Palepu, S. K., Wu, A. Q., Mannam, S. S., Gajjar, A. A., Jabarkheel, R., Ajmera, S., & Ahmad, H. (2025). Systematic evaluation of AI-based text-to-image models for generating medical illustrations in neurosurgery: A multi-stage comparative study. *Clinical Neurology and Neurosurgery*, 109039. https://doi.org/10.1016/j.clineuro.2025.109039

Fahrni, S., & Sabatasso, S. (2025). 3D modelling in anatomy teaching: State of the art and pilot investigations for its application. *Translational Research in Anatomy*, 100444. https://doi.org/10.1016/j.tria.2025.100444

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign. https://discovery.ucl.ac.uk/id/eprint/10139722/

Khlaif, Z. N., Ayyoub, A., Hamamra, B., Bensalem, E., Mitwally, M. A., Ayyoub, A., Hattab, M. K., & Shadid, F. (2024). University teachers' views on the adoption and integration of generative AI tools for student assessment in higher education. *Education Sciences*, *14*(10), 1090. https://doi.org/10.3390/educsci14101090

Lam, N. (2025). Interactive learning meets AI: Elevating student experiences in fashion and culture education. *EDULEARN25 Proceedings*. https://doi.org/10.21125/edulearn.2025.0609

Zenios, M. (2020). Educational theory in technology-enhanced learning revisited: A model for simulation-based learning in higher education. *Studies in Technology Enhanced Learning*, *1*(1). https://doi.org/10.21428/8c225f6e.1cf4dde8

Optimising AI for Chemistry Grading: Boosting Efficiency and Reducing Educator Workload in Foundation Courses

Johnson Chung Ming SHUM [2], Hung-Wing LI [1], To NGAI [1], Sam Chun Kit HAU [1]

[1] Department of Chemistry, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong SAR, People's Republic of China

[2] SimProBot Inc., Delaware, the United States

Abstract

In response to the growing challenges of grading in large STEM courses, this study introduces a hybrid human-AI grading platform designed for foundational chemistry education. By integrating large language models (LLMs) with human oversight, the platform automates the evaluation of diverse assignment types, including text responses, numerical calculations, and graphical chemical structures. The methodology combines structured rubrics, semantic matching, and iterative feedback loops to optimise grading accuracy, efficiency, and equity. Results demonstrate a significant improvement in GTA–AI (graduate teaching assistant–AI) agreement rates, rising from below 50% to over 80% after rubric refinements and prompt engineering. The platform streamlines grading workflows, reduces graduate teaching assistants' (GTAs') workloads, and provides timely, personalised feedback to students. Its scalable design benefits large, diverse classrooms and ensures equitable assessment standards. This work highlights the potential of AI-assisted grading to address global challenges in higher education, improving accessibility to quality education while reducing systemic faculty burdens.

Keywords: Chemical Education Research, Generative Artificial Intelligence, Chemistry Education, Grading Platform, Large-Enrollment Courses

1. Introduction

In the evolving landscape of higher education, particularly within STEM disciplines like chemistry, rising enrolments and the adoption of inquiry-based pedagogies have intensified the need for efficient, scalable grading systems. Traditional methods reliant on instructors and GTAs often struggle to keep pace in large classes, leading to delayed or superficial feedback that hinders student learning and contributes to educator burnout (Doe, 2013; Timmerman, 2011). This study investigates the integration of AI-assisted grading as a transforma-

tive solution, leveraging large language models (LLMs) and machine learning to automate the evaluation of diverse assignments (White, 2023). Our objective is to design and evaluate a hybrid human-AI platform for foundational chemistry courses, emphasising robust training processes and collaborative oversight to improve grading efficiency and equity.

The inefficiency of conventional grading in high-enrollment settings underscores the importance of timely, personalised feedback for fostering self-regulated learning. Delayed remediation correlates with weaker performance, while GTA workloads constrain in-person support (Doe, 2013; Abiad, 2025). Furthermore, human biases, such as grader leniency, introduce inconsistencies that undermine educational equity. Addressing these issues is particularly crucial as class sizes continue to grow, often exceeding hundreds of students in introductory chemistry courses (Doe, 2013). AI-assisted grading offers scalability and innovation; empirical evidence shows that AI reduces GTA workloads, doubles feedback detail, and improves student outcomes (Abiad, 2025). These advancements ensure greater accessibility to high-quality education, particularly benefiting underrepresented groups, while alleviating systemic faculty burdens (Timmerman, 2011; Abiad, 2025; Humphry, 2023).

Despite its promise, AI-assisted grading faces challenges, including resource intensity, limitations in nuanced assessments, and risks of bias or inconsistency without human oversight. To address these, we propose a novel hybrid human-AI platform for foundational chemistry courses. By integrating well-designed rubrics and collaborative human-AI loops, the platform achieves high reliability, equity, and efficiency. This work advances the field by optimising the synergy between AI and human expertise in grading systems, ensuring scalable, equitable, and high-quality education.

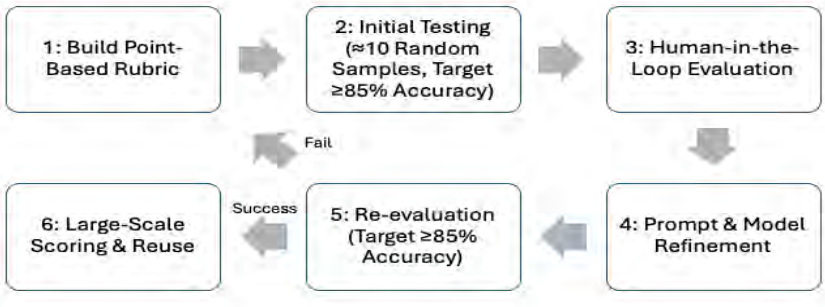
2. Methodology

Our large-enrollment undergraduate foundational chemistry courses, with over 350 students across two sections, present significant logistical and pedagogical challenges due to the intensive grading demands of four take-home assignments per semester, in addition to quizzes, a test, and a final exam. These assignments feature seven diverse question types: (i) multiple-choice, (ii) simple molecular formulas, (iii) basic calculations, (iv) calculations with units, (v) complex multistep calculations, (vi) text-based responses, and (vii) graphical chemical structures. Manual grading of these varied assessments is both time-consuming and labour-intensive for instructors and GTAs, limiting their ability to provide personalised support, address individual learning needs, and engage in meaningful classroom interactions. Furthermore, delayed feed-

back from manual grading prevents students from promptly identifying their strengths and weaknesses, hindering their ability to seek timely assistance—an essential component of success in foundational courses.

Training an AI system to grade written assignments using large language models (LLMs) involves a structured, semi-supervised approach that integrates computational intelligence with human expertise (Figure 1). The goal is to enable the AI to evaluate open-ended responses consistently and accurately according to predefined academic standards.

Figure 1. Schematic Diagram of the Training Flow for our AI model



The process begins with developing a point-based rubric, which serves as the ground truth for supervised fine-tuning. Each rubric defines key grading criteria—such as conceptual accuracy, completeness, clarity of reasoning, and correct domain-specific terminology—mapped to quantifiable score ranges. This structured rubric transforms qualitative assessment criteria into rule-based representations that the AI can interpret and apply effectively.

Next, the system undergoes a pilot evaluation phase, grading a small sample of student responses—typically around ten. These outputs form a training-validation dataset for assessing model performance. AI-generated scores are compared to human-assigned scores to calculate inter-rater reliability and identify discrepancies. This phase incorporates human-in-the-loop supervised learning, where human feedback informs iterative model refinement. Discrepancies guide prompt engineering and model calibration, refining instructions, enhancing domain-specific understanding, and encoding exceptions (e.g., recognising both British and American spelling variants). This iterative process, akin to reinforcement learning from human feedback (RLHF), optimises the model's scoring behaviour.

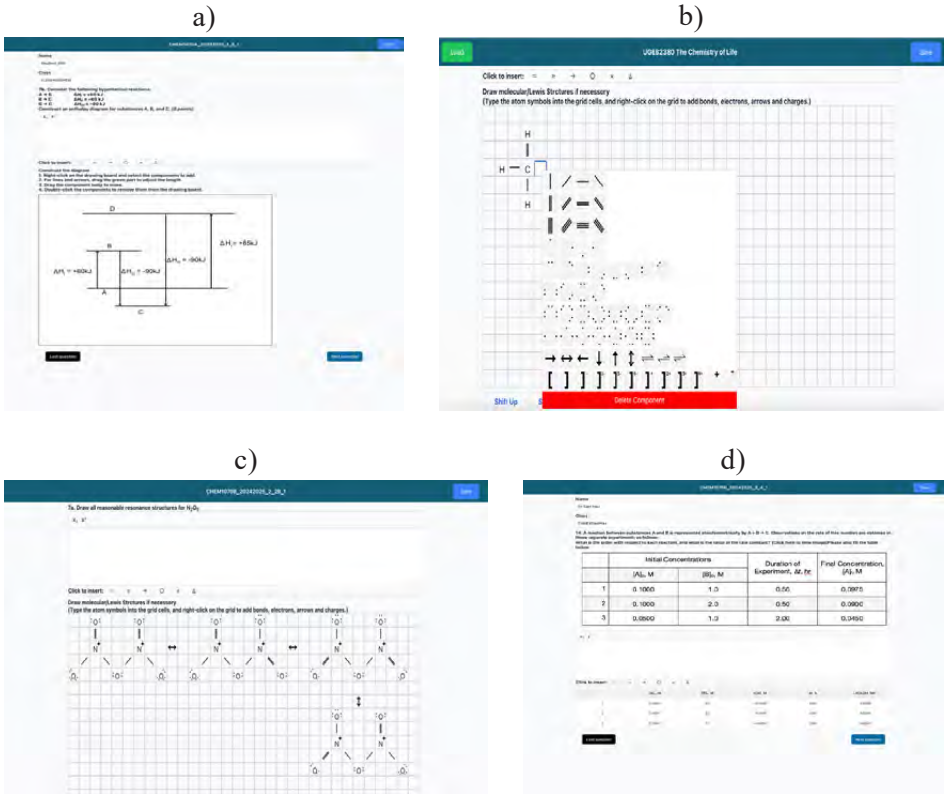
Once the rubric and prompts are refined, the initial sample set is re-evaluated using the updated configuration. Evaluation metrics such as accuracy, consistency, and correlation with human scores (e.g., Pearson correlation) are recalculated. The target is achieving at least 85% alignment with human grading. Once this benchmark is reached, the model is ready to scale across larger datasets and multiple assignment batches.

A key advantage of this process is the reusability of finalised rubrics and prompts, enabling standardised, repeatable assessment. Over time, the AI's meta-learning capabilities improve performance on similar tasks, reducing the need for extensive retraining.

Crucially, the finalised marking prompt and scoring rubric become reusable assets. They can be applied to subsequent instances of the same or similar questions, enabling standardised and repeatable assessment at scale. Over time, the AI's meta-learning capacity can accumulate domain-specific insights, further improving performance on related tasks without extensive retraining. In essence, this training methodology integrates principles of supervised learning, human feedback loops, and prompt-based adaptation. By systematically combining structured rubrics with data-driven refinement, the process ensures that the LLM evolves into a reliable and transparent automated grading assistant—capable of mirroring human evaluative judgement while significantly enhancing efficiency and scalability in academic assessment contexts.

The platform streamlines student access through batch account creation, providing secure login credentials for efficient onboarding. Students submit assignments using integrated rich text editors that support mathematical equations, chemical notations, and symbols. Specialised tools, such as molecular structure and formula builders, enable students to construct Lewis structures and diagrams (e.g., energy profiles or entropy diagrams). The platform supports four response types: text, molecular structures and formulae, drawings, and table completions (Figure 2).

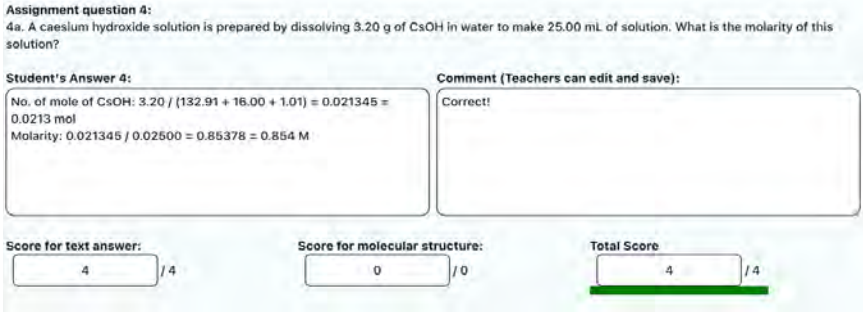
Figure 2. Diagram of the specialised tools available on our grading platform; a) Response includes energy profiles, b) Response includes molecular structures, c) Response includes molecular structures and formulae, d) Response includes table completions.



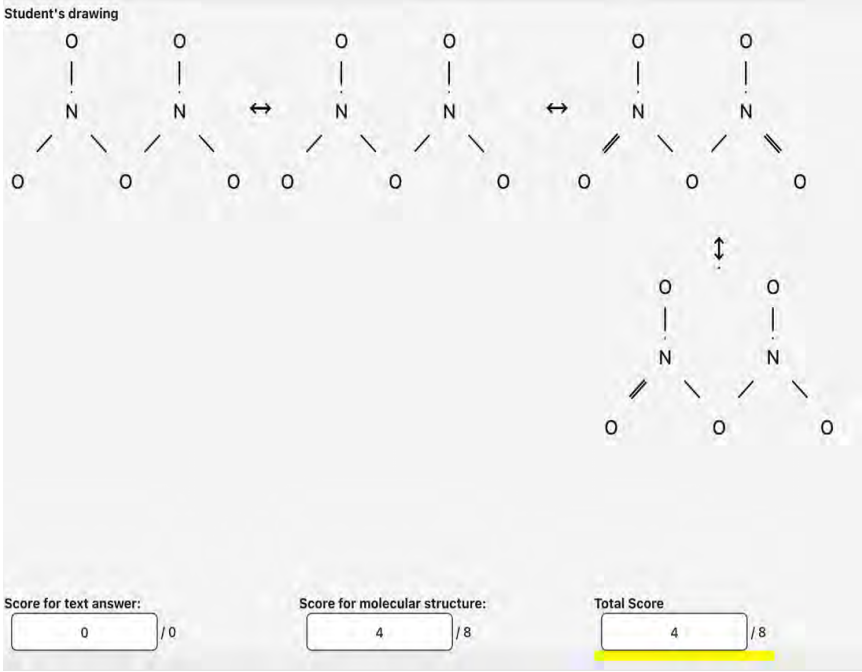
Text and table responses are processed directly by the GPT-4O LLM using grading prompts derived from rubrics. For molecular structures and drawings, submissions are converted to JPG images and evaluated by the vision-language model (VLM) component of GPT-4O, guided by grading prompts. The AI generates a numerical score and detailed feedback highlighting deviations from expected answers. Scores and feedback are stored on the application server for GTA review. A colour-coded review panel enhances efficiency: green for full scores, yellow for partially correct answers, and red for zero (Figure 3). GTAs can review and adjust scores or comments before saving changes. Once all assignments are reviewed, the status is updated to "release," allowing students to view their results. For further analysis, scores can be exported as .CSV files for integration with institutional systems. This platform enhances grading efficiency, supports diverse response formats, and ensures transparent, rubric-driven assessment with robust review mechanisms.

Figure 3. Diagram of the colour-coded review panel on our grading platform; a) Green for correct answers, b) Yellow for partially correct answers, and c) Red for incorrect answers with AI-generated feedbacks and comments.

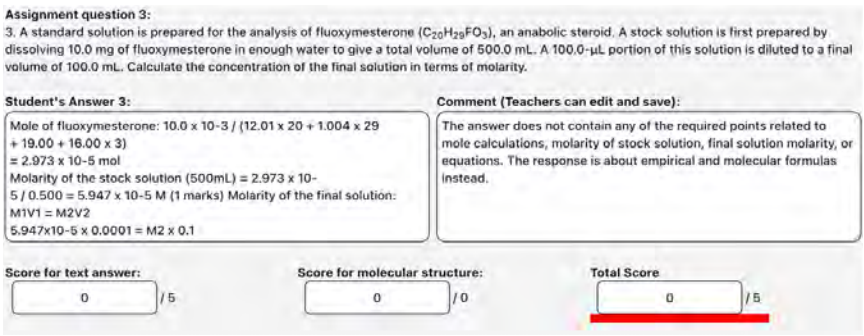
a)



b)



c)



3. Results and Discussion

For the academic year 2024–25, we implemented an AI-based pre-scoring system designed to evaluate student assignments and measure agreement between AI-generated scores and those assigned by human graders, either GTAs or instructors. This **agreement rate** serves as a practical estimate of the model's accuracy and reliability. Conceptually, the process mirrors supervised learning: human-assigned grades function as ground-truth labels against which AI predictions are systematically evaluated and refined.

Granular Rubric Design and Feature-Based Scoring

Because AI-based grading is inherently sensitive to ambiguity, precise and fully operationalised assessment criteria are foundational to the system's reliability. We therefore adopt a marking-scheme-based approach structured around a feature-based scoring model, in which the AI evaluates student responses by detecting specific answer components and assigning partial credit for each identifiable element. Individual grading points are kept intentionally small—typically ranging from 0.5 to 1 mark—with each unit corresponding to a narrowly defined criterion. For instance, 0.5 points may be awarded for correctly naming a concept, and a separate 0.5 points for accurately describing one of its properties. By decomposing the rubric into many fine-grained scoring units, the AI is guided to make a series of targeted, focused judgements rather than holistic assessments. This design substantially reduces subjectivity, improves inter-rater reliability between the AI and human GTAs, and renders the grading pipeline more robust, transparent, and interpretable.

A key principle underlying this design is to avoid overloading the AI with unnecessary complexity. Rather than functioning as an open-ended evaluator, the AI is constrained to a focused task: semantic equivalence checking. For each question, the required scoring points are defined at a semantic level. If the rubric specifies that a student must correctly express the concept "A and B," for example, the AI determines whether the response is semantically equivalent to this requirement—regardless of surface-level variation in wording, order, or formatting. As illustrated in Table 1, students may express the same concept in many ways, including paraphrasing, using synonyms, or reordering terms. The AI's role is not to evaluate style or presentation but to determine whether the substance of the response aligns with the intended rubric criterion. This approach moves decisively away from traditional string-matching systems, in which minor variations in word order or punctuation could cause correct answers to be marked incorrect. By treating the AI as a semantic matching component, we reduce noise, eliminate brittle matching behaviour, and produce a grading process that is more consistent and interpretable.

Table 1. Examples of Different Surface Forms in Student Responses

Form 1	Form 2	Form 3	Form 4	Form 5	Form 6
A and B	A, B	A B	A; B	B and A (order variation)	Using acceptable synonyms and equivalent phrases

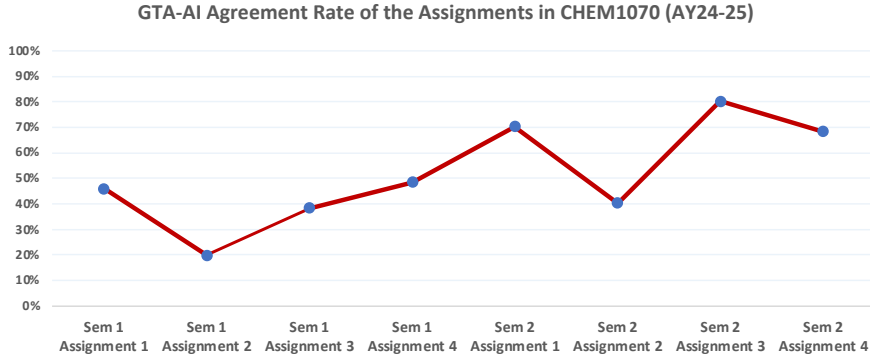
To operationalise this, we explicitly encode canonical phrasings and acceptable semantic variants for each rubric item, enabling robust content matching. Complementing this, we define a detailed, machine-interpretable error taxonomy that identifies common mistake types—such as missing key concepts, incorrect definitions, logical inconsistencies, or misuse of terminology—and assigns specific, predetermined penalties to each. This level of explicitness reduces ambiguity, decreases the AI's inferential burden, and further improves grading consistency and validity.

Calibration, Thresholding, and Iterative Refinement

The calibration process begins with a small validation set of approximately ten human-graded examples. Rather than serving as a definitive statistical validation, this set functions as an agile calibration tool—allowing instructors to rapidly identify systematic AI errors and iteratively refine prompts without placing undue burden on human graders. The instructor reviews AI-generated scores and feedback against ground-truth grades, assessing rubric consistency, accuracy, and recurring error patterns. We target an agreement threshold of 85%, a benchmark chosen because it meets or exceeds the typical inter-rater reliability observed between trained human graders. Once this threshold is achieved on the calibration batch, the configuration is deployed for full-class grading.

As shown in Figure 4, the GTA–AI agreement rate improved significantly from semester one to semester two. Iterative refinement proved essential to reaching this target: early configurations yielded agreement rates below 50%, largely because overly ambiguous prompts forced the AI to navigate excessive semantic noise and under-specified criteria. Through successive prompt engineering, agreement rates climbed to above 80%. Three key refinements drove this improvement: the adoption of micro-criteria decomposing rubrics into 0.5-mark increments; the integration of semantic matching enabling the AI to recognise synonyms and paraphrases rather than relying on rigid string matching; and the development of an error taxonomy assigning specific penalties to common mistakes, thereby preventing arbitrary point deductions.

Figure 4. GTA–AI Agreement Rates for CHEM1070 Assignments (Academic Year 2024–2025)



Vision-Language Model Pipeline for Diagram-Based Assessment

For grading chemical molecular structures and other diagram-based responses, our initial approach relied on a global image-similarity metric—comparing student drawings to a model answer and applying score deductions for detected differences. However, this holistic method proved too sensitive to minor, rubric-irrelevant variations and did not align well with criterion-based scoring. To address these limitations, we developed a structured, vision-based evaluation pipeline using a vision-language model (VLM). In this new approach, the VLM first processes the reference solution to extract and annotate fine-grained scoring elements—including specific atoms, bonds, functional groups, stereochemical indicators, labels, and spatial relationships—each treated as an explicit rubric scoring point. The same model then parses student-submitted drawings to determine whether each required element is present, correctly represented, and appropriately positioned. Scoring proceeds feature-by-feature, with missing or incorrectly rendered elements resulting in rubric-based deductions. This object-level approach minimises reliance on coarse global similarity metrics, substantially improves interpretability, and produces a grading process that is more robust, transparent, and aligned with structured assessment practices.

Following successful pilot testing, this VLM pipeline is currently being optimised for full-scale deployment and rigorous statistical validation in the coming semesters—representing a promising next step in extending AI-assisted grading to the visually complex assessment demands of STEM education.

4. Implications

This study offers a promising step toward addressing one of higher education's most pressing global challenges: the massification of STEM enrolment.

As universities worldwide grapple with maintaining assessment quality amid surging student populations and acute shortages of qualified grading personnel—disparities especially pronounced in under-resourced institutions—scalable and equitable grading solutions are increasingly needed. Our preliminary findings are encouraging: GTA-AI agreement rates surpassing 80% following iterative refinement suggest that this approach holds meaningful potential for improving accuracy, consistency, and efficiency in large-scale assessment. By working to minimise human bias and variability, the platform may also contribute to more equitable outcomes for underrepresented student populations across varied institutional contexts.

While further validation is needed, the framework's design built around structured rubrics, iterative human oversight, and adaptable workflows suggests it could be meaningfully extended across diverse linguistic, cultural, and institutional environments, moving beyond the resource-intensive models that have historically favoured well-funded institutions. Ultimately, we hope this work contributes, in a modest but meaningful way, to the broader global effort to deliver rigorous, personalised, and equitable STEM education at scale; redirecting educator capacity toward the high-value human interactions that technology, however capable, cannot replace.

Acknowledgments

We acknowledge the courseware development grant (CDGS-4171186) and teaching development grant (TDLEG-4171189) from The Chinese University of Hong Kong. Our deepest gratitude goes to Mr Shum, whose expertise, resources, and guidance were indispensable in developing this scalable, equitable grading platform.

Generative AI Policy

The author(s) used ChatGPT, a language model developed by OpenAI, to polish the language and enhance the clarity of this manuscript. After utilising the tool, the author(s) thoroughly reviewed and edited the content as necessary and assumed full responsibility for the publication's accuracy, originality, and overall content.

References

Abiad, A., Richter, B. S., Padilla, M. S. T. L., Flores, M. A., Thompson, J. C., Lechner, P. G., Edwards, K. D., & Link, R. D. (2025). AI-assisted grading increases quality feedback and grading efficiency in undergraduate chemistry courses. *Journal of Chemical Education*, 102(11), 4936-4943.

Doe, S. R., Gingerich, K. J., & Richards, T. L. (2013). An evaluation of grading and instructional feedback skills of graduate teaching assistants in introductory psychology. *Teaching of Psychology*, 40(4), 274–280.

Humphry, T.; Fuller, A. L. (2023). Potential ChatGPT Use in Undergraduate Chemistry Laboratories. *J. Chem. Educ.*, 100(4), 1434–1436.

Timmerman, B. E. C., Strickland, D. C., Johnson, R. L., & Payne, J. R. (2011). Development of a "universal" rubric for assessing undergraduates' scientific reasoning skills using scientific writing. *Assessment & Evaluation in Higher Education*, 36(5), 509–547.

White, A. D. (2023). The future of chemistry is language. *Nature Reviews Chemistry*, 7(7), 457–458.

How Perceived Learning Difficulties Shift in a Short-Term GenAI-Integrated Design Course

Henry MA [1] & Jinmei HUANG [2]
[1] School of Art & Design, The Hong Kong Polytechnic University, Hong Kong, China
[2] School of Art & Design, Guangzhou College of Commerce, Guangzhou, China

Abstract

This study examines shifts in students' perceptions of support and learning difficulties within a short-term Generative AI (GenAI)-integrated design course. A pre-post survey design was implemented in a four-week (32 contact-hour) undergraduate design course where GenAI was embedded as a pedagogical component. Matched responses (paired n = 36; post n = 39) assessed perceived efficiency support, creative support, self-efficacy, and future intention, alongside post-course mechanisms (guidance improvement, workflow formation, iteration intensity) and persistent bottlenecks. Pre-post changes were modest: efficiency support remained high but unchanged, creative support showed a small non-significant increase, and self-efficacy improved with marginal significance ($p = 0.058$). Most students reported an improved ability to guide GenAI and an emerging workflow, while common bottlenecks centred on controllability, consistency, and integration. The findings rely on perception-based measures and should be interpreted as indicators of students' experienced support and difficulty during the course. The results suggest that GenAI compresses early production barriers but shifts difficulty toward higher-order control and synthesis.

Keywords: Generative AI, Design education, Workflow formation, Self-efficacy, Learning difficulties

1. Introduction

Generative AI (GenAI) is rapidly entering higher education as both a productivity tool and a pedagogical force, prompting institutions to reconsider how learning, assessment, and integrity are designed (UNESCO, 2023; Kasneci et al., 2023; Cotton et al., 2024). In studio-based design education, this shift is especially consequential because learning develops through iterative making, critique, and reflective experimentation rather than one-shot correctness (Schön, 1983). Emerging evidence further suggests that GenAI can support ideation while also introducing new constraints such as fixation, making learning quality harder to infer from outputs alone (Wadinambiarachchi et al., 2024).

This study focuses on how GenAI redistributes difficulty across short-term design learning. Against this backdrop, it remains unclear which student perceptions shift over a four-week instructional cycle and which difficulties persist even when GenAI is integrated into teaching. This study addresses that gap by examining a short undergraduate design course using matched pre-post surveys to track changes in perceived support, strategy formation, and persistent bottlenecks around controllability, consistency, and integration.

To guide this investigation, the study was framed by the following research questions:

- **RQ1:** How do students' perceptions of efficiency support, creative support, self-efficacy, and future intention change from pre to post in a GenAI-integrated design course?
- **RQ2:** What process mechanisms emerge after the course (guidance improvement, workflow formation, and iteration intensity), and how are these mechanisms associated with perceived gains (efficiency gain, quality gain, self-efficacy, and success in realising ideas)?
- **RQ3:** What persistent bottlenecks do students report when using GenAI, and are these difficulties linked to pre-course controllability concerns and post-course self-efficacy differences?

2. Methodology

A pre-post survey design was employed to examine short-term shifts in students' perceptions of support and difficulty in a four-week required first-year undergraduate course, *Artificial Intelligence and Visual Communication*, totaling 32 contact hours. The course guided students through most stages of the design-thinking process, from identifying a problem and defining a concept to generating proposals, developing visual outcomes, and producing a final short video. Deliverables included logo design, IP character design, static and animated stickers, static and motion posters, a video script, and an approximately 40-second video. GenAI was integrated throughout these stages as part of ideation, visual exploration, refinement, and production support rather than being introduced as an optional add-on.

The pre-survey was administered after the first class, and the post-survey after the final class. The pre-survey captured students' baseline GenAI experience, expectations, and concerns. The post-survey measured perceived efficiency and task progression, perceived improvement in creative output quality, self-efficacy/controllability in using GenAI, and strategy formation in using GenAI. Perceived alignment is used here as a proximal indicator of students' felt ability to meet course demands. These measures are indirect proxies; therefore,

they should not be read as direct evidence that course learning outcomes were fully achieved.

The analysis tracks changes in perceived support, strategy formation, and difficulty during the course and helps locate where the pedagogy appeared to support progress or where problems remained from students' perspectives. It also collected process indicators such as iteration frequency, typical actions after initial AI outputs, encountered bottlenecks, and coping responses. Responses were matched across time points using the last four digits of student IDs to enable paired comparisons. Paired t-tests were used for comparable pre-post Likert items, while descriptive statistics summarised post-course distributions of strategy formation and bottleneck patterns. Non-significant changes are interpreted descriptively, not as confirmatory evidence, and are considered pedagogically relevant only when they align with post-course process indicators and recurring bottleneck patterns. Multi-select difficulty items were coded into binary variables and visualised to identify persistent challenges. Artifact-based or performance-based assessment was not included in this short paper and remains a priority for future work.

3. Results and Discussion

Matched pre-post responses were obtained for 36 students (paired n = 36), and 39 students completed the post-survey (post n = 39). Overall, the findings indicate modest shifts in students' perceptions of support alongside concentrated bottlenecks in controllability, consistency, and integration. Key quantitative findings across RQ1–RQ3 are summarised in Table 1. Taken together, these results support a cautious interpretation of how GenAI functions in a short-cycle design pedagogy.

Table 1 summarises the full quantitative pattern across the three research questions, including pre-post changes, post-course mechanisms, group differences linked to reported bottlenecks, and descriptive post-course distributions. Figure 1 corresponds to the pre-post comparison in RQ1; Figures 2a and 2b correspond to the post-course mechanism patterns in RQ2; and Figure 3 corresponds to the persistent bottlenecks highlighted in RQ3.

Table 1. Key findings on pre-post changes, post-course mechanisms, and persistent bottlenecks (paired n = 36; post n = 39)

Indicator / comparison	Key result (condensed)	p	Sig.
Panel A. Pre-post changes in key perception-based indicators (paired n=36)			
Efficiency support (Likert)	Pre 4.14±0.59; Post 4.14±0.87; t=0.00; dz=0.00	1	n.s.
Creative support (Likert)	Pre 3.81±0.86; Post 3.94±0.95; t=0.78; dz=0.13	0.443	n.s.
Self-efficacy (experience→effective use)	Pre 3.61±0.77; Post 3.89±0.82; t=1.96; dz=0.33	0.058	†
Future intention to use GenAI	Pre 4.11±0.62; Post 4.06±0.92; t=-0.39; dz=-0.06	0.701	n.s.
Panel B. Post-course mechanisms: guidance and strategy formation (matched n=36)			
Workflow formation ↔ Self-efficacy	ρ=0.559	<.001	***
Workflow formation ↔ Quality gain	ρ=0.491	0.002	**
Workflow formation ↔ Efficiency gain	ρ=0.468	0.004	**
Guidance improvement ↔ Efficiency gain	ρ=0.514	0.001	**
Guidance improvement ↔ Workflow formation	ρ=0.634	<.001	***
Workflow formation ↔ Success in realizing idea	ρ=0.443	0.007	**
Panel C. Persistent bottlenecks: group differences by reported difficulties (matched post n=36)			
Pre controllability concern → Post 'hard to integrate'	Issue=3.53 (n=15) vs No-issue=2.95 (n=21)	0.035	*
Post 'hard to integrate' ↔ Self-efficacy	Issue=3.60 (n=15) vs No-issue=4.10 (n=21)	0.082	†
Post 'unstable outputs' ↔ Efficiency gain	Issue=3.90 (n=21) vs No-issue=4.47 (n=15)	0.056	†
Panel D. Descriptive distributions and prevalence (post n=39)			
Efficiency gain ≥4/5	71.8% (28/39)		
Quality gain ≥4/5	66.7% (26/39)		
Self-efficacy ≥4/5	64.1% (25/39)		
Guidance improved (any)	97.4% (38/39)		
Workflow formed (clear/mostly)	79.5% (31/39)		
Iterated ≥4 times	59.0% (23/39)		
Bottleneck: unstable / hard to reproduce	61.5% (24/39)		
Bottleneck: mismatch with intention	53.8% (21/39)		
Bottleneck: aesthetics hard to refine	53.8% (21/39)		
Bottleneck: hard to edit/integrate	41.0% (16/39)		

Note:

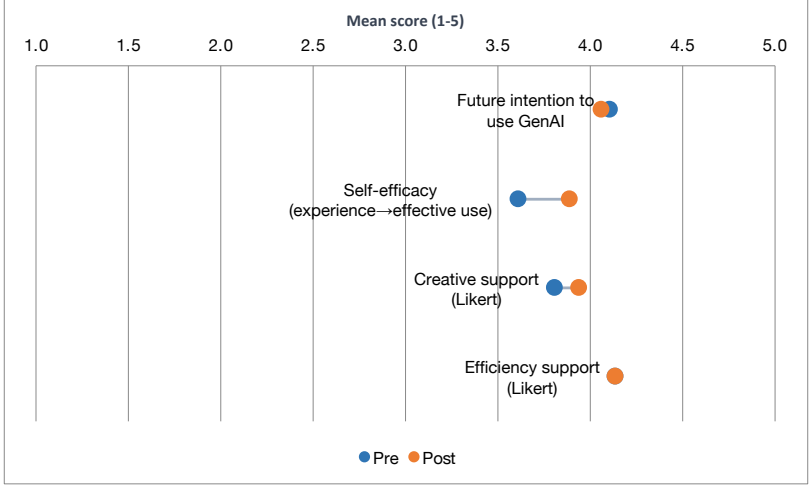
* *Pre-survey n = 43; Post-survey n = 39; matched paired n = 36 (via last four digits of student IDs).*

* *Panel A uses paired-samples t-tests (Cohen's dz).*

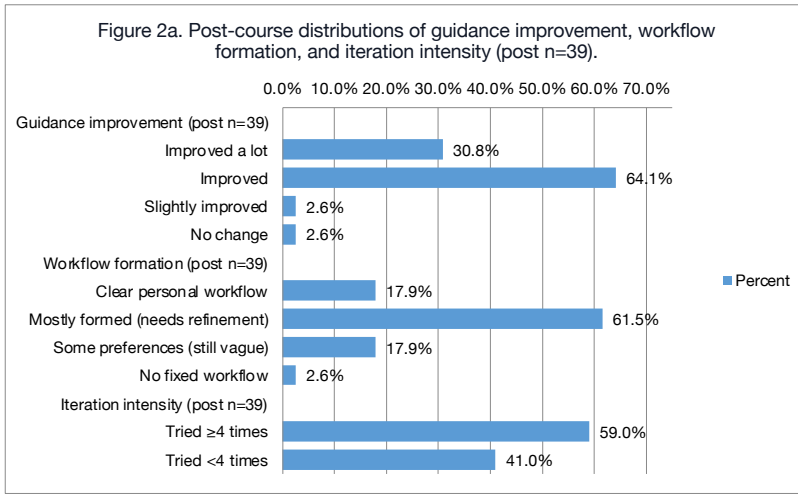
* *Panel B uses Spearman rank correlations (ϱ).*
* *Panel C uses Mann–Whitney U tests (two-tailed).*
* *Panel D reports descriptive percentages only.*
* *Significance codes: *** p<.001; ** p<.01; * p<.05; † p<.10; n.s. not significant.*

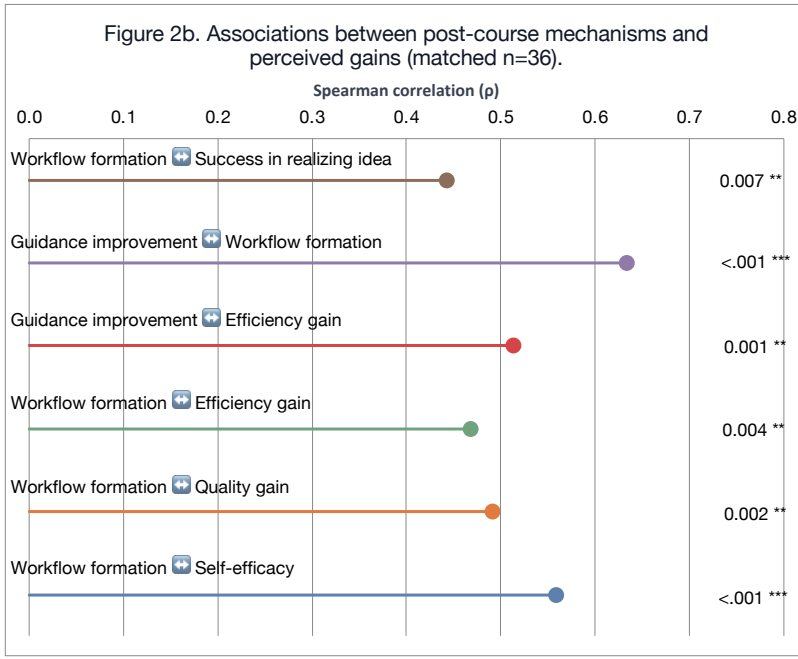
RQ1 (paired n = 36). Efficiency support remained high but unchanged (M = 4.14 → 4.14, t = 0.00, p = 1.000, n.s.). Creative support increased slightly (M = 3.81 → 3.94) but was not significant (p = 0.443, n.s.). Self-efficacy showed the largest gain (M = 3.61 → 3.89), reaching marginal significance (p = 0.058, †). Future intention to use GenAI remained stable (M = 4.11 → 4.06, p = 0.701, n.s.). Descriptively (post n = 39), 71.8% (28/39) rated efficiency gain ≥4/5, 66.7% (26/39) rated quality gain ≥4/5, and 64.1% (25/39) rated self-efficacy ≥4/5. These pre-post shifts are visualised in Figure 1. The pre-post movement is therefore modest rather than decisive.

Figure 1. Pre-post shifts in students' perceptions of key course-related indicators (paired n=36).

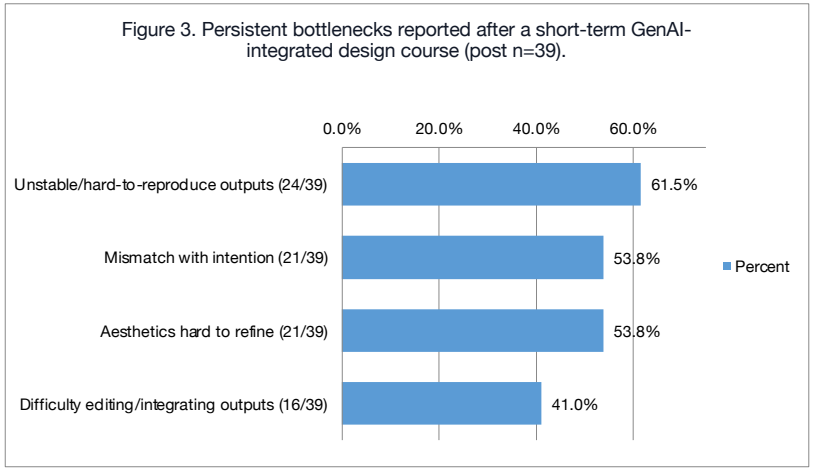


RQ2 (post n = 39; associations on matched n = 36). Most students reported an improved ability to guide GenAI (97.4%, 38/39) and an emerging workflow (clear/mostly formed: 79.5%, 31/39). Iteration was common: 59.0% (23/39) reported trying ≥4 times. Figure 2a displays the descriptive distributions of guidance improvement, workflow formation, and iteration intensity. Mechanism indicators were strongly associated (Table 1, Panel B): workflow formation correlated with self-efficacy (rho = 0.559, p < .001), quality gain (rho = 0.491, p = 0.002), and efficiency gain (rho = 0.468, p = 0.004); guidance improvement correlated with efficiency gain (rho = 0.514, p = 0.001) and workflow formation (rho = 0.634, p < .001). Figure 2b visualises these associations.





RQ3 (post n = 39; matched n = 36 for group tests). Frequent bottlenecks included unstable/hard-to-reproduce outputs (61.5%, 24/39), mismatch with intention (53.8%, 21/39), aesthetics hard to refine (53.8%, 21/39), and difficulty editing/integrating outputs (41.0%, 16/39). Group comparisons showed that students reporting "hard to integrate" had higher pre-course controllability concerns ($p = 0.035$, *) and tended to report lower self-efficacy ($p = 0.082$, †) (Table 1, Panel C). Figure 3 visualises the concentration of these persistent bottlenecks.



Across the three research questions, the results point to a clear redistribution of difficulty. Pre-post changes were modest. The clearest signal comes from the combination of workflow formation, improved guidance ability, and concentrated bottlenecks around control, consistency, and integration. This matters because classroom use of GenAI involves more than faster production; students still need to evaluate, direct, and refine what the system generates. The evidence here reflects students' perceptions of their ability to engage with course demands over a short cycle and should be read as an indirect indication of learning-related experience rather than a direct measure of achieved learning outcomes.

Studio-based design learning traditionally develops through iterative making, critique, and reflective experimentation (Schön, 1983). Difficulty remained present throughout the course, although it shifted from producing options to steering, refining, and integrating outputs into coherent design decisions. This pattern resonates with concerns that GenAI may encourage superficial product completion unless process evidence is made visible and valued (UNESCO, 2023; Xia et al., 2024). The persistent bottlenecks also suggest a more concrete pedagogical agenda. "Unstable or hard-to-reproduce outputs" points to a consistency problem: students may generate useful results once but cannot reliably regenerate or adapt them. "Mismatch with intention" points to weak controllability: students may describe goals broadly but lack methods for specifying constraints, references, and decision criteria. "Difficulty editing/integrating outputs" points to a synthesis problem: even acceptable AI outputs may remain hard to combine, refine, or translate into final design work. These recurring failure points help explain why short-term difficulty remains concentrated in control and integration despite high reported efficiency support.

4. Implications

For design educators, the practical implication is to teach control-oriented routines and embed GenAI within a broader design process. Three intervention areas appear especially important. First, **controllability strategies**: prompt iteration with explicit constraints, reference anchoring, and negative instructions can help students move from vague requests to steerable outputs. Second, **consistency and reproducibility routines**: students should document prompt versions, parameter choices, and selection criteria so outputs can be repeated, compared, and refined. Third, **integration and editing workflows**: courses should require students to combine AI outputs with manual editing, layout decisions, material judgment, and reflective rationale so that synthesis remains visible. These strategies respond directly to the main bottlenecks identified here and may better support studio-like learning processes in short-term GenAI use. Future work should triangulate these perception-based findings with artifact-based assessment and rubric-aligned judgments of integration quality and process documentation.

Acknowledgments

The authors thank the students from a design class at a university in Southern China for their participation in this study.

Ethical Declaration Statement

The project has been approved by the School of Design of The Hong Kong Polytechnic University (Reference Number: HSEARS20250217010).

Generative AI Policy

The authors used ChatGPT during the preparation of this work to refine language and check formatting; after utilising the tool, the authors thoroughly reviewed and edited the content as necessary and assumed full responsibility for the publication's content.

References

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61(2), 228-239. https://doi.org/10.1080/14703297.2023.2190148

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., ... Kasneci, G. (2023). ChatGPT for good?

On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.

UNESCO. (2023). *Guidance for generative AI in education and research*. https://unesdoc.unesco.org/ark:/48223/pf0000386693

Wadinambiarachchi, S., Kelly, R. M., Pareek, S., Zhou, Q., & Velloso, E. (2024). The effects of generative AI on design fixation and divergent thinking. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (CHI '24, Article 380, pp. 1-18). Association for Computing Machinery. https://doi.org/10.1145/3613904.3642919

Xia, Q., Weng, X., Ouyang, F., Lin, T. J., & Chiu, T. K. F. (2024). A scoping review on how generative artificial intelligence transforms assessment in higher education. *International Journal of Educational Technology in Higher Education*, 21(1), Article 40. https://doi.org/10.1186/s41239-024-00468-z

GenAI-Supported IELTS Speaking Preparation: Prompting, Strategy Use, and Learner Perceptions

Rui JI [1]
[1] The University of Nottingham Ningbo China

Abstract

This study investigates how Chinese IELTS candidates use generative artificial intelligence (GenAI) in self-directed IELTS test preparation and examines its influence on preparation strategies and learner perceptions. Adopting a sequential exploratory mixed-methods design, the study combined semi-structured interviews (n = 8) and a survey (n = 80). Findings show that learners strategically align GenAI use with task-specific demands across IELTS Speaking parts, employing chatbots for idea generation and discourse planning (especially in Parts 2 and 3 tasks) and automatic speech recognition (ASR) tools for pronunciation feedback. GenAI use was associated with increased planning, rehearsal, and self-monitoring, but also with test-wiseness strategies such as the memorisation of model responses for Part 2 tasks. While learners reported reduced anxiety and perceived improvement, trust in AI feedback—particularly regarding pronunciation scoring—was more limited, and evidence of long-term speaking development was weak. The study highlights the need for critical AI literacy, task-aligned pedagogical integration, and complementary human-mediated speaking practice in high-stakes assessment contexts.

Keywords: AI-mediated language learning; IELTS; speaking test preparation

1. Introduction

Preparing for the IELTS Speaking Test poses specific pedagogical challenges due to its task-based design and analytic scoring criteria. The test comprises three tasks with distinct demands: Part 1 involves short, spontaneous responses on everyday topics; Part 2 requires a time-constrained monologic turn; and Part 3 focuses on abstract reasoning through extended discussion (IELTS, 2025). Performance is assessed across four criteria: fluency and coherence, lexical resources, grammatical range and accuracy, and pronunciation (IELTS, 2025; Read, 2022).

With the rapid uptake of generative artificial intelligence (GenAI) in higher education (Jeon et al., 2023; Perrotta et al., 2025), learners increasingly use AI tools to support IELTS test preparation. However, research on AI-mediated

language learning has focused predominantly on writing, leaving speaking development in test-specific contexts underexplored (Jeon et al., 2023; Perrotta et al., 2025).

Against this background, this study investigates how Chinese IELTS candidates use GenAI tools in self-directed IELTS Speaking preparation, with particular attention to task-aligned prompting practices, preparation strategies, and learner perceptions of AI-mediated feedback (Cohen, 2014; Fernandez, 2018; Perrotta et al., 2025). In doing so, it contributes empirical evidence to ongoing discussions of AI literacy and test-oriented speaking pedagogy in higher education.

The study addresses the following research questions:

- **RQ1:** How do learners use GenAI tools when preparing for the IELTS Speaking Test?
- **RQ2:** How does GenAI influence learners' IELTS speaking preparation strategies?
- **RQ3:** How do learners evaluate the usefulness and limitations of GenAI tools for speaking development?

2. Methodology

2.1 Research Design

An embedded sequential mixed-methods design was adopted to investigate learner behaviour and perceptions in GenAI-mediated IELTS speaking preparation. The study began with in-depth qualitative interviews to explore learners' prompting practices for each speaking task, strategic shifts, and perceptions of effectiveness. Findings from this phase informed the design of a subsequent survey. Participants used their preferred GenAI tools in their self-directed IELTS preparation; no specific platform was required, as the study examined naturally occurring strategy use rather than platform effectiveness. This design prioritised ecological validity by examining how learners actually use these tools in self-directed test preparation, thereby enabling triangulation between qualitative and quantitative data.

2.2 Participants

Data were collected between June and August 2025. Firstly, eight participants were recruited for the qualitative phase using maximal variation sampling to capture diverse proficiency levels and AI usage frequencies (Table 1). Following the interviews, 80 learners (including the eight interviewees) from 20 provinces were recruited through convenience sampling from the researcher's university network and an online IELTS preparation community to complete the

survey. Screening questions at the beginning of the survey ensured eligibility (i.e., having taken the IELTS within the past 24 months or currently preparing for it, and having experience using GenAI tools for IELTS preparation). The respondents reported diverse English-learning backgrounds, varying EMI exposure, and prior IELTS experience (Table 2).

Table 1: Participant Profiles for Qualitative Interviews

Participants	Educational Level	Overall IELTS Score	Speaking Band Score	Test Date	AI Usage Frequency
Hans (Male)	Bachelor's	6.0	5.5	March 2024	Once per week
Dora (Female)	Bachelor's	6.5 6.5	5.5 5.5	August 2023 December 2023	3-5 times per week
Cathy (Female)	Master's	7.0	6.0	November 2023	Daily
Gloria (Female)	Master's	6.5 7	6.0 6.0	July 2023 October 2023	Daily
Anna (Female)	Master's	7.0 6.5	7.0 6.0	January 2024 July 2025	Daily
Bonnie (Female)	Master's	7.0	7.0	December 2023	Daily
Emily (Female)	Master's	7.5 7.5	7.5 7.5	July 2023 May 2025	3-5 times per week
Frank (Male)	Master's	7.5	8	January 2024	Once per week

Note: All participant names are pseudonyms. Shaded scores represent those achieved after using AI tools for IELTS Speaking preparation.

Table 2: Demographic Characteristics of Participants (*n* = 80)

	Category		
Variable		n	%
Gender	Female	50	62.5
	Male	28	35
	Prefer not to say	2	2.5
Education level	High school	1	1.25
	Undergraduate	22	27.5
	Master's	54	67.5
	Doctorate	3	3.75
EMI study experience	Yes	62	77.5
	No	18	22.5
IELTS taken before	Yes	77	96.25
	No	3	3.75
Attended IELTS speaking course	Yes	35	43.75
	No	45	56.25
AI usage frequency	Once per week 3-5 times per week Daily	17 44 19	21.25 55 23.75

2.3 Instruments

Two instruments were used. Semi-structured interviews examined task-specific prompting behaviours towards different IELTS speaking tasks, tool selection rationales, perceived strategy changes, and reflections on AI-mediated preparation. The survey measured GenAI usage patterns in IELTS speaking preparation, strategy use (cognitive, metacognitive, and test-wiseness), and perceived effectiveness and limitations. Most items were rated on a six-point Likert scale; additional multiple-choice and open-ended items captured tool preferences and usage frequency.

2.4 Data Analysis

Interview transcripts were analysed thematically through iterative coding guided by the research questions. Survey data (n = 80) were analysed using descriptive statistics to identify overall trends and support triangulation. The two datasets were analysed independently. Integration occurred at the interpretation stage, where qualitative findings were used to contextualise, rather than statistically reinforce, survey results.

3. Results and Discussion

3.1 RQ1: How learners use GenAI in speaking preparation

Learners' use of GenAI is closely aligned with the differentiated task demands of the IELTS Speaking Test. For Part 1, participants typically used single-turn prompts to generate brief, everyday responses, reflecting the requirement for spontaneous personal communication (IELTS, 2025). In contrast, preparation for Part 2 involved multi-turn prompting to refine discourse organisation, manage time constraints, and structure content within a one- to two-minute monologic turn. For Part 3, learners crafted prompts to elicit analytical and balanced responses mirroring the abstract reasoning expected in this section (Fernandez, 2018).

Tool differentiation was also evident. LLM-based chatbots were primarily used for idea generation and structural support, particularly for Part 2 tasks, while automatic speech recognition (ASR) tools were used to obtain feedback on pronunciation and fluency. Some learners alternated between tools depending on task demands, adopting a hybrid approach. This selective use reflects emerging prompting literacy and strategic alignment between GenAI affordances and IELTS speaking requirements (Perrotta et al., 2025), highlighting the pedagogical value of explicit training in prompt design and tool selection.

3.2 RQ2: Influence on speaking preparation strategies

GenAI use was associated with increased planning, rehearsal, and self-monitoring (Table 3)—strategies widely linked to effective IELTS speaking preparation (Cohen, 2014; Fernandez, 2018). At the same time, reliance on AI-generated templates raises concerns. Memorising model responses represents a form of test-wiseness strategy that may approximate band-descriptor expectations without fully engaging learners in independent language construction (Cohen, 2014).

Table 3: Descriptive Statistics for Strategy Use Among AI Users (*n* = 80)

Strategy Category	Strategy Item		M	SD
Cognitive	Q16	Practising IELTS speaking questions aloud	5.45	0.67
	Q19	Asking AI to suggest vocabulary or expressions	5.4	0.76
Meta-cognitive	Q21	Asking AI to correct vocabulary or grammar errors	5.21	0.86
	Q24	Asking AI to provide overall feedback and using it to improve	5.12	0.91
	Q22	Asking AI to correct pronunciation	3.76	1.42
	Q17	Recording and reviewing one's own speech	3.62	1.62
	Q26	Asking AI to help plan a study schedule	3.14	1.63
	Q25	Simulating IELTS speaking tests using AI	4.71	1.18
Test Wiseness	Q20	Memorising AI-generated Part 2 answers	3.43	1.54

While such strategies may enhance short-term performance readiness, overreliance on prefabricated structures risks limiting cognitive engagement and the development of transferable communicative competence (Cohen, 2014; IELTS, 2025). This tension reflects longstanding debates in test preparation research regarding the balance between strategic performance optimisation and genuine language development (He et al., 2024). GenAI therefore functions as a double-edged scaffold, supporting structured rehearsal while potentially encouraging surface-level learning if used uncritically.

3.3 RQ3: Learner perceptions of usefulness and limitations

Qualitative findings indicate that most participants perceived linguistic improvement and valued AI feedback as accessible and non-judgemental. Reduced anxiety during practice was frequently reported. However, concerns emerged regarding feedback reliability, particularly in relation to ASR-based pronunciation scoring. The quantitative results reflected this ambivalence: perceived improvement and affective support were rated relatively highly, whereas trust in AI feedback and reported long-term gains in fluency or test performance were more limited (Table 4).

Table 4: Perceived Effectiveness by Category

Category	Mean	SD
Affective/motivational	4.59	0. 17
Linguistic improvement	4.37	0.32
Trust in AI	3.88	0.65

Given that IELTS pronunciation criteria prioritise intelligibility and prosodic control rather than error-free production (IELTS, 2025), AI feedback emphasising surface accuracy may misalign with actual test performance. While GenAI can support confidence through low-stakes practice (Jeon et al., 2023), it does not guarantee sustained speaking development or reliable assessment outcomes (Chen et al., 2025). Without guidance, learners may overestimate readiness based on AI-generated feedback, revealing a perception–reality gap. This underscores the need for critical AI literacy and complementary human-mediated speaking practice in high-stakes assessment contexts.

4. Implications

The findings suggest that GenAI-mediated preparation is most pedagogically effective when explicitly aligned with the task demands of the IELTS Speaking Test. Research on IELTS speaking consistently underscores the role of task familiarity, strategic planning, and assessment awareness in shaping performance (Cohen, 2014; Fernandez, 2018).

From an instructional perspective, GenAI can be productively integrated as a pre-task planning resource, particularly for cognitively demanding tasks such as Parts 2 and 3. However, pedagogical guidance is essential to discourage memorisation of AI-generated responses, which risks privileging test-wiseness over independent language production (Cohen, 2014; He et al., 2024; Perrotta et al., 2025). Regarding pronunciation, ASR-based feedback should not be treated

as a proxy for IELTS assessment, as it does not fully capture the construct of intelligibility and communicative effectiveness (IELTS, 2025; Chen et al., 2025). Teachers therefore play a critical role in supporting learners' critical interpretation of AI feedback.

More broadly, as recent studies suggest, learners' prompting practices substantially shape the pedagogical value of GenAI use (Perrotta et al., 2025). Developing AI literacy for test-oriented speaking involves guiding learners to use GenAI as strategic scaffolding rather than a substitute for meaningful speaking practice.

Acknowledgments

The author gratefully acknowledges the supervisor, Ricky Jeffrey, for his continuous guidance and constructive support throughout this study. Without his invaluable mentorship, this work would not have been possible. Sincere thanks are also extended to all participants for contributing their time and insights.

Ethical Declaration Statement

This study complied with the Research Ethics Policy of the researcher's institution. Participants provided informed consent and were informed of their right to withdraw at any time. Confidentiality was ensured by using pseudonyms and removing identifying information. All data were securely stored and accessible only to the researcher. All data are deleted after the completion of the research.

Generative AI Policy

The author used Grammarly during the preparation of this work to ensure the accuracy of expression. After utilising the tool, the author thoroughly reviewed and edited the content as necessary and assumes full responsibility for its publication.

References

Allen, D. (2016). Investigating washback to the learner from the IELTS test in the Japanese tertiary context. *Language Testing in Asia*, 6(1). https://doi.org/10.1186/s40468-016-0030-z

Chen, X., Li, J., and Ye, Y. (2024). A feasibility study for the application of AI-generated conversations in pragmatic analysis. *Journal of Pragmatics*, 223, 14–30. https://doi.org/10.1016/j.pragma.2024.01.003

Cohen, A. D. (2014). *Strategies in learning and using a second language* (2nd ed.). London and New York: Routledge.

Fernandez, C. J. (2018). Behind a spoken performance: test takers' strategic reactions in a simulated part 3 of the IELTS speaking test. *Language Testing Asia*, 8(18), 1–20. https://doi.org/10.1186/s40468-018-0073-4

He, S., Sénécal, A.-M., Stansfield, L., and Suvorov, R. (2024). A scoping review of research on second language test preparation. *Language Testing*, 42(1), 11–47. https://doi.org/10.1177/02655322241249754 (Original work published 2025)

IELTS. (2025). *IELTS Academic format: Speaking*. IELTS. Retrieved from: https://ielts.org/take-a-test/test-types/ielts-academic-test/ielts-academic-format-speaking

IELTS. (2025). *IELTS scoring in detail*. IELTS. Retrieved from: https://ielts.org/take-a-test/your-results/ielts-scoring-in-detail

Jeon, J., Lee, S., and Choe, H. (2023). Beyond ChatGPT: A conceptual framework and systematic review of speech-recognition chatbots for language learning. *Computers and Education*, 206, Article 104898. https://doi.org/10.1016/j.compedu.2023.104898

Perrotta, C., Knoch, U., Selwyn, N., and Mohammadi, S. (2025). Exploring the possibilities of integrating communicative AI into the IELTS test preparation process. *IELTS Research Reports Online Series*, No. 2/25. British Council, IDP IELTS, and Cambridge University Press & Assessment. Available at: https://ielts.org/researchers/our-research/research-reports

Read, J. (2022). Test Review: The International English Language Testing System (IELTS). *Language Testing*, 39(4), 679–694. https://doi.org/10.1177/02655322221086211

Human-First Drafting and Scoped AI Feedback: An Integrated Workflow for Lower-Intermediate EFL Learners

Toshiyuki KANAMARU [1] & Hiroshi YOSHIZUKA[2]
[1] Institute for Liberal Arts and Sciences, Kyoto University, Japan.
[2] Institute for English Language Education, Seibido Publishing Co., Ltd., Japan.

Abstract

This paper presents a textbook-and-workflow model designed for elementary to lower-intermediate EFL learners. It integrates human-first drafting, scoped AI feedback via textbook-provided prompts, and process-oriented assessment through prompt-log excerpts and meta-reflection. Learners write an initial 150–200-word paragraph without AI assistance and request targeted feedback restricted to taught forms and organisational structures. They then revise the text, providing written reasons for their changes, before moving to oral production through read-aloud practice, transcription-based noticing, and a short follow-up conversation. We articulate a process-governance approach in which AI support is permitted only within the taught scope, AI-generated passages are not submitted as student work, and compact process artifacts support teacher review of revision decisions. The paper contributes design principles, a replicable workflow, and an assessable process design for integrating bounded AI support into ordinary EFL instruction. Educational implications, transferability, and next steps for pilot and multi-site evaluation are discussed.

Keywords: writing pedagogy, AI feedback, process assessment, speaking practice, learner autonomy

1. Introduction

The rapid proliferation of generative AI has shifted language education from the question of *whether* to integrate AI to the question of *how* to do so while preserving learner agency and instructional intent. As noted by Li et al. (2025) in their systematic review of empirical research from 2023 to 2024, the field has moved beyond exploratory inquiries to systematic investigations, confirming a maturing consensus that AI must be integrated with pedagogical rigor rather than mere novelty. For elementary to lower-intermediate writers, targeted support that clarifies organisation, unity, and taught language forms can lower cognitive load and reduce anxiety; however, unbounded AI use risks product substitution and stylistic inflation beyond the course scope.

We propose a writing-to-speaking framework that constrains AI to transparent roles. By requiring unaided drafting, scoped feedback, and meta-reflection, the approach maintains learner authorship while allowing instructors to verify decisions through auditable artifacts. In this approach, AI can be consulted during the process only within the taught scope, AI-generated passages are not submitted, and every assignment includes a short prompt-log excerpt and a meta-reflection. These design features are intended to keep authorship with the student while giving instructors compact and auditable evidence of decision-making. Technically, the approach relies on QR-linked prompts and familiar tools so that adoption may remain feasible across diverse contexts.

2. Methodology

This work is an educational design study whose objective is to develop a replicable instructional design and implementation guide for university EFL courses meeting weekly for ninety minutes across fifteen units. All design decisions were anchored to the target proficiency band of elementary to lower-intermediate learners, which informed model-text length, feedback scope, and vocabulary load. The study does not claim experimental effects or classroom implementation outcomes. Instead, it documents a curricular model and a governance approach that codify principled constraints on AI use and foreground artifacts that make the writing process visible and reviewable.

The contribution of the present paper therefore lies in its design rationale, workflow architecture, and evaluability through process evidence, while questions of feasibility, uptake, and effectiveness are reserved for subsequent pilot and multi-site research. At present, the framework has not yet undergone formal or informal classroom piloting and should therefore be understood as a design-based proposal rather than as evidence of implementation success.

The theme progresses from personal topics in early units to social topics in later units, with each unit culminating in a short essay that subsequently becomes material for oral production. The workflow consists of a human-first draft written without AI, followed by a round of feedback requested through QR-linked prompts that restrict the system to organisation, clarity, cohesion markers, and grammar points covered in class. Full rewrites and the introduction of new content are out of scope. Learners then revise and articulate the reasoning behind their choices in a concise meta-reflection. The revised text is transitioned to oral production through a speaking challenge that pairs read-aloud with speech transcription for pronunciation noticing and a short conversation that remains tied to the author's own ideas.

Each submission includes a prompt-log excerpt and the meta-reflection so that the instructor can sample processes rather than rely solely on product inspection. To support consistent uptake, the teachers' manual supplies model paragraphs aligned to taught forms, analytic rubrics for writing and speaking, compact templates for meta-reflection, and instructions for integrating oral production activities into regular lessons. The governance explanation clarifies that AI assistance is allowed only for micro-edits within the taught scope. Aligning with Mollick and Mollick's (2023) "Mentor" and "Coach" roles, the workflow prohibits the submission of AI-generated passages and incorporates oral checks intended to support ownership verification. The rationale is to centre evaluation on tangible and inspectable process evidence rather than on automated detection. For elementary to lower-intermediate learners, these boundaries keep language demands appropriate while still promoting noticing and reflective decision-making. Furthermore, by positioning AI feedback after human drafting, we leverage the "catalytic" potential identified by Xu et al. (2024), where AI interaction enhances foreign language enjoyment and subsequently boosts self-efficacy without bypassing the necessary cognitive struggle.

3. Design Outcomes and Discussion

The principal outcome of this design study is a comprehensive curricular framework paired with an implementation manual that specifies an end-to-end workflow from drafting to oral production for elementary to lower-intermediate university learners. The textbook comprises 15 units, each designed to fit a standard lesson. The workflow is characterised by a six-step sequence that governs where and how AI may be used within the instructional process. As detailed in Table 1, the architecture explicitly separates learner cognitive investment from AI-assisted feedback. Steps 2 and 3 prioritise human agency ("human-first"), while Steps 4 and 5 position AI in the roles of "Mentor" and "Coach."

In this section, "outcomes" refers to design outputs and pedagogical architecture rather than measured learner gains. Each unit follows an identical pedagogical arc so that students encounter a predictable sequence that gradually raises cognitive and linguistic demands while keeping authorship and process at the centre of assessment. Units open by activating background knowledge and modelling target discourse structures, allowing learners to observe how ideas are staged, supported, and closed. Immediately adjacent to the model, the textbook highlights reusable expressions as scaffolds for noticing and controlled imitation. These expressions remain within the taught scope but are combinable into longer stretches of argument and exemplification, which allows students to extend what they have already learned without stylistic inflation. This design choice is intended to maintain authorial integrity while preparing learners for the subsequent phase.

Table 1: Unit Workflow and AI Governance Protocol

Step	Activity	AI Permission	Theoretical Rationale
1. Input & Modeling	Analysis of model texts and confirmation of target expressions.	Prohibited	Schema formation through explicit instruction. Learners understand the target discourse structure without AI dependence.
2. Focused Thinking	Development of an outline (Situation, Position, Reasons).	Prohibited	Preservation of Agency. Ideation relies solely on the learner's cognitive resources to prevent the outsourcing of critical thinking.
3. Human-first Drafting	Drafting a 150–200-word paragraph.	Prohibited	Creation of the "substrate" for the catalytic effect (Xu et al., 2024); direct training of writing skills (Li et al., 2025).
4. Scoped AI Feedback	Requesting feedback via QR-linked prompts.	Permitted within scope	"Mentor" Role (Mollick & Mollick, 2023). Feedback is strictly scoped to organization and clarity; rewriting is disabled.
5. Revision & Reflection	Revising based on feedback and recording reasons.	Used for revision decisions	"Coach" Role (Mollick & Mollick, 2023). Activation of meta-cognition by requiring learners to articulate the rationale for accepted or rejected changes.
6. Oral Production	Read-aloud, Q&A, and dialogue using the revised text.	Not central to the step	"Shy to Fly" (Zhang et al., 2024). Leveraging the revised text as a scaffold to bridge the gap to confident oral communication.

The transition from reading and noticing into controlled production is managed through a focused thinking section that externalises the writer's decisions. Learners are guided to state the situation, articulate a position, select reasons and examples, signal a turn toward challenges or counter-considerations, and return to proposals and closure. Grammar work is embedded within this rhetorical progression rather than isolated from it. For instance, modal verbs that convey obligation, possibility, and feasibility are taught in context so that sentence-level choices reinforce paragraph-level moves and communicative intent. This integration keeps the focus on meaning while still providing clear

handles for language control appropriate to elementary to lower-intermediate proficiency.

Following independent writing, learners request scoped feedback via QR codes. Revisions are documented in a meta-reflection explaining accepted or rejected suggestions. The revised text then transitions to oral production. This transition is grounded in recent findings by Zhang et al. (2024), who demonstrated that interactive AI tools can significantly enhance EFL learners' Willingness to Communicate (WTC) by reducing anxiety compared to rigid evaluative systems. Leveraging this "from shy to fly" effect, students read it aloud with speech transcription and engage in a short dialogue seeded by their own ideas. At submission, students include a brief excerpt of the prompt-log together with the meta-reflection, which creates an auditable trail of decisions that links product and process.

This architecture has potential implications for assessment validity and instructional efficiency. Because drafting is evidenced through artifacts, evaluation can focus on how students navigated revisions and on whether their decisions reflect taught principles, rather than relying solely on surface features of the final product. When questions of ownership arise, short oral explanations of what changed and why provide a communicative check aligned with course goals. From a workload perspective, instructors can review a sample of logs and reflections and redirect time from line-by-line editing to higher-order coaching on claim clarity, support, and example fit, and discourse signalling. The unit-level rhythm that moves from warm-up to model and reusable expressions, then to planning, grammar in use, independent writing, scoped feedback, reasoned revision, and oral production, facilitates consistency across sections and instructors and makes programme-level quality assurance more tractable.

Finally, the design choices intentionally support portability beyond the original context. Because reusable expressions, grammar foci, and discourse patterns are taught as functions tied to communicative moves, they can be localised to other languages with adjustments to model texts, prompt wording, and target features such as agreement, aspect, word order, or cohesion devices. The same process evidence and integrity checks can be retained. The result is a coherent and repeatable workflow design that integrates AI in a bounded and auditable manner and that is intended to connect written development with oral production across the textbook as a whole.

4. Implications

The resulting framework is presented as a proposed curricular package

showing how human-first drafting, scoped feedback, and oral transitions may be integrated into ordinary classes without requiring new platforms. The textbook design aims to ensure that AI is used during the process only within the taught scope, that AI-generated passages are not submitted as student work, and that compact process artifacts such as prompt-log excerpts and meta-reflections keep evaluation focused on revision decisions rather than on black-box detection.

Although the present materials were designed for Japanese universities, the underlying production concept is intended to be language-agnostic and adaptable. The sequence of draft, bounded feedback, justified revision, and transition to oral production may map onto communicative goals across instructional settings, and the audit-ready artifacts may also be applicable to programmes with different scripts, typologies, or proficiency bands. Adapting the workflow to other languages would primarily involve localising model texts, feedback prompts, and taught features while retaining the same evidence structures and integrity checks.

For theory, the model offers one way of operationalising scaffolded development by placing AI after learner-generated text and before oral production, thereby encouraging noticing of organisation and cohesion while maintaining an appropriate level of challenge for novice writers. For practice, the framework offers a set of portable design principles and materials that could be localised with relatively low overhead and adapted to mobile-first or low-resource contexts. Future work should examine differential uptake by proficiency and genre and test the contribution of written and oral production tasks in broader contexts.

Several limitations should be noted. First, the present paper does not report controlled evidence of learning gains, and any claims regarding effectiveness must therefore remain tentative. Second, the feasibility of the workflow, including student compliance, prompt-log quality, and the classroom manageability of the writing-to-speaking transition, remains to be examined through pilot implementation. Third, although the framework is designed for adaptation across contexts, its transferability beyond the original instructional setting remains a proposition to be tested empirically.

Acknowledgments
This work was supported by JSPS KAKENHI Grant Number JP25K00482.

Generative AI Policy Statement

The authors used ChatGPT (GPT-5.4 Thinking) and Gemini 3 Flash (Paid Tier) to assist with language polishing during manuscript preparation. After utilising the tool, the authors thoroughly reviewed and edited the content and assume full responsibility for the publication's accuracy and originality.

References

Li, B., Tan, Y. L., Wang, C., & Lowell, V. (2025). Two years of innovation: A systematic review of empirical generative AI research in language learning and teaching. *Computers and Education: Artificial Intelligence, 9*, 100445. https://doi.org/10.1016/j.caeai.2025.100445

Mollick, E. R., & Mollick, L. (2023). Assigning AI: Seven approaches for students, with prompts. *The Wharton School Research Paper*. https://doi.org/10.2139/ssrn.4475995

Xu, S., Chen, P., & Zhang, G. (2024). Exploring the impact of the use of ChatGPT on foreign language self-efficacy among Chinese students studying abroad: The mediating role of foreign language enjoyment. *Heliyon, 10*(21), e39845. https://doi.org/10.1016/j.heliyon.2024.e39845

Zhang, D., Wu, J. G., & Fu, Z. (2024). From shy to fly: Facilitating EFL learners' willingness to communicate with an AI chatbot and an intelligent tutoring system. *System, 127*, 103501. https://doi.org/10.1016/j.system.2025.103501

Understanding Students' Perspectives on Using GenAI for Their Coursework in Science Higher Education

Dawn LO [1], Kim Hung LAM [2], Leo CHON [1], Xiaofeng HUANG [2]
[1] Educational Development Centre, The Hong Kong Polytechnic University, Hong Kong, China
[2] Department of Applied Biology and Chemical Technology, The Hong Kong Polytechnic University, Hong Kong, China

Abstract

The accelerated expansion of Generative AI (GenAI) is transforming education. While it offers notable teaching and learning benefits, concerns regarding academic integrity and plagiarism remain. The Hong Kong Polytechnic University (PolyU) has been promoting the integration of GenAI into its learning, teaching, and assessment strategy. To examine challenges and opportunities, the authors conducted a pilot study with 31 science students from the Department of Applied Biology and Chemical Technology (ABCT), exploring their perspectives and GenAI usage in coursework. About half used GenAI mainly for brainstorming, summarising, and information searching, while multimedia creation was less common (26%). Most students agreed on responsible practices, particularly tool selection and integrity, though only 45% recognised privacy risks. Key benefits included time-saving (M = 4.05) and obtaining insights (M = 3.85). Challenges such as unreliable outputs and ethical concerns highlight the need for AI literacy training and clear academic integrity guidelines.

Keywords: GenAI tools, Science education, Students' perception, Assessments

1. Introduction

The rapid development of generative artificial intelligence (GenAI) tools, such as ChatGPT, Gemini, Copilot, DeepSeek, and others, has significantly influenced higher education, particularly in assessment practices. Research studies explore the potential benefits of GenAI in higher education assessments. For instance, Alier et al. (2024) and Zaibout and Madrane (2024) suggested that AI can enhance assessment through personalised testing, instant feedback, and automated grading, supporting autonomous learning and reducing subjectivity. At the same time, worldwide educators are concerned about increased plagiarism and academic integrity problems due to the easily accessible GenAI tools (Alier et al., 2024).

Recent studies reveal diverse patterns of student adoption and perceptions of GenAI in academic contexts. For instance, Stranges and MacNutt (2025) surveyed 310 undergraduate health science students at UBC Okanagan, where GenAI was permitted to be used for reflective writing tasks. They found that only 33% of the respondents used GenAI for writing. Common reported GenAI usage included initiating drafts (36%) and revising existing drafts (37%), with perceived benefits such as improved efficiency (54%) and grade enhancement (38%). In contrast, Freeman (2025) reported widespread adoption among UK undergraduates, with 92% using GenAI for assessments and 18% incorporating AI-generated text. Similarly, Arbab et al. (2024) surveyed 145 students and found that 76% used AI primarily for idea generation, while nearly half considered it acceptable for AI to write. These findings highlight both the growing reliance on GenAI and the lack of clear institutional regulations governing its use.

The integration of GenAI into assessment has raised concerns about academic integrity. Moorhouse et al. (2023) emphasised the need for process-focused, staged assessments that promote critical thinking and AI literacy. Lodge et al. (2023) advocated for authentic engagement with AI through ethical practices such as prompt creation and reflective analysis. Chan and Colloton (2024) proposed metacognitive and self-regulated models that prioritise feedback and reflection over grades to address integrity challenges. Empirical studies further illustrate varied impacts of GenAI on learning outcomes. Smerdon (2024) found that 46 out of 69 students used AI for topic generation in a research proposal assignment, and the study showed no significant difference in scores compared to a control group. Conversely, Alimardani (2025) reported that law students often inserted fabricated citations and misleading facts when using GenAI, and recommended training to reduce integrity challenges.

According to many educational researches, redesigning assessments is necessary to safeguard academic integrity. Khlaif et al. (2025) introduced the "Against, Avoid, Adopt, and Explore" (AAAE) model, encouraging educators to move beyond traditional assessments and leverage AI for brainstorming, idea generation, and collaborative problem-solving. Similarly, Xia et al. (2024) warned that GenAI increases risks of cheating and high-scoring submissions, and they addressed the need for robust assessment strategies. Beyond assessment design, concerns extend to AI reliability and ethics. Khlaif et al. (2023) noted issues with source verification and authorship when using ChatGPT for research. Nikiforova and Georgiev (2024) found AI-text detectors unreliable for certain languages and hybrid content, and highlighted that face-to-face exams to be the most robust method to safeguard academic integrity. Ateeq et al. (2024) argued for a shift towards application-based, project-driven assessments

to reduce cheating risks, while Smolansky et al. (2023) and Gamage et al. (2023) highlighted the need for policies, training, and ethical guidelines to integrate GenAI responsibly.

In brief, these studies underscore the dual nature of GenAI in higher education; it is offering opportunities for innovation in education while posing challenges to academic integrity. Understanding students' perceptions and usage patterns is therefore critical for developing effective policies, alternative assessments, and pedagogical strategies that balance technological enhancement in education while maintaining ethical and educational goals.

To address these challenges, the first author collaborated with the University's Educational Development Centre (EDC) to conduct action research to investigate university students' perspectives and their usage of GenAI for their coursework. This paper will report our pilot study on a cohort of 31 Analytical Sciences for Testing and Certification (ASTC) students from the Department of Applied Biology and Chemical Technology (ABCT) and their use of GenAI tools for their assessments.

2. Methodology

The current study adopted a student survey developed by a larger project led by EDC. Informed by prior studies on university students' usage of Generative AI in higher education (e.g., Almassaad et al., 2024; Chan & Zhou, 2023; Wang et al., 2025), the survey collected students' views and experiences on using Generative AI (GenAI) in learning and completing assessment tasks in a science subject. It explored students' usage of GenAI in assessment, their perceived benefits, learning outcomes, and challenges, and their support needs. The study focused on the cohort subject ABCT4631 *Test Method, Measurement and Uncertainty*, which is a major core course for the Bachelor of Science in Analytical Sciences for Testing and Certification offered by the Department of Applied Biology and Chemical Technology. For the 2024/25 Semester 2, a total of 31 students enrolled in the subject. Student feedback was collected via an online survey using MS Forms.

2.1 Data Collection and Analysis

The student was set up online in MS Forms. The online survey was administered after students had completed their assessment tasks in the subject. Students were invited to answer the survey during a class session on 11 April 2025. Participation was voluntary. Data were cleaned and analysed using descriptive statistics (counts, percentages, means, and standard deviations).

2.2 Participants

A total of 20 undergraduate students from the same department participated in the survey, representing a response rate of 64.5% (20 out of 31 students enrolled in the subject). Table 1 shows the demographic characteristics of the participants. Most of them (95%, n=19) reported using GenAI tools in completing their assessment tasks.

Table 1. Demographic Characteristics of Participants

Characteristic	Full sample	
	n	%
Gender		
Female	7	35.0%
Male	13	65.0%
Year of Study		
3	19	95.0%
4 or above	1	5.0%
Use of GenAI		
Yes	19	95.0%
No	1	5.0%

Note: N = 20. Participants were on average 20.85 years old (SD = 1.09).

3. Results and Discussion

Frequency of Usage

Students used GenAI for brainstorming ideas (M = 3.53) and searching for information (M = 3.58) most frequently, with over 47–53% reporting "Often" or "Very often" (Table 2). Summarising information was also common (47%). Less frequent usage included multimedia creation (26%) and collaboration (26%).

Table 2. Frequency of Using GenAI Tools in Completing the Assessment Tasks (N = 19)

	Mean (M)	SD	% of 4 & 5
For brainstorming and generating ideas	3.53	0.84	52.6%
For searching information or literature	3.58	1.07	47.4%
For summarizing information	3.47	0.84	47.4%
For writing, proofreading and editing	3.00	1.05	36.8%
For translating	2.89	1.45	31.6%
For creating digital multimedia and presentations	2.74	1.24	26.3%
For analyzing data/ solving numerical problems/ coding/ programming	3.05	1.03	31.6%

For providing feedback to my work-in-progress	2.84	1.30	31.6%
For facilitating project collaboration	3.00	1.05	26.3%

Note: Question: How often did you use the GenAI tools to help you complete the assessment task(s)? Scale: 1=Never; 2 = Seldom; 3 = Sometimes; 4 = Often; 5 = Very often.

Responsible Use

Students generally agreed they have used GenAI responsibly, with the overall mean ≈ above the mid-point, 3, in all the related items (Table 3). In particular, selecting suitable tools and upholding academic integrity got the highest mean ratings (both M = 3.79). Additionally, more than half of the students agreed or strongly agreed that they cross-checked information, questioned underlying assumptions and biases, and chose appropriate solutions from GenAI outputs. Only about 42.1% of respondents were aware of privacy and information security issues when using GenAI tools. Our students have also informally shared that GenAI tools may not be reliable for calculations in this subject due to inconsistencies in the answers provided.

Table 3. Responsible Use of GenAI Tools (N = 19)

	Mean (M)	SD	% of 4 & 5
I can select the most suitable GenAI tools for various tasks.	3.79	0.71	63.2%
I always uphold academic integrity/honesty when using GenAI tools.	3.79	0.98	52.6%
I cross-check information provided by GenAI tools with other sources.	3.68	0.75	52.6%
I question the assumptions and biases underlying the information provided by GenAI tools.	3.63	0.76	57.9%
I can choose the appropriate solution from the various solutions provided by GenAI tools.	3.63	0.83	52.6%
I can evaluate the functionality and limitations of GenAI tools or products.	3.63	0.76	47.4%
I compare multiple sources of information before forming an opinion based on GenAI recommendations.	3.53	0.90	47.4%
I am always aware of the misuse of GenAI technology.	3.47	0.90	42.1%
I am always aware of privacy and information security issues when using GenAI tools.	3.42	0.84	42.1%

Note: Question: To what extent do you agree with the following statements when using GenAI tools to complete the assessment task(s)? Scale: 1 = Strongly Disagree, 2 = Disagree; 3 = Neutral; 4 = Agree; 5 = Strongly Agree.

Perceived Benefits and Learning Outcomes

Students also rated the extent to which GenAI tools have benefited them and their concern or difficulty when using GenAI tools in the assessment task(s), as summarised below:

Perceived benefits and learning outcomes: The highest-rated benefit was saving time finding information (M = 4.05; 70%), followed by obtaining insights (M = 3.85; 60%) and building confidence in preparing the assessment (M = 3.85; 55%) (Table 4). Other areas, such as reducing stress and clarifying expectations, were also positively rated (55%). Students reported moderate gains in problem-solving (M = 3.65; 50%) and judging GenAI output quality (55%) (Table 4). Improvements in creativity (25%) and communication skills (30%) were the least agreed upon. Only 30% of respondents reported that AI motivated their learning.

Table 4. Perceived Benefits (N = 20)

	Mean (M)	SD	% of 4 & 5
Helping me save time in finding information	4.05	0.94	70.0%
Obtaining new and useful insights and perspectives	3.85	0.81	60.0%
Becoming more confident in preparing the assessment task(s)	3.85	0.88	55.0%
Helping me save time in producing the work (e.g. reports, videos, graphics, powerpoint slides)	3.80	1.06	55.0%
Obtaining immediate and personalised feedback for improving my assessment task(s)	3.80	0.95	55.0%
Understanding the expectations and requirements of the assessment task(s)	3.75	0.91	55.0%
Becoming less stressed in preparing the assessment task(s)	3.75	0.91	55.0%
Preparing for what to include and how to complete my assessment task(s)	3.65	0.81	55.0%

Note: Question: To what extent do you agree that GenAI tools have assisted you in completing the assessment task(s) in the following ways? Scale: 1 = Strongly Disagree, 2 = Disagree; 3 = Neutral; 4 = Agree; 5 = Strongly Agree; Not applicable = have not used GenAI in the specific way.

Table 5. Perceived Learning Outcomes (N = 20)

The GenAI-expected assessment task(s)…	Mean (M)	SD	% of 4 & 5
enhanced my problem-solving skills	3.65	0.75	50.0%
enhanced my skills in judging the quality of GenAI output	3.60	0.75	55.0%

enhanced my understanding of subject knowledge and theories	3.55	0.76	50.0%
enhanced my critical thinking skills	3.50	0.83	40.0%
enhanced my ability in using GenAI tools (Al literacy/ competence)	3.50	0.83	40.0%
motivated me to learn this subject	3.45	0.76	30.0%
enhanced my communication skills	3.40	0.68	30.0%
enhanced my creativity	3.35	0.67	25.0%

Note: Question: To what extent do you agree the GenAI-expected assessment task(s) helped you in the following ways? Scale: 1 = Strongly Disagree, 2 = Disagree; 3 = Neutral; 4 = Agree; 5 = Strongly Agree.

Challenges and Support Needed

Top concerns from students were unreliable information from GenAI tools (M = 3.70) with 55% agreement (Table 6). Only 20% of students agreed/strongly agreed that they were not familiar with the GenAI tools. Other concerns involved ethical issues, academic integrity, and data privacy and security ($M \geq 3.4$; 45%). 40% of respondents reported over-relying on GenAI tools for completing their assessments. Most students requested training (60%) and guidelines/examples (55%), while 40% wanted additional tool access. Half of the students preferred using GenAI for assessments.

Table 6. Concerns/Challenges when using GenAI Tools (N = 20)

	Mean (M)	SD	% of 4 & 5
Unreliable/Inaccurate information/hallucinations	3.70	0.98	55.0%
Limited access to the GenAI tools	3.70	0.80	50.0%
Subscription fee for accessing advanced features	3.65	0.88	50.0%
Over relying on GenAI tools	3.45	1.05	40.0%
Privacy and data security	3.45	1.00	45.0%
More time to complete the assessment task	3.40	0.88	35.0%
Ethical concerns/academic integrity	3.40	0.88	45.0%
Lack of prompting skills	3.35	0.75	40.0%
Worries about being falsely accused of committing plagiarism	3.35	0.88	40.0%
Not familiar with the GenAI tool(s) (required by your teacher(s))	3.20	0.62	20.0%

Note: Question: To what extent do you agree the following is a concern or difficulty when using GenAI tools in the assessment task(s)? Scale: 1 = Strongly Disagree, 2 = Disagree; 3 = Neutral; 4 = Agree; 5 = Strongly Agree.

4. Implications

The rapid development of GenAI and students' over-reliance on the tools pose threats to academic integrity and plagiarism. In our preliminary study, about half of the respondents used GenAI mainly for brainstorming, summarising, and searching for information, while fewer students reported using it for data analysis, numerical problems, multi-media generation, or collaboration. This pattern suggests that Science students may strategically use GenAI tools as a supportive tool based on task demands, rather than relying on it to complete assessment tasks. Arguably, this reflects disciplinary norms in Science education, where numerical accuracy, methodological rigour, and verification are highly valued.

We found out that most students agreed/strongly agreed (52.6%) on responsible practices such as upholding academic integrity, cross-checking information produced by GenAI, but only 45% recognised the issues of data privacy and security. Challenges such as unreliable outputs and ethical concerns highlight the need for AI literacy training and clear academic integrity guidelines. Teaching strategies that explicitly model effective prompting, source verification, and critical evaluation of GenAI output may help students to use GenAI tools more reflectively and meaningfully. Incorporating scaffolded GenAI activities into coursework, such as comparing AI-generated explanations with authoritative scientific sources like peer-reviewed journals, can further support students in developing their scientific reasoning and evaluative judgement.

Owing to the small sample size, further investigation in the future will be conducted to confirm our findings. Although the small sample size limits the generalisability of its findings, this study offers a valuable starting point for rethinking how GenAI can be more intentionally integrated into Science education in ways that enhance learning while safeguarding academic and ethical standards in the age of GenAI.

Acknowledgments

This study was part of a larger project titled, *Exemplary learning, teaching and assessment practices in UGC-funded institutions in the age of GenAI*, supported by the Fund for Innovative Technology-in-Education (FITE) of the University Grants Committee (UGC). We gratefully acknowledge this support, as well as the assistance provided by the Educational Development Centre and colleagues for their assistance in data collection and analysis, and for the valuable feedback on the survey design. The authors would also like to thank the FLTI seed fund project "Developing a Flipped Learning Signature Showcase" (Ref. no.: Invite-FS-ABCT) for ABCT to support this study.

Ethical Declaration Statement

Ethical approval was obtained from the Institutional Review Board (IRB) of The Hong Kong Polytechnic University (Reference Number: HSEARS20240522003).

Generative AI Policy

The author(s) used MS Co-pilot during the preparation of this work to enhance the clarity of the text and streamline the drafting process. 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

Alier, M., García-Peñalvo, F. J., & Camba, J. D. (2024). Generative artificial intelligence in education: From deceptive to disruptive. *International Journal of Interactive Multimedia and Artificial Intelligence*, 8(5), 5-14. https://doi.org/10.9781/ijimai.2024.02.011

Alimardani, A. (2025). Borderline disaster: An empirical study on student usage of GenAI in a law assignment. *IEEE Transactions on Technology and Society*, 6(2), 210–219. https://doi.org/10.1109/TTS.2025.3540978

Almassaad, A., Alajlan, H., & Alebaikan, R. (2024). Student Perceptions of Generative Artificial Intelligence: Investigating Utilization, Benefits, and Challenges in Higher Education. *Systems*, 12(10), 385. https://doi.org/10.3390/systems12100385

Arbab, A. N., Dhuhli, B. A., Krishnan, Y., & Crisostomo, A. S. (2024). Student's utilization and assistance of AI tools in assessment completion: Perceptions and implications. *International Linguistics Research*, 7(3), 1. https://doi.org/10.30560/ilr.v7n3p1

Ateeq, A., Alzoraiki, M., Milhem, M., & Ateeq, R. A. (2024). Artificial intelligence in education: Implications for academic integrity and the shift toward holistic assessment. *Frontiers in Education*, 9, Article 1470979. https://doi.org/10.3389/feduc.2024.1470979

Chan, C. K. Y., & Colloton, T. (2024). Redesigning assessment in the AI era. In *Generative AI in higher education: The ChatGPT effect* (pp. 87–126). Routledge. https://doi.org/10.4324/9781003459026-4

Chan, C. K. Y., & Zhou, W. (2023). An expectancy value theory (EVT) based instrument for measuring student perceptions of generative AI. *Smart Learning Environments*, 10(1), 64. https://doi.org/10.1186/s40561-023-00284-4

Freeman, J. (2025). HEPI student generative AI survey. Higher Education Policy Institute. https://www.hepi.ac.uk/wp-content/uploads/2025/02/HEPI-Kortext-Student-Generative-AI-Survey-2025.pdf

Gamage, K. A. A., Dehideniya, S. C. P., Xu, Z., & Tang, X. (2023). ChatGPT and higher education assessments: More opportunities than concerns? *Journal of Applied Learning & Teaching*, 6(2), 358–369. https://doi.org/10.37074/jalt.2023.6.2.32

Khlaif, Z. N., Mousa, A., Hattab, M. K., Itmazi, J., Hassan, A. A., Sanmugam, M., & Ayyoub, A. (2023). The potential and concerns of using AI in scientific research: ChatGPT performance evaluation. *JMIR Medical Education*, 9(1), Article e47049. https://doi.org/10.2196/47049

Khlaif, Z. N., Alkouk, W. A., Salama, N., & Abu Eideh, B. (2025). Redesigning assessments for AI-enhanced learning: A framework for educators in the generative AI era. *Education Sciences*, 15(2), 174. https://doi.org/10.3390/educsci15020174

Lodge, J., Howard, S., Bearman, M., Dawson, P., & Associates. (2023). Assessment reform for the age of artificial intelligence. Tertiary Education Quality and Standards Agency (TEQSA). https://www.teqsa.gov.au/sites/default/files/2023-09/assessment-reform-age-artificial-intelligence-discussion-paper.pdf

Moorhouse, B. L., Yeo, M. A., & Wan, Y. (2023). Generative AI tools and assessment: Guidelines of the world's top-ranking universities. *Computers and Education Open*, 5, 100151. https://doi.org/10.1016/j.caeo.2023.100151

Nikiforova-Ilieva, K., & Georgiev, T. (2024). AI-generated text detection and prevention of unethical AI use in student exams. In *2024 5th International Conference on Communications, Information, Electronic and Energy Systems (CIEES)* (pp. 1–4). IEEE. https://doi.org/10.1109/CIEES62939.2024.10811427

Smerdon, D. (2024). AI in essay-based assessment: Student adoption, usage, and performance. *Computers and Education: Artificial Intelligence*, 7, Article 100288. https://doi.org/10.1016/j.caeai.2024.100288

Smolansky, A., Cram, A., Raduescu, C., Zeivots, S., Huber, E., & Kizilcec, R. F. (2023). Educator and student perspectives on the impact of generative AI on assessments in higher education. In ACM (Ed.), *L@S 2023 - Proceedings of the 10th ACM Conference on Learning @ Scale* (pp. 378–382). ACM. https://doi.org/10.1145/3573051.3596191

Stranges, T. N., & MacNutt, M. J. (2025). Reflective writing assignments in the era of GenAI: Student behavior and attitudes suggest utility, not futility. *Advances in Physiology Education*, 49(2), 582–592. https://doi.org/10.1152/advan.00241.2024

Wang, C., Wang, H., Li, Y., Dai, J., Gu, X., & Yu, T. (2025). Factors Influencing University Students' Behavioral Intention to Use Generative Artificial Intelligence: Integrating the Theory of Planned Behavior and AI Literacy. *In-*

ternational Journal of Human–Computer Interaction, 41(11), 6649–6671. https://doi.org/10.1080/10447318.2024.2383033

Xia, Q., Weng, X., Ouyang, F., Lin, T. J., & Chiu, T. K. (2024). A scoping review on how generative artificial intelligence transforms assessment in higher education. *International Journal of Educational Technology in Higher Education*, 21(1), 40. https://doi.org/10.1186/s41239-024-00468-z

Comparing Generative AI and Human Grading of Undergraduate Students' Personal Development Plans

Man Kin LEUNG[1], Sui Leung Walter FUNG[2], Weihang LIAO[3], Yagiz TORCUK[4]

[1] Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China

[2] Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China

[3] Department of Data Science & Artificial Intelligence, The Hong Kong Polytechnic University, Hong Kong, China

[4] Department of Mathematics, Technische Universität Dresden, Dresden, Germany

Abstract

The use of Generative AI (GenAI) in assignment grading has brought convenience to teachers but remains controversial due to concerns regarding its accuracy and consistency in providing nuanced grades and feedback. This study compares the scores of Personal Development Plans from 50 students in a Freshman Seminar course, with each plan graded by three human graders and three AI graders. Through quantitative and qualitative analyses, we examine the grading patterns of different GenAI models and evaluate the feasibility of AI-assisted grading.

Keywords: GenAI, assessment, automated grading, reflective writing

1. Introduction

Generative Artificial Intelligence (GenAI), based on Large Language Models (LLMs), s hows potential in various aspects of education, including automated assessment (Chan & Hu, 2023). With comprehensive prompts, GenAI can evaluate students' written assessments efficiently, providing detailed scores and feedback. However, studies report inconsistencies and inaccuracies in GenAI-assisted grading compared to human graders (Jauhiainen & Garagorry Guerra, 2024; Usher, 2025). Considering these potential benefits and limitations, the authors have developed an AI grading system to assist in the grading of written assessments. The results from human graders and AI graders based on different LLM models are evaluated and compared in terms of consistency and similarity.

The objectives of this study are addressed through the following research questions:

- **RQ1:** To what extent do grade distributions by AI graders align with human graders?
- **RQ2:** What are the differences in grading styles between AI graders and human graders?

2. Methodology

This study is based on a Freshman Seminar course offered by the Faculty of Computer and Mathematical Sciences at The Hong Kong Polytechnic University. Students were required to submit Personal Development Plans reflecting their motivation for study, inspirations from the seminars, and career goals. Grading criteria were defined on five aspects with a total of 40 marks:

- **C1:** Self-assessment and reflection (10)
- **C2:** Clarity and specificity (10)
- **C3:** Addressing societal needs (5)
- **C4:** Actionable plan (10)
- **C5:** Structure and writing skills (5)

An AI grading system had been developed prior to its offering in Fall 2025. To provide a standard for the system, these criteria were further elaborated into rubrics describing the quality of work required to attain different points in each component. The system was also prompted to provide feedback under each component for each student, followed by a conclusive comment.

A convenience sampling of 50 submissions was selected by chronological order of submission, balancing the ratio of students from three departments. Each submission was graded independently by three teachers from different departments (the first three authors). It was also graded by three AI graders: **M1** (GPT-4.1-mini), **M2** (Grok-4-Fast), and **M3** (Claude-Haiku-4.5), deployed on Poe. Apart from the system prompt, the course introduction document and rubrics of the assignment were uploaded as a knowledge base for the models.

Studies in the literature have verified that the consistency of grading using the same model is high upon multiple trials (Jauhiainen & Garagorry Guerra, 2024). However, inconsistency might be significant across different AI models (Wetzler et al., 2025). In this study, the distance between score distributions of the graders was analysed using Kullback-Leibler divergence (Belov & Armstrong, 2011). By comparing the results to those of human grading, the submissions with the greatest difference between the medians of human graders and AI graders were selected as cases for qualitative analysis to provide in-depth

explanation for the difference, based on the detailed feedback provided by the AI grading system.

3. Results and Discussion

We first present the quantitative analysis of the overall score distribution (3.1), followed by a qualitative analysis of a few cases with high inconsistency (3.3).

3.1 Descriptive results of the overall grading

The descriptive statistics for the scores assigned by three human graders (H1 to H3) and three AI graders (M1 to M3) are presented in Table 1. In general, human graders assign higher mean scores (30.22 to 31.76) compared to the AI graders (26.80 to 29.78). Notably, human graders exhibit lower score variability than AI graders, as indicated by the standard deviations. The AI graders, particularly M2 and M3, utilise a broader scoring range, assigning lower minimum scores (14) than human graders (21). Additionally, the descriptive figure suggests greater consistency among human graders. Their medians (30.5, 32, 31) and 75% quartiles (34, 34, 33.75) are closely clustered. In contrast, AI graders exhibit larger differences in their central tendencies and spreads. This suggests that AI graders have a stricter grading tendency and less consensus among themselves compared to human graders.

Table 1: Descriptive Statistics for all Graders (out of 40 marks)

Grader	Mean	Std	Min	25%	50%	75%	Max
H1	30.22	4.82	21	26.25	30.5	34	38
H2	31.76	3.16	24	31	32	34	38
H3	30.62	3.87	21	28	31	33.75	38
M1	29.78	4.25	18	27	30	33	37
M2	27.34	5.31	14	24	28	30.75	37
M3	26.80	5.83	14	23	26	31.75	38

3.2 Consistency and comparison among graders

To further investigate the consistency among graders, we analyse the divergence between score distributions and the confidence intervals for the population mean. The symmetric Kullback-Leibler divergence matrix quantifies the dissimilarity between score distributions of each pair of graders, presented in Table 2. A larger number indicates more dissimilarity, while 0 indicates an identical distribution. The divergence values among human graders are relatively low (all smaller than 1), confirming their scoring alignment. Conversely, higher divergence is observed between certain AI graders (e.g., M1-M3: 1.39) and between some hu-

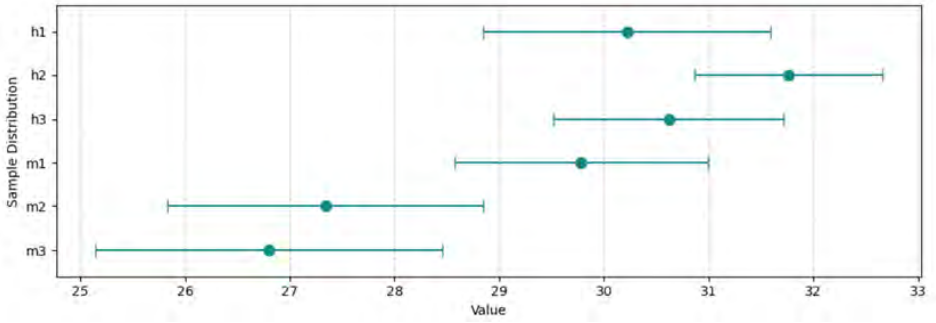
man-AI pairs (e.g., H2-M3: 1.52), indicating significant differences in their grading patterns. Interestingly, the distribution of M1 is closer to H1-H3 rather than M2 and M3, suggesting a similar grading profile with human graders.

Table 2: Symmetric KL Divergence Matrix of Results by all Graders

	H1	H2	H3	M1	M2	M3
H1	0	0.87	0.15	0.69	0.13	0.23
H2		0	0.29	0.21	1.22	1.52
H3			0	0.24	0.24	0.42
M1				0	0.95	1.39
M2					0	0.10
M3						0

The 95% confidence intervals for the estimated population mean are shown in Figure 1. The intervals for human graders show substantial overlap, implying they are likely estimating a similar mean score for the student population. The intervals for the AI models are more dispersed. Notably, the CIs for M2 and M3 are wider, reflecting their higher score variance, and they do not overlap with the CI for H2. This suggests a statistically significant difference between the mean scores assigned by these AIs and the most generous human grader.

(Figure 1: 95% Confidence Intervals for the Population Mean)



3.3 Cases with significant differences between human and AI grading

For each submission, the score given by each AI grader is compared to the median of scores by human graders. This difference is regarded as significant if it exceeds 4 out of 40 marks. The frequency of such cases is shown in Table 3. Three cases with the greatest difference are selected for case study. M2 was instructed to regrade and provide detailed explanation for the scores of each component. The case study is based on M2's feedback, reviewed by a human.

Table 3: Submissions with Significant Difference between AI Graders and Median of Human Graders

AI Grader	Significantly above		Significantly below		Average difference	
	n	%	*n*	%	-*m*	%
M1	5	10	14	28	-1.24	-3.1
M2	1	2	23	46	-3.68	-9.2
M3	1	2	26	52	-4.22	-10.6

- **Case 1:** (H1: 28; H2: 26; H3: 27; M1: 18; M2: 16; M3: 17)
  M2 gives fair scores for C1, C2, and C3 but low scores in C4 and C5. It explains that the essay lacks a clear actionable plan and is below the word limit. Human graders see that the student has put the actionable plan throughout different parts of the essay but not on a specific paragraph. Also, the word count barely reaches the minimum word limit of 500. Without titles and subtitles, it would be under 500 words.

- **Case 2:** (H1: 28; H2: 30; H3: 34; M1: 26; M2: 18; M3: 19)
  M2 gives fair to good scores for C1, C3, and C5 but low scores for C2 and C4. It explains that although the student has listed a few goals with a realistic actionable plan, they are too broad and generic, lacking clarity, prioritisation, and measurable parameters.

- **Case 3:** (H1: 29; H2: 34; H3: 25; M1: 19; M2: 14; M3: 14)
  M2 gives low scores on all criteria. It points out that the student fails to address own strength and motivation related to major study. The points show vagueness and irrelevance to address disciplinary knowledge and the needs in society and industry. The writing style is also unsatisfactory with various parts not being connected smoothly.

Observations from these three cases demonstrate that AI graders adopt a different grading style from human graders. AI graders tend to adhere strictly to the rubrics provided, with careful evaluation of the whole structure and fulfilment of all components and grading criteria for each submission. In contrast, human graders rely on a general impression when grading after getting an overall understanding of the student's plan, allowing flexibility in the focus of content and writing style. In addition, AI graders show independence in grading for different submissions. Human graders may unintentionally use previous submissions as reference, hence having bias dependent on the order of grading. The personal experience and knowledge of human graders are also accounted for in the deviation.

4. Implications

Our study shows that an AI grading system developed with precisely defined rubrics can efficiently facilitate grading. However, its consistency compared with human graders varies with different models. Among the three models used, GPT-4.1-mini performs most similarly to human graders. In general, AI graders are stricter in grading and more rigid in adhering to the rubrics. This lack of flexibility results in lower grades compared with human graders. If the context of assessment is highly subjective, such as career development, human experience is essential to accurately assess students' work and provide feedback. Nonetheless, AI provides an objective ground for human graders, especially in large courses with multiple graders. While this study provides an evaluation of AI grading for practitioners in education, its generalisability is limited by the diversity in assessment format and contextual knowledge. The results are more usable for reflective writings in text format of generic knowledge.

Ethical Declaration Statement

This project has been approved by the Institutional Review Board of The Hong Kong Polytechnic University (Reference number: HSEARS20251024002).

Generative AI Policy Statement

The author(s) used POE, GPT-4.1 mini, Grok 4 Fast, and Claude Haiku 4.5 during the preparation of this work to generate AI grading results. 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

Awidi, I. T. (2024). Comparing expert tutor evaluation of reflective essays with marking by generative artificial intelligence (AI) tool. *Computers and Education: Artificial Intelligence, 6*, Article 100226. https://doi.org/10.1016/j.caeai.2024.100226

Belov, D. I., & Armstrong, R. D. (2011). Distributions of the Kullback-Leibler divergence with applications. *British Journal of Mathematical and Statistical Psychology, 64*, 291–309. https://doi.org/10.1348/000711010X522227

Chan, C.K.Y., & Hu, W. (2023). Students' voices on generative AI: perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education, 20*, 43. https://doi.org/10.1186/s41239-023-00411-8

Flodén, J. (2024). Grading exams using large language models: A comparison between human and AI grading of exams in higher education using ChatGPT. *British Educational Research Journal, 51*(1), 201–224. https://

doi.org/10.1002/berj.4069

Jauhiainen, J. S., & Garagorry Guerra, A. (2024). Generative AI in education: ChatGPT-4 in evaluating students' written responses. *Innovations in Education and Teaching International, 62*(4), 1377–1394. https://doi.org/10.1080/14703297.2024.2422337

Usher, M. (2025). Generative AI vs. instructor vs. peer assessments: A comparison of grading and feedback in higher education. *Assessment & Evaluation in Higher Education, 50*(6), 912–927. https://doi.org/10.1080/02602938.2025.2487495

Wetzler, E. L., Cassidy, K. S., Jones, M. J., Frazier, C. R., Korbut, N. A., Sims, C. M., Bowen, S. S., & Wood, M. (2024). Grading the graders: comparing generative AI and human assessment in essay evaluation. *Teaching of Psychology, 52*(3), 298–304. https://doi.org/10.1177/00986283241282696

Lost in Translation? GenAI Promotion and Design Educators' Preparedness in Hong Kong Higher Education

Henry MA [1] & Yuanyi LIAO [1]
[1] School of Design, The Hong Kong Polytechnic University, Hong Kong, China

Abstract

Generative Artificial Intelligence (GenAI) has rapidly become a strategic priority in higher education, with universities promoting its adoption to enhance innovation, efficiency, and assessment practices. This study explores how design educators in Hong Kong experience institutional GenAI promotion and how these experiences shape their willingness and perceived capability to integrate GenAI into studio teaching and assessment. Utilising the Theory of Planned Behaviour (TPB) as an analytical framework, the research conducted qualitative interviews with fifteen design educators across communication, product, interaction, media, and social design disciplines in Hong Kong higher education. Analysis through TPB constructs was followed by inductive identification of design-specific themes; cross-comparison revealed three integration patterns: *enthusiastic but under-supported adopters*, *compliance-driven integrators*, and *principled sceptics*. These patterns are distinguished by the relative strength and alignment of attitudes, norms, and perceived control. These findings can inform discipline-oriented policy and professional development.

Keywords: Generative AI (GenAI), Design Education, Theory of Planned Behaviour (TPB), Artificial Intelligence in Education (AIEd), Artificial Intelligence (AI)

1. Introduction

The higher education sector increasingly frames GenAI as a strategic issue for innovation, efficiency, and rethinking assessment, with sector bodies calling for planned adoption in teaching and learning (Higher Education Authority, 2025; Association of Pacific Rim Universities, 2025). In Hong Kong, universities have organised GenAI resource hubs to help faculty stay informed about AI in education (HKUST Center for Education Innovation, 2023), while government guidelines have promoted AI adoption and recommended integrity measures such as prior approval for GenAI use in coursework and human review of AI-assisted grading (Legislative Council Secretariat, 2025; Hong Kong SAR Government, 2025). These developments create a policy environment that encourages educators to reconsider curriculum design amid shifting student learning practices.

The Theory of Planned Behaviour (TPB) conceptualises behaviour as shaped by attitudes, subjective norms, and perceived behavioural control (Ajzen, 1991). This study applied TPB to analyse how Hong Kong design educators interpret institutional promotion, peer and student norms, and their perceived ability to use GenAI for studio teaching and assessment. It addresses two research questions:

1. How do institutional norms and policies shape Hong Kong design educators' intentions to integrate GenAI?
2. How do attitudes, norms, and perceived behavioural control interact to produce distinct integration patterns?

Evidence on how studio-based design curricula within Asian higher education interpret GenAI promotion remains limited. This study contributes practical insights from design educators' first-hand perspectives across multiple design disciplines.

2. Methodology

A qualitative, interpretive research design was used. Participants were fifteen design educators working in Hong Kong higher education institutions, including communication, product, interaction, media, and social design disciplines. Data were collected through semi-structured interviews of approximately 45–60 minutes, conducted online or face-to-face. The interview protocol addressed four areas aligned with TPB:

1. Attitudes towards GenAI in design teaching;
2. Perceived institutional norms and policies;
3. Perceived peer and student norms; and
4. Perceived behavioural control.

Analysis followed a two-stage thematic process informed by TPB. First, transcripts were coded deductively using TPB constructs: attitudes (beliefs about GenAI's pedagogical outcomes), subjective norms (perceived expectations from institutions, peers, students, and industry), and perceived behavioural control (perceived ease of integrating GenAI, including access to tools, training, time, and assessment guidance). Second, within each TPB category, inductive coding identified design-specific themes such as process authenticity, aesthetic homogenisation, close reading, local contextual inquiry, and employment preparedness. These themes were synthesised by studying how attitude reactiveness, normative pressures, and control barriers combined within each integration pattern. Each pattern represents a distinct combination of TPB constructs that clarifies why educators with similar attitudes may demonstrate differing behaviours.

3. Results and Discussion

3.1 Attitudes: Efficiency, Exploration, and Ambivalence

Interviewees generally valued GenAI for practical efficiency and visual exploration. Product and interaction design educators described using image-based GenAI to produce rapid design variations and prototypes. One product design educator mentioned that students with limited visualisation skills produced "phenomenal" outcomes using Vizcom, while an interaction design educator redesigned their entire UI workflow around GenAI, placing AI at the brainstorming and prototyping stages with human refinement in between. Educators also valued GenAI for lesson preparation, summarising readings, drafting assignment descriptions, and generating teaching materials.

However, strongly conflicting views emerged regarding the compatibility of GenAI with studio-based teaching methodology. Social design staff emphasised that contextual fieldwork and situated immersion—observing how individual street cleaners design their own tools, or understanding each elderly person's unique circumstances—"simply cannot be generated by AI." Communication and product design educators worried that premature reliance on GenAI bypasses drawing, iterative modelling, and critical judgement, leading to homogenised aesthetics. One educator observed that AI-generated outputs "all look the same," while a media design educator said that students explicitly resisted AI-generated visuals, stating they "draw faster" and "draw better." These findings align with broader concerns about preserving process authenticity and student effort in an era of generative AI (Sabzalieva et al., 2024).

3.2 Norms: Institutional Promotion, Peer Division, and Student Expectations

Interviewees described institutional GenAI promotion through several concrete channels: university-wide communications, teaching fund criteria requiring AI components, surveys tracking AI adoption, programme leader meeting agendas, and resource hubs such as the institutional AI Playground. One social design educator mentioned that "when applying for teaching funds, AI is a key requirement... it's frequently on the agenda," while another described university surveys creating "both pressure and a reminder." However, educators consistently interpreted these initiatives as superficial or misaligned with design disciplines. An advertising educator described workshops as focused on "productivity stuff," not discipline-specific application. A media design educator found institutional AI sessions "not suitable for our design education... very superficial." A product design educator described available training like "learning software... I have not been particularly impressed."

Peer norms were divided. One communication design educator described "two different schools of thought" among colleagues, one embracing AI, the other preferring manual craftsmanship. Industry norms created additional pressure: product design practitioners expect graduates to use AI for speed and experimentation, and media industry employers reportedly will not hire graduates unfamiliar with AI tools. Student responses were equally mixed; product design students enthusiastically adopted AI for visualisation, while media design students rejected AI imagery in personal projects, concerned about appearing "too AI." Social design students selectively used GenAI for literature reviews and grammar but relied on human-led approaches for contextual projects. These findings suggest normative influences operate differentially across design disciplines, reflecting global patterns in which high-level institutional AI principles outpace what educators need to integrate GenAI into specific disciplinary contexts (Association of Pacific Rim Universities, 2025; Higher Education Authority, 2025).

3.3 Perceived Behavioural Control: Tools, Time, and Assessment

Perceived behavioural control emerged as the primary constraint on GenAI integration. Educators expressed uncertainty about which tools are reliable, with the rapidly expanding landscape creating decision fatigue. One media design educator indicated: "there are so many tools and limited time. How to choose which tool to use is a learning challenge for teachers." An interaction design educator said subscribing to approximately five platforms simultaneously, spending at least one hour daily tracking developments while paying out of pocket. Time constraints were universal: a communication design educator had not implemented AI despite strong intentions because of "lack of time," and a media design educator attributed limited adoption to having no time to learn what AI tools can do.

Assessment concerns further undermined perceived behavioural control. Educators pointed to inconsistent practices regarding GenAI disclosure requirements, criteria for evaluating AI-mediated visual work, and strategies for ensuring that close reading, analytical abilities, and manual craftsmanship are genuinely developed. One foundation-year educator required students to submit prompts with final work, while other disciplines had no such requirement. These challenges resonate with policy-oriented findings internationally, underscoring the need for clearer governance frameworks to support responsible GenAI adoption (Higher Education Authority, 2025; Sabzalieva et al., 2024).

3.4 Patterns of GenAI Integration

Through analysing attitudes, norms, and perceived behavioural control

within interviewees' narratives—not merely coding each construct independently—cross-comparison revealed three indicative integration patterns. The analytic criterion for pattern assignment was the configuration of TPB constructs: the relative strength and alignment of attitude reactiveness (positive, moderate, or cautious), normative pressures (institutional, peer, industry, student), and perceived control barriers. Educators were grouped into patterns based on which TPB constructs most strongly shaped their integration behaviour and whether constructs reinforced or contradicted one another.

Table 1: GenAI Integration Patterns

Pattern Label	Attitudes to GenAI	Key Norms	Perceived Behavioural Control	Classroom Practices
Enthusiastic but under supported	Positive; see GenAI as integral to future design practice and student employability	Perceive institutional encouragement and industry demand; some peer support	Low–moderate; lack design-specific training, pay out of pocket for tools, reply on case-by-case judgement	Experimenting with GenAI in visualisation, prototyping, or coding; developing custom workflows
Compliance-driven integrators	Moderately positive; views GenAI as a practical but not central to pedagogy	Sense institutional and industry expectations to incorporate AI; mixed student responses	Moderate; can use basic tools but uncertain about long-term implications and documentation practices	Integrate GenAI in limited ways (e.g., research support, mood boards, grading reference) primarily to align with expectations
Principled sceptics	Ambivalent or cautious; priorities process authenticity, local context, and deep reading/creation	Aware of institutional promotion but perceive misalignment with core disciplinary values	Variable; technically capable but restrict use to protect key learning processes	Delay or limit GenAI use, especially in foundational courses or tasks focused on fieldwork, close reading, or hand skills

These patterns show that educators with similar positive attitudes may behave differently when perceived control is low (enthusiastic adopters struggling with tool access) or when disciplinary norms conflict with institutional promotion (principled sceptics who are technically capable but deliberately restrict use). This finding indicates that interventions need to address attitudes, subjective norms, and perceived behavioural control together rather than focusing solely on promotion or tool availability.

4. Implications

Co-developing GenAI guidelines with design educators can translate high-level policy into contextualised principles applicable across stages of the design process, including norms for GenAI use in creative ideation and visual presentation, expectations for process documentation and prompt transparency, and criteria for evaluating AI-mediated outputs in critiques, portfolios, and work-based assessment. Such guidance addresses the gap this study identified between institutional promotion and disciplinary practice.

Institutions should move beyond generic AI literacy courses and support scenario-based training centred on authentic design briefs, as multiple educators described existing workshops as too superficial for studio application. This approach demonstrates how GenAI can scaffold experimentation, reflection, and ethical awareness without supplanting foundational skills or contextual inquiry.

The TPB-informed patterns identified in this study can help institutions recognise whether primary bottlenecks are attitudinal (e.g., concerns over homogenisation), normative (e.g., pressure to showcase AI amid unclear expectations), or related to perceived behavioural control (e.g., insufficient design-oriented training). Using these insights to shape discipline-oriented policies can support more context-sensitive GenAI adoption that respects disciplinary diversity and foregrounds pedagogy, fairness, and student agency.

Ethical Declaration Statement

The project has been approved by the human ethics approval of The Hong Kong Polytechnic University (Reference Number: HSEARS20251014001).

Declaration of Use of AI in the Writing Process

The authors used Perplexity during the preparation of this work to assist with language refinement. After utilising the tool, the authors thoroughly reviewed and edited the content as necessary and assumed full responsibility for the publication's content.

References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Association of Pacific Rim Universities. (2025). *Generative AI in higher education*. https://www.apru.org/wp-content/uploads/2025/01/APRU-Generative-AI-in-Higher-Education-Whitepaper_Jan-2025.pdf

Higher Education Authority. (2025). *Generative AI in higher education teaching and learning: Sectoral perspectives*. https://hea.ie/assets/uploads/2025/09/Gen-AI-in-Higher-Education-Teaching-and-Learning-Sectoral-Perspectives.pdf

HKUST Center for Education Innovation. (2023). *Generative AI & education*. https://cei.hkust.edu.hk/en-hk/education-innovation/generative-ai-education

Hong Kong SAR Government. (2025). *Hong Kong generative artificial intelligence technical and application framework*. https://www.digitalpolicy.gov.hk/en/our_work/data_governance/policies_standards/ethical_ai_framework/

Legislative Council Secretariat. (2025). *Measures to support development of AI education tools in the United Kingdom and Singapore* (Research Brief 2025ISE11). https://www.legco.gov.hk/research-publications/english/essentials-2025ise11-measures-to-support-development-of-AI-education-tools-in-the-united-kingdom-and-singapore.htm

Sabzalieva, E., Valentini, A., & Cardoso, S. (2024). A global perspective of institutional adoption policies and guidelines for generative AI in higher education. *arXiv*. https://arxiv.org/abs/2405.11800

The Role of AI in Transforming Education in Spatial Data Science

Wu CHEN [1], Xintao LIU [1], Zhongyu LAI [1]
[1] Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China

Abstract

Rapid advancements in Artificial Intelligence (AI) are revolutionising education, particularly within Geomatics fields such as land surveying, remote sensing, and Geographic Information Systems (GIS). This work explores the integration of AI tools into educational frameworks to enhance spatial understanding, data analysis, and immersive learning. We introduce **GeoAI Mentor**, an intelligent tutoring system powered by the **GeoDataGPT** engine, which integrates Large Language Models (LLMs) with dynamic knowledge graphs and Retrieval-Augmented Generation (RAG). Functioning as a 24/7 intelligent partner, the system provides personalised guidance, equipping students with essential skills in geospatial processing and predictive modelling. Pilot evaluations using Hong Kong scenarios demonstrate that **GeoAI Mentor** effectively dismantles technical barriers by achieving high-precision semantic retrieval and automating complex multi-step spatial reasoning. By bypassing the steep learning curve of software syntax, this approach reallocates cognitive resources toward higher-order analytical intent, preparing graduates for the complexities of modern geospatial industries and the demands of the digital era.

Keywords: Geomatics, Generative Artificial Intelligence, Education

1. Introduction

The digital transformation of the geoinformation industry demands professionals proficient in spatial data science (Tu et al., 2021). However, education in this field faces a significant barrier: the steep learning curve associated with professional GIS software and programming languages. Learners often struggle with complex operational details and syntax debugging, diverting their attention from the underlying logic of spatial analysis (Zhu et al., 2021). This creates a "technical barrier" where students may master software tools but lack the ability to apply spatial thinking to solve real-world problems.

Artificial Intelligence (AI), particularly Large Language Models (LLMs), offers a transformative solution by enabling natural language interaction with

complex data systems (Rudolph et al., 2023). While AI has been explored for isolated geomatics tasks, current tools often function merely as general coding assistants rather than "spatial-aware" mentors. They lack the domain-specific pedagogical frameworks necessary to guide students through the complete data-to-insight workflow, failing to effectively bridge the gap between theoretical concepts and technical implementation (Li & Ning, 2023).

To address this gap, this study introduces **GeoAI Mentor**, an intelligent tutoring system designed for the Hong Kong Common Spatial Data Infrastructure (CSDI). Built upon the **GeoDataGPT** core engine, this platform integrates a dynamic knowledge graph and adaptive learning mechanisms. By functioning as a 24/7 intelligent partner, **GeoAI Mentor** democratises access to spatial analytics, transforming spatial data science education from a tool-centric to a logic-centric model and equipping students with the critical skills needed in the digital era.

2. Methodology

This study employs a technical development and educational integration approach to construct **GeoAI Mentor,** a comprehensive intelligent tutoring system for geomatics education. The system architecture is hierarchical: it utilises **GeoDataGPT** as the underlying computational engine for spatial data processing, while integrating a Dynamic Domain Knowledge Base and an Adaptive Learning Module to fulfil pedagogical objectives.

2.1 Core AI Framework: GeoDataGPT for Spatial Intelligence

The **GeoDataGPT** framework is an AI-driven system designed to democratisate geospatial data analysis and enhance learning outcomes in geoinformation education. Centred on the integration of an LLM with domain-specific knowledge augmentation, the framework translates natural language queries into executable geospatial workflows. This reduces the reliance on specialised programming languages and GIS tool proficiency.

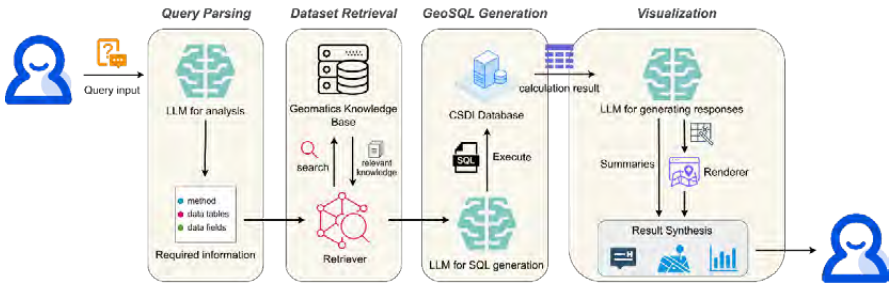
As illustrated in Figure 1, **GeoDataGPT** integrates four core modules to achieve context-aware spatial intelligence, which are operationalised through two primary functional phases: **Semantic Parsing and Knowledge Retrieval**, followed by **Computational Execution and Visualisation**.

- **Semantic Parsing and Knowledge Retrieval:** The system utilises the reasoning capabilities of the DeepSeek-R1 LLM agent to decompose user queries into structured sub-tasks. To ensure domain-specific accuracy, a Retrieval-Augmented Generation (RAG) engine employs BGE-M3-Embedding to perform semantic searches across a geospatial

knowledge base containing over 1,000 CSDI dataset schemas and 500 annotated educational Q&A records.

- **Computational Execution and Visualisation:** By synthesising user intent with retrieved metadata, the LLM agent generates validated GeoSQL scripts. These scripts undergo iterative refinement via a rule-based validator to ensure compliance with CSDI schema constraints, such as coordinate system compatibility, before execution in a spatial database. Finally, results are transformed into multi-modal outputs, including natural language explanations and interactive spatial visualisations rendered through ArcGIS JS, to help students connect results with geospatial theories.

Figure 1: Technology architecture of the GeoDataGPT.



Note: The architecture of the GeoDataGPT framework consists of four sequential modules. The system first parses a user's query and enriches it with a knowledge retriever. An LLM then generates an executable GeoSQL script, and a final module synthesises the database results into summaries and visualisations.

2.2 GeoAI Mentor

To enhance geomatics education, we are exploring the expansion of the **GeoDataGPT** framework into a comprehensive educational platform, "**GeoAI Mentor**", by integrating custom teaching modules (as shown in Figure 2). It will serve as a virtual mentor for geomatics students, answering their questions about the field, guiding them through data analysis, and providing personalised learning paths. This section details the proposed educational extension functions, focusing on modules beyond the inherited AI infrastructure.

Dynamic Domain Knowledge Base

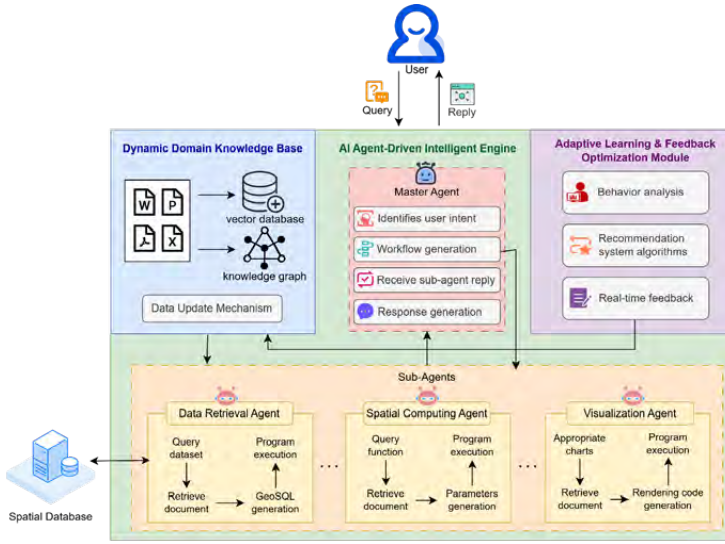
To link theoretical knowledge with practical applications, **GeoAI Mentor** introduces a semantic knowledge graph that organises geospatial concepts into an interactive, explorable network. Built using Neo4j, the graph integrates existing educational resources (i.e., textbooks, publications, lecture slides, exam papers) to link entities (e.g., "Coordinate System") with hierarchical subtopics

(e.g., "HK1980 Grid vs. WGS84") and cross-domain applications (e.g., "GNSS in building monitoring"). Unstructured content (e.g., research articles, textbooks) is encoded into semantic vectors using the BGE-M3 model, enabling context-aware retrieval through similarity searches. An automated update process periodically ingests the latest academic publications via APIs and incorporates educator-curated case studies, ensuring knowledge currency. This dynamic infrastructure transforms static textbooks into an evolving resource that adapts to technological advancements.

Adaptive Learning and Feedback Optimisation

GeoAI Mentor employs a machine learning-driven system to optimise individual learning trajectories. By analysing interaction patterns such as query frequency and task completion rates, the platform identifies specific knowledge gaps to construct detailed student profiles. A Long Short-Term Memory (LSTM) network is utilised to predict learning trajectories, allowing for dynamic resource allocation and content delivery strategies that align with Cognitive Load Theory. This creates a closed-loop optimisation process where real-time student feedback fine-tunes the system's recommendations.

Figure 2: Technology architecture of the GeoAI Mentor.



Note: GeoAI Mentor consists of three functional modules. The AI Agent-Driven Intelligent Engine, also known as GeoDataGPT, is the core of the system. The Dynamic Knowledge Base stores mapping domain knowledge and provides a reliable knowledge reference for the LLM. The Adaptive Learning and Feedback Optimisation Module provides personalised learning routes and optimises the knowledge base based on student feedback.

3. Results and Discussion

To assess the pedagogical efficacy of **GeoAI Mentor** prior to large-scale classroom deployment, we conducted a systematic pilot evaluation using real-world scenarios from the Hong Kong CSDI. This assessment focuses on how the system resolves three foundational learning barriers in geomatics education: data discovery, analytical execution, and complex spatial reasoning.

3.1 Evaluation of Educational Outcomes and Barrier Mitigation

Mitigating the Data Discovery Barrier

The first formidable obstacle in spatial data science education is the "technical barrier" of identifying appropriate datasets among thousands of options. In our pilot test regarding "road network density", the system successfully bypassed the limitations of traditional keyword-based searches by demonstrating high precision in semantic retrieval. Unlike standard tools, the knowledge augmentation engine mapped the user's high-level intent to specific CSDI schemas, identifying the 'Road Network' and 'District Boundary' datasets. Crucially, by providing a validity assessment (see Figure 3a) and advising on data limitations, the system functions as a digital scaffold, mimicking an expert instructor's guidance to ensure students understand data quality before analysis.

Bypassing Technical Operations to Accelerate Insight

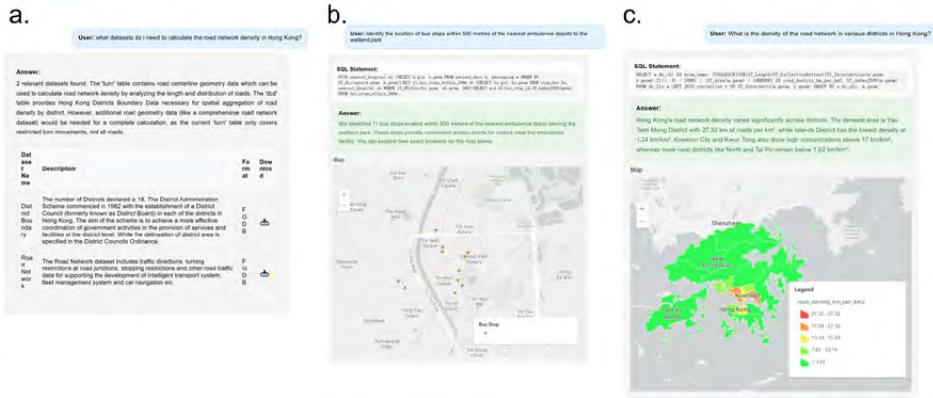
Moving to quantitative analysis, the system effectively abstracted complex technical operations that typically divert student attention from spatial logic. When tasked with calculating road density across districts, the system automated the spatial aggregation and normalisation process (km/km^2), bypassing the steep learning curve of GIS software and programming languages. As shown in Figure 3b, the output was not merely a data table but an instantly rendered choropleth map. This "Code-to-Visualisation" capability significantly reduced the "time-to-insight" and enabled learners to immediately allocate their cognitive resources to identifying and interpreting spatial patterns (e.g., the high density in Yau Tsim Mong) rather than getting bogged down in procedural debugging.

Scaffolding Multi-step Spatial Reasoning

The most significant finding is the system's ability to structure chained logic for novices, a skill that typically requires years of practice. The scenario "finding bus stops within 500m of ambulance depots near a wetland park" (Figure 3c) requires a complex sequential workflow. By utilising Common Table Expressions, the system successfully translated a high-level cognitive problem into a structured computational workflow. This confirms that **GeoAI Mentor**

provides the necessary ‘Chain-of-Thought’ scaffolding, allowing students to master complex spatial reasoning.

Figure 3: Demonstration of GeoAI Mentor capabilities.



Note: The user’s request in the blue block was input into the system, which deconstructed the complex query and generated the feedback in the grey block. This output includes the sophisticated SQL for the chained analysis, a text summary of the findings, and an interactive map visualising the results.

3.2 Pedagogical Impact and Limitations

The progression from data retrieval to complex reasoning illustrates a substantial reduction in extraneous cognitive load (Sweller, 2010). By automating the procedural “how”, the system reallocates student working memory to the “why”, aligning with Cognitive Load Theory by focusing on germane load. This pilot success validates a shift toward prioritising logical spatial reasoning over rote software proficiency (Bearman et al., 2016; Zhu et al., 2021). Students are challenged to move beyond the “Remembering” level of Bloom’s Taxonomy, focusing instead on the “Analyzing and Evaluating” levels by articulating logical spatial questions (Anderson et al., 2001).

However, reliance on AI introduces new challenges. The accuracy of the system is strictly bound by the quality of the CSDI metadata; ambiguous data descriptions can lead to retrieval failures. Furthermore, the risk of “automation bias” necessitates that “AI verification” becomes a core competency in the modern geomatics curriculum, transforming students into critical auditors of AI-generated spatial data (Mollick & Mollick, 2023).

4. Implications

The emergence of **GeoAI Mentor** signifies more than just a technological milestone; it heralds a fundamental paradigm shift in the pedagogical landscape of spatial data science by dismantling the technical barrier that has long teth-

ered students to the minutiae of software syntax and procedural operations. In this redefined educational ecosystem, the role of the learner is radically transformed: students are no longer passive consumers of GIS tools but are empowered as architects of inquiry who utilise AI as a cognitive scaffold to engage in higher-order problem-solving and "dialogic geospatial analysis".

This transition necessitates a comprehensive overhaul of our training directions, prioritising new literacies such as AI verification, cartographic ethics, and the ability to critically synthesise AI-generated outputs with grounded geographical reality. Consequently, the societal demand for professional talent is shifting away from the traditional GIS technician toward the spatial data strategist—individuals who possess the "Spatial-AI Hybridity" required to navigate complex data infrastructures and address the wicked problems of the 21st century.

Ultimately, the democratisation of spatial intelligence through such mentors ensures that geographic analysis becomes a universal language, empowering a more informed citizenry to leverage spatial data as a strategic asset for social and environmental resilience.

Ethical Declaration Statement

The project has been approved by the Department of Land Surveying and Geo-Informatics of The Hong Kong Polytechnic University (Reference Number: 88JT).

Generative AI Policy

During the preparation of this work, the authors used "DeepL" and "ChatGPT" to improve the writing style. After utilising the tool/service, the authors thoroughly reviewed and edited the content as necessary and assumed full responsibility for the publication's content.

References

Anderson, L. W., Krathwohl, D. R., & Bloom, B. S. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives* (Complete ed.). Longman.

Bearman, N., Jones, N., André, I., Cachinho, H. A., & DeMers, M. (2016). The future role of GIS education in creating critical spatial thinkers. *Journal of Geography in Higher Education, 40*(3), 394–406. https://doi.org/10.1080/03098265.2016.1144729

Li, Z., & Ning, H. (2023). Autonomous GIS: the next-generation AI-powered

GIS. *International Journal of Digital Earth, 16*(2), 4668–4686. https://doi.org/10.1080/17538947.2023.2278895

Mollick, E., & Mollick, L. (2023). *Assigning AI: Seven Approaches for Students, with Prompts*. https://doi.org/10.48550/arxiv.2306.10052

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning & Teaching, 6*(1), 342–363. https://doi.org/10.37074/jalt.2023.6.1.9

Sweller, J. (2010). Element Interactivity and Intrinsic, Extraneous, and Germane Cognitive Load. *Educational Psychology Review, 22*(2), 123–138. https://doi.org/10.1007/s10648-010-9128-5

Tu, W., Xia, J., Wang, C., Lu, M., & Yue, Y. (2021). Exploration and Practise of Education Reform of Spatial Computing and Analysis Courses Towards Smart City. *Journal of Geomatics, 47*(S1), 14–17. https://doi.org/10.14188/j.2095-6045.2021911

Zhu, A.-X., Zhao, F.-H., Liang, P., & Qin, C.-Z. (2021). Next generation of GIS: must be easy. *Annals of GIS, 27*(1), 71–86. https://doi.org/10.1080/19475683.2020.1766563

AI-Powered Clinical Role-Play Simulators: Healthcare Education Pilot

Kai Loon Caron ONG [1]
[1] School of Humanities & Social Sciences, Temasek Polytechnic, Singapore

Abstract

This paper presents findings from a multi-disciplinary pilot implementation of four AI-powered role-play chatbot simulators in Psychology and Gerontology education at Temasek Polytechnic. 226 students across gerontology and psychology engaged with domain-specific simulators designed for clinical skills practice, receiving personalised feedback and scenario-based learning experiences. Survey data (n=127$, 56.2% response rate) revealed high educational value, with 96.1% rating the simulators as "good or better" and 74.8% reporting faster learning compared to traditional methods. All four simulators achieved satisfaction rates exceeding 90%, demonstrating promising potential for AI-powered role-play in clinical skills development. Findings are reported as a Quality Assurance (QA) programme evaluation of routine curriculum delivery (descriptive, self-report; no comparison group), rather than a generalisable effectiveness study.

Keywords: generative AI, clinical simulation, role-play pedagogy, healthcare education, skills development

1. Introduction

Healthcare education depends on authentic practice before real patient or client encounters, but limited time, staffing, placements, and cost restrict opportunities, especially for interpersonal competencies such as rapport, questioning, and diagnostic interviewing. Peer role-plays and standardised patients can help, but they can be inconsistent, resource-intensive, and difficult to scale. Recent advances in generative AI enable chatbots to simulate patient or client interactions and provide immediate, personalised formative feedback, creating new options for scalable practice (Kasneci et al., 2023). When designed for low-stakes trial-and-error learning aligned to course outcomes, such role-play can reflect principles of cognitive apprenticeship and deliberate practice (Collins et al., 1991; Ericsson et al., 1993). However, implementation-focused evidence in higher education remains limited, and questions remain about functional fidelity, learner engagement, feedback quality, and cost-effective integration into lessons and assessments.

To inform institutional decision-making, we piloted four AI role-play simulators embedded in lessons (Dementia Role-Play, Frailty Assessment, Counselling Client, Mental Health Clinical). We report descriptive QA evidence (survey perceptions and thematic feedback) to inform programme improvement, not generalisable effectiveness claims.

We asked: (i) how students perceived educational value and learning efficiency across contexts; (ii) which design and implementation factors supported positive perceptions; and (iii) what considerations arise for pedagogy, assessment, and scaling.

2. Methodology

2.1 Context and Participants

The pilot was conducted during one semester in AY2025/26 across healthcare diploma modules. Four simulators were embedded into scheduled lessons: Dementia Role-Play (n=57), Frailty Assessment (n=42), Counselling Client (n=74), and Mental Health Clinical Simulator (n=53). Students typically completed one to three sessions in class, with optional self-practice.

2.2 Simulator Design and Implementation

Simulators were deployed on an institutionally approved GenAI platform. Prompts were co-developed with subject teams to model authentic personas and scenario progression, and to deliver structured formative feedback aligned to intended competencies. Lecturers briefed students on objectives, demonstrated usage, and facilitated structured reflection; conversation logs informed iterative prompt tuning. Prompts were tuned for local communicative norms, and activities were embedded within lesson and assessment workflows.

2.3 Data Collection and Analysis

An anonymous, voluntary 15-item online survey captured student perceptions across three domains: learning outcomes, user experience, and learning efficiency (5-point Likert scales), plus three open-ended prompts (most valuable aspects, skills gained, suggestions). Response rates varied by simulator (13–95%), yielding n=127 responses (56.2%). We report descriptive statistics and thematic summaries; no inferential tests were conducted. Implementation cost was monitored via API usage.

2.4 Ethical Considerations

This activity was conducted as part of routine curriculum delivery and programme-improvement QA. Survey participation was voluntary and anonymous;

no grades, demographics, or personally identifying data were collected. A local SBER ethics committee was consulted and confirmed that this activity met QA/ Exempt criteria as routine programme evaluation.

3. Results and Discussion

3.1 Overall Outcomes

Across implementations (n = 127 survey respondents), students reported high educational value (96.1% "good or better", mean 3.95/5) and faster learning compared with traditional methods (74.8%). Total delivery cost for the four simulators was US$162.11 (~US$0.72 per student).

Table 1: Implementation Outcomes Summary Across Simulators

Simulator	Students	Survey Respondents (n)	Response Rate	Good or Better	Learned Faster	Cost per Student
Dementia Role-play	57	25	44%	96%	72%	US$0.85
Frailty Assessment	42	40	95%	92.5%	52.5%	US$0.33
Counselling Client	74	55	74%	98.2%	90.9%	US$1.00
Mental Health Clinical	53	7	13%	100%	85.7%	US$0.48
TOTAL	226	127	56%	96.1%	74.8%	US$0.72

3.2 Outcomes by Simulator

All four simulators demonstrated consistently positive perceptions, each exceeding 90% "good-or-better" ratings.

- **Counselling Client:** High perceived value (good+ 98.2%) and efficiency (learned faster 90.9%); comments emphasised realistic practice and actionable feedback.
- **Dementia Role-play:** High perceived value (good+ 96%) and efficiency (learned faster 72%); comments cited behavioural markers and person-centred strategies.
- **Frailty Assessment:** Positive value (good+ 92.5%) with mixed efficiency (learned faster 52.5%); some requested sharper realism and more consistent guidance.
- **Mental Health Clinical:** Positive ratings (good+ 100%; learned faster 85.7%) but based on $n=7$; suggestions focused on difficulty calibration and avoiding diagnosis reveals.

3.3 Perceived Success Factors

Across implementations, positive student perceptions tended to coincide with the following observed features:

- **Functional realism (authentic behavioural modelling):** Students valued realism in both structured and open-ended responses. On the Likert item *"The scenarios presented felt realistic and authentic"*, 100/127 (78.7%) selected Agree/Strongly agree. In free-text responses, realism was frequently cited as the "most valuable aspect," e.g., *"Simulating real life scenarios."* Students also suggested enhancements to strengthen realism (e.g., *"Add a touch of realism, such as not accommodating to everything that the user says."*).
- **Actionable formative feedback:** Feedback emerged as a key success driver. On *"The feedback I received was helpful for improving my performance"*, 116/127 (91.3%) selected Agree/Strongly agree. Students highlighted personalised feedback and improvement guidance (e.g., *"The personalised feedback provided was useful…"*), alongside suggestions for more concise feedback.
- **Low-stakes repeated practice:** Students described the simulator as a low-stakes way to practise skills and iterate, for example, *"Practice counselling without feeling awkward with a real person."* Students also requested features that support repetition and continued access (e.g., *"Allow us to have multiple chat histories…"*).

3.4 Implementation Considerations

Some students noted scenario repetition when attempting multiple cases and occasional mis-calibration of difficulty. Expanding persona and case pools and refining difficulty may reduce repetition and support stepwise challenge (Ericsson et al., 1993; Issenberg et al., 2005).

3.5 Limitations

This work reports programme evaluation QA using self-report data without a control group. Response rates varied (13–95%). One implementation (Mental Health Clinical) had a small respondent pool (n=7), limiting inference. We did not assess objective performance gains; future ethics-approved studies should triangulate perceptions with performance measures and conversation-analysis data.

4. Implications

4.1 Pedagogical Implications

Students' emphasis on explicit guidance and opportunities to retry maps onto cognitive apprenticeship processes, making expert strategies visible, coaching during performance, and gradually building independence (Collins et al., 1991), and reflects deliberate practice through repeated, feedback-rich

attempts at calibrated challenge (Ericsson et al., 1993). From a simulation perspective, the pattern of results is consistent with functional fidelity mattering more than surface realism: what students valued was alignment of conversational demands and formative guidance to intended competencies, not theatrical authenticity (Issenberg et al., 2005). GenAI appears to expand access to such practice at low cost, while raising design and implementation questions that educators should evaluate locally (Kasneci et al., 2023).

4.2 Future Directions

Future ethics-approved studies could examine: (i) comparative effectiveness versus peer role-play, including cost-effectiveness; (ii) retention and transfer to placements and authentic assessments; and (iii) transferability of role-play design principles to other chatbot types, informing decisions about guiding use of general GenAI versus building custom tools (Kasneci et al., 2023). As a next step, we plan an ethics-approved study of the Dementia Role-Play simulator, using objective performance measures alongside learner-reported perceptions to evaluate learning outcomes.

5. Conclusion

Within routine curriculum delivery, AI role-play simulators yielded consistently positive QA feedback on educational value and perceived learning efficiency across four healthcare contexts at low cost per student. Students highlighted authentic practice and actionable feedback as the most valued features, pointing to two practical priorities for educators adopting similar tools: prioritise competency-aligned formative feedback with concrete next steps, and calibrate behavioural authenticity, difficulty, and case diversity to sustain repeated practice without fatigue (Ericsson et al., 1993). While results are descriptive, they offer a basis for scaling and for designing ethics-approved studies that triangulate learner perceptions with objective performance measures.

Acknowledgments

We thank the subject teams from Abnormal Psychology, Counselling Psychology, Mental Health & Dementia Management, and Programme Planning for Health Promotion for their engagement in this pilot, and Deputy Director Arthur Foo of the School of Humanities & Social Sciences, Temasek Polytechnic, for his support of teaching and learning innovation. We are also grateful to the team at the National University of Singapore AI Centre for Educational Technologies for their expertise in prompt development and implementation support, and to all students who participated.

Ethical Declaration Statement

This project reports a curriculum innovation and routine programme evaluation. Student feedback was voluntary and anonymous; no grades or assessments were affected, and no personally identifying information was collected. A local SBER ethics committee was consulted and confirmed that this activity met QA/Exempt criteria as routine programme evaluation.

Generative AI Policy Statement

The author used Claude (Anthropic) during the preparation of this work to organise analysis and refine language clarity. After utilising the tool, the author thoroughly reviewed and edited the content as necessary and assumes full responsibility for the publication's content.

References

Collins, A., Brown, J. S., & Holum, A. (1991). Cognitive apprenticeship: Making thinking visible. *American Educator, 15*(3), 6-11, 38-46.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*(3), 363-406. https://doi.org/10.1037/0033-295X.100.3.363

Issenberg, S. B., McGaghie, W. C., Petrusa, E. R., Gordon, D. L., & Scalese, R. J. (2005). Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Medical Teacher, 27*(1), 10-28. https://doi.org/10.1080/01421590500046924

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., … Kasneci, G. (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

Students' Perception Survey Results

Note: Total number of students = 226; total number of respondents = 127.

Table 2A: Likert Scale Responses

Question	Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree		Strongly Agree & Agree	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
The chatbot helped me develop practical skills relevant to this subject.	0	0	2	1.6	13	10.2	71	55.9	41	32.3	112	88.2
The feedback I received was helpful for improving my performance	0	0	2	1.6	9	7.1	55	43.3	61	48.0	116	91.3
I can apply what I learned to real-world situations.	0	0	2	1.6	16	12.6	63	49.6	46	36.2	109	85.8
My confidence in this subject area has increased.	0	0	1	0.8	21	16.5	59	46.5	46	36.2	105	82.7
The chatbot was engaging and held my attention	0	0	0	0	7	5.5	50	39.4	70	55.1	120	94.5
The chatbot was easy to use and navigate.	0	0	1	0.8	3	2.4	58	45.7	65	51.2	123	96.1
The scenarios presented felt realistic and authentic.	0	0	9	7.1	18	14.2	56	44.1	44	34.6	100	78.7
The chatbot provided appropriate guidance when needed.	1	0.8	5	3.9	20	15.7	57	44.9	44	34.6	101	79.5

Table 2B: Learning Efficiency Comparison

Question	Much more slowly		Somewhat more slowly		About the same speed		Somewhat faster		Much faster		Somewhat faster & much faster	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
Compared to traditional methods, this chatbot helped me to learn concepts/skills…	0	0	2	1.6	30	23.6	76	59.8	19	15.0	95	74.8

Table 2C: Overall Educational Value Rating

Question	Poor		Fair		Good		Very good		Excellent		Good or better	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
Overall, how would you rate the educational value of this chatbot?	0	0	5	3.9	32	25.2	54	42.5	36	28.3	122	96.1

Improving Education Through Educational Material Development: The Combination of Generative Artificial Intelligence and Human Insight

Farzad SABETZADEH [1], Fanfei MENG [1], Chong TANG [1]
[1]Faculty of Business, City University of Macau

Abstract

This study investigates how generative artificial intelligence (AI) and human expertise can be combined to improve the development of educational materials. Using a mixed-methods design, it integrates expert questionnaires and interviews with comparative experiments on courseware produced by humans and AI in Management, World History, and Mathematics. Results show that AI excels in efficiency, structural organisation, and visual presentation, particularly in social science and humanities content, but performs less well in mathematically demanding tasks requiring precise reasoning. Experts and learners value AI for proofreading, resource integration, and real-time support, yet remain cautious about its creativity, reliability, and ethical risks. Trust is markedly higher in human teachers, especially for personalised and high-stakes learning. The study proposes a task-based human–AI collaboration model that allocates routine and scalable work to AI while retaining human leadership in design, interpretation, and learner-facing decisions.

Keywords: Generative artificial intelligence; Educational materials; Textbook development; Human–AI collaboration; Knowledge co-creation

1. Introduction

Generative AI tools, such as ChatGPT, have rapidly expanded within education, offering scalable support for content creation, question generation, and translation (Rahman & Watanobe, 2023). At the same time, the production of textbooks and other core learning materials remains predominantly manual, demanding extensive time, expert labour, and editorial resources. This mismatch widens the gap between fast-changing knowledge and relatively slow textbook revision cycles, potentially limiting the responsiveness of curricula to learners' diverse needs (Yin et al., 2022).

Human-authored materials are grounded in subject expertise, pedagogical judgement, and contextual understanding; however, they are vulnerable to subjective bias, inconsistency, and resource constraints (Xiao & Sun, 2025).

Conversely, generative AI can process large corpora, recognise patterns, and deliver consistent structures, yet it is prone to hallucinations, shallow creativity, and embedded biases (Feuerriegel et al., 2023). These contrasting strengths and weaknesses raise a practical and theoretical question: how should responsibilities in educational material development be shared between humans and AI?

Recent studies comparing AI-generated and human-authored educational materials have reported mixed findings (Wang et al., 2025). Some studies have shown that AI-generated quiz or assessment items can be produced more efficiently and with acceptable structural quality, but they often lack the contextual sensitivity and pedagogical nuance found in teacher-designed materials (Jalgaonkar et al., 2025). Furthermore, AI-generated reading or instructional materials may approach human-authored texts in readability and coherence while still falling short in narrative engagement, disciplinary depth, and alignment with specific learning objectives (Jain & Braden, 2026). Comparative evidence also indicates that the relative performance of AI-generated materials may vary substantially across task types and subject domains, suggesting that the effectiveness of AI is highly context-dependent rather than uniformly superior or inferior to human work (Arcuri & Fraser, 2013).

This paper examines this question by combining expert perspectives with learner evaluations of AI-generated and human-generated courseware. It aims to compare the perceived advantages and limitations of generative AI across textbook development stages and to empirically assess learner responses to AI versus human materials in different disciplines. Additionally, this study proposes a scenario-based collaboration model that organises the division of labour between humans and AI in educational publishing. The goal is not to replace human educators, but to design a hybrid approach that improves efficiency, quality, and equity in educational materials (Su & Yang, 2023).

2. Methodology

This study adopts a mixed-methods design combining expert surveys and interviews with comparative experiments involving learners. The focus is on the development of structured teaching materials, such as textbooks and slide-based courseware.

2.1 Expert Interview Design

Education experts, mainly university instructors from Tianjin and Macau, were invited to complete a structured questionnaire regarding generative AI in teaching and textbook development. Background variables (discipline, rank, textbook-writing experience, and AI usage) were collected as controls.

The questionnaire, themed around trust and concerns regarding the use of AI in different processes within textbook development scenarios, linked typical textbook-development scenarios (e.g., audience analysis, structural design, content generation, editing, customisation, and updating) to specific AI advantages and risks.

Advantages included high automation and efficiency, diverse and up-to-date resources, and high accuracy and stability. Limitations covered algorithm and dataset constraints, oversimplification and lack of creativity, and data privacy and ethical concerns. For each scenario, experts rated the degree of AI involvement and their confidence or concern on Likert-type scales. They also rated the frequency of AI use in different teaching activities.

Figure 1: Features-Situation Corresponding Quantitative responses were converted into weighted indicators capturing perceived benefits and risks of AI participation in each stage of material development.

	Features / Situation	Basic concept understanding	Data collection	Resource/material search	Text output	Translation	Psychological/ emotional support	Guidance on learning/teaching methods	Chart/graph drawing
Advantages of AI	High automation and efficiency		1	1	1	1			1
	Diverse and continuously updated resource coverage	1	1	1		1		1	
	Intelligent filtering and personalized output			1	1	1	1	1	
	High accuracy and stability of output content	1	1		1	1			1
	Discover potential correlations and predictive capabilities	1	1				1	1	1
	Timely and targeted feedback	1	1	1			1	1	
	Diverse output formats to meet different needs	1			1		1	1	1
	Reduce interference from human subjective biases			1	1	1	1		1

In addition, semi-structured interviews with experts in English and History teaching in Macau provided qualitative insights into domain-specific expectations, concerns about creativity and authenticity, and the perceived impact of AI on student learning habits.

2.2 Comparative Experiments Design

To compare learner perceptions of AI- and human-generated materials, courseware samples were collected or generated in three subject areas: Principles of Management, World History, and Mathematics. High-rated slide decks of human-authored samples were selected from Baidu Wenku, an open platform

where users can upload and share their courseware, papers, industry reports, and other materials. 'High-rated' was defined strictly as materials possessing a user rating of 4.5 stars or above and a minimum of 500 downloads, which ensured recognised pedagogical quality.

Matching topics were then generated using Baidu Wenku AI, a mainstream generative AI tool with the GenFlow intelligent agent system as its core, powered by the ERNIE large model for underlying capabilities. The prompting procedure utilised a structured prompt based on selected human samples: 'Based on the selected human sample, maintain consistency in topic, fundamental concepts, and subheadings to generate a slide deck outline, and then produce the full slides accordingly. Ensure the difficulty level is readable to a layperson.'

Default parameters were maintained to replicate the typical, out-of-the-box usage pattern of an average educator who may not possess advanced prompt-engineering skills. Default parameters were retained to maximise ecological validity, aiming to reflect the conditions that most educators would realistically use generative AI tools without specialised prompt-engineering expertise.

Key concepts were chosen for each field—for example, Maslow's Hierarchy of Needs in Management, the Battle of Verdun in World History, and derivative rules in Mathematics. Equivalence between the human and AI materials was strictly controlled by maintaining consistency in each aspect: setting identical sub-topics in terms of topic coverage, ensuring a layperson-readable difficulty level, similar length, a format of bullet points accompanied by visual placeholders, and the same intended teaching purpose.

AI and human samples covered equivalent content but were not labelled by origin. Learners recruited via an online survey platform evaluated both materials for each topic. Across the three disciplines, a total of 168 valid responses were obtained. 58.33% of the participants had attained a graduate degree or higher, indicating a generally higher level of education. Among them, there were 76 responses for the management principles courseware, 61 for world history, and 31 for mathematics.

They were asked to use six dimensions to rate each sample: content accuracy, structural clarity, comprehensiveness, logical coherence, relevance to learning objectives, and visual design. Paired t-tests were conducted to compare AI and human samples in each discipline, after checking the reliability and validity of the scales. The questionnaire further asked participants about their preferences for AI-assisted versus teacher-led support in various scenarios,

such as real-time interaction, resource organisation, after-class tutoring, course customisation, and personalised guidance. Trust in AI versus teachers was also measured. Quantitative findings from learners were integrated with expert perspectives to design a practical framework for human–AI collaboration in educational material development.

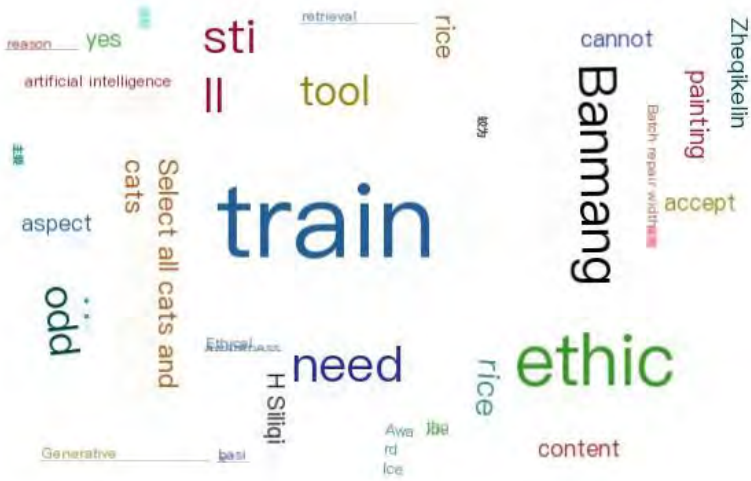
3. Results and Discussion

3.1 Expert Perspectives on Advantages and Risks

Experts reported increasingly frequent use of generative AI in teaching-related tasks, but adoption in high-stakes content creation remains cautious. Quantitative analysis showed that "high automation and efficiency," "diverse and continuously updated resources," and "high accuracy and stability" are the most valued benefits, especially for content proofreading and editing. AI was regarded as less suitable for full content generation, reflecting doubts about its depth of understanding and originality.

Among limitations, "algorithm and training data constraints" and "content oversimplification and lack of creativity" were rated as the most serious concerns. Issues such as data privacy were noted but perceived as less immediately restrictive. Despite recognising risks, more than half of the experts indicated that their institutions lacked formal AI-related risk management measures, relying instead on manual checking or informal guidance to students. Overall, experts see AI as an efficient assistant rather than an autonomous author, particularly in disciplines requiring rich interpretation and nuanced argumentation.

Figure 2: Special Training Word Cloud



3.2 Learner Evaluations of AI and Human Materials

According to the statistical results, internal consistency was acceptable across all dimensions (Cronbach's α range: Management = 0.908; World History = 0.934; Mathematics = 0.923). The comparative experiments reveal a nuanced pattern across disciplines (*Figure 3*).

In **Management**, AI-generated slides slightly outperformed human materials in structural organisation ($t = 3.153$, $p < .05$, $d = 0.002$**) and visual design ($t = 3.088$, $p < .05$, $d = 0.003$**), while maintaining comparable content accuracy.

In **World History**, AI samples showed clear advantages in structure ($t = -3.152$, $p < .05$, $d = 0.003$**), comprehensiveness ($t = -2.726$, $p < .05$, $d = 0.008$**), and visual appeal ($t = -6.119.153$, $p < .05$, $d = 0.000$**), suggesting that generative models handle narrative, fact-rich content well when guided by existing templates.

In **Mathematics**, however, AI-generated courseware lagged behind human-created materials in logical coherence ($t = -0.441$, $p < .05$, $d = 0.662$) and relevance to learning objectives ($t = -2.160$, $p < .05$, $d = 0.039$*). Learners perceived explanations as less clear and connections between formulas and concepts as weaker. This indicates that generative AI can struggle in domains that demand stepwise reasoning and precise conceptual scaffolding, despite its ability to generate correct formulas.

Regarding learning support, 94.05% of participants expressed openness to using AI in some stage of learning. AI was preferred for resource organisation (78.57%) and real-time interaction (58.13%), and moderately favoured for after-class tutoring. Yet, for course customisation and personalised tutoring, learners leaned towards teachers, and an overwhelming 79.76% reported greater trust in teachers compared to only 4.77% who trusted AI more. This trust gap underscores that empathy, responsibility, and long-term guidance remain firmly associated with human educators (Martin & Dowson, 2009).

Figure 3: Summary of t-test results

Discipline	Dimension	AI M	AI SD	Human M	Human SD	t	df	p	d
Management	CA	4.14	0.60	4.05	0.67	1.306	75	0.196	0.150
	SC	4.11	0.64	3.76	0.89	3.153	75	0.002**	0.362
	CP	3.95	0.83	3.79	0.88	1.300	75	0.198	0.149
	LC	4.07	0.81	3.86	0.89	1.634	75	0.106	0.187
	RLO	3.99	0.77	3.87	0.87	1.118	75	0.267	0.128
	VD	3.79	0.79	3.36	1.24	3.088	75	0.003**	0.354
World History	CA	3.80	0.81	3.93	0.81	-1.588	60	0.117	0.203
	SC	3.46	1.10	3.89	0.78	-3.152	60	0.003**	0.404
	CP	3.61	1.05	3.97	0.84	-2.726	60	0.008**	0.349
	LC	3.61	0.92	3.82	0.83	-1.751	60	0.085	0.224
	RLO	3.61	0.94	3.84	0.92	-1.843	60	0.070	0.236
	VD	2.95	1.27	3.97	0.95	-6.119	60	0.000**	0.783
Mathematics	CA	3.84	0.86	3.94	0.63	-0.828	30	0.414	0.149
	SC	3.55	0.96	3.77	0.84	-1.754	30	0.090	0.315
	CP	3.71	0.78	3.58	0.99	0.941	30	0.354	0.169
	LC	3.74	0.96	3.81	0.98	-0.441	30	0.662	0.079
	RLO	3.61	0.99	3.97	0.87	-2.160	30	0.039*	0.388
	VD	3.00	1.32	3.13	1.57	-0.597	30	0.555	0.107

Note: M = mean; SD = standard deviation; df = degrees of freedom; d = Cohen's d. AI = artificial intelligence. CA = content accuracy; SC = structural clarity; CP = comprehensiveness; LC = logical coherence; RLO = relevance to learning objectives; VD = visual design.

3.3 Synthesising Findings

Together, the expert and learner findings reveal a complementary pattern in educational material development. Generative AI performs best in structured, repetitive, and resource-intensive tasks, whereas human expertise remains essential in activities requiring contextual judgement, creativity, ethical reflection, and personalised engagement with learners. This overall pattern provides the empirical basis for the collaborative framework summarised in Table 1.

Table 1: Proposed Human-AI Collaboration Modes between Humans and Generative AI in Educational Material Creation

	Human	Generative AI
Confirming target audience needs	Lead	Assist
Structural design	Collaborate	Collaborate
Text content generation	Lead	Assist
Multimedia content generation	Assist	Lead
Personalized design	Lead	Assist
Content proofreading and editing	Assist	Lead
Assessment and feedback	Lead	Assist
Content updating	Collaborate	Collaborate

4. Implications

The findings support a task-based approach to human–AI collaboration in educational material development for educational institutions and publishers. Structurally repetitive tasks—such as layout optimisation, preliminary proofreading, and the generation of illustrative media—can be delegated to AI to shorten development cycles and reduce authors' workload, while human experts should retain primary responsibility for conceptual design, curriculum alignment, pedagogical adaptation, and responsiveness to diverse learner needs.

For policymakers and school leaders, the findings underline the need for clear institutional guidelines on AI use, including quality assurance, transparency in AI-assisted materials, and procedures for error correction. Investing in professional development that combines ethical awareness, data management, and technical skills can support teachers in using AI tools responsibly while preserving academic integrity and student agency.

In the global context, where education systems face pressure to expand access and keep pace with rapid knowledge growth, a hybrid human–AI framework offers a practical pathway. It can help reduce the lag between emerging knowledge and textbook updates, support multilingual and localised materials, and promote more inclusive access to high-quality resources—provided that human judgement, critical reflection, and pedagogical purpose remain at the centre of design and decision-making.

Despite the contributions of this study, several limitations should be acknowledged: First, our survey samples were divided into three disciplines, causing the sample size for the mathematics discipline to be relatively small (n = 31), which may limit the statistical power and generalisability of the findings within that subgroup. In addition, the study relied on a single AI platform (Baidu Wenku) and employed only default parameter settings, restricting our ability to examine how variations in platform design or parameter tuning might influence outcomes. Third, the research was conducted within a China-specific context, so the results may not fully extend to other cultural or educational settings without further validation.

References

Arcuri, A., & Fraser, G. (2013). Parameter tuning or default values? An empirical investigation in search-based software engineering. *Empirical Software Engineering, 18*(3), 594-623. https://doi.org/10.1007/s10664-013-9249-9

Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2023). Generative AI. *Business & Information Systems Engineering, 66*(1), 111-126. https://

doi.org/10.1007/s12599-023-00834-7
Jain, N. U., & Braden, E. (2026). The Incorporation of AI-Enabled Culturally Relevant Narrative-Based Learning Pedagogy to Engage Students in a Diverse STEM Classroom. *Reading Research Quarterly, 61*(1), e70086. https://doi.org/10.1002/rrq.70086
Jalgaonkar, Z., Gaikwad, H., & Laddha, M. (2025). Empowering Educators Through Large Language Models: A Systematic Review. *2025 OITS International Conference on Information Technology (OCIT)*.
Martin, A. J., & Dowson, M. (2009). Interpersonal relationships, motivation, engagement, and achievement: Yields for theory, current issues, and educational practice. *Review of Educational Research, 79*(1), 327-365. https://doi.org/10.3102/0034654308325583
Rahman, M. M., & Watanobe, Y. (2023). ChatGPT for Education and Research: Opportunities, Threats, and Strategies. *Applied Sciences, 13*(9). https://doi.org/10.3390/app13095783
Su, J., & Yang, W. (2023). Unlocking the Power of ChatGPT: A Framework for Applying Generative AI in Education. *ECNU Review of Education, 6*(3), 355-366. https://doi.org/10.1177/20965311231168423
Wang, R., Ding, Y., Shen, Y., Liu, H., Wang, P., & Gao, Z. (2025). Comparative Evaluation of Teaching Plans on Prostate Cancer Generated by Various Large Language Models and a Human Expert. *Engineering Reports, 7*(8), e70303. https://doi.org/10.1002/eng2.70303
Xiao, Z., & Sun, J. (2025). Video Production and Analysis Method of Ideological and Political Education in Colleges and Universities. *International Journal of Interdisciplinary Telecommunications and Networking (IJITN), 17*(1), 1-20. https://doi.org/10.4018/IJITN.392501
Yin, A., Chuck, C., Yuqi, S., Guang, L., Jun, P., Feng, R., Zheng, W., & Jie, Z. (2022). Identifying and monitoring students' classroom learning behavior based on multisource information. *Mobile Information Systems, 2022*(1), 9903342. https://doi.org/10.1155/2022/9903342

Cultivating Critical Collaboration: GenAI as a Co-Pilot in Equity Valuation Education

Jianfu SHEN [1] & Yi ZHANG [1]
[1] Department of Building and Real Estate, The Hong Kong Polytechnic University, Hong Kong, China

Abstract

This study explores the integration of Generative AI (GenAI) as a multifunctional pedagogical tool within an equity valuation course, demonstrating its role in bridging theoretical finance with practical application. Framed by distributed cognition, the instructor employed GenAI as a "co-pilot" to generate cases, automate financial data processing, and execute valuation models. This collaborative approach transformed students from passive calculators into active, critical analysts, enabling deeper strategic engagement by offloading routine tasks. Findings indicate that while GenAI enhances analytical efficiency and provides useful scaffolding, it remains limited in strategic thinking, qualitative assessment, and assumption development within equity valuation, highlighting the irreplaceable role of human judgment. The course cultivated critical GenAI collaboration literacy, equipping students to command, validate, and synthesise AI-generated insights, thereby preparing them for the evolving demands of the finance profession.

Keywords: Generative AI, Equity Valuation, Business Education, Pedagogical Innovation, AI Literacy

1. Introduction

Equity valuation, a cornerstone of finance education, demands that students synthesise accounting principles, financial theory, and strategic analysis to estimate a company's intrinsic value. Traditionally, pedagogy has emphasised manual data extraction from financial statements, the application of predetermined models, and theoretical case studies. While foundational, this approach can be time-intensive and may fall short of replicating the dynamic, iterative analysis characteristic of professional practice (Wang et al., 2025).

Generative artificial intelligence (GenAI) offers a transformative means to bridge this gap. With advanced capabilities in natural language processing and data analysis, GenAI can perform valuation-relevant tasks, such as extracting key figures from reports, generating competitive analyses, explaining model nu-

ances, and drafting sections of a valuation, thereby reframing the student's role from passive executor to active, critical analyst. Yet, without deliberate instructional design, such integration risks fostering superficial learning. This raises two pivotal questions: How can generative AI be meaningfully embedded in equity valuation pedagogy to deepen learning? And what is its practical impact on the learning process?

This study adopts a pragmatic, experience-based research approach, investigating the integration of generative AI as an instructional tool within a live corporate valuation course. The methodology centres on deploying AI as a "co-pilot" for students, actively assisting them in tasks such as data extraction, financial modelling, and report drafting. The findings indicate that this guided partnership can enhance the efficiency and depth of the analytical process, allowing learners to focus on higher-order strategic interpretation rather than manual computation. Crucially, the study underscores that instructor scaffolding remains indispensable: it frames the technology's use, promotes critical engagement with AI outputs, mitigates over-reliance, and safeguards the integrity of learning objectives. Ultimately, this work contributes a practical framework for developing essential human-AI collaborative literacy, providing an evidence-based example of how to embed such competencies authentically within global business education.

2. Methodology

This study employed a mixed-methods case study design, centred on a corporate valuation course with 51 undergraduate students enrolled in a finance subject at a research-intensive university. Data were triangulated from three sources to provide a holistic evaluation of the pedagogical innovation.

First, the instructional team maintained reflective journals to document the design and implementation of AI-assisted learning activities, including the development of prompts for tasks such as financial data extraction and model application. Second, a post-course survey was administered to the entire cohort, utilising Likert-scale items to quantify perceptions and open-ended questions to capture detailed feedback on the AI-augmented learning experience. Third, semi-structured interviews were conducted with a purposive sample of five students, selected to represent a range of academic performance and different levels of familiarity with GenAI, as reflected in the survey responses. The aim was to investigate the development of their prompting strategies and their critical engagement with AI-generated outputs.

The qualitative data from reflections, open-ended responses, and inter-

views were analysed thematically to identify recurring patterns and insights. These findings were integrated with descriptive statistics from the survey to construct a comprehensive understanding of both the teaching practice and its impact on student learning.

3. Results and Discussion

3.1 Reflections and Observations from the Lecturer

Generative AI served as a versatile pedagogical instrument in the equity valuation course, bridging theory and practice by acting as a co-pilot. This approach, aligned with distributed cognition, shifted students from passive learning to active, critical analysis. The instructor used GenAI to develop real-world cases, which students then employed for tasks such as summarising reports, extracting data, and applying valuation models.

The technology excelled in automating procedural tasks such as data processing, complex calculations, and the generation of analytical frameworks. This student-centred method reallocated class time towards higher-order strategic interpretation and financial analysis, enhancing both teaching efficiency and analytical depth. A critical observation confirmed AI's limitations, particularly its struggle with nuanced strategic interpretation, the development of forward-looking assumptions, and qualitative assessment, which underscored the irreplaceable need for human oversight and expertise.

The experiential model thus achieved two key goals: developing technical valuation competence and fostering a realistic understanding of AI's possibilities and constraints. Ultimately, this exercise served a dual purpose: it functioned as a powerful tool for achieving technical proficiency and, more importantly, as a practical demonstration of its own boundaries, thereby training students to command AI as a critical assistant while cultivating the sophisticated professional judgment required to validate its outputs, synthesise information, and make final investment decisions.

3.2 Survey from Students

Thirty-eight surveys were received (α=0.875). Most students knew GenAI, although their familiarity with these tools varied across the cohort, as reflected in the survey results. About 36.8% had medium skill; 26.3% were advanced, indicating that students entered the course with different levels of prior experience using GenAI tools.

Use was common: 47.4% used it 1–2 times weekly, and another 34.2% used it over five times. Most spent 30 minutes to 2 hours weekly (65.8%),

while 26.3% spent more than 2 hours. This variation suggests that students approached GenAI as a learning support tool to different degrees during the course. Appendix 1 gives the summary of background information.

Table 1 summarises student perceptions of using GenAI as a learning tool. Overall, feedback was positive: a majority of students reported that GenAI was helpful for understanding core valuation concepts (71.1%) and for analysing the relationships between different models (73.7%). However, a significant portion also indicated drawbacks, noting that AI-generated explanations sometimes led to confusion (68.4%) or reduced clarity (60.6%). This suggests that GenAI outputs can be incomplete or lack necessary context. Consequently, GenAI is most effective as a supplemental tool rather than a stand-alone resource, highlighting the importance of guided instruction in helping students interpret and validate AI-generated outputs. And its use should be coupled with guided instruction to help students interpret, validate, and contextualise its outputs.

Table 1: Using GenAI to Learn Equity Valuation

Item	1	2	3	4	5	Mean	SD
GenAI helped me better understand core valuation concepts (e.g., DDM, FCF, WACC).	3 (7.9%)	0 (0%)	8 (21.1%)	18 (47.4%)	9 (23.7%)	3.79	1.06
GenAI was helpful in explaining complex financial terminology.	2 (5.3%)	3 (7.9%)	6 (15.8%)	18 (47.4%)	9 (23.7%)	3.76	1.06
I used GenAI to clarify topics and cases discussed in class.	2 (5.3%)	2 (5.3%)	6 (15.8%)	22 (57.9%)	6 (15.8%)	3.74	0.96
GenAI helped me understand the relationships among different valuation models.	3 (7.9%)	2 (5.3%)	5 (13.2%)	22 (57.9%)	6 (15.8%)	3.68	1.05
GenAI's explanations sometimes confused me.	2 (5.3%)	2 (5.3%)	8 (21.1%)	19 (50.0%)	7 (18.4%)	3.71	1.00
Some concepts became less clear when I relied on GenAI for explanations.	1 (2.6%)	4 (10.5%)	10 (26.3%)	18 (47.4%)	5 (13.2%)	3.58	0.94

Note: N = 38. The survey utilises a 5-point Likert scale; 1 = Strongly Disagree, 5 = Strongly Agree.

Table 2 reports the student feedback on using GenAI in their equity valuation project, a key course assessment in the subject. Students widely perceived GenAI as a helpful tool throughout their equity valuation project, with mean

usefulness scores ranging from 3.37 to 3.76 on a 5-point scale. GenAI was rated most effective for structuring and drafting the final report (Mean = 3.76) and reliably useful for core technical tasks, such as data extraction, ratio analysis, model application, and key input estimation, where scores clustered between 3.50 and 3.58. However, its perceived utility was comparatively lower for broader, contextual tasks like company and industry background research (Mean = 3.37). This pattern indicates that GenAI serves best as a supplemental co-pilot for well-defined, quantitative analysis, while complex, nuanced interpretation still benefits from guided instruction and critical human oversight.

Table 2: GenAI's Application in the Equity Valuation Project

Item	1	2	3	4	5	Mean	SD
Company and industry background research	1 (2.6%)	6 (15.8%)	12 (31.6%)	16 (42.1%)	3 (7.9%)	3.37	0.93
Extracting and summarizing key information from financial statements	2 (5.3%)	3 (7.9%)	10 (26.3%)	20 (52.6%)	3 (7.9%)	3.50	0.94
Performing ratio analysis and interpreting results	0 (0%)	4 (10.5%)	13 (34.2%)	16 (42.1%)	5 (13.2%)	3.58	0.85
Applying valuation models (e.g., DDM, FCF, REM)	0 (0%)	6 (15.8%)	11 (28.9%)	16 (42.1%)	5 (13.2%)	3.53	0.91
Estimating key inputs such as growth rates and terminal values	0 (0%)	5 (13.2%)	12 (31.6%)	16 (42.1%)	5 (13.2%)	3.55	0.88
Structuring and drafting the valuation report	1 (2.6%)	0 (0%)	9 (23.7%)	25 (65.8%)	3 (7.9%)	3.76	0.70
Reviewing and validating my own valuation assumptions	2 (5.3%)	2 (5.3%)	9 (23.7%)	18 (47.4%)	7 (18.4%)	3.68	1.00

Note: N = 38. The survey utilises a 5-point Likert scale; 1 = Strongly Disagree, 5 = Strongly Agree.

The survey also inquired about the primary benefits and main challenges students encountered when using GenAI. The survey results highlight a clear contrast in how students perceived GenAI. The primary benefits were practical: most students cited increased speed and efficiency, along with quick access to explanations. Conversely, the main challenges were analytical, with many reporting that outputs were often superficial and difficult to verify for reliability. Issues like factual inaccuracies and the risk of over-reliance were also commonly noted, while few found the technical process of prompting or using features to be a significant hurdle.

3.3 Semi-structured Interviews

Semi-structured interviews were conducted with five students regarding their use of GenAI in equity valuation. Thematic analysis of the transcripts revealed several prominent trends. As illustrated in Appendix 2, the findings are organised into five central themes, supported by direct student quotations.

Thematic analysis of student interviews revealed that Generative AI integration fundamentally reshapes learning approaches in equity valuation. Students developed prompt engineering as a core skill, learning to structure tasks for effective human-AI collaboration, which provided cognitive scaffolding aligned with guided learning. They also established personal critical evaluation frameworks, often using a three-stage process to verify outputs by comparing AI models, checking sources, and assessing logic. Encountering AI errors was frequently viewed as an opportunity for deeper investigation, reinforcing core competencies.

A notable learning paradox emerged: while GenAI boosted efficiency, students acknowledged it could sometimes shortcut the deeper cognitive engagement required for mastery, reflecting a tension highlighted in cognitive load theory. Ultimately, learners evolved from viewing AI as a mere tool to recognising it as a bounded strategic partner with specific capabilities and limitations. This shift underscored the need for transformed pedagogy, with students requesting instructors to act as guides and workflow designers to foster mindful, self-regulated collaboration.

4. Implications and Conclusions

This mixed-methods study demonstrates that integrating Generative AI into equity valuation education significantly enhances analytical efficiency and refocuses student effort on higher-order interpretation and strategy. However, its educational value is contingent upon the parallel development of critical GenAI literacy—the disciplined ability to command, validate, and critique GenAI outputs—rather than passive acceptance.

From a pedagogical perspective, the findings suggest several implications for course design. First, instructors may integrate structured human–AI collaboration tasks into finance education, requiring students not only to generate AI-assisted outputs but also to critically evaluate their validity. Second, assessment practices can emphasise transparent reasoning processes by encouraging students to explain how AI-generated results were interpreted, validated, or revised. Such approaches help ensure that GenAI enhances learning efficiency without replacing essential analytical thinking.

At a broader level, universities may consider developing guidelines for responsible GenAI integration in business education. These may include promoting AI literacy among students, clarifying expectations regarding AI-assisted academic work, and supporting faculty in designing effective human–AI collaborative learning activities.

The findings necessitate a pedagogical transformation. Instruction must evolve beyond introducing GenAI as a simple tool to explicitly cultivate competencies in prompt engineering, output verification, and assumption defence. This requires redesigning assessments to include tasks that mandate structured reasoning and justification without GenAI assistance.

Theoretically, this study contributes by operationalising critical digital literacy into a teachable framework of professional skills for finance. It provides empirical evidence for the dual effect of cognitive offloading: while GenAI saves time on computation and data collection for finance education, it can inadvertently reduce the cognitive engagement necessary for deep conceptual understanding unless activities are deliberately designed to counteract this. The central pedagogical challenge thus becomes balancing efficiency with effortful thinking to achieve durable learning.

In summary, this work responds to the global imperative for AI-ready graduates. By demonstrating how GenAI can function as a structured "co-pilot" in finance learning, the study provides a practical example of how AI tools can be integrated into specialised business education while preserving the central role of human judgment and critical thinking.

Acknowledgments

The team acknowledges students that participated in the survey and received interviews.

Ethical Declaration Statement

The project has been approved by the Research and Innovation Office of The Hong Kong Polytechnic University (Reference Number: HSEARS20251118002).

Generative AI Policy

The authors used GPT and DeepSeek during the preparation of this work to instruct students to use GenAI, design and prepare initial survey and interview questions, and proofread the draft. After utilising the tool/service, the au-

thors thoroughly reviewed and edited the content as necessary and assumed full responsibility for the publication's content.

References

DeLuca, C., Volante, L., & Holden, M. (2025). The AI3 model: Future directions for artificial intelligence, assessment innovation, and academic integrity. *Educational Researcher*, https://doi.org/10.3102/0013189X251385537

Doulougeri, K., Vermunt, J. D., Bombaerts, G., & Bots, M. (2024). Challenge-based learning implementation in engineering education: A systematic literature review. *Journal of Engineering Education*, 113(4), 1076-1106. https://doi.org/10.1002/jee.20588

Seel, N. M. (2017). Model-based learning: A synthesis of theory and research. *Educational Technology Research and Development*, 65(4), 931-966. https://doi.org/10.1007/s11423-016-9507-9

Wang, C., Chen, Y., Hu, Z., Li, Y., & Gu, X. (2025). The journey of challenges and victories: exploring the transformation action framework in the GenAI era from multifaceted policies: The journey of challenges and victories: exploring the transformation action framework. *Educational Technology Research and Development*, 1-43. https://doi.org/10.1007/s11423-025-10535-5

Yin, R. K. (2018). *Case study research and applications* (Vol. 6). Thousand Oaks, CA: Sage. https://uk.sagepub.com/en-gb/eur/case-study-research-and-applications/book250150

Appendix 1: Summary of Background Information and GenAI Usage

Item / Category	Frequency	Percentage (%)
A1. Prior experience with GenAI		
1 = No prior experience	3	7.9
2 = Beginner	9	23.7
3 = Intermediate	14	36.8
4 = Proficient	10	26.3
Total (Valid)	36	94.7
Missing	2	5.3
A2. Frequency of use during the course		
1 = Less than once per week	2	5.3
2 = 1–2 times per week	18	47.4
3 = 3–5 times per week	3	7.9
4 = More than 5 times per week	13	34.2
Total (Valid)	36	94.7
Missing	2	5.3
A3. Weekly time spent using GenAI		
1 = Less than 30 minutes	2	5.3
2 = 30 minutes–1 hour	13	34.2
3 = 1–2 hours	12	31.6
4 = More than 2 hours	10	26.3
Total (Valid)	37	97.4
Missing	1	2.6

Appendix 2: Key Findings and Supporting Quotes from Interviews

Theme	Key Finding	Illustrative Evidence
Professionalization of prompt engineering	Students evolved from basic queries to structured, role-specific prompt design.	S1: "Now I specify the task context and requirements more concretely."
Autonomous critical evaluation	Students developed multi-source verification strategies for AI outputs.	S2: "I ask several different AIs simultaneously. If they arrive at consistent conclusions based on different logic, the credibility is higher."
Efficiency vs. cognitive reliance	GenAI boosted efficiency but raised concerns about over-reliance and shallow processing.	S2: "I became more reliant on the AI's logical framework, and my own ability to independently conceive and write a complete report actually declined somewhat."
Evolving understanding of GenAI	Student perceptions shifted from a generic tool to a nuanced understanding of its specific capabilities and limits.	S5: "I originally thought AI could only answer simple questions, but the new generation of models performs like a 'high-EQ teacher'..."
Structured pedagogical appeal	Students requested systematic training in prompting and demonstrations of complete human-AI workflows.	S2: "I suggest using a full tutorial session to demonstrate a complete AI workflow from start to finish, rather than explaining it in fragments."

AI-First vs Human-First Workflows in Journalism Education: Student Choices and Effects on Learning Outcomes**

Ming Hang SO [1]
[1] Department of Journalism, Hong Kong Baptist University, Hong Kong, China

Abstract

This exploratory mixed-method case study examines how journalism students engage with generative AI in a financial news reporting assignment. Students selected between an AI-first, human-edited workflow and a human-first, AI-assisted workflow when reporting on actual corporate financial results. The study found that 86.7% of students preferred the AI-first approach. Assignment scores and qualitative analysis from reflective papers and the instructor's observations revealed that higher AI involvement did not guarantee better performance. Learning outcomes depended on students' ability to critically evaluate AI-generated content and apply disciplinary knowledge when collaborating with AI. Some students, influenced by AI-generated drafts, tended to adopt corporate language or subjective phrasing, while others effectively refined AI drafts to produce accurate and neutral reporting. The findings highlight both the risks of over-reliance on AI and its value in supporting students as they navigate unfamiliar tasks, offering practical insights for instructors integrating generative AI into discipline-specific coursework.

Keywords: Journalism education; AI-assisted learning; Pedagogy and technology integration; Student AI literacy

1. Introduction

Universities are generally taking a more proactive approach to introducing GenAI into teaching, learning, and assessment. There is a growing body of research regarding how students use GenAI in various assignments, such as writing, coding, and collaborative projects; however, there are calls for further study to understand how GenAI is used within specific disciplines by students and the quality of student work produced at different levels of collaboration with AI (Kim et al., 2025; Qu et al., 2024).

In the field of journalism and communication, both academics and practitioners urge more AI training in university journalism curricula, as the news media is being rapidly transformed by emerging AI technology. Major interna-

tional news organisations such as the Associated Press and Bloomberg began implementing AI-powered automated journalism about a decade ago. They use machines to generate complete news reports from raw inputs, such as press releases and company documents, with human journalists acting as the final gatekeepers (Graefe, 2016). This approach can be regarded as an AI-first, human-edited approach. For other news media organisations, which often operate with limited resources and adopt a prudent approach towards new technology, AI is mainly used as an assistant for specific journalistic tasks, such as brainstorming story ideas, general background research, proofreading, or headline suggestions (Cools & Diakopoulos, 2024). In other words, this constitutes a human-first, AI-assisted approach.

This exploratory study aims to examine journalism students' perceptions of and preferences for the above two news writing approaches, which involve different levels of AI involvement. It also investigates how their choice of approach may affect the quality of their assignments. The study is relevant not only to journalism and communication but also to higher education research, as it explores how different levels of teacher-permitted AI involvement may influence students' academic performance and learning outcomes.

2. Methodology

This exploratory study adopted a mixed-method case study design, which provides a more comprehensive way to understand students' learning experiences with technology (Swart, 2017). An undergraduate financial journalism news writing class with 15 students was selected. As one of the assignments for the course, the students were required to write a news report on the third-quarter financial results of a Hong Kong listed company. The aim was to help the students learn how to report on corporate financial results accurately and professionally, while also exploring the effective and responsible use of AI tools to assist their journalism process.

Three lessons regarding financial result news reporting and the university principles for the Use of Generative AI Tools in teaching and learning were provided before asking the students to complete the assignment. The selected third-quarter financial results of a Hong Kong listed company (WH Group, the largest pork company in the world and a Chinese multinational meat and food processing company) formed the basis of a real-life assignment. This meant that the students acted as practicing journalists who wrote financial news reports, with the help of AI, on the actual company result announcement date.

The students could choose one of the two approaches to complete the assignment:

1. **Human-first, AI-assisted:** The news report was written by the student. Generative AI was used for polishing, grammar checking, correcting factual errors, summarising data, and/or conducting background research. The final version must be manually reviewed, edited, and verified by the student.
2. **AI-first, human-edited:** The initial version of the news report was generated by generative AI using raw materials such as the listed company's press release and its financial results reports. The news report was then fact-checked and revised by the student. The final version must be manually reviewed, edited, and verified by the student.

All students were required to write a reflection paper documenting the approach they used and reflecting on their AI use. The students were informed that the choice of approach (AI-first or human-first) would not affect their grades. The news report would be assessed solely on the quality and accuracy of the final output based on a rubric in terms of content and presentation. When the news writing assignments were graded, the instructor was blind to the approach chosen by each student. After the marking of the assignments was completed, the instructor then reviewed the reflection papers to identify which approach each student had taken. The instructor has taught this course for more than 10 years, and the same assignment was given to his students; however, this was the first time that AI was allowed for the student to complete this assignment.

This case study incorporates the instructor's observational insights alongside quantitative data from assignment scores and qualitative data from reflective papers to evaluate student performance.

3. Results and Discussion

There were altogether 15 students in this undergraduate financial journalism news writing class. Thirteen of them (86.7%) chose the AI-first approach to complete the assignment, while the remaining two students chose the human-first approach. This high level of willingness to incorporate AI tools into academic assignments aligns with previous research demonstrating that university students are rapidly increasing their adoption of generative AI tools in teaching and learning contexts due to perceived efficiency gains (Jeong & Yun, 2025).

The mean score for human-first approach assignments was 7.50 (out of 10), while the AI-first approach averaged 7.89, with scores ranging between 6.5 and 9. These results indicate that higher involvement of AI to finish an assignment does not guarantee better output quality, nor does it necessarily lead to bad performance. Rather, the quality depends on how students employ AI tools.

Qualitative analysis of student writing and reflection papers provides deeper insights into how AI use may influence academic performance. Professional journalistic writing should avoid replicating the jargon and wording found in corporate documents, as such language may be overly technical or too favourable to the company. It was observed that AI-first students' writing was generally more susceptible to influence from source materials, including the specific language used by listed companies. For example, professional norms among financial journalists do not require including "Limited" when reporting company names, as all listed companies are limited by definition. However, the official announcement always uses the full name of the company with the "Limited" word, and thus AI-generated articles were likely to adopt the same term. The human-first student did not make this mistake. Among AI-first students, some successfully identified and corrected this issue by applying knowledge from class instruction, while others failed to recognise the discrepancy.

AI generally demonstrates notable weakness in distinguishing between financial news reporting, commentary, and public relations writing styles. When asked to write a financial news report, it frequently inserted subjective commentary following factual data. For example, one AI-generated passage stated: "Basic earnings per share increased from 8.43 cents to 9.10 cents, demonstrating stable profitability." One student noted that the initial AI draft was "biased," mimicking the "corporate announcement style" with subjective adjectives like "pleased to announce" and "bright performance". The student had to explicitly prompt the AI to "maintain neutrality" and remove subjective adjectives to meet journalistic standards. However, some students failed to address this issue, resulting in submissions containing excessive subjective language.

However, when compared to previous cohorts, the instructor observed improved accuracy in numerical reporting and unit notation with the use of AI in this cohort. Without the help of technology, because financial reporting requires handling substantial numerical data, students frequently made careless errors when transcribing figures in the past. AI assistance appeared to enhance precision. Formatting consistency also improved. Previously, students inconsistently applied units (such as USD) to figures, sometimes including them and sometimes omitting them, despite the requirement that all figures from financial statements include appropriate units in news writing.

An additional benefit of AI use in this assignment, reported by the students in their reflection papers, is that they were more confident in dealing with their "mathematics anxiety" as journalism students. In general, journalism students are more interested in textual and artistic subjects rather than mathematics and

numbers. One student described how AI helped reduce the learning curve and increased her willingness to engage with number-heavy financial statements: “As someone not very familiar with earnings reports, I initially felt a bit lost when looking at earnings reports alone, unable to find the key points. So I think using AI first and then manually revising the method can help me analyse the main highlights of WH Group’s earnings report.” (Student 11)

4. Implications

This study revealed that journalism and communication students exhibit strong enthusiasm for incorporating AI into their academic assignments. When permitted by instructors, the majority select the AI-first, human-edited approach, whereby they request AI to generate an initial draft which they subsequently revise and refine. Previous studies pointed out that over-dependency on AI by students could cause “metacognitive laziness”, meaning that students habitually fail to engage in the learning process and monitor their own learning (Fan et al., 2025). Journalist practitioners also emphasise maintaining human editorial control and judgment in the news production process and are cautious about the AI-first automation approach in the newsroom (van Dalen, 2024). This study suggests these concerns are partially justifiable. Some students, influenced by AI-generated drafts, produced poor news writing, such as using subjective language and unnecessary technical jargon. However, other students employed the AI-first approach effectively by integrating knowledge acquired during the course and refining their prompts to produce high-quality financial news reporting. These results suggest that increased AI involvement is not inherently problematic. Rather, instructors should focus greater attention on guiding students to apply their disciplinary knowledge to uphold professional standards in their work with AI involvement.

Furthermore, this study identifies AI’s potential to serve as a supportive tool helping students navigate unfamiliar subject matter. In this case, journalism students who acknowledged relative weakness in quantitative analysis reported increased willingness to engage with financial statements and extract key information with AI assistance, rather than merely completing assignments mechanically. Instructors from other disciplines worldwide can consider allowing students to have higher levels of AI involvement in assignments addressing content areas students find particularly challenging.

As an exploratory case study, this research has its own limitations. The investigation focused on a single assignment type, and further research could examine different types of journalistic tasks. Also, larger-scale quantitative studies surveying broader student populations would help to understand general

patterns of AI adoption across disciplines and the effects on students' learning outcomes with different levels of AI involvement.

Ethical Declaration Statement

The project has been approved by the Research Ethics Committee of Hong Kong Baptist University (Reference Number: REC/24-25/0622).

Generative AI Policy

The author(s) used GPT-5 and Claude-4.5 for improving language and grammar. The author(s) also used Consensus, a university-subscribed AI-powered research tool, to locate and understand relevant scholarly literature. After utilising these tools and services, the author(s) thoroughly reviewed, verified, and edited all content as necessary and assumed full responsibility for the publication's accuracy, originality, and overall quality.

References

Cools, H., & Diakopoulos, N. (2024). Uses of generative AI in the newsroom: Mapping journalists' perceptions of perils and possibilities. *Journalism Practice*. Advance online publication. https://doi.org/10.1080/17512786.2024.2394558

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*, 56(2), 489-530. https://doi.org/10.1111/bjet.13544

Graefe, A. (2016). *Guide to Automated Journalism*. Tow Center for Digital Journalism. https://doi.org/10.7916/D80G3XDJ

Jeong, Y., & Yun, H.-J. (2025). An analysis of university students' experiences and perceptions of using text-generating AI in assignment writing: Focusing on the perspective of the theory of education-as-the-principle. *The Journal of Educational Principles*, 30(1), 67–110. https://doi.org/10.19118/edp.2025.30.1.67

Kim, J., Yu, S., Detrick, R., & Li, N. (2025). Exploring students' perspectives on Generative AI-assisted academic writing. *Education and Information Technologies*, 30(1), 1265-1300. https://doi.org/10.1007/s10639-024-12878-7

Qu, Y., Tan, M. X. Y., & Wang, J. (2024). Disciplinary differences in undergraduate students' engagement with generative artificial intelligence. *Smart Learning Environments*, 11(1), Article 51. https://doi.org/10.1186/s40561-024-00341-6

Swart, R. (2017). Critical thinking instruction and technology enhanced learning from the student perspective: A mixed methods research study. *Nurse Education in Practice*, 23, 30-39. https://doi.org/10.1016/j.nepr.2017.02.003

van Dalen, A. (2024). Revisiting the algorithms behind the headlines: How journalists respond to professional competition of generative AI. *Journalism Practice*. Advance online publication. https://doi.org/10.1080/17512786.2024.2389209

AI Teammates in Group Discussion: Supporting Collaborative Problem Solving and L2 Speaking

Huong Nam VU [1] & Selim Ben SAID [1]
[1] Foreign Languages and Literatures, National Sun Yat-sen University, Taiwan, China

Abstract

This study investigates how multi-persona conversational AI can support collaborative problem-solving (CPS) and L2 speaking development in higher education. A web-based system, *CollaboSpeak*, was designed to simulate a four-person group discussion in which learners interacted with three AI "teammates" who embodied distinct collaborative roles. Fifteen Taiwanese undergraduates completed a five-week cycle consisting of a human-based pre-test, three AI-mediated sessions, and a human-based post-test. Participants' CPS moves were evaluated using the PISA 2015 CPS framework, and linguistic development was analysed using the Complexity, Accuracy, and Fluency (CAF) framework. Results revealed a significant increase in question-asking behaviour and the emergence of team-organisation moves that were absent at pre-test. Participants also demonstrated increased syntactic complexity alongside stable or improved fluency and accuracy. Survey responses indicated strong acceptance of AI teammates as a low-anxiety speaking practice environment. Findings suggest that multi-persona AI can provide an effective, ecologically valid environment for developing CPS-oriented L2 speaking skills.

Keywords: Multi-persona AI, L2 speaking, collaborative problem-solving, polyadic interaction

1. Introduction

Group discussion is a core component of communication, yet it presents persistent challenges for L2 learners, who must manage turn-taking, interpret social cues, and navigate the complex demands of multi-party discourse. In Taiwanese higher education, English-medium group tasks are increasingly required under national bilingual initiatives. However, many learners experience foreign language anxiety, which often leads to reduced participation and limited opportunities to develop interactional competence (Chou, 2018; Horwitz et al., 1986). This creates a pedagogical paradox: to develop group communication skills, one must speak in groups, yet the anxiety related to group performance often prevents them from doing so.

Although recent advances in generative AI enable convenient, on-demand speaking practice, most AI tools adopt a dyadic (one-on-one) design. Such interactions do not reflect the complexity of multi-party communication, where speakers must negotiate meaning, evaluate alternatives, and manage group organisation. As a result, learners may improve individual fluency but remain underprepared for collaborative speaking tasks.

This study addresses this pedagogical gap by exploring whether multi-persona AI can simulate group dynamics and support collaborative speaking development. A web-based system, *CollaboSpeak*, was developed to allow learners to interact with three distinct AI personas (Supportive, Creative, and Skeptical), each designed to elicit specific collaborative behaviours aligned with the PISA 2015 Collaborative Problem Solving (CPS) framework (OECD, 2017).

The purpose of this study is twofold. First, it examines how AI-mediated polyadic interaction supports learners' CPS behaviours and L2 speaking development. Second, it investigates learner perceptions of using AI teammates for speaking practice. This work contributes to emerging research showing that AI-mediated environments can reduce anxiety and promote willingness to communicate, while extending the field beyond dyadic chatbot interaction toward ecologically valid group communication.

Against the rapid adoption of generative AI in higher education, this study responds to growing interest in how AI can support not only individual practice but also collaborative and interactional learning processes.

2. Methodology

Participants

Fifteen undergraduate L2 English learners at a Taiwanese university participated in a five-week intervention. All fifteen students completed both a pre-test and post-test with three scripted participants (human), as well as three weekly AI-mediated discussion sessions. An additional group of 13 students completed AI-only sessions and contributed perception data but were not included in the pre–post comparison.

System Design

A custom web application, *CollaboSpeak*, was developed to simulate multi-party discussion. Three AI personas—Supportive, Creative, and Skeptical—were implemented using ElevenLabs' Conversational Agent and Gemini Flash 2.0. Each persona embodied specific collaborative roles to scaffold CPS behaviours:

- **The Supportive Teammate:** Tends to agree, build rapport, and acknowledge ideas.
- **The Creative Teammate:** Proposes creative or distracted suggestions, prompting the learner to evaluate and refine ideas.
- **The Skeptical Teammate:** Asks follow-up questions, requests clarification, or raises concerns, providing opportunities for explanation and elaboration.

To encourage listening comprehension and reduce reliance on reading, the interface masked transcripts during real-time interaction. Each session lasted between five and ten minutes and concluded with a short written reflection and AI-generated feedback.

Procedure

The study followed a pre-test, AI intervention, post-test structure. Pre- and post-test tasks were adapted from PISA CPS tasks and conducted with three scripted participants. The three AI sessions presented weekly scenarios. All scenarios focus on consensus building (e.g., choosing a location for a school trip). This task type was selected for the present study because it naturally elicits polyadic discourse: the learner must solicit opinions, evaluate conflicting views, and drive the group toward a final agreement.

Data Analysis

Three main dimensions were analysed:

a. Collaborative Problem-Solving

Transcripts were segmented into AS-units and analysed using the PISA 2015 CPS framework. Codes captured moves related to building shared understanding, taking action, and maintaining team organisation. In addition, the proportion of AS-units functioning as questions was calculated to capture proactive collaborative behaviour.

b. Linguistic Development

Linguistic performance was examined using the CAF framework (Skehan, 2009):

- **Complexity:** Mean Length of AS-Unit (MLAS).
- **Accuracy:** Errors per 100 words.
- **Fluency:** Words per minute, fillers per minute, and percentage of speaking time.

c. Perception

Post-study surveys and reflections were analysed thematically to understand learner attitudes toward AI teammates.

3. Results and Discussion

3.1 Collaborative Problem-Solving

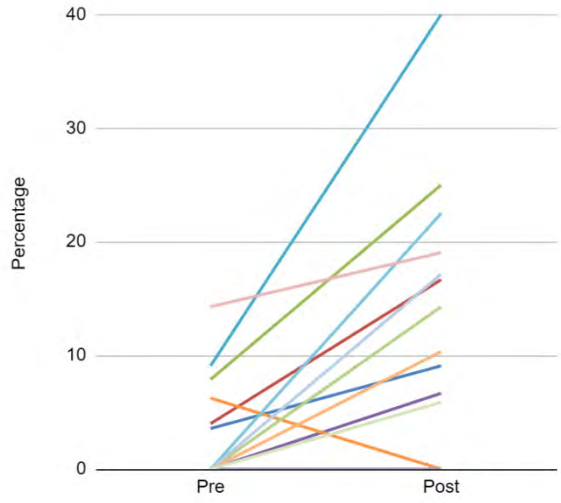
The results indicate a measurable, though modest, improvement in overall CPS proficiency. This shift indicates a transition from Level 2 (collaborating in medium-difficulty spaces) to Level 3 (orchestrating roles and solving complex problems). Notably, the most striking finding from the data is the emergence of Competency 3 (Establishing and maintaining team organisation). In the pre-test, this competency was completely non-existent. This indicates that participants initially functioned mainly as "responders" who answered questions posed by the scripted participants. They contributed to the solution but rarely managed the team.

Following the AI intervention, approximately two-thirds of participants produced at least one organisational move. Participants were observed checking for others' opinions or prompting other members to find a specific piece of information. Although the absolute numbers remain small, the emergence of this previously absent competency indicates that students became more aware of team management and engaged in more agentive participation during group discussion. This shift suggests that interacting with multi-persona AI encouraged learners to adopt more proactive collaborative roles.

Increase in Question-Asking

Figure 1 illustrates a clear upward shift in question production from pre-test to post-test. In fact, the proportion of AS-units coded as questions increased from 3.01% at pre-test to 12.44% at post-test. The increase was statistically significant ($p = .0023$). More importantly, learners used questions not only to seek clarification but also to evaluate alternatives and prompt contributions from peers. This shift suggests an increased willingness to engage collaboratively and to use inquiry as a tool for moving the discussion forward.

(Figure 1: Question Production From Pre-Test To Post-Test)



3.2 Linguistic Development

Learners showed clear signs of linguistic development across the complexity, accuracy, and fluency dimensions. In terms of complexity, many learners began producing longer and more elaborated utterances in the post-test. Their speech contained more subordinated clauses, more explanations, and comparisons, whereas pre-test responses were often short and minimal. Learners more frequently compared options, stated conditions, and elaborated on their choices. This indicates greater willingness to expand their contributions.

Regarding accuracy, learners maintained control of grammar and vocabulary even as they attempted more linguistically demanding utterances. In terms of fluency, learners spoke more smoothly during the post-test discussions. Their speech contained fewer unnecessary pauses or filler words, and they were more comfortable sustaining turns without losing momentum. Students also showed more balanced participation within the group. These patterns suggest that repeated exposure to the AI teammates allowed learners to gain confidence in speaking continuously and managing the flow of multi-party interaction.

Overall, the CAF-related changes point to the potential for polyadic AI-mediated discussions to provide meaningful, low-pressure practice that helps learners develop more complex, accurate, and fluent spoken language.

3.3 Learner Perceptions

Survey results showed strong acceptance of AI teammates: a large majority of participants felt that the three personas were distinct, that the AI generally understood what they were saying, and that the discussions felt natural. Many learners also felt that the AI helped move the discussion forward. This perception of a multi-person team is very visible in the reflections.

One participant wrote: *"Very nice tool to train speaking skills. I think the three AIs have different roles, Fiona is the host, Eli is an amiable person, and Clara has a lot of thoughts. That creates a good conversation!"* Or even showed emotion for a specific persona: *"Alex is energetic." "Clara isn't an enjoyable persona"*

Many expressed excitement and surprise at how AI understands them: *"I'm surprised by how fast and precise they are right after I finished each sentence, but it is a cool experience, especially the part where I randomly brought up jerky."* Others described the AI as colleagues or teammates rather than tutors: *"Be more prepared, my colleagues (AI) are well knowledgeable able to ask proper questions [I] didn't think of."*

Learners highlighted reduced anxiety, psychological safety, and enjoyment: *"One big advantage is that AI doesn't judge me. I can make mistakes, ask simple questions, or repeat without feeling embarrassed." "We don't feel guilty at all to discuss anything because they aren't real human beings."*

Although human discussions were perceived as more natural, the AI sessions were viewed as more efficient and less judgmental, ideal for low-stakes speaking practice. The result clearly shows a strong positive preference for using AI teammates for practice, as two-thirds of participants (67.9%) agree or strongly agree they would prefer to use the AI app instead of doing a similar task with human classmates. However, a significant minority still disagreed or strongly disagreed. This indicates that efficiency does not fully translate to preference; a substantial number of users still prefer or value practising directly with human classmates. After all, 100% would recommend the system to peers.

4. Implications

The findings suggest several pedagogical implications. First, polyadic AI teammates can supplement classroom interaction by giving learners additional opportunities to practice managing group discussions without the pressure of performing in front of peers. This is particularly useful in contexts where learners experience speaking anxiety or where classroom time is limited.

Second, the AI-mediated environment appears to scaffold CPS moves, such as asking questions, organising the team, and evaluating alternatives. These skills are critical not only for language learning but also for academic success and workplace collaboration.

Finally, this study highlights the potential for combining CPS frameworks with CAF measures to understand how learners develop both interactional and linguistic competencies. As generative AI becomes increasingly integrated into educational contexts globally, systems like *CollaboSpeak* offer a promising direction for creating practice environments that are both ecologically meaningful and psychologically supportive.

Several limitations should be noted. The study involved a small sample size and a short intervention period, which limits generalisability. In addition, while the AI-mediated sessions provided repeated practice opportunities, longer-term effects on collaborative speaking development remain to be examined. Future research could examine longer intervention periods, compare different AI persona configurations, or explore how AI-mediated group discussion complements human-only interaction over time. Such work would help clarify how AI teammates can be integrated sustainably into classroom practice.

Declaration of AI Use

Generative AI technologies were used as part of the research design in this study, specifically to implement the multi-persona conversational agent described in the Methodology section.

References

Chou, M.-H. (2018). Speaking Anxiety and Strategy Use for Learning English as a Foreign Language in Full and Partial English-Medium Instruction Contexts. *TESOL Quarterly*, 52(3), 611–633. https://doi.org/10.1002/tesq.455

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign Language Classroom Anxiety. *The Modern Language Journal*, 70(2), 125–132. https://doi.org/10.2307/327317

OECD. (2017). *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematics, Financial Literacy and Collaborative Problem Solving*. OECD. https://doi.org/10.1787/9789264281820-en

Skehan, P. (2009). Modelling Second Language Performance: Integrating Complexity, Accuracy, Fluency, and Lexis. *Applied Linguistics*, 30(4), 510–532. https://doi.org/10.1093/applin/amp047

Reforming Interdisciplinary Postgraduate Teaching with GenAI: A Case Study of the "Laser Irradiation Effects" Course

Dapeng ZHAO [1], Enlong WANG [1], Xi WANG [1]
[1] College of Electronic Engineering, National University of Defence Technology, China

Abstract

The trend toward interdisciplinary postgraduate education presents a dual challenge: integrating students' diverse disciplinary backgrounds with the inherent complexity of cross-disciplinary content. This study examines a Generative AI (GenAI)-assisted pedagogical reform within a postgraduate "Laser Irradiation Effects" course, a field spanning physics, chemistry, materials science, optics, and mathematics. Supported by an Anhui Province education reform initiative, we implemented a student-centred framework integrating GenAI across four key stages: AI-enhanced learner analysis, personalised instructional design, facilitated cross-disciplinary knowledge exploration, and intelligent assessment. Our findings indicate that GenAI provides an effective cognitive framework, helping bridge knowledge gaps, personalise learning pathways, and shift instructional emphasis toward higher-order competencies such as synthesis and critical problem-solving. The approach led to measurable improvements in teaching efficiency, student engagement, and the quality of interdisciplinary project outcomes. This model offers a replicable and scalable paradigm for enhancing both teaching and learning in interdisciplinary higher education contexts.

Keywords: GenAI, interdisciplinary education, postgraduate course, laser irradiation effect

1. Introduction

Postgraduate education increasingly adopts interdisciplinary frameworks to address complex scientific challenges. Courses like "Laser Irradiation Effects," which integrates physics, chemistry, materials science, optics, and mathematics, are essential for developing adaptable researchers. This approach, however, introduces significant pedagogical dilemmas. Instructors must reconcile students' diverse academic backgrounds with the course's integrative aims while maintaining a learner-centred ethos. Concurrently, students face the difficulty of efficiently identifying and bridging knowledge gaps across unfamiliar

disciplines to achieve core competencies.

While established interdisciplinary pedagogies—such as team-teaching, project-based learning, case-based instruction, and flipped classrooms—have demonstrably enhanced collaborative and applied learning, they share a common limitation: the inability to offer personalised, real-time scaffolding that adapts to each learner's unique knowledge trajectory (Gao et al., 2020). For instance, team-teaching integrates multiple expert perspectives yet often struggles to address individual student gaps during live instruction. Project-based learning, though effective for fostering collaboration, typically advances at a group pace, leaving little room for real-time customisation of content to suit varying levels of prior knowledge (Guo et al., 2020). Similarly, case-based and flipped approaches rely heavily on students' self-regulated preparation, which may not adequately support those from significantly divergent disciplinary backgrounds.

The emergence of Generative Artificial Intelligence (GenAI) introduces a paradigm-shifting complement to these methods, offering dynamic support that is both adaptive and immediate (Baidoo-Anu & Ansah, 2023). For instructors, GenAI serves as a co-design partner, significantly accelerating the development of differentiated instructional materials and enabling responsive curriculum adjustments (Mollick & Mollick, 2022). For learners, it functions as an always-available interdisciplinary tutor, capable of translating concepts across domains and facilitating real-time knowledge synthesis. It provides a level of just-in-time support that traditional methods structurally cannot provide (Kasneci et al., 2023).

This paper details a comprehensive teaching reform, supported by an Anhui Province Graduate Education Quality Project, that embedded GenAI throughout the instructional lifecycle of a postgraduate "Laser Irradiation Effects" course. We demonstrate that a systematic GenAI-enhanced framework can effectively address the core challenges of interdisciplinary education, fostering a more efficient, inclusive, and intellectually stimulating learning environment.

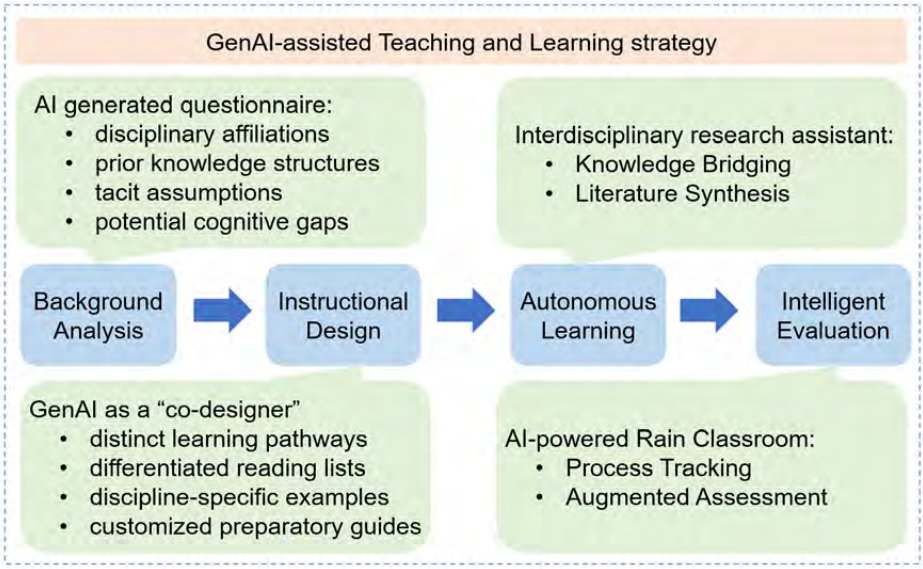
2. Methodology

2.1 GenAI-assisted Teaching and Learning Framework

To address the dual challenges of student diversity and interdisciplinary complexity, we designed and implemented a closed-loop pedagogical framework that strategically embeds GenAI across four interconnected stages. This framework aims to shift from a uniform, instructor-paced delivery to a dynamic, responsive, and personalised learning ecosystem. Each stage leverages GenAI's

capabilities to diagnose, design, facilitate, and assess, thereby creating a coherent support structure for interdisciplinary mastery (see Figure 1 for a conceptual overview).

Figure 1: GenAI-assisted Teaching and Learning Strategy.



2.1.1 Stage 1: AI-Enhanced Diagnostic Analysis of Learner Background

Prior to the commencement of the course, we conducted a foundational diagnostic analysis to move beyond simplistic demographic categorisations. Utilising advanced large language models (LLMs) such as DeepSeek, we developed a sophisticated diagnostic instrument. Rather than employing generic surveys, we engineered specific prompts to guide the AI in generating a contextualised questionnaire. These prompts were based on the course's core syllabus and its interdisciplinary knowledge map. For example, we instructed the AI: "Generate questions that probe a student's conceptual understanding of photothermal and photochemical effects from the distinct perspective of their stated major."

This approach yielded a diagnostic tool that assessed not only students' disciplinary affiliations but also their prior knowledge structures, tacit assumptions, and potential cognitive gaps. The aggregated results provided the instructional team with a nuanced "learner profile matrix," offering critical data to inform subsequent personalised design.

2.1.2 Stage 2: Dynamic and Student-Centred Instructional Design

Informed by the diagnostic profiles from Stage 1, the instructor employed

GenAI as a "co-designer" to translate a monolithic syllabus into flexible, personalised learning pathways. This stage operationalises the principles of differentiated instruction. For each core topic, GenAI was prompted to generate multiple, parallel instructional narratives tailored to major student background clusters.

For instance, for the topic "Laser-Induced Breakdown Spectroscopy (LIBS)," the prompt was: "Develop two distinct learning pathways for a single lecture module: Pathway A for materials science students, framing LIBS primarily as an elemental analysis tool and connecting it to thin-film characterisation challenges. Pathway B for physics students, focusing on the fundamentals of plasma diagnostics and the physical mechanisms underlying spectral line broadening."

This co-design process efficiently produced tailored resources, including differentiated reading lists, discipline-specific illustrative examples, and customised preparatory guides, to create a multi-threaded curriculum adaptable to diverse learner entry points.

2.1.3 Stage 3: GenAI-Facilitated Autonomous Learning

To empower students as active agents in bridging their own knowledge gaps, we integrated GenAI as a sanctioned "interdisciplinary research assistant" into the learning process. Students received structured training in effective prompt engineering to maximise productive interactions. Guided prompting frameworks were emphasised for two key tasks:

1. **Knowledge Bridging:** For example, "I am a biology student. Explain the concept of 'laser ablation threshold' using analogies from cellular membrane rupture or enzyme kinetics."
2. **Literature Synthesis:** For example, "Compare and summarise the last three years' key findings on ultrafast laser surface structuring from the journals *Applied Physics Letters*. Present the summary in a comparative table format."

This sanctioned use of GenAI allowed students to obtain on-demand, contextualised explanations and synthesise literature across fields in real-time, thereby accelerating their transition from foundational knowledge acquisition to applied, collaborative problem-solving in complex group projects.

2.1.4 Stage 4: Intelligent Process Evaluation and Feedback

The final stage focused on implementing a holistic, data-informed as-

sessment strategy leveraging the AI-powered "Rain Classroom" platform. This moved evaluation beyond summative grading to embrace formative, process-oriented feedback.

- **Process Tracking:** Involved using AI algorithms to analyse patterns in discussion forum posts and Q&A interactions, automatically identifying trending misconceptions and visualising the evolution of interdisciplinary dialogue among student groups.
- **Augmented Assessment:** While the final evaluative judgement remained with the instructor, GenAI provided preliminary, analytic scaffolding. For student submissions like research proposals, GenAI was tasked with highlighting strengths and weaknesses in logical coherence across disciplines and accuracy of terminology use. This augmented analysis enabled instructors to deliver more targeted, substantive, and efficient feedback, closing the loop between learning activities and instructional refinement.

2.2 Participants

The study involved 16 students enrolled in the course and two instructors. The student participants were drawn from two consecutive academic years, forming a comparative cohort: six students received traditional instruction, while ten students engaged in a GenAI-assisted instructional approach. The instructor participants comprised one educator who delivered the traditional curriculum and a newly appointed instructor who integrated GenAI into the course design to support pedagogical innovation. Owing to the limited sample size, all participants were included in the study, and each submitted the required reports.

2.3 Survey Instrument

A student questionnaire was developed with the support of GenAI and administered via the Rain Classroom platform to all 16 student participants across the two instructional cohorts. The instrument consisted of ten closed-ended items organised into three key dimensions: efficacy in bridging knowledge gaps, quality of interdisciplinary synthesis, and promotion of higher-order cognitive engagement. An open-ended section was also included to elicit additional comments and suggestions.

To complement the survey data, students' learning analytics—including performance records from the Rain Classroom platform and final course grades—were incorporated into the analysis. In addition, semi-structured interviews were conducted with the two instructors. The interviews explored three primary themes: efficiency in lesson preparation, instructional precision, and

the evolving role of the teacher in the context of GenAI-supported teaching.

3. Results and Discussion

Our analysis of data from course analytics, final project evaluations, and student surveys indicates that the GenAI-enhanced pedagogical model yielded significant positive outcomes, measurable in both instructional efficiency and student learning.

3.1 Instructional Efficiency and Precision

A primary quantitative outcome was a substantial reduction in the time investment required for instructional preparation. The newly appointed instructor, who integrated GenAI into the course design, reported a substantial reduction in time devoted to developing differentiated instructional materials. Specifically, the instructor estimated a 35–40% decrease in preparation time compared to the counterpart delivering traditional instruction. This efficiency gain was attributed to GenAI's capacity to rapidly generate tailored examples, construct explanatory analogies, and design customised learning pathways, thereby streamlining the material development process.

Qualitatively, the AI-generated learner profiles enabled a shift from broad, assumption-based planning to data-informed, pinpoint instructional interventions. This allowed for the precise identification and targeted addressing of cohort-wide challenging concepts such as non-linear optical effects, and facilitated the provision of individualised resource recommendations with markedly greater accuracy.

3.2 Student Learning Outcomes and Engagement

Student-reported and performance-based data confirmed substantial improvements in learning processes and outputs.

(1) **Efficacy in Bridging Knowledge Gaps:** Survey results revealed that over 80% of students found using GenAI for clarifying cross-disciplinary concepts to be "significantly faster" than traditional search methods, which directly contributed to reducing initial learning anxiety and lowering barriers to entry into unfamiliar domains.

(2) **Quality of Interdisciplinary Synthesis:** A comparative analysis of final group project rubrics demonstrated a measurable enhancement in the depth of interdisciplinary integration. Compared to a previous cohort taught using traditional methods, the intervention cohort's

average scores for “multidisciplinary synthesis” and “innovative approach” increased by 22% and 18%, respectively. This suggests that the model effectively supported students in integrating diverse knowledge strands into coherent, novel solutions.

(3) **Promotion of Higher-Order Cognitive Engagement:** By providing foundational knowledge support, AI accelerated the cognitive progression of classroom work, enabling a quicker shift to evaluative and creative tasks. A key example was a project by physics and electronic engineering students, who designed an experiment for real-time correlated analysis of laser-induced damage in photodetectors. This involved monitoring both transient optical responses and subsequent electrical degradation. The concept originated from GenAI-assisted dialogue, which facilitated rapid translation of concepts between laser-matter interaction physics and detector performance principles, allowing students to synthesise a coherent interdisciplinary research question.

3.3 Discussion

This practice demonstrates that, when strategically embedded, GenAI has the potential to fundamentally reshape interdisciplinary education. The instructor’s role transitions from that of a sole knowledge authority to a designer of learning experiences and a facilitator of critical inquiry. Concurrently, the student’s role evolves from a passive recipient of information to an active architect of knowledge, leveraging AI as a tool for exploration and interdisciplinary synthesis.

Owing to the small sample size, statistical reliability analysis was not conducted. To enhance the trustworthiness of the findings, the study employed methodological triangulation, integrating survey data with semi-structured interviews and learning analytics from the Rain Classroom platform. Nevertheless, this study represents a single-case implementation, which limits the generalisability of the findings. Future research should include longitudinal studies across multiple disciplines and institutional contexts to validate and extend these observations. The development of domain-specific, fine-tuned AI tutors for interdisciplinary subjects holds promise for enhancing both the accuracy and relevance of AI-generated support. Moreover, establishing open repositories of effective pedagogical prompts and AI-integrated assignment designs would contribute to the cultivation of a collaborative community of practice.

4. Implications

This study confirms the potential of GenAI as a transformative agent in interdisciplinary postgraduate education. The proposed "Analysis-Design-Learning-Assessment" framework provides a concrete methodology for leveraging AI to address the dual challenge of diverse student backgrounds and complex interdisciplinary content. By acting as both a scaffold for knowledge integration and a catalyst for autonomous learning, GenAI can significantly enhance teaching efficiency, improve learning quality and satisfaction, and ultimately cultivate the agile, synthetic thinkers required to solve tomorrow's complex problems. The insights and model presented here offer a valuable reference for educators aiming to reform advanced courses across the spectrum of interdisciplinary science and engineering.

Acknowledgements

This project is supported by the Anhui Province Graduate Education Quality Project (No. 2024jyjxggyjY270).

Ethical Declaration Statement

The project has been approved by the Ethics Committee of the College of Electronic Engineering, National University of Defence Technology (Reference Number: JIAO20260002).

Declaration of Use of AI in the Writing Process

The authors used DeepSeek, Meta Sota and Rain Classroom during the conduction of this project to generate questionnaires, assist teaching and student learning, and for assessment. After utilising the tool/service, the authors thoroughly reviewed and edited the content as necessary and assumed full responsibility for the publication's content.

References

Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning. *Journal of AI*, 7(1), 52-62. https://doi.org/10.61969/jai.1337500

Gao, X., Li, P., Shen, J. & Sun, H. (2020) Reviewing assessment of student learning in interdisciplinary STEM education. *International Journal of STEM Education* 7, 24. https://doi.org/10.1186/s40594-020-00225-4

Guo, P., Saab, N., Post, L. S., & Admiraal, W. (2020). A review of project-based learning in higher education: Student outcomes and measures. *International Journal of Educational Research*, 102, Article 101586. https://doi.

org/10.1016/j.ijer.2020.101586

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., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Mollick, E. R., & Mollick, L. (2022). New modes of learning enabled by AI chatbots: Three methods and assignments. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4300783

Designing Authentic Assessments: Mitigating Academic Integrity Risks Posed by Generative AI Tools

Chun Fan WONG [1]
[1] Graduate School, Vern' University, Croatia

Abstract

The rapid proliferation of sophisticated Generative Artificial Intelligence (GenAI) tools, such as large language models (LLMs), presents an unprecedented challenge to traditional assessment paradigms in higher education. Current methods relying heavily on summative written submissions are increasingly vulnerable to sophisticated academic misconduct, rendering detection mechanisms an unsustainable "arms race." This paper argues that the sustainable solution lies not in prohibition, but in pedagogical transformation—specifically, the rigorous design of authentic assessments that necessitate human-centric skills, contextual synthesis, and demonstrable process. A conceptual framework is proposed, built upon four core principles: Contextual Specificity, Process Over Product, Multimodal Integration, and Stakeholder Justification. Through practical examples across various disciplines, the paper further illustrates how these principles facilitate the creation of assessments where GenAI can serve as a collaborator in drafting or ideation, rather than a substitute for genuine intellectual engagement, thereby ensuring the integrity of learning outcomes in the age of hybrid intelligence.

Keywords: Incorporating GenAI in Learning, Teaching and Assessment; Ethics and Integrity

1. Introduction

Higher education stands at a critical juncture where technological advancement fundamentally challenges established practices. The integration of technology into teaching and learning is no longer optional; yet, the emergence of powerful Generative AI tools has immediately exposed the fragility of conventional assessment strategies. The core promise of higher education—that grades accurately reflect a student's mastery of complex concepts and skills—is now under direct threat when unsupervised submissions can be generated instantaneously by non-human entities.

The conference theme, "Integrating Pedagogy and Technology in Higher Education to Transform Teaching, Learning and Assessment," provides the necessary lens through which to view this crisis. While technology offers immense potential for personalised learning and efficiency, its misuse demands an imme-

diate pedagogical response. This paper focuses specifically on the intersection of incorporating GenAI in learning, teaching and assessment and ethics and integrity.

Traditional assessments, often relying on summarising existing knowledge or constructing standard argumentative essays, are now easily outsourced to LLMs. Detection software, while improving, operates in a perpetual state of reaction, chasing the latest model update. Therefore, this paper posits that mitigation of academic integrity risk must be achieved through proactive assessment design, shifting the focus from product verification to process demonstration.

The aim of this paper is twofold: first, to critically analyse why existing assessment structures fail in the presence of sophisticated GenAI; and second, to introduce and elaborate upon a robust framework for designing authentic, AI-resistant assessments that uphold rigorous academic standards while acknowledging the reality of hybrid intelligence in the learning ecosystem.

2. Methodology

Given the need to develop a robust pedagogical framework grounded in current evidence regarding GenAI capabilities and established learning theories, the methodology employed is a Systematic Literature Review (SLR) followed by Conceptual Synthesis. This approach moves beyond a general literature review by employing rigorous search and synthesis methods to justify the proposed framework.

2.1 Systematic Literature Review Protocol

The SLR focused on peer-reviewed articles and scholarly reports published between 2021 and 2025 concerning Generative AI, Large Language Models (LLMs), academic integrity policies, and assessment redesign strategies.

Search Strategy and Keywords: The search utilised databases including Google Scholar and ERIC. Key search strings combined terms related to technology and assessment: (“Generative AI” OR “LLM”) AND (“Academic Integrity” OR “Misconduct”) AND (“Assessment Redesign” OR “Authentic Assessment”) AND (“AI” OR “ChatGPT”) AND (“Assessment” OR “Evaluation”) AND (“Process Documentation” OR “Iterative Drafting”).

Inclusion and Exclusion Criteria:

- **Inclusion:** Studies published in English, empirical or conceptual papers directly addressing assessment challenges or solutions related to GenAI

in higher education, and foundational texts on authentic assessment (Wiggins, 1990).
- **Exclusion:** Opinion pieces without empirical grounding, technical papers solely focused on detection algorithms, and literature predating 2021 unless foundational (e.g., Wiggins, 1990).

Data Extraction and Synthesis: Data extraction focused on identifying documented GenAI capabilities that threaten assessment, limitations of current detection methods, and proposed pedagogical adaptations. The synthesis process involved thematic analysis to identify converging recommendations for assessment modification, which directly informed the development of the conceptual framework.

2.2 Conceptual Framework Development

The synthesis directly informed the framework development by integrating established principles of authentic assessment (Wiggins, 1990) with findings on AI limitations. The goal was to develop actionable design principles that maximise the cognitive load required for genuine human input, thereby rendering generic AI output insufficient.

3. Results and Discussion

The literature review yielded two significant, interconnected results. First, there is a near-universal consensus that current AI detection methods are insufficient and prone to false positives, making them an unreliable foundation for academic integrity enforcement (Kim et al., 2025). Second, the most effective pedagogical response involves shifting assessment focus away from the final product toward the process of learning and the context of application.

3.1 Conceptual Framework: Principles of AI-Resilient Assessment Design

Authentic assessment, in this context, moves beyond simple knowledge recall and requires students to apply knowledge and skills in ways that mimic real-world professional practice. To specifically counter the capabilities of GenAI, our framework integrates four guiding principles designed to maximise the cognitive load required for genuine human input (Giray et al., 2025).

3.1.1 Principle 1: Contextual Specificity and Local Knowledge Integration

GenAI models excel at generalised knowledge synthesis but struggle with highly specific, localised, or proprietary information. Therefore, assessments must be anchored to unique local contexts, requiring analysis of specific institutional policies, internal departmental reports, or datasets accessible only within the course environment. Additionally, assessments should mandate the integra-

tion of live, in-class discussions, guest speaker insights, or unique field observations that occur during the assessment window. By requiring synthesis based on specific, non-public case studies provided exclusively for a given cohort, students are compelled to engage directly with course material. This renders generic AI output easily recognisable as insufficient or irrelevant.

3.1.2 Principle 2: Process Over Product: Mandatory Iterative Documentation

If the final product is the only artefact evaluated, the risk to integrity remains high. An AI-resilient design emphasises the importance of the creation journey. This includes implementing mandatory checkpoints to document thought progression, such as requiring annotated drafts, early outlines, or partial drafts that demonstrate structural planning. Students should also maintain AI usage logs to transparently document how and where they applied AI, thereby shifting the ethical burden from hiding usage to transparent integration. Additionally, requiring reflective memos with each submission facilitates a detailed account of challenges encountered and intellectual breakthroughs made during the drafting process. This focus on the process allows educators to identify genuine struggles and learning moments, distinguishing them from instantaneous, unearned output (Cotton et al., 2023).

3.1.3 Principle 3: Multimodal Integration and Representation

Given that large language models (LLMs) primarily excel in text generation, assessments demanding the integration of diverse media formats increase complexity beyond what current text-based LLMs can manage. Tasks should require students to translate complex ideas across various communication modes. For instance, students might be asked to take a written analysis and transform its core findings into an infographic, conceptual map, or data visualisation using tools like Tableau. Alternatively, requirements for a short recorded presentation or podcast segment can compel students to verbally defend the nuances of their written arguments, promoting spontaneous articulation rather than rehearsed recitation. This approach forces students to demonstrate mastery through multiple cognitive pathways, which AI cannot easily navigate or authentically simulate in a single submission.

3.1.4 Principle 4: Stakeholder Justification and Real-World Application

Authentic assessment simulates professional tasks where work must be defended, justified, or presented to an audience with vested interests. Thus, assignments should be framed as professional deliverables instead of academic exercises. For example, instead of writing an essay on sustainable urban planning, students could draft a policy brief addressed to the local City Council, adopting

the tone, structure, and persuasive language appropriate for that audience. Similarly, rather than offering a theoretical critique, students might design a training module for new employees on an ethical dilemma, requiring them to anticipate stakeholder reactions and justify their pedagogical choices. When students are aware that their work must withstand scrutiny from a simulated client or stakeholder, the incentive to submit unverified AI content diminishes significantly.

The effectiveness of these principles in mitigating GenAI risk is best understood by distinguishing between discursive and structural changes to assessment design. While current institutional responses often rely on discursive changes—communicating rules about permissible AI use (like traffic light systems or declarations)—these methods create an 'enforcement illusion' because they depend entirely on unenforceable student compliance. Conversely, the four principles outlined above constitute structural changes, as they fundamentally reshape the mechanics of the assessment task itself, building validity into the architecture rather than imposing it via rules (Corbin et al., 2025).

Furthermore, this structural redesign must align with evolving learning outcomes; research indicates that GenAI's impact necessitates a focus on career-driven competencies and lifelong learning skills, such as critical thinking and creativity (Weng et al., 2024), which these principles are designed to foster by increasing GenAI susceptibility through criteria like Intellectual Task Complexity and Collaborative Authorship (Bannister et al., 2025).

3.2 Implementation Strategies

These four principles are designed to be integrated collectively within a single, complex assessment task to maximise AI resistance. While an instructor could apply one principle to one assignment (e.g., Principle 1 to a case study) and another principle to a different assignment (e.g., Principle 3 to a final presentation), the highest level of integrity assurance is achieved when a single major deliverable incorporates all four elements: it uses contextually specific data (P1); it requires submission of process documentation including AI logs (P2); it demands output in both written and visual/oral formats (Multimodal, P3); and it is framed as a stakeholder deliverable requiring defence (P4). This integrated approach creates a multi-layered cognitive barrier that GenAI cannot easily overcome without significant, verifiable human input at each stage.

To illustrate the practical application of these principles, three common assessment types and how they can be redesigned for AI resilience were examined.

3.2.1 Case Study Analysis (Discipline Agnostic)

In traditional settings, AI summarises facts. The AI-resilient redesign introduces proprietary data points. The process requires submission of a "Decision Rationale Matrix" (P2) showcasing three competing courses of action, explicitly referencing the proprietary data (P1). Final justification requires a 5-minute video presentation (P3) summarising the decision and addressing a pre-assigned "what if" scenario to a simulated executive board (P4).

3.2.2 Literature Review/Synthesis (Humanities/Social Sciences)

The redesign focuses the review on a niche area, requiring integration of two specific, recent articles discussed only in the last two weeks of class (P1). The process mandates submission of a "Conceptual Mapping Document" (P3) visualising argument interaction, accompanied by a written reflection on intellectual evolution (P2). The final paper must include a section titled "My Intervention," where students critique an established argument using a framework discussed early in the semester, ensuring integration across the term and providing a defence point (P4).

3.2.3 Coding/Problem Solving (STEM Fields)

Problems utilise specific, institutionally licensed software libraries or unique hardware configurations only available in the lab (P1). Students submit a Git repository documenting every commit, including comments explaining rationale, particularly when debugging AI-assisted attempts (P2). Final assessments involve a mandatory 15-minute one-on-one session where the instructor asks students to explain the most complex function in their code, compelling spontaneous articulation (P3/P4).

The adoption of AI-resilient assessment design necessitates a fundamental shift in the instructor's role—moving from a gatekeeper focused on detecting fraud to a facilitator focused on coaching intellectual development. When assessments are designed around process, context, and justification, the use of GenAI shifts from being an act of cheating to a potentially valuable, albeit monitored, drafting tool. This approach supports the broader conference goal of Integrating Pedagogy and Technology. If students learn to use AI responsibly—to brainstorm, summarise background material, or check initial syntax—they are developing the Hybrid (Human-AI) Intelligence crucial for the future workplace. The ethical challenge is managed not by banning the technology, but by structuring tasks where the human contribution (contextualisation, critical evaluation, justification) remains the non-negotiable core of the learning outcome. Furthermore, this shift inherently supports Student AI Literacy and

AI-Readiness. By requiring transparency (Principle 2), educators teach students the critical skill of evaluating AI output for bias, accuracy, and relevance—a skill far more valuable than simple content generation proficiency.

While implementing these changes requires significant upfront investment in course redesign and potentially more time spent on personalised feedback during process checkpoints, the long-term benefit is the preservation of academic rigor and the validation of earned credentials.

4. Implications

The overall conclusion drawn from this synthesis is that the integration of Generative AI into academic life is permanent, requiring a fundamental shift in assessment philosophy. The major finding—that authentic, process-focused assessment is the most robust defence against misuse—has significant implications for global educational practice. For theory, this reinforces the enduring value of constructivist assessment models that value situated performance over decontextualised testing (Lave & Wenger, 1991). For practice, institutions must invest heavily in faculty development focused on designing assessments that require students to do something unique with knowledge, rather than merely recite it. This pedagogical transformation is crucial for ensuring that academic credentials retain their value in a future where information synthesis is automated, but critical judgment remains uniquely human.

Generative AI Policy

The author(s) used Gemini during the preparation of this work to enhance the clarity of paragraphs. 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

Bannister, P., Urbieta, A. S., & Alvira, N. B. (2025). Appraising higher education assessment validity: Development of the PANDORA GenAI Susceptibility Rubric. *Journal of Applied Learning and Teaching*, 8(1), 41–55. https://doi.org/10.37074/jalt.2025.8.1.20

Corbin, T., Dawson, P., & Liu, D. (2025). Talk is cheap: Why structural assessment changes are needed for a time of GenAI. *Assessment & Evaluation in Higher Education*, 1–11.

Cotton, D. R. E., Cotton, P. A., & Shipway, R. J. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 60(6), 800–811. https://doi.org/10.1080/14703297.2023.2190148

Giray, L., Sevnarayan, K., & Madiseh, F. R. (2025). Beyond policing: AI writing detection tools, trust, academic integrity, and their implications for college writing. *Internet Reference Services Quarterly*, 29(1), 83–116. https://doi.org/10.1080/10875301.2024.2437174

Kim, J., Kelly, S., & Prahl, A. (2025). Navigating AI in education: Foundational suggestions for leveraging AI in teaching and learning. *Communication Education*, 74(1), 104–113. https://doi.org/10.1080/03634523.2024.2447235

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press. https://doi.org/10.1017/CBO9780511815355

Weng, X., Qi, X. I. A., Gu, M., Rajaram, K., & Chiu, T. K. (2024). Assessment and learning outcomes for generative AI in higher education: A scoping review on current research status and trends. *Australasian Journal of Educational Technology*, 40(6), 37–55. https://doi.org/10.14742/ajet.9540

Wiggins, G. (1990). The case for authentic assessment. *Practical Assessment, Research, and Evaluation*, 2(2), 1–2. https://doi.org/10.7275/ffb1-mm19

Transparent Rubrics with GenAI: Faster Feedback and Fairer Grading in Large Classes

Yang YANG [1], Kai XU [1], Junyao HOU [1], Miao ZHANG [1], Zhengqiu ZHU [1], Quanjun YIN [1]
[1] College of Systems Engineering, National University of Defence Technology

Abstract

Large classes often experience slow feedback and uneven grading, even when rubrics exist. This paper presents a rubric-first workflow that integrates GenAI to deliver faster feedback and support fairer grading in large classes. We implement transparent rubrics as machine-actionable criteria with clear performance descriptors and required evidence anchors. Given a student submission and the transparent rubric, GenAI produces (i) criterion-level score suggestions, (ii) cited excerpts that justify each judgement, and (iii) actionable, rubric-aligned feedback. Human graders retain final authority through a lightweight calibration loop. We evaluate the workflow in a large undergraduate class using turnaround time, feedback coverage, inter-rater reliability, deviation from final human scores, and students' perceptions of transparency and fairness. Results show faster feedback, more consistent criterion-level judgements, and higher perceived transparency, supporting fairer grading in large classes.

Keywords: Transparent rubrics, Generative AI (GenAI), Grading consistency and fairness

1. Introduction

Large classes have become a defining feature of higher education worldwide, yet assessment practices have not scaled at the same pace (Lee, 2023). In many high-enrolment courses, instructors and teaching assistants (TAs) must grade large volumes of student work while meeting tight turnaround expectations. The outcome is familiar: feedback is returned late, often after students have already progressed to the next task, and grading quality varies across graders, sections, and time. These challenges persist even when courses adopt rubrics. Although rubrics can clarify expectations, they are often interpreted inconsistently, applied unevenly, or used without explicit links to evidence in the student submission (Taylor et al., 2024). For students, this can translate into feedback that is generic or poorly grounded in their submission, making it difficult to identify concrete next steps and undermining confidence in grading fairness.

Recent advances in generative artificial intelligence (GenAI) have renewed interest in scalable feedback and assessment support (Memarian & Doleck, 2023). GenAI systems can produce written feedback quickly and summarise performance, which is appealing in large classes where instructor time is limited. In grading, however, GenAI raises concerns about accountability, transparency, and bias, and may produce generic, overconfident, or weakly grounded feedback if unconstrained (Bhullar et al., 2024; Shen et al., 2024). For GenAI to play a credible role in assessment, it must be embedded in a pedagogy- and rubric-driven process that preserves human authority, supports auditability, and makes the basis for judgements clear to learners (Deng et al., 2025).

This paper addresses slow feedback and uneven grading in large classes by proposing a rubric-first workflow that integrates GenAI to deliver faster feedback and support fairer grading. A core element of the workflow is the concept of "transparent rubrics," which we operationalise as machine-actionable criteria with clear performance descriptors and required evidence anchors. Rather than treating a rubric as a static scoring sheet, we require each criterion-level judgement to be supported by specific excerpts from the student submission. Given a submission and the transparent rubric, GenAI generates three structured outputs: criterion-level score suggestions, cited excerpts that justify each judgement, and actionable feedback aligned with the rubric and focused on improvement. Human graders retain final authority through a lightweight calibration loop. We evaluate the workflow in a large undergraduate class. Our analysis considers both operational and educational outcomes, including turnaround time, feedback coverage, inter-rater reliability, deviation from final human scores, and students' perceptions of transparency and fairness. By aligning GenAI assistance with transparent rubrics and selective human oversight, this study offers a practical and auditable approach to scaling assessment without compromising trust and consistency.

In summary, this work makes three contributions:

1. We introduce **transparent rubrics**, a machine-actionable rubric representation that requires criterion-level evidence anchors, enabling auditability and consistent application.
2. We propose a **rubric-first GenAI-assisted assessment workflow** that produces score suggestions, cited justifications, and actionable feedback, coupled with a lightweight human calibration loop.
3. We report an *in-situ* evaluation in a large undergraduate course, measuring operational efficiency and grading consistency as well as students' perceptions of transparency and fairness.

2. Methodology

2.1 Context and Participants

The study was conducted in a large-enrolment undergraduate course in which written assignments contributed substantially to summative assessment. Marking was distributed across the instructor and a team of teaching assistants (TAs) using a shared rubric and a fixed turnaround schedule. The focal assessment task was a rubric-graded written submission (e.g., a structured report or short essay) requiring students to present a clear argument supported by evidence.

The study population comprised submitting students and the graders (the instructor and TAs). Where feasible, records used for analysis were anonymised prior to reporting. The course enrolled 150 students; 1,500 submissions were analysed by 5 TAs and 1 lead instructor.

2.2 Implementation Procedure

The workflow was implemented in four stages:

1. **Rubric preparation:** The course rubric was converted into a transparent, evidence-anchored format.
2. **GenAI assistance:** For each submission, GenAI generated structured outputs aligned to the rubric, including criterion-level score suggestions, cited evidence anchors, and concise improvement-oriented feedback drafts. GenAI outputs were produced using GPT-4o (OpenAI) with a fixed rubric-first prompt template (rubric + submission text + criterion-by-criterion instructions for score suggestions, evidence anchors, and feedback); generation parameters used default API settings unless otherwise noted.
3. **Human oversight:** For each submission, graders reviewed the GenAI rubric-aligned draft, accepted or edited it, and finalised rubric scores and feedback.
4. **Release and data capture:** Final grades and feedback were released to students, and data were captured for analysis from three sources:
   (i) Process timestamps and/or marking logs (submission deadline, grading activity where available, and feedback release) to compute turnaround time and grading time;
   (ii) Rubric score records at both the criterion and overall levels, including the final adjudicated scores and a subset of submissions that received independent double-marking by two graders for reliability estimation. For this double-marked subset, we computed weighted agreement (weighted κ) on the 1–5 ordinal scale at the

criterion level, and we also calculated score deviation between GenAI suggestions and final human-adjudicated scores;

(iii) A post-feedback survey was administered after grade release to assess perceived transparency, fairness, and usefulness, supplemented by optional open-ended comments.

2.3 Transparent Rubrics: Construction and Operationalisation

To support auditability and reduce interpretive drift, we operationalised transparent rubrics as a machine-actionable specification in which each criterion includes (i) observable performance descriptors and (ii) an explicit evidence-anchoring rule. Each criterion in the template contains four fields: criterion name, level descriptors (e.g., High/Mid/Low), required evidence type, and feedback focus. This makes rubric application inspectable because every criterion-level judgement must be linked to text in the submission. Descriptors were written in behavioural and text-visible terms (e.g., "states a clear claim and supports it with at least two course-relevant sources") rather than broad evaluative adjectives (e.g., "excellent argument").

For evidence anchors, we required graders (and GenAI) to cite short, located spans such as paragraph numbers, sentence quotations, section headings, or line ranges. Where feasible, students submitted work with headings and numbered sections/paragraphs to reduce ambiguity and enable fast retrieval.

Example (excerpt): For the criterion *Use of Evidence*, a "High" descriptor required claims to be supported with specific citations and interpretation of sources; the evidence anchor rule required quoting the claim sentence(s) and the corresponding cited support from the submission. A "Low" descriptor captured unsupported assertions or evidence not connected to the claim; anchors were required for both the claim and the weak/missing evidence. This anchoring rule was applied across submissions and used for both GenAI drafts and human review.

2.4 GenAI-Assisted Scoring and Feedback Generation

For each submission, GenAI received the student work and the transparent rubric template. Prompts constrained GenAI to: follow rubric criteria only, provide anchors for each judgement, and provide improvement guidance. GenAI produced three structured outputs:

1. **Criterion-level score suggestions:** Recommended level/score per criterion (not final grades).
2. **Cited excerpts (evidence anchors):** Specific passages supporting each judgement; at least one anchor per criterion. If relevant evidence could

not be located, GenAI was instructed to state "insufficient evidence" rather than infer or fabricate a justification.

3. **Actionable, rubric-aligned feedback:** Brief, criterion-specific revision steps phrased using rubric descriptors.

3. Results and Discussion

Table 1 summarises outcomes for the rubric-first GenAI workflow, using a 1–5 rubric scale for criterion-level scoring.

Table 1: Outcomes of the GenAI workflow across efficiency, consistency, and student perceptions.

Outcome (RQ)	Baseline	GenAI	Change
Turnaround time (days) (RQ1)	10.2	6.4	-3.8 (-37%)
Grading time per submission (min) (RQ1)	14.8	10.9	-3.9 (-26%)
Feedback coverage: full criterion-level comments (RQ1)	62%	91%	+29%
Inter-rater reliability (weighted κ, criterion-level) (RQ2)	0.54	0.68	+0.14
Deviation from final human scores (MAE) (RQ2)	-	0.32	-
Borderline cases escalated for lead review (RQ2)	-	12%	-
Perceived transparency (1–5 Likert, mean) (RQ3)	3.4	4.2	+0.8
Perceived fairness (1–5 Likert, mean) (RQ3)	3.5	4.1	+0.6

Note. Baseline is a prior comparable assignment graded under the same rubric and schedule without GenAI suggestions. MAE = mean absolute error between GenAI suggestions and final human-adjudicated scores, reported in points on the 1–5 rubric scale. Survey results are based on 80 responses (53.3% response rate; denominator = 150 enrolled students); weighted κ was computed on a double-marked random subset (n = 50).

RQ1 (Efficiency) results indicate faster feedback and reduced marking effort. Turnaround time decreased from 10.2 to 6.4 days (−37%), and average grading time per submission fell from 14.8 to 10.9 minutes (−26%). The workflow also increased feedback coverage: the proportion of submissions receiving full criterion-level comments rose from 62% to 91% (+29 percentage points). These changes are practically meaningful for large-class operations because they shorten the feedback loop while expanding the amount of rubric-aligned guidance students receive.

RQ2 (Grading Consistency as a Proxy for Fairer Grading) was improved at the criterion level. Weighted agreement across graders increased from $\kappa = 0.54$ to $\kappa = 0.68$, suggesting a more stable interpretation and application of rubric descriptors. GenAI score suggestions were generally close to final human-adjudicated scores (MAE = 0.32 points on the 1–5 scale), with larger

discrepancies concentrated in a small subset of submissions. Consistent with the workflow design, 12% of cases triggered borderline review, typically due to weak or missing anchors, atypical structure, or ambiguous evidence. The evidence-anchoring rule supported moderation: when a judgement could not be justified by located excerpts, graders revised the score or flagged insufficient evidence, reducing the risk of ungrounded scoring.

RQ3 (Student Perceptions) was positive. Students reported higher perceived transparency (3.4 → 4.2) and fairness (3.5 → 4.1). Open-ended comments frequently attributed these gains to excerpt-linked justifications, which clarified why a score was assigned and highlighted concrete revision targets.

Overall, the results suggest that a rubric-first GenAI workflow can improve efficiency, feedback coverage, and grading consistency in large classes while keeping final decisions with human graders. Evidence anchors strengthened accountability by making criterion-level judgements inspectable and by focusing human review on borderline cases. Students' higher ratings of transparency and fairness further indicate that excerpt-linked justifications and rubric-aligned next steps made feedback easier to understand and act on.

4. Implications

This study shows that a rubric-first GenAI workflow can scale assessment in large classes by improving turnaround time, expanding criterion-level feedback coverage, and supporting more consistent rubric application, while preserving human authority over final grades. The key implication for assessment theory is that "transparency" can be operationalised as an auditable linkage between rubric descriptors and evidence anchors, making judgement formation inspectable rather than implicit. Practically, the workflow offers a viable model for institutions facing global pressures of massified higher education, limited instructional staffing, and expectations for timely, high-quality feedback.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grants No. 62103420, and in part by the China Postdoctoral Science Foundation under Grant No. 2025M784474.

Ethical Declaration Statement

The project has been approved by the College of Systems Engineering of National University of Defence Technology.

Declaration of Use of AI in the Writing Process

The authors used ChatGPT during the preparation of this work to identify grammatical issues and improve clarity and phrasing in the manuscript text. After utilising the tool/service, the authors thoroughly reviewed and edited the content as necessary and assumed full responsibility for the publication's content.

References

Bhullar, P. S., Joshi, M., & Chugh, R. (2024). ChatGPT in higher education: A synthesis of the literature and a future research agenda. *Education and Information Technologies*, 29, 21501–21522. https://doi.org/10.1007/s10639-024-12723-x

Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education*, 227, 105224. https://doi.org/10.1016/j.compedu.105224

Lee, A. V. Y. (2023). Supporting students' generation of feedback in large-scale online course with artificial intelligence-enabled evaluation. *Studies in Educational Evaluation*, 77, 101250. https://doi.org/10.1016/j.stueduc.101250

Memarian, B., & Doleck, T. (2023). ChatGPT in education: Methods, potentials, and limitations. *Computers in Human Behavior: Artificial Humans*, 1(2), 100022. https://doi.org/10.1016/j.chbah.100022

Shen, Y., Ai, X., Raj, A. G. S., John, R. J. L., & Syamkumar, M. (2024). Implications of ChatGPT for data science education. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education (SIGCSE TS 2024)*, Volume 1 (pp. 1230–1236). Association for Computing Machinery. https://doi.org/10.1145/3626252.3630874

Taylor, B., Kisby, F., & Reedy, A. (2024). Rubrics in higher education: An exploration of undergraduate students' understanding and perspectives. *Assessment & Evaluation in Higher Education*. Advance online publication. https://doi.org/10.1080/02602938.2023.2299330

Integrating AI-Student Paired Programming with a Multi-dimensional Pedagogy for Enhanced Computer Programming Learning

Hai Wei FENG [1], Qi ZHANG [1], Shuo YANG [1], YingMei WEI [1]
[1] College of Systems Engineering, National University of Defense Technology, China

Abstract

The rapid advancement of generative AI has posed significant challenges to computer programming education. Its code generation capabilities enable students to complete programming exercises with minimal effort, thereby undermining their motivation, creativity, and skill development. To address this, our study focuses on enhancing the effectiveness of in-class programming exercises, which are central to course learning. Specifically, we designed a set of AI-student paired programming exercises, each constrained to a 10-minute completion time. We also introduced a structured three-dimensional "read-write-comment" pedagogical framework to foster critical engagement and assess genuine learning during these activities. This model emphasises active AI collaboration, analytical code review, and reflective commentary, shifting the students' focus from outcome-oriented to process-oriented learning. Implemented over one semester, results show that classes using this approach achieved consistently and statistically higher average scores on final exams compared to classes taught traditionally.

Keywords: Generative AI, programming exercise, read-write-comment

1. Introduction

Programming ability is an essential foundational skill for college students to meet the demands of future information-based positions and to undertake complex software development tasks. For a long time, computer programming courses have served as core courses in the talent cultivation plans for engineering majors at colleges, bearing the critical responsibility of solidifying students' basic programming skills. In computer programming course teaching, undertaking adequate programming exercises and maintaining critical rethinking are the most important ways to steadily improve students' programming capability.

However, with the rapid advance and wide adoption of Generative AI techniques, Large Language Models (LLMs) have been widely adopted by col-

lege students, particularly in scenarios such as programming assistance, subject knowledge Q&A, and academic writing (Manorat et al., 2025). Observations from practical course teaching experiences reveal that many students directly input programming exercise tasks into LLMs and submit the automatically generated code as their exercise answers without incorporating their own thought processes, learning the code design logic, or commenting on the code's correctness. Long-term misuse and over-reliance on LLMs could lead to weakened foundational programming skills and a decline in critical thinking regarding code quality among students (Garcia, 2025).

Despite these challenges, this technological shift also presents a revolutionary opportunity for AI-assisted hybrid teaching pedagogy (Ghimire & Edwards, 2024; Shen, 2025). This study therefore aims to harness generative AI not as a substitute for student effort, but as a structured collaborative partner within a controlled and guided pedagogical framework, which could lead students to use LLMs with critical thinking. We posit that by strategically integrating AI into time-constrained, in-class exercises guided by a multi-dimensional "read-write-comment" pedagogy, we can counteract passive misuse and promote deeper, process-oriented learning. Our study contributes a novel and practical instructional model that transforms a major threat to programming education into a catalyst for enhancing genuine student competency and critical engagement.

2. Methodology

2.1 Research Question and Framework

To address the challenge of students tending to passively acquire AI-generated code, this study investigates the primary research question: *How can we design a pedagogical intervention that transforms generative AI from a "substitute tool" into a "collaborative partner" to effectively promote deep learning and genuine competency growth during in-class programming exercises?*

To address this question, we proposed and implemented an integrated model of "AI-student paired programming" with a "read-write-comment" multi-dimensional pedagogy. The study employs a methodological framework combining design-based research and quasi-experimental research, iteratively implemented and evaluated over one semester. This framework follows a "design-implementation-evaluation-iteration" cycle, aiming to validate the effectiveness of the proposed model through concrete teaching practices and systematically answer the research question.

2.2 Research Design and Implementation

2.2.1 Course Setting and Participants

The computer programming course opened seven parallel classes for engineering majors. We applied this study in one class, while the other six classes used the traditional teaching approach. The students in these classes showed no statistically significant differences in their entrance scores or performance in prerequisite courses. The selected class was randomly assigned as the experimental group, adopting the proposed AI-paired programming and three-dimensional pedagogy. The other classes served as the control group, continuing with the traditional "practice-submission" teaching model without structured AI collaboration. The research spanned one full semester (12 weeks in total) to observe the teaching effects.

2.2.2 AI-Student Paired Programming and the Three-Dimensional Framework

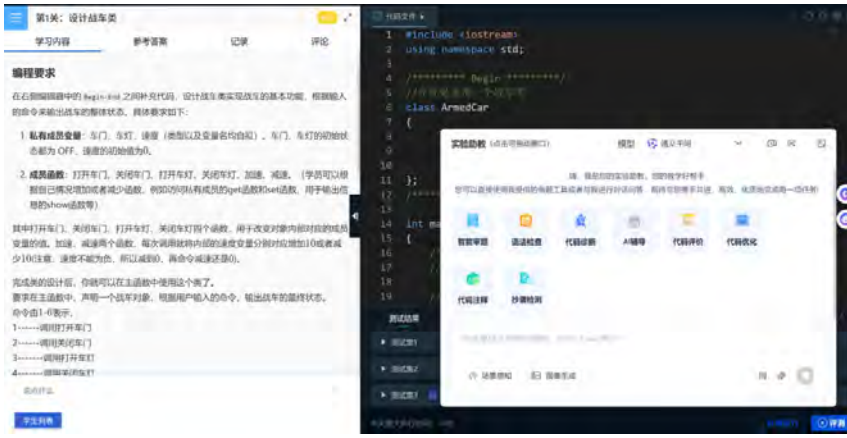
The core intervention for the experimental group consisted of two tightly coupled components: time-constrained AI-student paired programming tasks and a structured "read-write-comment" instructional process.

Design of AI-Student Paired Programming Tasks

We designed a series of in-class programming exercises synchronised with the course schedule. Each exercise was accompanied by clear requirements to be completed within a strict 10-minute time frame. As shown in Figure 1, students were required to work directly with pre-configured LLM assistance within an integrated development environment to complete the task. This strict time limit aimed to simulate realistic programming pressure, prevent students from engaging in aimless, lengthy AI conversations, and force efficient, goal-oriented collaboration.

Compared with traditional syntax-focused simple exercises, the scale and difficulty of these AI-paired exercise tasks were significantly increased. Each exercise typically included a clear application context and a simple code framework, enabling students to leverage AI for problem abstraction, code implementation, and review/refinement, rather than merely practising syntax. The exercises were carefully designed to be of reasonable yet challenging complexity, making the AI generate error-prone or incomplete answers. Traps and open-ended questions were intentionally incorporated to encourage students to critically evaluate and comment on the AI's suggestions, guiding the students to rationally use LLMs.

Figure 1: The AI-student paired programming exercise demonstration



The "Read-Write-Comment" Three-Dimensional Instructional Process

During each 10-minute exercise cycle, students were required to follow an explicit three-phase protocol to structure their engagement:

1. **Read & Refactor:** Students first prompted the AI to generate an initial solution. Their subsequent task was not to submit it directly but to deeply read and comprehend the AI-generated code. They needed to ask the AI clarifying questions (e.g., "Explain the logic of line X") or request the AI to refactor the code using a different algorithm. This phase aimed to cultivate students' ability to parse and understand the logic of code written by the AI.
2. **Write & Iterate:** Based on their understanding, students had to personally modify or rewrite key parts of the code. This could include fixing potential bugs, optimising specific functions, or adjusting the code structure according to their own understanding. This phase emphasised the transition from "reading comprehension" to "hands-on practice", ensuring the active application of coding skills.
3. **Comment & Reflect:** Upon completing the code, students were required to provide written commentary in code comments or a separate reflection area, evaluating the strengths and weaknesses of the AI's initial solution, justifying their own modifications, and summarising key lessons learned. This phase aimed to develop meta-cognitive and critical evaluation skills.

Along with the in-class exercising process, the teacher acted as the facilitator and monitor that circulated the classroom to ensure students adhered to the

protocol and did not merely copy-paste AI output.

2.2.3 Control Group Instruction

The control group received the same syllabus, theoretical instruction, and completed in-class exercises on the same topics. The key difference was their method: after the teacher explained examples, students completed exercises independently or through peer discussion, allowing conventional web searches but not encouraging or structuring the use of advanced AI code generation tools. They submitted the final code directly for instructor grading. The evaluation focus of the in-class exercises was on the correctness of the final product.

2.3 Data Collection and Analysis Method

To evaluate the pedagogical effectiveness, we collected learning outcome data for quantitative and qualitative analysis.

Learning Outcome Data

The two-stage exam scores of both classes served as the core metric for measuring ultimate learning achievement. The exam was uniformly designed by the course team, covering programming concept understanding, algorithm design, and code implementation, ensuring fair assessment for both groups.

Data Analysis Methods

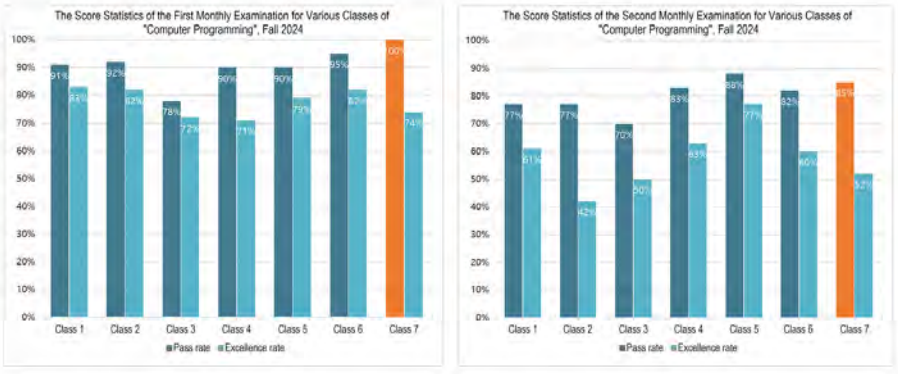
A quantitative analysis was conducted on the two-stage exam scores of the experimental and control groups to determine if there was a statistically significant difference in mean scores. This methodological design aims to rigorously validate whether the proposed AI-student paired programming with a multi-dimensional pedagogy can effectively enhance students' programming learning outcomes by controlling instructional variables, collecting objective data, and applying statistical analysis. It provides an empirically supported solution to the educational challenges posed by generative AI.

3. Results and Discussion

3.1 Key Quantitative Results

The primary quantitative outcome of this study was the comparative performance on summative assessments. As detailed in the methodology, the two-stage exam scores of the experimental group and the control group were compared. As shown in Figure 2, in the 2024 fall semester's computer programming teaching practice, Class 7 (the pilot class for the "AI-student" experimental method) demonstrated a statistically significant higher pass rate on the exams.

Figure 2: The score statistics of two stage exams of the pilot Class 7 and the other six classes



3.2 Discussion

The discussion hinges on several key mechanisms derived from our findings:

From Passive Consumption to Active Critique

The mandatory "read & refactor" phase disrupts the passive copy-paste cycle. By forcing students to interrogate the AI's output, the model cultivates the critical skill of code review and comprehension, which is foundational for programming expertise but often neglected when AI is used unguided. This active manipulation of code transforms abstract concepts into tangible skills and reinforces learning through doing.

The Reflection Phase

The reflection phase is crucial for developing higher-order thinking. Articulating justifications for changes and lessons learned requires students to synthesise their process, transforming an experiential task into explicit knowledge. This meta-cognitive practice is a significant driver of long-term skill retention and adaptability.

Critical Evaluation and Limitation

This study has limitations. The research was conducted within a specific institutional context. The long-term retention of these benefits beyond the course duration was not measured. Furthermore, the model requires reliable AI tool access and instructor training for effective facilitation. Despite these limitations, the convergence of positive quantitative outcomes strongly supports the efficacy of the proposed model. It successfully addresses the core challenge of

AI misuse by redesigning the learning activity itself, making productive, critical collaboration with AI the central goal rather than an avoidable shortcut.

4. Implications

This study demonstrates that the perceived threat of generative AI to programming education can be strategically mitigated and even reversed through deliberate pedagogical redesign. The AI-student paired programming with a multi-dimensional pedagogy model moves beyond merely policing AI use and instead integrates it as a catalyst for active, critical, and reflective learning. By shifting the focus from code product to cognitive process, the model fosters the very skills that passive AI misuse erodes.

The results confirm that this approach leads to statistically significant improvements in student learning outcomes compared to traditional instruction. The challenge of integrating generative AI into education is a pressing global issue. This study offers a proactive and positive response that aligns with worldwide educational goals of fostering skills of critical analysis, adaptability, and responsible technology use. As AI literacy becomes essential, this model teaches students not just to use AI, but to interrogate and collaborate with it effectively. By demonstrating how to turn a disruptive technology into a partner for deepening human learning, this work provides a significant reference point for educational institutions worldwide navigating the integration of AI into their curricula. It advocates for a future where AI enhances, rather than diminishes, the development of foundational human intellect and creativity.

Acknowledgments

This work was supported by National University of Defense Technology Education and Teaching Research Project "Integrating Large Models into Programming General Education Curriculum Reform to Enhance Engineering Practice Ability in Simulation—A Case Study of 'Programming and Data Structures'", and the Hunan Provincial Key Project for Teaching Reform Research in Regular Undergraduate Universities (No. 202502000002).

Ethical Declaration Statement

The study has been approved by the College of Systems Engineering of National University of Defense Technology.

Generative AI Policy

The authors used the DeepSeek service during the preparation of this work to polish the language. After using the tool/service, the authors thoroughly re-

viewed and edited the content as needed and assume full responsibility for the content of the publication.

References

Garcia, M. B. (2025). Teaching and learning computer programming using ChatGPT: A rapid review of literature amid the rise of generative AI technologies. *Education and Information Technologies*, 1–25. https://doi.org/10.1007/s10639-025-13452-5

Ghimire, A., & Edwards, J. (2024). Coding with AI: How are tools like ChatGPT being used by students in foundational programming courses. In *International Conference on Artificial Intelligence in Education* (pp. 259–267). Springer. https://doi.org/10.1007/978-3-031-64299-9_20

Manorat, P., Tuarob, S., & Pongpaichet, S. (2025). Artificial intelligence in computer programming education: A systematic literature review. *Computers and Education: Artificial Intelligence*, 100403. https://doi.org/10.1016/j.caeai.2025.100403

Shen, F. (2025). A study of human-AI interaction patterns in artificial intelligence-assisted second language writing. *International Journal of English Literature and Social Sciences, 10*(2), 595–617. https://doi.org/10.22161/ijels.102.8

A Multiple LLM Approach to Automated Essay Scoring: Comparing Seven Open-Source Models with Human Graders

Edmond YEUNG [1], Jialing Juley XIAO [1], Jonathan LEE [2], Marco BETTONI [2], Rachel WONG [1]

[1] Department of Management and Marketing, The Hong Kong Polytechnic University, Hong Kong, China

[2] Industrial Centre, The Hong Kong Polytechnic University, Hong Kong, China

Abstract

This study evaluates whether open-source large language models (LLMs) can support automated essay scoring (AES) of reflective writing in higher education. Forty-nine reflective essays from a foundation entrepreneurship and innovation course were scored by lecturers and independently graded by seven open-source LLMs (Llama 3 8B, DeepSeek R1 32B, Ministral 3 8B, Qwen 3 8B, GPT-OSS 20B, Gemma 3 27B, and Phi 3 14B) using the same rubric. Model performance was assessed via exact match, mean absolute error (MAE), Pearson and Spearman correlations, and paired t-tests/Wilcoxon tests against human scores. Ministral produced complete outputs and demonstrated the most balanced alignment with human grading, showing no significant difference in mean or median scores. Other models exhibited larger deviations, conservative bias, or missing outputs, raising scalability concerns. Results suggest that LLMs may serve as decision-support tools but do not yet replace human judgement for reflective assessment.

Keywords: Automated Essay Scoring (AES), reflective essay scoring, open-source LLM, higher education assessment, AI grading

1. Introduction

Recent advances in artificial intelligence are reshaping education by automating routine tasks, thereby allowing educators to focus on mentoring and curriculum design (Awidi, 2024). Essay grading is one of the most time-intensive tasks; in large classes, it may invite fatigue and inconsistent ratings (Awidi, 2024). In our undergraduate management courses, summative assessment relies on reflective essays. With cohorts numbering in the thousands, this creates a substantial marking load. If AI could assist scoring at scale while preserving validity and fairness, it could alleviate stress and release time for formative feedback.

Automated Essay Scoring (AES) has long promised rapid, consistent scoring; however, most evidence stems from K–12 settings or standardized English tests (e.g., TOEFL/IELTS), rather than discipline-specific university writing. Where higher education is concerned, research often emphasises surface language accuracy rather than higher-order thinking and reflective integration (Uyar & Büyükahıska, 2025; Ouyang et al., 2022). Large Language Models (LLMs) such as ChatGPT have renewed interest in AES because they can interpret context-rich text and produce human-like rationales. Yet, whether they can validly and reliably evaluate reflective essays—requiring meaning-making, disciplinary alignment, and critical self-assessment—remains uncertain (Awidi, 2024; Yeung et al., 2025).

Automated Essay Scoring (AES) and ChatGPT

ChatGPT is arguably the most widely discussed LLM since its introduction, and its application to AES has attracted considerable attention. Numerous studies have explored ChatGPT's ability to score essays and compared its outputs with human raters. Findings, however, are mixed. For instance, Awidi (2024) examined ChatGPT's performance in grading short reflective essays in a first-year engineering course at an Australian university. Significant discrepancies persisted between AI-generated and human scores, leading Awidi (2024) to recommend human moderation and to position AI as a supplementary tool rather than a replacement. Similarly, Uyar and Büyükahıska (2025), using IELTS Writing Task 2 descriptors in an EFL context, reported substantial differences between ChatGPT and human ratings across genres, with correlations ranging from weak to moderate—insufficient for practical deployment. While some studies echo these limitations (e.g., Liu et al., 2023; Yoon et al., 2023), others suggest potential improvements under certain conditions (e.g., Tate et al., 2023), creating a contradictory evidence base that underscores the need for new approaches to AES.

A Multiple LLMs Approach

Recent research has shifted from evaluating a single proprietary model to benchmarking multiple LLMs against human graders to assess AES feasibility across diverse contexts. For example, Abujadallah et al. (2025) analysed a large dataset of grade 6–12 argumentative essays to evaluate five open-source LLMs. Their findings indicate that rubric-aligned prompting significantly enhances performance, with models achieving correlations above 0.83 and F1 scores near 0.93. In contrast, Gaggioli et al. (2025), studying psychology essays written in Italian in higher education, found that LLMs struggle to assess tasks requiring "disciplinary insight and contextual sensitivity" (p. 1). These contexts differ substantially from ours: a higher education setting where English is the primary

language of instruction and reflective essays serve as evidence of learning. This scenario offers high transferability to similar global contexts. However, no existing research compares multiple LLMs' performance in evaluating reflective essays within higher education.

Addressing this gap, we pose the following research questions:

1. Which open-source LLMs demonstrate the strongest performance in evaluating reflective essays?
2. How does their accuracy compare to human raters' grades?

2. Methodology

The primary aim of this research is to assess the effectiveness of LLMs in scoring reflective essays within a higher education context. We evaluated seven open-source LLMs available to us. We selected the latest versions: **Llama 3 8B, DeepSeek R1 32B, Ministral 3 8B, Qwen 3 8B, GPT-OSS 20B, Gemma 3 27B, and Phi 3 14B**. Models were selected for diverse representation, while their open-source nature ensures transparency and reproducibility in AES research.

Sample: The Reflective Essays

The dataset comprises 49 reflective essays written by students and previously graded by human raters. The essays were submitted as the summative assessment for a foundation-level university course in Hong Kong introducing concepts related to entrepreneurship and innovation. Students were instructed to reflect on (1) what they learned from the course and (2) how that learning connects to their academic pathway and/or career plans.

The grading rubric contains three components: (i) reflection, (ii) language, and (iii) citation. Each LLM was prompted to assign a single overall score using the same rubric applied by human graders. Scores range from 1 to 12, corresponding to letter grades from F to A+. Human grading was completed by two course lecturers, who also designed the rubric and taught the course.

Data Analysis

All analyses were conducted using SPSS. Each model's score distribution is summarised using descriptive statistics such as mean, standard deviation, minimum, maximum, median, and quartiles. To evaluate scoring accuracy relative to human grades, we calculated:

1. **Exact Match:** The proportion of essays where the model grade equaled the human grade.

2. **Mean Absolute Error (MAE):** To quantify the model's average grading deviation relative to human scores; here, *deviation* refers to the differences between model and human grades (ignoring whether the model over- or under-estimated the score).
3. **Pearson's correlation coefficient (*r*)** and **Spearman's rank correlation (ρ):** To examine linear association and rank-order consistency.
4. **Paired sample t-test (*p*)** and **Wilcoxon signed-rank test (*p*):** To test whether the differences between the means and medians were statistically significant.

3. Results and Discussion

Table 1 summarises the descriptive statistics of grades assigned by the human raters and the seven LLMs. Three models—DeepSeek, GPT-OSS, and Phi—produced output errors and therefore graded fewer than the full 49 essays: DeepSeek failed to score nine essays (18.37%), with smaller shortfalls for GPT-OSS and Phi. Although these error rates appear modest in a small sample, they could become a substantial reliability issue when grading thousands of essays.

In terms of score distributions, **Ministral** appears most aligned with human grading: it successfully graded all essays, and its mean and SD closely match the human raters', with identical quartiles, minimum, and maximum values. Phi shows a comparable mean but a notably larger SD and wider range, indicating greater variability than the human raters. LLaMa displays a compressed distribution with a smaller SD and narrower range, suggesting limited differentiation among essays relative to humans. Qwen, GPT-OSS, and Gemma show lower average scores than humans, consistent with a tendency toward more conservative grading. Finally, DeepSeek exhibits the greatest dispersion—more than double the human SD—alongside missing outputs, indicating both stability and reliability concerns for large-scale deployment.

Table 1. Descriptive statistics of all seven models

	Human	LLaMa	DeepSeek	Ministral	Qwen	GPT	Gemma	Phi
n	49	49	40	49	49	46	49	48
Mean	9.5102	9.0612	9.175	9.6939	8.9796	8.3478	8.5714	9.3333
SD	0.7671	0.3168	1.5834	0.7693	1.2160	1.8405	0.7906	1.4192
Min	8	8	5	8	5	4	7	5
Q1	10	9	9	10	9	9	9	9
Median	10	9	10	10	10	9	9	10
Q3	12	10	12	12	12	12	10	12
Max	12	10	12	12	12	12	10	12

Table 2 provides converging evidence that Ministral is the most promising model overall in this dataset. Ministral achieves the lowest Mean Absolute Error (MAE = 0.6327), indicating the smallest average deviation from human grades. Although both correlation coefficients indicate only moderate correlation, Ministral shows the strongest linear alignment with human scores (Pearson's $r = 0.4114$) and a comparatively high rank-order association (Spearman's $\rho = 0.3140$). Importantly, both the paired-samples t-test ($p = 0.130$) and Wilcoxon signed-rank test ($p = 0.127$) are non-significant for Ministral, implying no statistically significant difference from human graders in terms of mean or median scores.

Table 2. Performance metrics of all seven models

	LLaMa	DeepSeek	Ministral	Qwen	GPT	Gemma	Phi
Exact Match	0.3265	0.4286	0.2653	0.3265	0.2653	0.3673	0.3061
MAE	0.6939	2.8367	0.6327	0.8980	2.0408	1.0612	1.2245
Pearson r	0.2975	0.2156	0.4114	0.2124	0.3905	0.2307	0.1101
Spearman ρ	0.2260	0.0600	0.3140	0.2440	0.3640	0.0970	0.0670
t-test p	<0.001	0.2890	0.1300	0.006	<0.001	<0.001	0.4030
Wilcoxon p	<0.001	0.3350	0.1270	0.0080	<0.001	<0.001	0.6350

Although several models outperform Ministral on individual indicators, these advantages do not translate into an overall profile consistent with a reliable and valid automated grader. For instance, DeepSeek and Gemma show higher Exact Match rates than Ministral, and GPT-OSS exhibits a higher Spearman's ρ. However, these apparent strengths are offset by substantive limitations in other metrics and in operational robustness. DeepSeek's performance is undermined by substantial missing outputs (reduced n), markedly poorer error-based accuracy (high MAE), and agreement indices, raising concerns about scalability and consistency in real-world, high-volume grading. GPT-OSS, despite stronger rank correlation, shows large absolute deviations from human scores (elevated MAE) and a statistically significant difference between its mean and median compared with the human raters'. Therefore, they are less likely to be considered to have sufficient reliability, consistency, and validity for AES purposes.

4. Implications

The findings suggest that **Ministral 3 8B** is the most viable candidate among the evaluated open-source models for supporting automated essay scoring (AES) of reflective writing in a higher education context. Across the seven models, Ministral demonstrated the most balanced performance profile. These

results support the implication that Ministral can approximate human scoring patterns more closely than the other models in this study, and therefore has potential as an AES component for reflective essays when appropriately constrained and monitored.

However, the study also indicates clear boundaries on what the current approach can justify:

1. **Moderate Correlation:** The observed correlations remain moderate rather than strong. While Ministral generally moves in the same direction as human grades, it does not reliably reproduce fine-grained distinctions between adjacent grade bands at the individual-essay level. As a result, Ministral cannot yet replace human grading but may serve as a decision-support tool. Human graders should retain responsibility for final decisions, while the model's value lies in reducing routine workload and enabling instructors to redirect effort toward richer formative feedback.
2. **Need for Replication:** The results underscore the need for replication with a larger dataset before claiming generalisability. With 49 essays from one foundation course, the estimates of accuracy and association are necessarily sample-dependent and may shift with changes in cohort composition, topic prompts, or rater expectations. A larger dataset would allow more stable evaluation of performance and would better reflect the real operational environment where reflective essays vary widely in content and citation practices.

Ethical Declaration Statement

The project has been approved by the Departmental Research Committee of The Hong Kong Polytechnic University (Reference Number: HSEARS20251203008).

Generative AI Policy

The authors used Copilot during the preparation of this work to proofread and generate ideas for the title of the research. After utilising the tool/service, the authors thoroughly reviewed and edited the content as necessary and assumed full responsibility for the publication's content.

References

Abujadallah, M., Saad, M., & Abudalfa, S. (2025). Evaluating Open-Source LLMs for Automated Essay Scoring: The Critical Role of Prompt Design. https://www.preprints.org/frontend/manuscript/90bf5ac43bbcb975d13a26f61ce24d7e/download_pub

Awidi, I. T. (2024). Comparing expert tutor evaluation of reflective essays with marking by generative artificial intelligence (AI) tool. *Computers and Education: Artificial Intelligence, 6*, 100226. https://doi.org/10.1016/j.caeai.2024.100226

Gaggioli, A., Casaburi, G., Ercolani, L., Collovà, F., Torre, P., & Davide, F. (2025). Assessing the Reliability and Validity of Large Language Models for Automated Assessment of Student Essays in Higher Education. https://arxiv.org/pdf/2508.02442

Liu, Y., Moosavi, N. S., & Lin, C. (2024, August). LLMs as narcissistic evaluators: When ego inflates evaluation scores. In *Findings of the Association for Computational Linguistics: ACL 2024* (pp. 12688-12701). https://doi.org/10.18653/v1/2024.findings-acl.753

Seßler, K., Fürstenberg, M., Bühler, B., & Kasneci, E. (2025, March). Can AI grade your essays? A comparative analysis of large language models and teacher ratings in multidimensional essay scoring. In *Proceedings of the 15th International Learning Analytics and Knowledge Conference* (pp. 462-472). https://dl.acm.org/doi/pdf/10.1145/3706468.3706527

Tate, T. P., Steiss, J., Bailey, D. H., Graham, S., Ritchie, D., Tseng, W., Moon, Y., & Warschauer, M. (2023). Can AI Provide Useful Holistic Essay Scoring? https://doi.org/10.31219/osf.io/7xpre

Yeung, W. E., Xiao, J. L., & Wong, F. R. (2025). Using ChatGPT for Automated Essay Scoring of Reflective Essays in University Management Courses. In *Academy of Management Proceedings 2025*. https://cdmcd.co/vaA8wJ

Yoon, S. Y., Miszoglad, E., & Pierce, L. R. (2023). Evaluation of ChatGPT Feedback on ELL Writers' Coherence and Cohesion. https://arxiv.org/pdf/2310.06505

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/bmjoq-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

A Process-Based Evaluation of EFL Learners' Course Papers in the GenAI Era

Jianbin ZENG[1], Rong ZHAO[2], Xuhua ZHANG[2]
[1] School of Foreign Languages and Literature, Fudan University, Shanghai, China
[2] College English Centre, Fudan University, Shanghai, China

Abstract

This paper presents an innovative course design for evaluating course papers, developed in response to the challenges posed by AI-empowered writing technologies for EFL graduate students in China. To address these challenges, a process-based approach is proposed that integrates continuous assessment of EFL graduate students' class performance, in-class tests, and the writing process of their course papers. Drawing on a case study within an English Academic Paper Writing course, our process-based approach generates traceable records of student performance. This method addresses issues of reliability, validity, and fairness in course paper evaluation while mitigating AI-related risks. Data confirms that this process-based evaluation yields a more reliable, valid, and objective measure of EFL learners' academic performance and achievements. Furthermore, it provides a foundation for updating guidelines on AI-empowered course paper writing and assessment, and proactively exploring strategies to leverage AI for enhancing teaching, learning, and assessment in higher education.

Keywords: process-based evaluation; course paper assessment; EFL learners; AI-empowered writing; academic integrity

1. Introduction

The widespread use of course papers for assessment in higher education is increasingly challenged by AI-empowered writing technologies, which undermine the reliability, validity, and fairness of such assessments, particularly among EFL graduate students in China. On one hand, the take-home, open-book nature of the writing process involves an extended timeline and substantial workload. This often clashes with other intensive academic tasks at the end of the semester, leading to student stress and increasing the likelihood of academic misconduct. On the other hand, product-focused grading criteria for course papers lack the clear-cut feasibility characteristic of in-class, closed-book examinations. Consequently, it is not uncommon for different instructors—or even

the same instructor under varying circumstances—to assign divergent grades to identical course papers, thereby undermining the objectivity and fairness of assessment (Zeng & Zhang, 2025).

Currently, the integration of AI into academic writing is exacerbating long-standing issues of academic misconduct and subjective evaluation in course paper assessment, especially for EFL graduate students in China. Hence, we propose a process-based course paper evaluation as a pedagogical innovation to address persistent issues regarding reliability, validity, and fairness. This approach aims to proactively explore innovative models of AI-empowered teaching, learning, and assessment in higher education.

2. Methodology

According to a survey conducted within an English Academic Paper Writing course at a leading Chinese university, the application rate of AI tools was as high as 97% among the 92 registered EFL graduate students. These tools included DeepSeek, ChatGPT, Grammarly, DeepL, Kimi, Doubao, Wenxin Yiyan, AI Taiyan, Chemdraw, GitHub Copilot, Connected Papers, Research Rabbit, SemiKong, LaTeX, Writefull, Sider, Claude-Sonnet, and others. However, controversy and debate remain prevalent regarding the extent and standards of AI application in the academic writing process. All respondents (100%) urged the establishment of clear AI policies to ensure that the use of AI-assisted writing tools aligns with academic norms and ethical standards, thereby enhancing the reliability, validity, and fairness of course paper assessments.

In a follow-up interview with the faculty of the English Academic Paper Writing course for EFL graduate students, all 12 instructors reported witnessing increasingly severe challenges posed by AI writing tools to course paper assessment. On one hand, the quality of course papers submitted by EFL graduate students over the past two years has shown unusual improvement, particularly in structure and readability. On the other hand, instructors find it more challenging to objectively and fairly assess EFL learners' genuine English academic writing proficiency based on these unusually high-quality papers. This difficulty has led to inconsistencies in grading standards and, in some cases, arbitrary grading based on subjective evaluation among instructors. Moreover, current evaluation criteria for course papers have been outpaced by breakthroughs in AI technology; they lack clear guidelines regarding the ethical boundaries of AI assistance and exhibit institutional gaps in AI policies (Moorhouse et al., 2025; Smit et al., 2025; Zou et al., 2025). Addressing these evaluation issues while mitigating AI risks is therefore critical, particularly for EFL graduate students in China. In Section 3, we explore a concrete solution through a case study and a process-based approach.

3. Results and Discussion

3.1 A Case Study of the English Academic Paper Writing Course

The *English Academic Paper Writing* course is a two-credit EAP module offered at a leading research university in China. It is designed to support EFL graduate students in their academic studies, research projects, and international communication. The course employs a process-based assessment framework for its final paper, utilising a comprehensive, multi-faceted, and multi-dimensional formative approach. (See Table 1)

Table 1

A Formative Assessment of Course Papers in English Academic Paper Writing

Component	Content	Criteria	Weight(%)
Class performance	Discussion	eLearning discussion	30
	Group work	Collaboration in presentation	
	Reflection	Course paper reflection	
Mid-term test	Material-based question	Sentence reordering & justification	30
		Data analysis from visual to verbal	
		Draft revising in condensed or expanded versions	
	Short-answer question	Analysis of course paper writing process	
		AI-assisted course paper writing norms	
		Evaluation of English Academic Paper Writing	
Course paper	Self check	Originality statement & AI disclosure	40
	Peer review	Feedback & responses	
	Tutorial	Instructor guidance	
Overall assessment			100

Table 1 illustrates the overall assessment structure for the *English Academic Paper Writing* course, which comprises class performance (30%), a midterm test (30%), and the final course paper (40%).

First, **class performance** accounts for 30% of the total assessment. This component includes topic discussions, group work, and paper presentations. Topic discussions require students to share comments on course topics via the eLearning platform. Group work engages students in collaboration and the presentation of their academic writing topics in class. Paper presentations involve individual reflections on the course paper writing process.

Second, the **closed-book midterm test** accounts for another 30%. It consists of material-based and short-answer questions designed to evaluate students' genuine writing competence and achievements.

Third, the **course paper** constitutes the final 40% of the assessment. This component includes self-check reports, peer reviews, and tutorial feedback. The self-check report primarily comprises an originality statement and an AI disclosure/disclaimer. Peer reviews encompass comments on course papers submitted by peers and the EFL learners' responses to such feedback. Tutorials provide personalised guidance from instructors, alongside students' revisions based on this guidance.

This course paper assessment represents a comprehensive, multi-faceted, and multi-dimensional formative approach.

First, it covers the entire process of course paper writing, spanning the preparation phase prior to drafting, the initial drafting phase, and the final revision phase. Pre-writing preparation primarily records EFL learners' performance, including their participation and contribution in both online and offline learning and teaching activities. Group work involves team presentations and collaborative reflections on disciplinary or interdisciplinary academic writing. During the initial drafting phase, data such as topic selection, organisation of information, outline design, and draft writing are collected and recorded. For instance, paper presentations require learners to reflect on their progress in writing their course papers. The midterm test primarily assesses students' mastery of course content and writing skills. The revision and finalisation phase focuses on tracking self-assessment, peer review, AI evaluation, tutorial feedback, and the iterative optimisation of initial drafts.

Second, the multi-faceted assessment involves a diversity of participants, including EFL learners themselves, peers, AI evaluation assistants, and instructors. Self-assessment, an internal evaluation by learners, includes eLearning topic discussions, midterm tests, personal paper reflections, drafts, revisions, and final self-assessment reports. These elements aim to motivate EFL learners' independent learning interest and foster a sense of responsibility regarding their own evaluation. External evaluators—such as peers, AI evaluation assistants, and instructors—integrate diverse external resources to provide more comprehensive and authentic multi-faceted perspectives. For instance, peers offer simulated peer reviews from fellow EFL learners of equal standing; AI evaluation assistants provide modular feedback and recommendations empowered by AI technology; and instructors, based on course objectives and requirements, offer targeted and personalised guidance on course paper norms and formats.

Third, the multi-dimensional evaluation tracks EFL learners' dynamic performance throughout the entire learning process, covering stages from class

participation and midterm tests to the writing of the course paper. For example, teamwork presentations, paper reflections, and tutorial feedback assess EFL learners' listening and reading comprehension, as well as their collaboration and communication skills. Midterm tests, self-assessment reports, peer review comments, originality statements, and AI disclosures accurately record EFL learners' understanding and expression in reading, writing, and academic integrity. Online discussions on course topics via the eLearning platform and interactive dialogues with AI evaluation assistants reflect the digital and intelligent capabilities required for course paper writing and evaluation in the GenAI era.

In summary, our process-based approach offers a comprehensive, multi-faceted, and multi-dimensional formative assessment of course papers, thereby enhancing the reliability, validity, and fairness of course paper assessment. As shown in Table 2, the overall student evaluation score is 98.73%, compared with 97.78% for the course team, 97.96% for the college, and 98.01% for the graduate school. These figures demonstrate high popularity and recognition among EFL learners.

Table 2

Student Evaluation across Classes, Team, College & School

Semester	Classes	Students	Mean Scores (avg. across classes)	MS (team)	MS (college)	MS (school)
Spring 2025	3	96	99.31, 99.46, 99.7 (99.49)	97.71	98.21	98.07
Fall 2024	3	100	98.84, 99.17,99.4 (99.14)	98.24	98.46	98.32
Spring 2024	3	92	98.01, 97.93, 97.7 (97.88)	97.88	97. 67	97.82
Fall 2023	3	96	98.15, 98.14, 97.64 (97.98)	97.46	97.9	97.93
Spring 2023	3	91	99.59, 99.25, 98.64 (99.16)	97,62	97.57	97.92
Total	15	475	98.73	97.78	97.96	98.01

Note. N = 15 classes (475 students). N = 12 team members. N = 111-146 college faculty. N = 2191-3324 graduate school faculty.

3.2 A Process-Based Approach to Course Paper Evaluation

As presented in the case study in Section 3.1, our process-based approach focuses not only on the final outcome but also on the dynamic process involving class performance, eLearning discussions, teamwork presentations, midterm tests, peer reviews, self-assessments, AI evaluations, and tutorials. This dynamic strategy tracks the progress of course paper writing, regularly monitoring and intervening at critical stages (such as outlining, drafting, AI integration,

personal reflection, peer and tutorial feedback, editing, proofreading, and final submission). It enables timely adjustments to writing strategies, allowing instructors to monitor the extent and scope of AI involvement in the writing process. Consequently, this approach facilitates a more comprehensive and objective understanding of the actual writing process. Furthermore, it transforms EFL graduate students from passive recipients into active participants in evaluation; self-reflection, peer collaboration, AI evaluation, and tutorial feedback offer diverse perspectives that address the limitations of traditional product-focused assessment. This multi-dimensional formative assessment allows for the evaluation of various aspects of the course paper, such as content, structure, style, design, and ethics, enabling a layered examination of critical thinking, research design, data processing, logical construction, and AI assistance throughout the writing process.

Additionally, the process-based approach adopts an evidence-based evaluation model to reduce unfair practices, such as arbitrary grading based on subjective impressions or personal preferences. This requires the dynamic collection and recording of observable, traceable evidence during the assessment process, including participation in discussions, group and individual presentations, in-class tests, self-assessments, peer reviews, AI evaluations, and tutorial feedback at key stages such as drafting, revising, and finalising the course paper. It also allows for breaking down the writing process into multiple stages (e.g., topic selection, literature review, outline drafting, initial draft, revisions, and final draft) and designing specific checklists and evaluation criteria for each stage. Evidence is collected at critical points to enhance the feasibility of formative assessment. Moreover, AI technology provides digital annotation and templated feedback tools[1], making it easier for EFL graduate students to annotate evaluation feedback and ensure traceability. This improves the transparency and consistency of the assessment process.

4. Implications

Our course design innovation proposes a paradigm shift from product-focused to process-oriented evaluation through a formative framework encompassing class performance, in-class tests, and scaffolded course paper drafting. This approach addresses issues of reliability, validity, and fairness while mitigating AI-related risks in course paper evaluation. Course evaluation data confirms that this formative evaluation yields a more reliable, valid, and objective measure of EFL graduate students' academic performance in the GenAI era. Future efforts could optimise the assessment process, technological tools, and compositional ratios; update policies for AI-assisted course paper writing and evaluation; and explore innovative models of AI-empowered teaching, learning,

and assessment within higher education.

Generative AI Policy: The authors utilised DeepSeek V-3.2 during the preparation of this work for language polishing. After utilising the AI service, the authors thoroughly reviewed and edited the content as necessary and assume full responsibility for the publication's content.

Note 1: Although a wide range of AI tools (e.g., DeepSeek, ChatGPT, Grammarly, DeepL, Claude-Sonnet) were used by our students, the AI evaluation assistant was exclusively DeepSeek V-3.1 via institutional access, focusing on the pedagogical integration of AI within our process-based evaluation. DeepSeek V-3.1 functioned as a structured feedback generator rather than an autonomous grading system. Students were required to submit draft versions to the AI assistant, obtain structured feedback, and provide written reflections explaining how they incorporated, modified, or critically evaluated the AI suggestions. AI-generated feedback was treated as advisory and subject to instructor oversight, with instructors reviewing AI comments during tutorial sessions. Final grading authority remained exclusively human-controlled, and discrepancies between AI-generated feedback and instructor evaluation were documented to refine prompt calibration. This ensured the transparency, traceability, and pedagogical accountability of generative AI technologies.

References

Moorhouse, B. L., Wan, Y., Wu, C., Wu, M., & Ho, T. Y. (2025). Generative AI tools and empowerment in L2 academic writing. *System, 133*, Article 103779. https://doi.org/10.1016/j.system.2025.103779

Smit, M., Wagner, R. F., & Bond-Barnard, T. J. (2025). Ambiguous regulations for dealing with AI in higher education can lead to moral hazards among students. *Project Leadership and Society, 6*, Article 100187. https://doi.org/10.1016/j.plas.2025.100187

Zeng, J., & Zhang, X. (2025). Ethical principles for generative AI in research paper writing for EFL graduate students in China. In Y. Wang, P. Moore, T. Hsieh, & H. Tran (Eds.), *Proceedings of the International CALL Research Conference* (Vol. 2025, pp. 191-196). https://doi.org/10.29140/97817637116240-25

Zou, M., Kong, D., & Lee, I. (2025). Doctoral student's strategy use in GAI chatbot-assisted L2 writing. *Journal of English for Academic Purposes, 76*, Article 101521. https://doi.org/10.1016/j.jeap.2025.101521

Design a Flipped Lesson to Help Students Learn Effectively with Gen AI

Paul NG [1], Steven LIM [2]
[1] Centre for Learning and Teaching Excellence, Ngee Ann Polytechnic, Singapore
[2] School of Business and Accountancy, Ngee Ann Polytechnic, Singapore

Abstract

Participants will explore the confluence of Generative AI (Gen AI) for learning assistance, analytics, and student AI literacy and cognitive development in this workshop. They will role-play first-year Polytechnic students in a digital flipped lesson to experience a Gen AI Retrieval-Augmented Generation (RAG) learning assistant, first asynchronously and then in-person in real-time, with or without voice. The lesson will quickly prepare them for an in-person authentic learning activity—a debate on a case—in an introductory business law module. They will then explore ways to respond to diverse learning needs through analysis of the logs of students' text-chats. Finally, they will consider in-person lesson activities, such as debates with peers, to help avoid over-reliance on Gen AI.

Outcomes-based Learning Objectives

Participants will be able to:

- Design lessons to educate students in the proper use of Gen AI to assist their own learning, leveraging the affordances and operating within the constraints of digital flipped learning.
- Achieve, in their learning design, an ethical, tech-enabled balance between respect for students' privacy and the pedagogical benefits of knowing how students learn, both on their own and in-person.

Target Audience

- Lecturers; module (subject) leaders, especially but not only in law.
- Learning designers, education/faculty developers.

Topic Relevance

This workshop is relevant to:

- AI tools and Pedagogy
- Ethics and Integrity
- Learning Analytics

- Incorporating GenAI in Learning, Teaching and Assessment
- Student AI Literacy, AI-Readiness, and Cognitive Development

Workshop Design

Experiential: Participants will learn by doing (role-playing) and reflective discussion.

Workshop Format and Activities:

1. Introduction

a) Welcome and introduction to the institution.
b) Workshop overview including learning outcomes and plan for the workshop.
c) Introduction to a sample lesson plan extracted from the business law module and tools.
   a. The tools including a RAG tool will be accessible online.

2. Main Content Delivery

Topics

a) What Digital Flipped Learning and Gen AI RAG Are.
b) Experience a digital flipped lesson as students
c) Analyse the chatlog as a tutor
d) Reflective "discuss-design-and-critique" group discussions on designing Digital Flipped Learning to deter students from taking inappropriate shortcuts with Gen AI.

Frameworks

Kolb's Experiential Learning Theory underpins the design of this workshop. The workshop aims to develop educators to at least Level 2 in Dimension 5 of the Digital Education Council (DEC) AI Literacy Framework, where the domain in question is education.

Facilitators expect to refer during reflective discussions to frameworks applied within the institution to help students develop AI literacy, namely Oguz Acar's PAIR framework for developing transferable Gen AI competencies and the 3A's (an institution-specific framework) for the systematic incorporation of Gen AI into institutional curricula.

Instructional Method

Experiential. Participants will role-play as students and a tutor in the module teaching team, reflect and discuss the experience, then plan and critique peers' plans.

3. Interactive Activities

After the introduction (5 minutes), the anchor activity shall be a breakout session for participants to experience a 20-minute digital flipped lesson—both the online asynchronous and in-person segments—leveraging on a specialised Gen AI assistant to learn. The digital flipped lesson replicates the facilitators' real-life module's design. It structures students' learning as a Kolb's Experiential Learning cycle.

Students who are new to law iteratively progress from individual Gen-AI-supported, own-time, own-target preparatory experiences to deeper, peer- and tutor-mediated, time-bound sense-making. Abstract conceptualisations of the law and Gen AI, initially developed via concrete but self-paced learning experiences, are further developed in subsequent in-person debates. Debates deepen students' grasp of both the law and appropriate Gen AI use by compelling peers to challenge each other's initial conceptualisations with feedback and guidance from the tutor.

Note that although the workshop will only provide enough time for one Kolb cycle, the module develops students' sense of the law and Gen AI via multiple such cycles on different law topics but the same Gen AI assistant. On the DEC AI Literacy framework, Kolbian iterative experiential learning design, implemented in digital flipped learning cycles with a RAG Gen AI assistant, offers a good balance of aid and impetus to develop all four dimensions of the framework—understanding how AI works, evaluating AI output, using AI ethically and responsibly, and ensuring humans remain at the core—while learning how to apply AI in a specific context, that is, in a foundational law module (the fifth dimension of the framework).

On PAIR, participants may use other tools and not just the module's RAG tool for Gen AI assistance. Again, this reflects the facilitators' actual module design. Participants will therefore be free during the workshop's simulated lesson to address their legal problem(s) by interacting with their choice(s) of AI tools and reflecting on the outcomes.

After the lesson, participants will gather to briefly (5 to 10 minutes) analyse the data in the role of "tutors", with guidance from facilitators and the assistance of Gen AI tools. This would be followed by another breakout of 10 to 15 minutes for facilitated discussions in small groups where participants would be encouraged to elicit their lessons learnt and then to draft a simple lesson plan for peers to critique (with the help of Gen AI if they so desire).

The closing segment would be a critique of selected plans, a summary of key takeaways, and an open question-and-answer session (10 to 15 minutes).

Participants would be offered as a concrete takeaway from this workshop a digital flipped learning lesson plan template to develop all five dimensions of the DEC's AI literacies through experiential learning and PAIR.

4. Discussion and Q&A

- "What do you think are key assumptions or intents behind the design of the lesson you just experienced?"
- "What was your actual experience like?"
- "In what ways did this lesson assist or hinder you from achieving the lesson's outcomes?"
- "What does the data tell us about how the class used the tool? Frequency of use, by table (group). Top 3 most popular subtopics, learning challenges based on the nature of the interactions – by table/group and in the class as a whole."

The tool provides users with options to use it in different ways—for quizzes, as a tutor, and to practice analysing cases. Rank the options by frequency of use and summarise how groups and the class as a whole performed/what they did on that option.

Please download a copy of this template and take a few minutes to create a simple Gen AI-assisted lesson plan. Selected plans can be shared for feedback from the rest of the class.

5. Closing Remarks

Key takeaways from this workshop include:

- Tutors should analyse students' interactions with Gen AI in order both to improve the Gen AI learning assistant and to improve students' learning by addressing key learning issues, for example, in subsequent in-person lessons.
- Some learning designs work better than others to develop critical thinking, communication, and other human-specific competencies and deter passivity in the face of pervasive Gen AI.

Acknowledgments

Colleagues within the institution who assisted with the development and implementation of the Gen AI learning assistant and the associated learning design that will be experienced in this workshop will be given credit.

Ethical Declaration Statement

The institution approved this workshop application for this conference. The proposed workshop design briefly replicates a learning design that was approved by the institution.

Generative AI Policy

The author(s) used Microsoft Copilot during the preparation of this work to conduct searches within the institution's databases and on the Internet for relevant resources, and to compile the references list below. After utilising the tool/service, the author(s) thoroughly reviewed and edited the content as necessary and assume full responsibility for this publication's content.

References

Acar, O. A. (2023, June 15). *Are your students ready for AI? A 4-step framework to prepare learners for a ChatGPT world*. Harvard Business Publishing Education. https://hbsp.harvard.edu/inspiring-minds/are-your-students-ready-for-ai

Digital Education Council. (2025). *DEC AI literacy framework*. https://www.digitaleducationcouncil.com/post/digital-education-council-ai-literacy-framework

King's College London. (n.d.). *PAIR (Problem, AI, Interaction, Reflection) framework guidance*. https://www.kcl.ac.uk/about/strategy/learning-and-teaching/ai-guidance/pair-framework-guidance

Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Prentice Hall.

Open University. (n.d.). *Learning, thinking and doing: The experiential model of learning*. OpenLearn. https://www.open.edu/openlearn/science-maths-technology/engineering-technology/learning-thinking-and-doing/content-section-3.3

Build Your Own Customised AI-Powered Virtual Scenario-Based Simulations in GPTutor

Jessica NEUVILLE [1], Richard Wing Cheung LUI [2]
[1] School of Optometry, The Hong Kong Polytechnic University, Hong Kong, China
[2] Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China

Abstract

This workshop will guide participants step-by-step through creating virtual scenario-based simulations to enhance learners' critical thinking and communication skills. The simulations will be developed using the innovative GPTutor platform, an AI-powered tutoring system that provides learners with immediate and personalised feedback. Participants will be instructed on how to customise simulations, design avatars, and adapt assessment rubrics to align with discipline-specific learning outcomes.

Outcomes-based Learning Objectives

1. Participants will build their own customised patient/customer/assistant chatbots for scenario-based teaching.
2. Participants will use AI to develop customised grading rubrics that align with course learning outcomes.

Target Audience

This workshop is designed for professionals in healthcare fields or other academic disciplines where effective communication and interaction with people is a crucial part of the professional role.

Topic Relevance

This workshop aims to train participants on a customisable GenAI tool to enhance discipline-specific teaching, learning, and assessment.

Workshop Design

1. Workshop Format and Activities: Introduction

In this hands-on workshop, participants will be guided through the development of a scenario-based virtual simulation using the GPTutor platform developed at The Hong Kong Polytechnic University. GPTutor is a GenAI-powered virtual environment for personalised and active learning (Lui, 2024; Bai, 2025).

The workshop will focus on the use of GPTutor to enhance specific learning outcomes, such as developing students' professional competence, critical thinking, effective communication, and a self-motivated lifelong learner mentality. The workshop will start with an overview of the platform features. Then, participants will trial the platform in small groups with a pre-built healthcare professional simulation to experience the simulation capabilities from the student perspective. Following this initial introduction, participants will be given a few minutes to discuss potential use cases specific to their discipline within their small groups. Afterward, participants will build their own simulations.

2. Main Content Delivery

While GenAI tools for teaching are widely available nowadays, there are still significant barriers to adoption, including a lack of training for educators, access to resources, gaps between theory and practical applications, and time constraints associated with learning new technology (Huang & Yu, 2025; Taheri et al., 2025). Many teacher professional development (TPD) activities are a one-way flow of information, lacking detail or the necessary software for implementation, and do not result in a change of practice (Haras, 2018; Beer et al., 2016). Conversely, studies on effective teacher professional development programmes show that key characteristics include "collaborative learning environments, hands-on digital training, ongoing mentorship, and institutional support" (Amemasor et al., 2025).

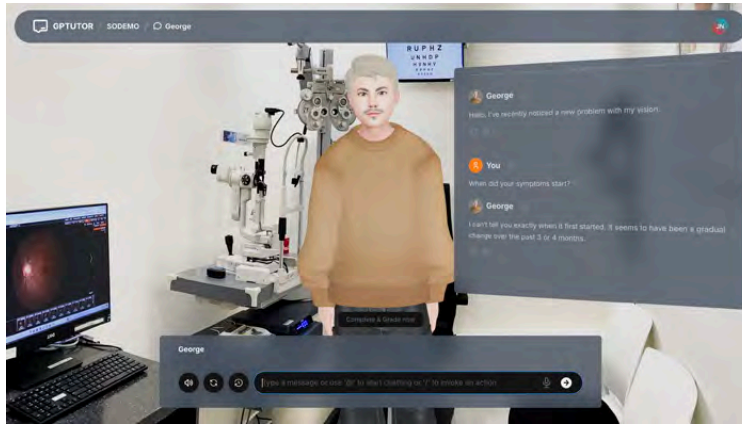
The main instructional method for this workshop will be a hands-on guided training session leading to the development of a customised simulation by the end of the workshop, which participants can deploy in their teaching the very next day. The workshop will utilise the GPTutor platform, which is developed with a simplified interface so that educators do not need any prior knowledge of computing or coding to use it. The workshop will be conducted in a collaborative learning space with round tables to allow participants to interact during the session. For those who join the workshop, ongoing mentorship and technical support will be offered after the in-person session to ensure that the tool can be effectively integrated into teaching.

3. Interactive Activities

For the hands-on guided training session, participants can work individually or collaborate in teams with others from their discipline.

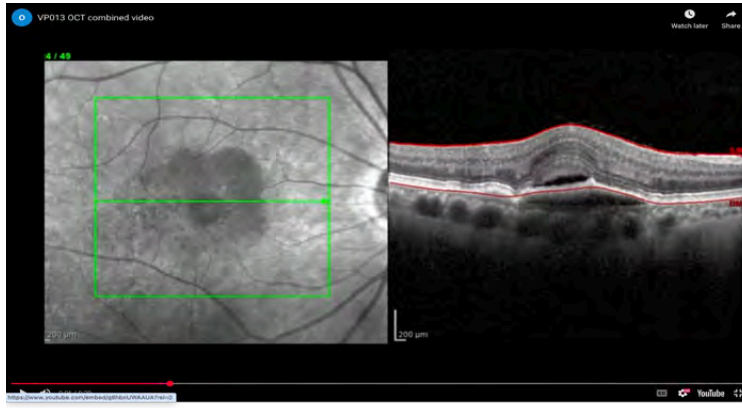
- **Step 1:** We will start by having participants create a profile for the patient/customer/person chatbot. Then, they will customise the chatbot's

avatar appearance and add a background image to set the scene. Afterward, participants will learn how to test the chatbot and further tailor the prompt to suit the scenario.



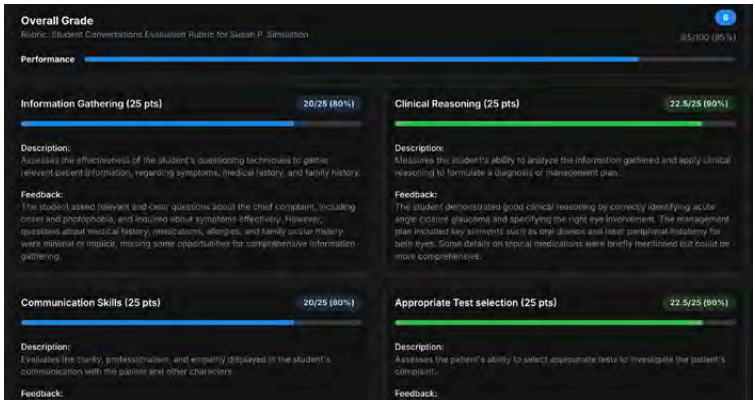
(Fig 1: Example of a virtual patient avatar in an eye examination room scene.)

- **Step 2:** Participants will be guided on how to add tests or tools to the scenarios. These are predefined outputs that can be in text, image, or video format. For example, in a medical use case, these outputs might include test results or clinical data that learners need to analyse in order to make decisions about the virtual patient's condition.



(Fig 2: Example of a predefined output in video format.)

- **Step 3:** Next, participants will be instructed on how to use GenAI to create a grading rubric for the simulation, including customising the criteria and marking scheme. They will then test the rubric to evaluate the AI-generated personalised feedback and the accuracy of the marking.



(Fig 3: Example of an AI-generated assessment report with personalised feedback.)

4. Discussion and Q&A

The discussion will focus on discipline-specific applications and suggestions for improvement to facilitate wider adoption of the platform and GenAI-powered tutors in general.

1. How could GPTutor virtual scenario-based simulations be incorporated into your teaching or discipline?
2. What suggestions do you have to improve the platform to facilitate its use in your teaching?
3. What barriers do you foresee in implementing virtual simulations into your teaching?

5. Closing Remarks

This workshop will provide hands-on digital training on using GPTutor to create discipline-specific simulations. By the end of the workshop, participants will have a custom-designed scenario-based simulation for their own teaching. Attendees will be encouraged to pilot the simulation with students and explore other features of the platform. Mentorship and technical support will be provided after the workshop to minimise barriers to implementation. A survey will be conducted at the end of the workshop and 2–4 weeks after the workshop to evaluate change in teaching practice and the effectiveness of the workshop format.

Acknowledgments

This project is supported by the Fund for Innovative Technology-in-Education (FITE) and the Teaching Development and Language Enhancement Grant, Inter-institutional Collaborative Activities (TDLEG-IICA 25-28).

Ethical Declaration Statement

The project has been approved by The Hong Kong Polytechnic University Institutional Review Board ("PolyU IRB") or its delegates. [Reference Number: HSEARS20240530005].

Generative AI Policy

The author(s) used Grok during the preparation of this work to search for relevant references. 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

Amemasor, S. K., Oppong, S. O., Ghansah, B., Benuwa, B.-B., & Essel, D. D. (2025). A systematic review on the impact of teacher professional development on digital instructional integration and teaching practices. *Frontiers in Education*, 10, Article 1541031. https://doi.org/10.3389/feduc.2025.1541031

Bai, H., Lui, W. C., & Khiatani, P. V. (2025). Promoting student engagement with GPTutor: An intelligent tutoring system powered by generative AI. *International Journal of Educational Technology in Higher Education*, 22, Article 77. https://doi.org/10.1186/s41239-025-00571-9

Beer, M., Finnström, M., & Schrader, D. (2016). *The great training robbery* (Working Paper No. 16-121). Harvard Business School. https://www.hbs.edu/faculty/Pages/item.aspx?num=50953

Haras, C. (2018, January 17). Faculty development as an authentic professional practice [Blog post]. *Higher Education Today*. https://www.higheredtoday.org/2018/01/17/faculty-development-authentic-professional-practice/

Huang, W., & Yu, W. (2025). Barriers to integrating artificial intelligence education: Implications from five early adopters in South Korea. *Computer Science Education*, 1–24. https://doi.org/10.1080/08993408.2025.2565199

Lui, R. W. C., Bai, H., Zhang, A. W. Y., & Chu, E. T. H. (2024). GPTutor: A Generative AI-powered Intelligent Tutoring System to Support Interactive Learning with Knowledge-Grounded Question Answering. In *2024 International Conference on Advances in Electrical Engineering and Computer Applications (AEECA)* (pp. 702–707). IEEE. https://doi.org/10.1109/AEECA62331.2024.00124

Taheri, R., Nazemi, N., Pennington, S. E., Clark, J. A., & Dadgostari, F. (2025). Factors influencing educators' AI adoption: A grounded meta-analysis review. *Computers and Education: Artificial Intelligence*, 9, 100464. https://doi.org/10.1016/j.caeai.2025.100464

Exploring Generative AI Use in Interactive Oral Assessments: A Case Study from a Company Law and Secretarial Practice Module

Paul NG
Centre for Learning and Teaching Excellence, Ngee Ann Polytechnic, Singapore

Abstract

This is a study of the implementation of Interactive Oral Assessments (**IOA**) in a second-year Company Law and Secretarial Practice (**CLSP**) module within the Accountancy diploma at Ngee Ann Polytechnic, Singapore.

1. Introduction

Traditionally, law is assessed through tests and assignments where students apply legal principles to address case- or scenario-based problems. CLSP implemented IOA - live, adaptive conversations assessed by rubric in which students are assigned trigger cases or scenarios to help them learn content before the assessment that will subsequently - during the assessment - be tested by application to novel situations. CLSP also purposefully-developed students to use Generative AI (**Gen AI**) as a learning assistant to prepare for their IOA.

2. Research Questions

a) To what extent was Gen AI effective as a preparatory tool for IOA?
b) What do students perceive are the challenges and benefits of IOA?

3. Methodology

a) **Participants:**
All 70 October 2025 semester CLSP students underwent the IOA. 66 students (94.3%) responded to a voluntary, anonymous postassessment survey.
b) **Design**: for details see **Appendix 1.**
c) **Data, Findings, Discussions and Limitations**: see **Appendix 2.**

4 Conclusion

IOA are a relatively new form of assessment, being about a decade old. CLSP positioned Gen AI as a preparatory aid rather than a substitute for performance in IOA. This case shows that IOA with Gen AIassisted preparation offers an efficient and integrityrich alternative to traditional tests and assignments to assess key human skills in the era of pervasive Gen AI including critical thinking, communication and adaptability.

(Readers can scan QR Codes to open and zoom in on poster visuals shared via an online depository).

Appendix 1: Design of the IOA in CLSP

Format, including steps to improve fairness, validity and reliability and a sample scenario:

IOA are the final assessment component in CLSP, and only cover end-of-life issues in company law including corporate recovery (schemes of arrangement, judicial management), receivership and liquidation.

Figure 1: The assignment brief (screenshots):

Brief Document

Module: Company Law – Secretarial Practice
Assessment Type: Oral Interview (Scenario-Based)
Duration: 10–15 minutes per student
Units Covered:

- Unit 7: Corporate Restructuring _ Scheme of Arrangement, Judicial Management and receivership
- Unit 8: Liquidation and Winding Up

Assessment Format and Expectations

Each student will be assessed individually on Unit 7 and Unit 8 through a scenario-based oral interview. The assessment will test your ability to apply legal principles to practical corporate scenario-based situations, articulate legal reasoning, and demonstrate a sound understanding of statutory frameworks under Singapore law.

A scenario-based question stem will be released before the scheduled interview. This advance notice is intended to give students sufficient time to prepare for the oral assessment using Gen AI tools. Students will be informed of their respective time slots before the oral exams.

Based on the scenario, the student prepares an answer on the priority of creditors in the event of winding up, where there are sufficient funds to pay all creditors under Section 203(1) IRDA 2018.

1. Student will have to fill their particulars in **Appendix A** and insert their answer on priority of debt in the box provided.
2. **Appendix A** will be printed and submitted to the examiner during the assessment.
3. The student is not allowed to bring in the textbook, notes, or any electronic devices during the examination.
4. During the actual oral assessment, the student will be asked questions (including "what if questions").
5. Student will be **present at least half hour before the allocated time and leave the waiting/examination area and not mingle with their classmates after the oral assessment.** The student should **refrain** from discussing the questions or answers with other students. This is an assessment and therefore **the tutor reserves the right to penalise any student found discussing or sharing his/her/examiners' assessment questions and answers with other students**.

Preparation Guidelines

- Review **Units 7 and 8** thoroughly.
- Read Chapters 13, 14, 15 in Gopalan, C. (2022). *Company law in Singapore* (6th ed.). McGraw-Hill Education (Asia).
- For avoidance of any doubt, take note of guidance on **S203 (1) IRDA 2018 & S203(2) IRDA 2018 in respect of the current cap for the amount payable under S203 (1) (d), (e) and (h) IRDA 2018 is SGD13K or 5 months salary whichever is lower.**
- Please also take note of S203 (6) IRDA 2018 in respect of **postponement of Priority of Floating Charge** and S101(1) IRDA 2018 in respect of **rescue financing and Payment of Super Priority Debts** at page 257 of the Textbook.
- Use Gen AI tools to simulate interviews and refine responses.
- Practice articulating answers clearly and confidently.
- Structure responses using Law, Application, Conclusion.
- Be prepared to respond to follow-up questions during the interview, including "what if" speculative questions.

Use of Generative AI Tools

Students are **required** to use **Generative AI tools** to support their preparation for the oral assessment. These tools should be used to:

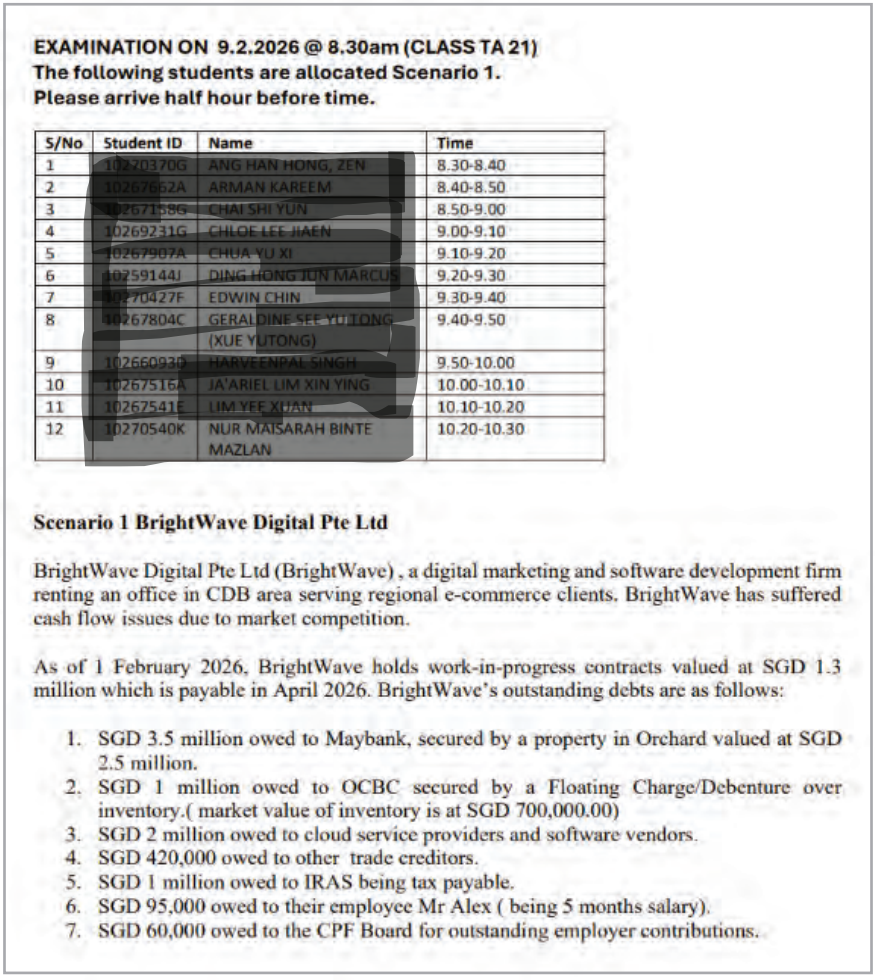
- Simulate mock interviews or oral assessments on the scenario-based stem provided.
- Clarify legal concepts and statutory provisions.
- Generate and refine structured responses based on law, application and conclusion.
- Prepare to respond to follow-up questions during the interview.

Important: Students must verify all AI-generated advice and other AI content with authoritative legal sources. Students may refer to Gopalan, C. (2022). *Company law in Singapore* (6th ed.) McGraw-Hill Education (Asia), the PPTs and Video lectures for Unit 7and 8 in the Brightspace. These materials should be used to confirm:

- Statutory references and legal principles.
- Case law and judicial interpretations.
- Procedural accuracy and terminology.
- Students should also prepare for and expect to be asked "what if" questions during the actual oral assessment.

Each CLSP class was split into two groups. Each group received a different scenariobased question one week before the IOA. Assigning varied scenarios across classes provided students with opportunities to practise applying legal principles to new facts by sharing their scenarios and improved assessment validity by deterring students from "test-smarts" - trying to maximise the universe of questions and answers to anticipate and memorize across sessions since the scenarios were different.

Figure 2: Screenshot of a scenario assigned to a group to prepare for their IOA (students' particulars anonymized):



EXAMINATION ON 9.2.2026 @ 8.30am (CLASS TA 21)
The following students are allocated Scenario 1.
Please arrive half hour before time.

S/No	Student ID	Name	Time
1	10270370G	ANG HAN HONG, ZEN	8.30-8.40
2	10267662A	ARMAN KAREEM	8.40-8.50
3	10267158G	CHAI SHI YUN	8.50-9.00
4	10269231G	CHLOE LEE JIAEN	9.00-9.10
5	10267907A	CHUA YU XI	9.10-9.20
6	10259144J	DING HONG JUN MARCUS	9.20-9.30
7	10270427F	EDWIN CHIN	9.30-9.40
8	10267804C	GERALDINE SEE YU TONG (XUE YUTONG)	9.40-9.50
9	10266093B	HARVEENPAL SINGH	9.50-10.00
10	10267516A	JA'ARIEL LIM XIN YING	10.00-10.10
11	10267541E	LIM YEE XUAN	10.10-10.20
12	10270540K	NUR MAISARAH BINTE MAZLAN	10.20-10.30

Scenario 1 BrightWave Digital Pte Ltd

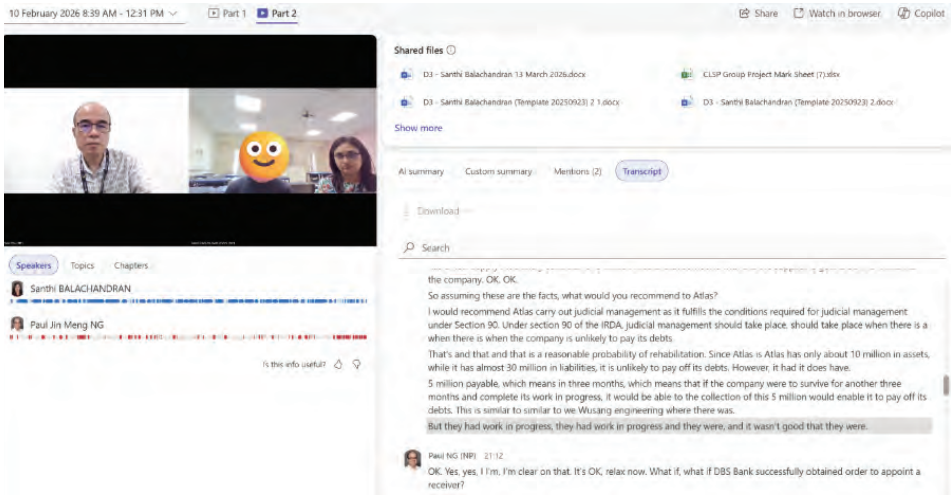
BrightWave Digital Pte Ltd (BrightWave) , a digital marketing and software development firm renting an office in CDB area serving regional e-commerce clients. BrightWave has suffered cash flow issues due to market competition.

As of 1 February 2026, BrightWave holds work-in-progress contracts valued at SGD 1.3 million which is payable in April 2026. BrightWave's outstanding debts are as follows:

1. SGD 3.5 million owed to Maybank, secured by a property in Orchard valued at SGD 2.5 million.
2. SGD 1 million owed to OCBC secured by a Floating Charge/Debenture over inventory.(market value of inventory is at SGD 700,000.00)
3. SGD 2 million owed to cloud service providers and software vendors.
4. SGD 420,000 owed to other trade creditors.
5. SGD 1 million owed to IRAS being tax payable.
6. SGD 95,000 owed to their employee Mr Alex (being 5 months salary).
7. SGD 60,000 owed to the CPF Board for outstanding employer contributions.

Each student was assessed by two coassessors, including one who had not taught the student, to enhance assessment validity. All IOA were recorded via MS Teams, with recordings used to support grading reliability through moderation across sessions.

Figure 3: screenshot from an MS Teams IOA recording with the student's face anonymized, showing one co-assessor (the module leader) in the physical venue with the student, and a part of the transcript of the conversation including the other co-assessor's "what if" question:



Duration:

Students underwent three weeks of preparation before the IOA, which was conducted in the final week and weighted at 25% of the module grade.

Figure 4: screenshot of the IOA portion of the CLSP teaching and learning schedule.

14	19-Jan	**Unit 7:** Corporate recovery schemes 7.1 Scheme of arrangement 7.2 Judicial management	Discuss & present answers to tutorial questions in Unit 7	*Tutorial Topic – Unit 7* Feedback on Common Test
15	26-Jan	**Unit 8:** Corporate insolvency proceedings 8.1 Receivership 8.2 Liquidation	Discuss & present answers to tutorial questions in Unit 8	*Tutorial Topic – Unit 8*
16	02-Feb	Revision for quiz & Scenario-based oral assessment		**Class Quiz 2 -5%** **[Unit 6-8]**
17	09-Feb	**SCENARIO-BASED ORAL ASSESSMENT (Unit 7 & 8)**		

The IOA were conducted in Week 17 and administered in halfday sessions, with each class allocated just over four hours and up to ten minutes per student. Assessment was guided by a rubric aligned to the Polytechnic's grading scale.

Figure 5: The grading rubric:

Criteria	A+	A	B	C	D	F	Max Marks
Description of the Relevant Law	Description of the law which is entirely accurate with detailed references.	Description of the law which includes mostly accurate and detailed referencing; minor gaps may exist.	Description which is fundamentally accurate but fails to provide detailed references to some of the relevant law.	Description which although fundamentally accurate, fails to provide detailed references to most of the relevant law.	Description of the law which although fundamentally accurate, fails to provide detailed references to all of the relevant law.	Incorrect description of the law.	25
Application (including Conclusion)	Comprehensive application of the law to given scenarios without any errors.	Application of the law to given scenarios with only minor errors in the analysis.	Application of the law to given scenarios with a few significant errors in the analysis.	Application of the law to given scenarios with some significant errors in the analysis.	Application of the law to given scenarios with many significant errors in the analysis.	No or entirely erroneous application or analyses.	25
Total Score	**Grade**						
42.5 – 50	A+						
40 – **42**	A						
35 – **39**	B						
30 – **34**	C						
25 – **29**	D						
0 – **24**	F						

Figure 6: Screenshots from the co-assessors' shared gradebook showing the mapping of some students' grades to their marks and some co-assessor's notes – made contemporaneously or immediately after every conversation - for a flavour of the conversations that are a hallmark of IOA. Students' particulars have been anonymized:

Student ID	NAME	STUD ACAD PROG	CLASS SECTION	CLASS NBR	Law (grade)	Law (25 marks)	Application (Grade)	Application (25 Marks)	MARKS	Comments
[redacted]		ACC	TA21	7	A+	25	A+	25	50	Vision affected. Given an A3 size paper scenario sheet to refer to, and extra time to read more slowly. Receivership only resolves the problem with the secured charges. Like earlier students he thinks SOA and creditor vote won't work and compared to JMO winding up is the more likely remedy. Able to cite s.125 rule on not paying a debt in 3 weeks (presumption of insolvency). Able to point out the cap on Alex's salary of 13k, remaining under unsecured debts. Answering the what if about in progress value $5m - if WIP is $5m since (some hesitation, looks at the numbers again) - if Orchard property is sold it's still no go... but settled ultimately on JMO; business sustainable, all creditors can be paid. Pro-rated etc. - enough to pay creditors - probable rehab of company, asked for time to name all the law on JMO including section numbers and cases.
[redacted]		ACC	TA21	8	A+	25	A	21	46	The numbers don't meet SOA criteria - 75% rule cited and applied, but couldn't figure out why Maybank may NOT oppose an SOA with moratorium. Said JMO the better option. Poss rehab - gave the correct s.91 - s.89 IRDA - s.210CA - can cite the law but application weak in than students before. She has the 1.3m - moratorium for it. Not enough to pay all debts but potential is there. Digital marketing is a growing market. Overall, the weakness is cannot reason out why she herself said secured creditors wouldn't vote for SOA. A bit shallow reasoning there.
[redacted]		ACC	TA21	9	A	21	B	19.5	40.5	He believed SOA is better. Pointed out revenue coming in April. They have 2 secured loans, JMO - lower value, more shortfall - so go for SOA (reasoning unclear to me). Alm for longer repayments. OCBC & Maybank agreeing to SOA bec realise they will eventually lose out (advantageous realisation). On Law: cases on SOA? section 210 CA. Receivership by 1. and 2 on the list. He did NOT know the others cannot apply for receivership. Winding up - he insisted on the 1.3m WIP. W/up as last option. On Law on w/up, or JMO, he gave Wusang Engineering.

Student ID	NAME	STUD ACAD PROG	CLASS SECTION	CLASS NBR	Law (grade)	Law (25 marks)	Application (Grade)	Application (25 Marks)	MARKS	Comments
[redacted]		ACC	TA23	2	A+	25	A	21	46	Was unable to reason out that OCBC would not benefit from JMO unless interest was factored in for the moratorium (delay) on repayment; OCBC can get 100% recovery by appointing a receiver
		ACC	TA23	3	A	21	C	17	38	poorer grasp of the specifics e.g. section numbers or specific words, compared to the earlier 2 students, but he got Wusang as well, and had certainly memorised the law. Application: it's actually possible for OCBC to get back what it is owed, without SOA. [redacted] considered UOB and pointed out correctly that UOB cannot get everything back via receivership. He went through characteristics of SOA and JMO but found it difficult to figure out why JMO preferable to SOA. His answers indicated he cannot link the law to the facts to e.g. explain why UOB necessarily would object to a SOA. Couldn't answer the question about fair and equitable - his answers.
[redacted]		ACC	TA23	4	A	21	A	21	42	[redacted] chose JMO to start with and didn't go through SOA, which was a very efficient way to address my question on next steps for Lionfech. She came to SOA later in her analysis. She was able to analyse OCBC potential objection to JMO correctly. She was able to figure out vaguely that Lionfech would have to offer OCBC something in return for not objecting to JMO. She offered shares - unlikely - but she also pointed out that OCBC is safe in any event, they always have the recovery of $1million even with JMO (it's just delayed). [redacted] did point out that liquidation doesn't benefit OCBC i.e. it's not a good remedy in any case, from OCBC point of view.

Fairness: steps to Prepare students using Gen AI

Learning activities in the 3 weeks before the assessment:

Students were prepared through endofclass roleplay activities conducted in small groups, where they alternated between the roles of assessor and candidate. Using a case or scenario, students used Gen AI to generate and practise responding to potential scenariobased and followup or alternative "what if" questions. Through iterative prompting, some students generated a large pool of potential questions and answers, allowing them to discuss their assigned case or scenario from multiple angles. Multiple Gen AI tools (e.g. Gemini, ChatGPT, Copilot, and Google NotebookLM) were explored.

Figure 7: Screenshots showing a scenario-based question students used to prepare for their IOA, and from a student showing the student role-playing an "Examiner" querying the Gen AI "adviser" (Gemini):

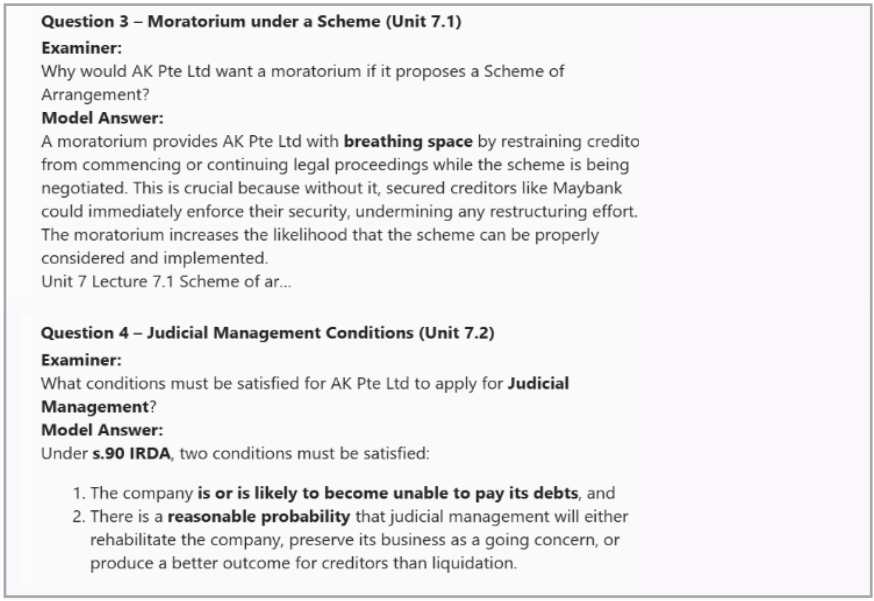
Scenario-Based Analytical Question 1:
AK Pte Ltd, a long-established Singaporean manufacturer of high-precision metal components for the automotive industry, is facing significant financial distress due to the global transition to electric vehicles (EVs) and increased automation. These industry disruptions have led to declining demand, outdated machinery, and mounting financial distress.
As of 30 July 2025, the company holds SGD 1 million in cash.
AK Pte Ltd has several outstanding financial obligations:

1. It owes SGD 6 million to Maybank Singapore, secured by a fixed charge over its freehold factory land and building (market value: SGD 7 million).
2. It has a floating charge over its inventory (market value: SGD 1 million) securing a SGD 1 million loan.
3. It owes SGD 3 million to JP Pte Ltd (raw material supplier)
4. It owes SGD 500,000 to TSpeed Pte Ltd (transportation costs).
5. The company is in arrears of SGD 100,000 in unpaid salary owed to five employees (SGD 20,000 each, representing five months of unpaid salary).
6. It is also liable for SGD 36,000 in retrenchment benefits (equivalent to two months' salary) following the termination of its Finance Manager.

[redacted] /ACC 20/01 12:26 AM

Question 1 – Identifying Financial Distress (Unit 7 Intro)
Examiner:
Based on the facts, is AK Pte Ltd financially distressed under Singapore law?
Model Answer:
Yes, AK Pte Ltd is financially distressed and likely insolvent. The company has only **SGD 1 million in cash** but total liabilities exceeding **SGD 10 million**, including secured bank loans, trade creditors, and unpaid employee wages. The company is also in **arrears of salaries**, which is a strong indicator that it is **unable to pay its debts as and when they fall due**, satisfying the insolvency test under the IRDA. This places AK Pte Ltd squarely within the scope of corporate recovery or insolvency proceedings.
Unit 7 Lecture 7.1 Scheme of ar...

Question 2 – Suitability of a Scheme of Arrangement (Unit 7.1)
Examiner:
Would a **Scheme of Arrangement** be suitable for AK Pte Ltd?
Model Answer:
A Scheme of Arrangement could be suitable if AK Pte Ltd believes it can **rehabilitate its business** through debt restructuring. Under **s.210 Companies Act and ss.63–72 IRDA**, the company can propose a compromise to creditors, such as deferring loan repayments or partially extinguishing debts.
However, the scheme will only succeed if **a majority of creditors representing at least 75% in value approve it**, and if the court finds it **fair and reasonable**. Given the presence of multiple creditors with competing interests, especially secured creditors like Maybank, achieving the required approval may be challenging but still possible.
Unit 7 Lecture 7.1 Scheme of ar...



Question 3 – Moratorium under a Scheme (Unit 7.1)

Examiner:
Why would AK Pte Ltd want a moratorium if it proposes a Scheme of Arrangement?

Model Answer:
A moratorium provides AK Pte Ltd with **breathing space** by restraining credito from commencing or continuing legal proceedings while the scheme is being negotiated. This is crucial because without it, secured creditors like Maybank could immediately enforce their security, undermining any restructuring effort. The moratorium increases the likelihood that the scheme can be properly considered and implemented.
Unit 7 Lecture 7.1 Scheme of ar...

Question 4 – Judicial Management Conditions (Unit 7.2)

Examiner:
What conditions must be satisfied for AK Pte Ltd to apply for **Judicial Management**?

Model Answer:
Under **s.90 IRDA**, two conditions must be satisfied:

1. The company **is or is likely to become unable to pay its debts**, and
2. There is a **reasonable probability** that judicial management will either rehabilitate the company, preserve its business as a going concern, or produce a better outcome for creditors than liquidation.

As part of preparation, students were also trained to use the podcast function in Google NotebookLM for 2-way verbal discussion. Students prompted the podcast to conduct a questionandanswer session. The podcast posed questions one at a time, students responded verbally, and were then instructed to prompt the tool to critique, correct, or refine their responses.

This was aligned to IOA as IOA require verbal structured thinking and communication skills. This also exposed students to the use of innovative Gen AI tools.

Figure 8: A link to a video of the module leader's experiment with Google Notebook LM as a sparring partner to prepare for the IOA:
https://ivid2.np.edu.sg/media/From+Santhi+Balachandran.+Google+NotebookLM+audio+podcast+radio+talkshow/1_wwgcuovt

The module leader showed a video and conducted a live demonstration for students to pose questions to the NotebookLM podcast. She also demonstrated how to adjust the speed of the conversation and guided students to set up NotebookLM independently, including limiting source materials to reduce hallucinations.

Appendix 2: Findings and Discussion

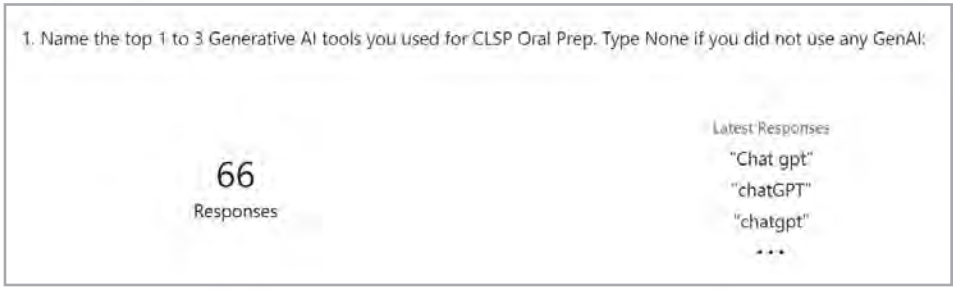
Survey details:

66 students out of 70, or 94.3% of the cohort, responded to the module's anonymous post-assessment survey. At a 95% confidence level and 5% margin of error, the minimum sample size for statistically-valid conclusions to be drawn about the module's student population was 60. Therefore, valid conclusions can be drawn from this survey about October 2025 semester CLSP students as a whole, with a margin of error of up to 5%.

Survey Findings:

Survey item #1 asked students which tools they used the most often to prepare for their IOA.

Figure 9: survey item #1 on usage of Gen AI tools:



Responses exceed 66 because students could name their top 3 tools.

Usage of Generative AI Tools (n = 66)

Generative AI Tool	Number of Students
ChatGPT	**48**
NotebookLM (Google)	**28**
Gemini (Google)	**25**
Copilot	4
None (Did not use GenAI)	4
DeepSeek	2
Perplexity	1
Clue	1

Quick observations (useful for reporting / reflection)

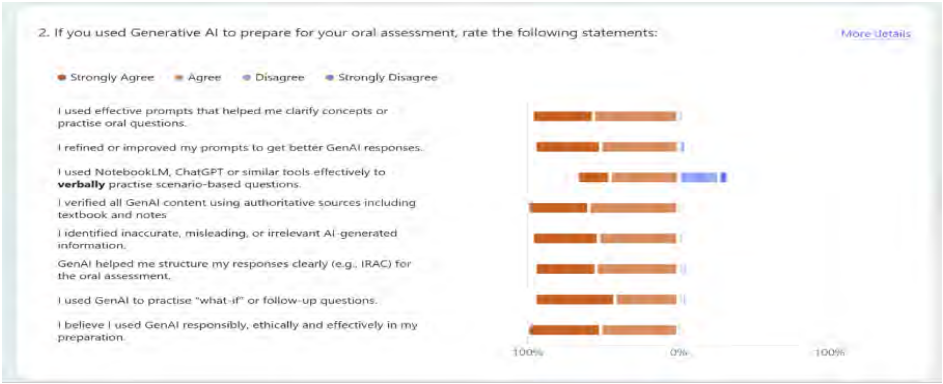
- **ChatGPT** is clearly the dominant tool (used by ~73% of respondents).
- **NotebookLM and Gemini** are also heavily used, showing strong uptake of **Google ecosystem tools**.
- Very **few students did not use any GenAI** (only 4).
- Alternative tools (DeepSeek, Perplexity, Clue) had **minimal adoption**

The most popular tool was ChatGPT. This was followed by NotebookLM and Gemini. CoPilot, DeepSeek, Perplexity and "Clue" (possibly a reference to Cluely.ai?) experienced much less usage than the first-mentioned three tools.

Survey Item #2:

Most students self-assessed their use of Gen AI to prepare for the IOA favourably, except for "I used Notebook LM, ChatGPT or similar tools effectively to verbally practice scenario-based questions."

Figure 10: survey item #2 on how students used Gen AI to prepare for IOA



Discussion:

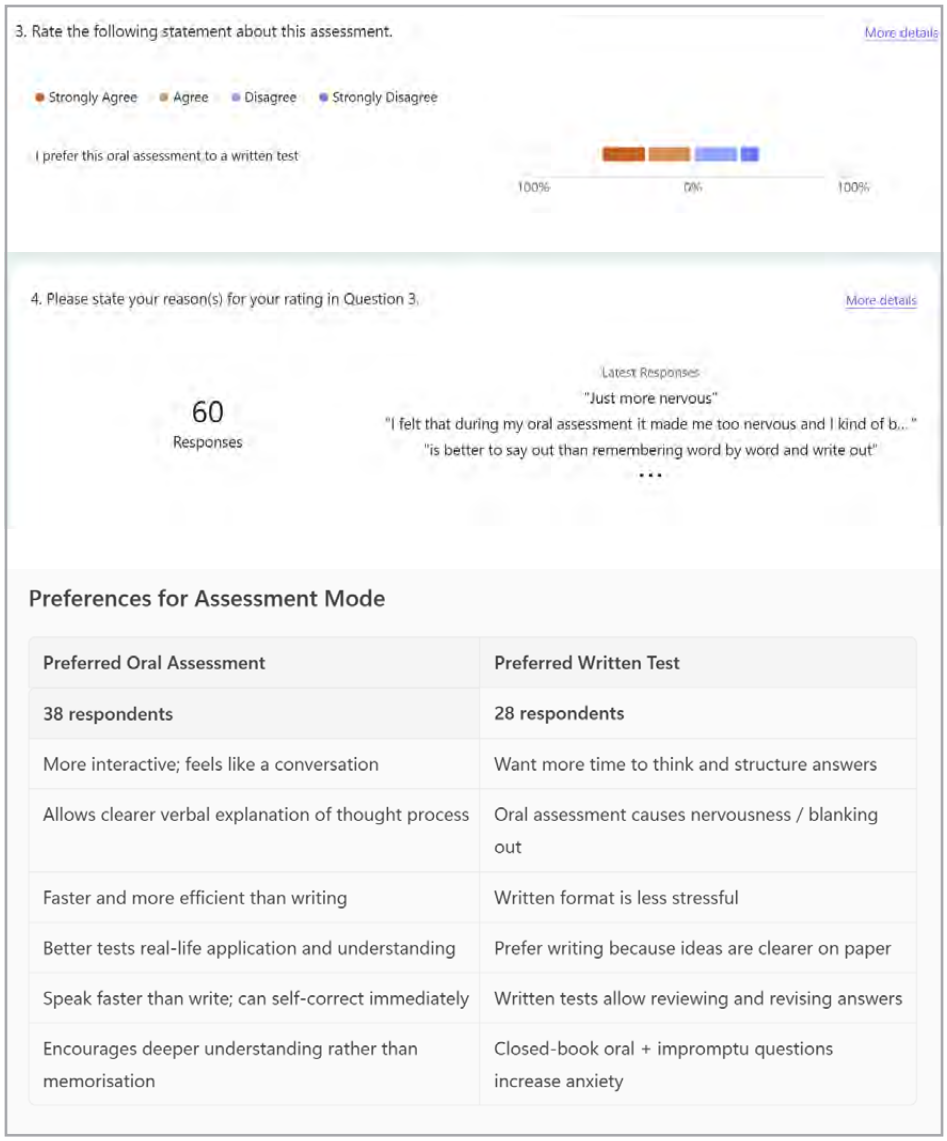
Survey items #1 and #2 address the Research Question: To what extent was Gen AI effective as a preparatory tool for IOA? Answer: from the survey results, it was effective to a very great extent, with nearly 100% Strongly Agree and Agree responses on all sub-items except verbal practice.

Students had been introduced to Gen AI verbal practice features so lack of familiarity was not the reason for their poor self-rating. These tools however require patience as the verbal interactions take time to generate and there is also some lag in response. Further, a quiet environment is necessary for such practice which prevents students from practising wherever such an environment does not exist. More work seems necessary on the one hand to monitor and leverage on improvements in Gen AI that facilitate verbal interactivity and on the other, to help students improve, in practical terms, their use of Gen AI to practice verbal interactions prior to IOA.

Survey Items #3, #4 and #5:

CLSP students prefer, by a majority of 58% to 42%, the IOA to a written test (survey item #3). Reasons students gave in survey item #4 for their responses to item #3 are thematically summarized below.

Figure 11: survey items #3 and #4 on students' preferences for oral rather than written test, and their reasons for their preferences, organized by theme.

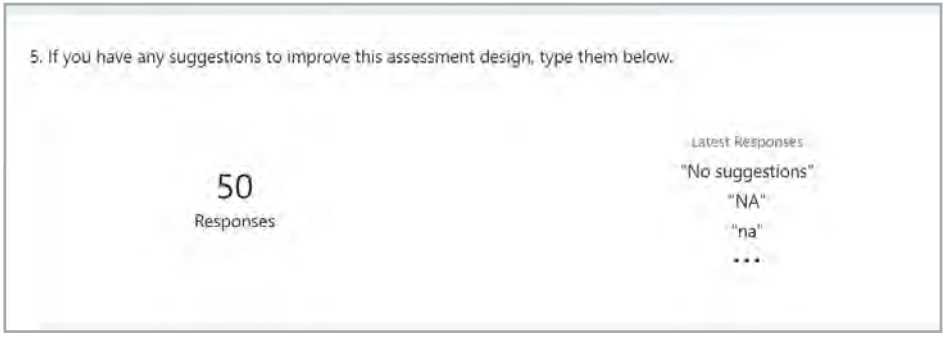


Preferences for Assessment Mode

Preferred Oral Assessment	Preferred Written Test
38 respondents	**28 respondents**
More interactive; feels like a conversation	Want more time to think and structure answers
Allows clearer verbal explanation of thought process	Oral assessment causes nervousness / blanking out
Faster and more efficient than writing	Written format is less stressful
Better tests real-life application and understanding	Prefer writing because ideas are clearer on paper
Speak faster than write; can self-correct immediately	Written tests allow reviewing and revising answers
Encourages deeper understanding rather than memorisation	Closed-book oral + impromptu questions increase anxiety

Finally, **survey item #5** asked students for suggestions to improve the assessment:

Figure 12: survey item #5 on suggestions to improve, organized by theme:



Theme	Count	Representative student comment
Nil / NA / No suggestions to improve	34	"nil" "i think it is a good assessment to test concepts though"
Openbook / reference / cheatsheet support	8	"If we can bring in references will be better."
More time / preparation support	2	"Maybe give the students 5 mins to prepare before they answer."
Examiners' clarity or consistency	3	"Questions should be in detail. Like one of the questions asked was look at the list and say which creditors would be affected most. I wasn't sure if they wanted it judged of the assumption that there was sufficient finds for all creditors or assuming strictly the facts of the case. The answer will defer"
Practice, feedback, or alternative formats	3	"Maybe possibly do a video reflection instead where the chers can assign different questions to the students 1 day beforehand like DDA video assignment"

Discussion:

Survey items #3, #4 and #5 address the Research Question: What do students perceive are the challenges and benefits of IOA?

Answer: IOA were well received by a majority (58%) of CLSP students despite their novelty and the ubiquity of written tests. Students who offered suggestions for improvement were primarily hoping to bring in materials and have quiet time to refer to them during the IOA, as opposed to having to engage with assessors in free-flowing conversations based in part on memory - a pre-requisite for good-quality verbal reasoning. A few students may also have felt some lines of conversation were unclear. Their survey responses suggest that affective factors and cognitive load are significant challenges to students in IOA.

After due consideration, the CLSP module leader decided fundamentally to continue with IOA in its current form. IOA and traditional written or typed tests – even open book tests - are NOT mere substitutes for one another. IOA conversationally enable deeper dives than written tests into each student's quality of learning. Nearly 6 in 10 CLSP students seem to have appreciated the benefits of IOA in what is only its second semester of implementation. IOA's acceptance should grow over time, aided by increasing staff expertise gained through experience and correspondingly, better preparation of students.

Limitations of this study:

The simple post-assessment survey in this study failed to directly obtain student feedback on their perceptions of the fairness, validity, manageability and reliability of the IOA. Those perceptions have had to be inferred from responses to more general survey items.

Acknowledgments: The authors acknowledge with thanks the support of the management of the School of Business and Accountancy and the Centre for Learning and Teaching Excellence for the implementation of IOAs in CLSP.

Ethical Declaration Statement

The institution approved this application.

This proposed poster briefly describes a study of an assessment the implementation of which was approved by the institution.

Generative AI Policy

The author(s) used Microsoft CoPilot to create the reference list below. The author(s) used Microsoft CoPilot and other Gen AI tools to implement the

assessment, including preparing students for it. Gen AI tools were used to assist with survey creation and other aspects of the study. After utilising the Gen AI tools/services for the study, the author(s) thoroughly reviewed and edited the content of this submission as necessary and assume full responsibility for the content of the poster that shall be displayed at the conference.

References:

Della Vedova, C., Faulkner, M., Chanana, V., et al. (2025). Reimagining assessment: Interactive oral assessments as authentic, integrity-rich alternatives to traditional evaluation methods. *ASCILITE 2025 Conference Proceedings.* https://doi.org/10.65106/apubs.2025.2744

Griffith University. (2020). *Interactive oral assessments: A viable solution to mitigate global threats to higher education.* https://app.secure.griffith.edu.au/exlnt/uploads/attachments/09569/493e94a-4b8454848830a005e9397a32c614f745b.pdf

Ryan, J., Ziguras, C., & Mulder, R. (2025). *Interactive oral assessments: Design and implementation guide.* Melbourne Centre for the Study of Higher Education. https://melbourne-cshe.unimelb.edu.au/resources/topics/assessment-and-feedback/specific-help/interactive-oral-assessments

Stephenson, Z., Johnson-Glauch, N., & Cruchley, S. (2025). Interventions and facilitators of oral assessment performance in higher education: A systematic review. *Assessment & Evaluation in Higher Education,* 50(7), 1140–1153. https://doi.org/10.1080/02602938.2025.2504621

Ward, M., O'Riordan, F., Logan-Fleming, D., Cooke, D., Concannon-Gibney, T., Efthymiou, M., & Watkins, N. (2024). Interactive oral assessment case studies: An innovative, academically rigorous, authentic assessment approach. *Innovations in Education and Teaching International,* 61(5), 930–947. https://doi.org/10.1080/14703297.2023.2251967

Influence of ChatGPT's Feedback on Student Motivation and Learning: The Moderating Role of Trust

Cong QI [1] & Yue LEI [1]
[1] Department of Management and Marketing, The Hong Kong Polytechnic University, Hong Kong, China

Abstract

This study investigates the influence of ChatGPT's feedback on students' intrinsic motivation, engagement, and learning outcomes in higher education, with a focus on the moderating effect of trust. Drawing from Self-Determination Theory (SDT) and Self-Regulated Learning (SRL) theory, this study examines how AI-generated feedback affects learners' cognitive learning status and academic performance. A survey was conducted among 300 programming students; data analysis results revealed that feedback from ChatGPT enhances intrinsic motivation, which leads to a higher level of affective and behavioural engagement, ultimately boosting learning outcomes. However, contrary to our assumption, trust towards ChatGPT negatively (instead of positively) moderates the relationship between feedback and intrinsic motivation. These findings suggest that integrating AI-driven feedback mechanisms like ChatGPT can greatly enhance educational experiences by promoting intrinsic motivation, active engagement, and academic outcomes.

Keywords: Feedback, ChatGPT, Intrinsic Motivation, Engagement, Learning Outcomes

1. Introduction

Recently, generative AI has been used to provide reliable, personalised feedback to students with less teacher effort and intervention (Cavalcanti et al., 2021). ChatGPT has become a powerful tool to facilitate students' learning and improve educational outcomes. This study employs Self-Determination Theory (SDT) (Ryan & Deci, 2000) to explain how feedback enhances intrinsic motivation and uses Self-Regulated Learning (SRL) theory (Nicol & Macfarlane-Dick, 2006) to clarify how self-regulated learners engage affectively and behaviourally to improve learning outcomes.

Trust has been frequently studied as an antecedent of ChatGPT adoption (Choudhury & Shamszare, 2023); however, its moderating role in the relationship between ChatGPT's feedback and students' intrinsic motivation is yet to be explored. We proposed that trust towards ChatGPT would positively moderate the relationship between feedback and intrinsic motivation; thus, when the perceived trust level is high, students are more likely to view the feedback as helpful and reliable, thereby enhancing their enjoyment and pleasure in learning.

In sum, we examine how feedback shapes motivation, engagement, and learning outcomes, as well as the moderating effect of trust towards ChatGPT. Our research questions are:

1. Will high-quality feedback from ChatGPT influence students' intrinsic motivation? Will this relationship be positively moderated by perceived trust in ChatGPT?
2. Will intrinsic motivation positively influence two forms of engagement (affective and behavioural), and eventually lead to enhanced learning outcomes?

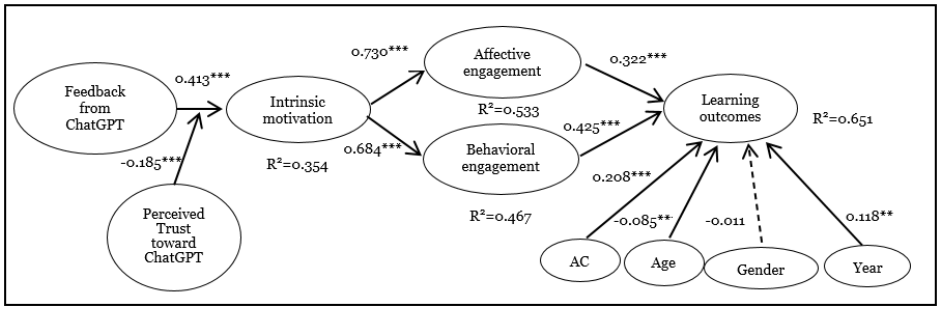
2. Methodology

Measures were adopted from past literature with minor adjustments to the ChatGPT context. Feedback from ChatGPT was measured by timeliness, accuracy, detailedness, and more (Lipnevich & Panadero, 2021). Feedback from ChatGPT is usually structured and standardised in response to programming-related prompts. Intrinsic motivation involves satisfaction, achievement, and confidence; learning outcomes are a direct measure of self-reported course grades.

A cross-sectional survey was conducted at a Hong Kong university. The participants were undergraduate students who had taken programming-related subjects in the past two years. A total of 1,000 questionnaires were sent out, with 410 valid responses, making the response rate 41%. We then excluded those with incomplete responses (20) and those who never used ChatGPT in their schoolwork (90). Finally, 300 data points became the final dataset for analysis. SPSS (v. 29) and SmartPLS (v. 4) were used to analyse the data.

3. Results and Discussion

Figure 1: Structural Model Data Analysis Results



(Note: H2 indicates the hypothesis regarding the moderating role of trust)

- **Hypothesis Testing:** All other main hypotheses were supported except H2.
- **Model Fit:** R-squared values demonstrated good explanatory power for the model.

The data analysis results revealed the following:

1. High-quality feedback from ChatGPT triggers intrinsic motivation, which influences learning outcomes through the mediating effects of affective and behavioural engagement.
2. The findings support SDT and SRL, indicating that effective feedback motivates self-regulated learners to engage emotionally and behaviourally in learning, thereby enhancing their performance.
3. Trust towards ChatGPT was not proven to positively influence the relationship between high-quality feedback and intrinsic motivation. Instead, it negatively moderated the relationship between feedback and intrinsic motivation. This may be because over-reliance on GenAI increases laziness and reduces students' intrinsic motivation to learn (Sarabadani & Akbari Ghatar, 2025).

4. Implications

The most interesting finding is the negative moderating effect of trust in the relationship between feedback from ChatGPT and intrinsic motivation. This result provides practical implications for universities to pay special attention to over-reliance issues regarding the use of ChatGPT. University policymakers should warmly embrace students' normal use of ChatGPT, as feedback from it can greatly motivate students to enjoy studying and further engage emotionally

and behaviourally with learning tasks, ultimately leading to enhanced learning outcomes. However, universities should also guide students to avoid over-reliance on the use of ChatGPT, as it could potentially diminish students' motivation to learn.

References

Cavalcanti, A. P., Barbosa, A., Carvalho, R., Freitas, F., Tsai, Y.-S., Gašević, D., & Mello, R. F. (2021). Automatic feedback in online learning environments: A systematic literature review. *Computers and Education: Artificial Intelligence*, 2(100027), 100027. https://doi.org/10.1016/j.caeai.2021.100027

Choudhury, A., & Shamszare, H. (2023). Investigating the Impact of User Trust on the Adoption and Use of ChatGPT: Survey Analysis. *Journal of Medical Internet Research*, 25(1), e47184. https://doi.org/10.2196/47184

Lipnevich, A. A., & Panadero, E. (2021). A Review of Feedback Models and Theories: Descriptions, Definitions, and Conclusions. *Frontiers in Education*, 6(1). https://doi.org/10.3389/feduc.2021.720195

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative Assessment and Self-regulated learning: a Model and Seven Principles of Good Feedback Practice. *Studies in Higher Education*, 31(2), 199–218. https://doi.org/10.1080/03075070600572090

Ryan, R. M., & Deci, E. L. (2000). Self-determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-being. *American Psychologist*, 55(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Sarabadani, J., & Akbari Ghatar, P. (2025). Unpacking Students' Irresponsible Use of Generative AI in Higher Education: A Review of Antecedents and Consequences. *Proceedings of AMCIS 2025*, (12). https://aisel.aisnet.org/amcis2025/is_education/is_education/12

SmartFeedback: AI-Powered Grading and Feedback for Data Analysis Reports

WU Tian [1] & WANG Ho, Simon [2]
[1] Department of Mathematics, Hong Kong Baptist University, Hong Kong, China
[2] Language Centre, Hong Kong Baptist University, Hong Kong, China

Abstract

SmartFeedback is an AI-assisted system designed for rubric-based grading and feedback on data analysis reports. Built upon the Poe App-Creator platform (utilising Claude Sonnet 4.5), it comprises two modules: an instructor grader that produces a rubric-aligned table of deductions with linked evidence for review, and a student summariser that delivers concise, actionable takeaways. In testing, SmartFeedback reduced grading time by over 30% (from approximately 21 to approximately 14 minutes per report) while maintaining comparable effectiveness. This effectiveness was assessed via student surveys regarding feedback fairness, clarity, consistency, and actionability. The workflow incorporates human oversight, anonymised data, rubric-constrained prompts, and mandatory evidence citations to mitigate hallucinations and bias. The framework is scalable to other data-centric courses and supports student-centred pedagogy.

Keywords: AI-assisted grading, data analysis report, rubric-based assessment, human-in-the-loop

1. Introduction

Grading data analysis reports is a time-intensive, deadline-driven task; yet students require consistent, actionable feedback to develop their statistical reasoning and scientific communication skills. Prior work has demonstrated efficiency gains from AI-supported assessment while underscoring the critical role of trust, human oversight, and feedback quality in promoting learning (Alfredo et al., 2024; AlGhamdi et al., 2025; Venter et al., 2025; Zhang et al., 2025). This study introduces SmartFeedback, a rubric-constrained, human-in-the-loop workflow that reduces grading time by over 30% while preserving comparable perceived effectiveness. By addressing practical challenges in scaling high-quality feedback, our approach holds pedagogical significance for integrating AI in transparent, ethical, and student-centred ways.

2. Methodology

Participants and Setting

Seventy master's students in a statistics-related course submitted reports as Jupyter Notebooks.

Design and Procedure

Students were randomised into two groups: an AI-assisted group (38) and an instructor-only grading group (n = 32). In both arms, instructors finalised scores and approved all feedback. A post-grading questionnaire regarding feedback effectiveness yielded 65 complete responses (AI-assisted: n_1 = 36; instructor-only: n_2 = 29). To reduce expectancy effects, students were blinded at the time of the questionnaire survey. After the study, all participants were debriefed that AI may have been used under instructor oversight, consistent with guidance on trust and bias in AI-in-education (Alfredo et al., 2024; Zhang et al., 2025).

Intervention

SmartFeedback operates in two steps:

1. **Generation:** Generate a rubric-aligned analysis with linked evidence for instructor review and customisation (Figure 1). This alignment is reinforced by de-identified high/mid/low exemplars with brief annotations during the setup phase.
2. **Summarisation:** Transform verified feedback into a concise, student-facing summary (Figure 2). Final feedback can be downloaded and archived for future reference.

Measures

- **Efficiency:** Minutes per report. Manual grading was measured from the start of reading to completion of feedback typing; AI-assisted grading was measured from uploading to SmartFeedback to final saving.
- **Feedback Quality:** 5-point Likert ratings of perceived effectiveness.

Analysis

Welch's t-test was employed for one-sided comparison of grading time and two-sided comparison of perceived feedback quality.

3. Results and Discussion

The AI-assisted group demonstrated a significantly shorter average grading time (mean 13.66 minutes, SD 7.05) compared to the manual-grading group (mean 20.84 minutes, SD 7.49; $p < 0.0001$), representing a 34.45% reduction (Figure 1). Over time, average grading time declined in the manual-grading

group—likely reflecting growing familiarity and time pressure—whereas the AI-assisted group showed no such trend (Figure 2). Although the instructor-only group exhibited slightly higher mean ratings, there were no statistically significant differences (raw p-values > 0.05) between groups on any aspect of perceived feedback effectiveness (Table 1).

4. Implications

SmartFeedback streamlines grading while maintaining feedback quality, allowing instructors to redirect time towards mentoring and formative support. By translating rubric criteria into concise, actionable guidance, it helps students understand what to improve and how. The human-in-the-loop design is portable across report-based, data-centric courses. In a broader context, it provides a replicable, transparent workflow that supports equitable, consistent feedback at scale while upholding ethical standards for AI in education.

Ethical Declaration Statement

The project was approved by the REC of Hong Kong Baptist University (Reference Number: REC/25-26/0067).

Declaration of Use of AI in the Writing Process

The author(s) used GPT-5 during the preparation of this work to polish English writing. 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

Alfredo, R., Echeverria, V., Jin, Y., Yan, L., Swiecki, Z., Gašević, D., & Martinez-Maldonado, R. (2024). Human-centred learning analytics and AI in education: A systematic literature review. *Computers and Education: Artificial Intelligence*, 6, 100215. https://doi.org/10.1016/j.caeai.2024.100215

AlGhamdi, E., Li, Y., Gašević, D., & Chen, G. (2025). Leveraging prompt-based LLMs for automated scoring and feedback generation in higher education. *Computers & Education*. Advance online publication. https://doi.org/10.1016/j.compedu.2025.105511

Venter, J., Coetzee, S. A., & Schmulian, A. (2025). Exploring the use of artificial intelligence (AI) in the delivery of effective feedback. *Assessment & Evaluation in Higher Education*, 50(4), 516–536. https://doi.org/10.1080/02602938.2024.2415649

Zhang, A., Gao, Y., Suraworachet, W., Nazaretsky, T., & Cukurova, M. (2025, April 15). Evaluating trust in AI, human, and co-produced feedback among undergraduate students [Preprint]. *arXiv*. https://doi.org/10.48550/arXiv.2504.10961

Figure 1. A screenshot of the SmartFeedback prototype – Step 1.

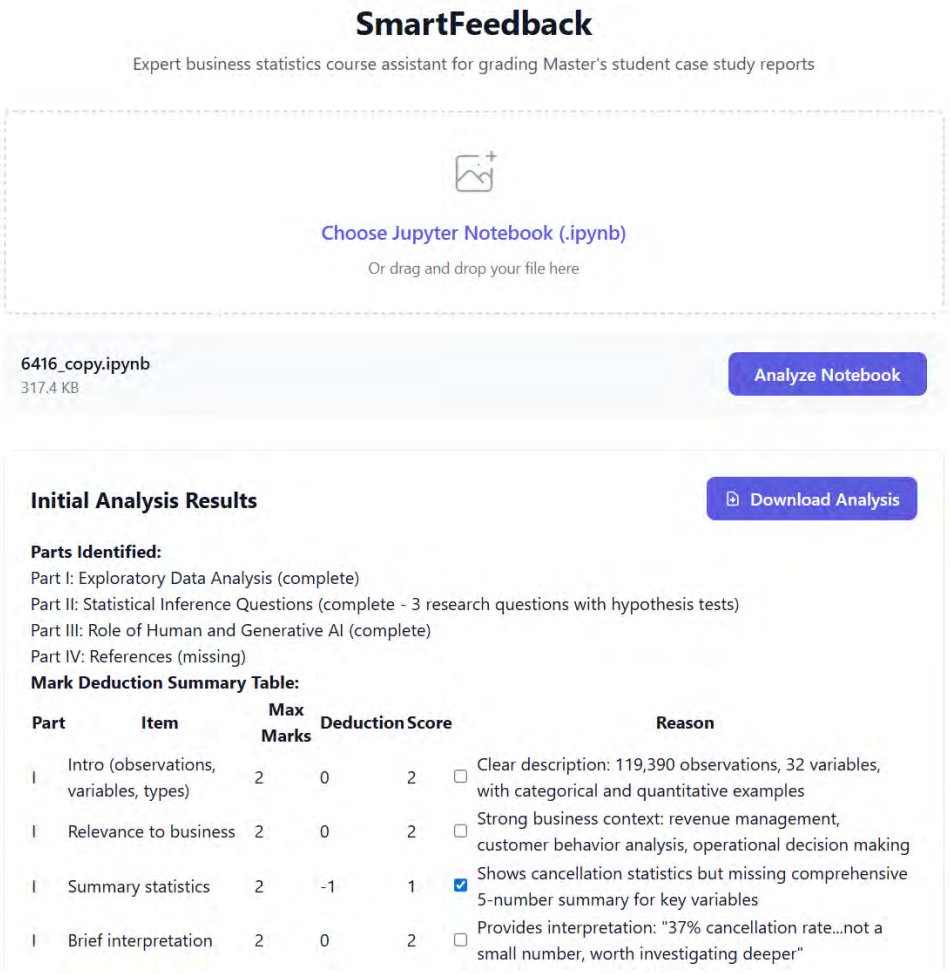


Figure 2. A screenshot of the SmartFeedback prototype – Step 2.

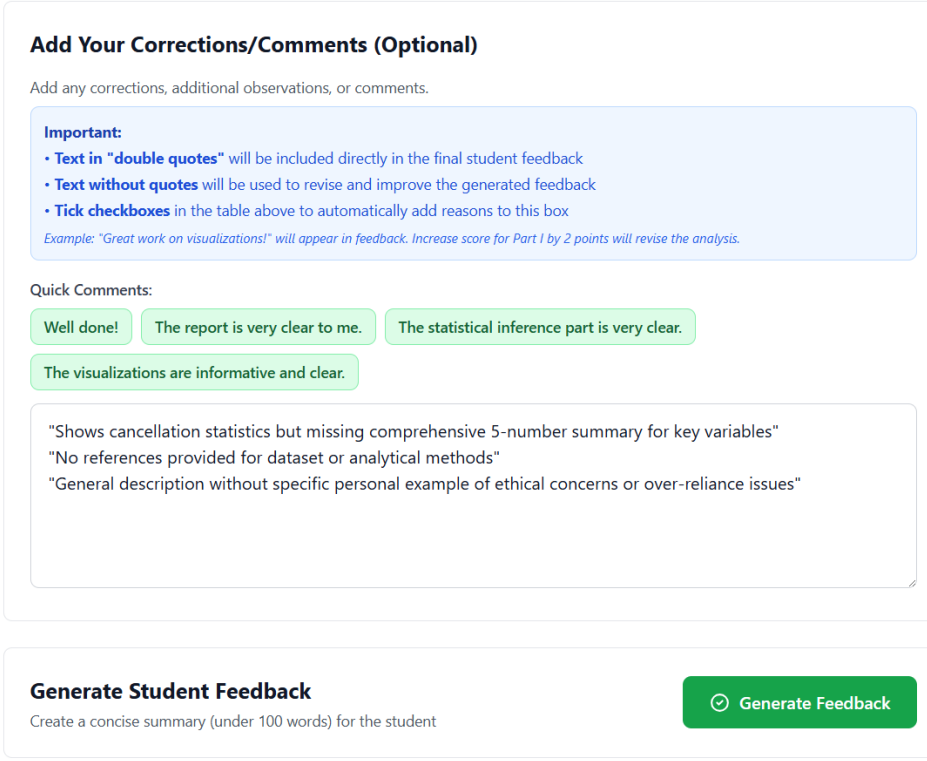


Figure 3. Average grading time by group (95% Confidence Interval).

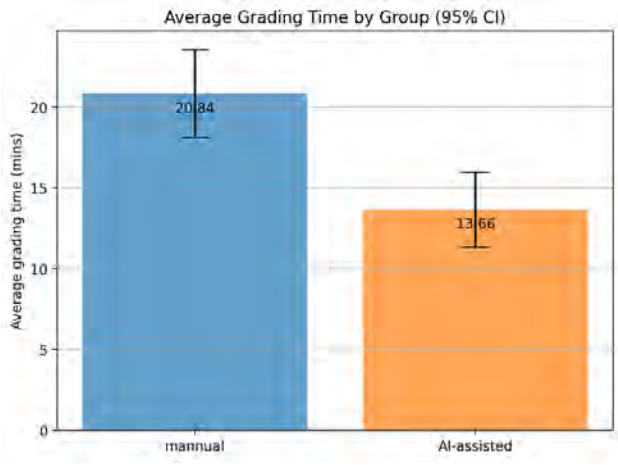


Figure 4. Average grading time by group over time.

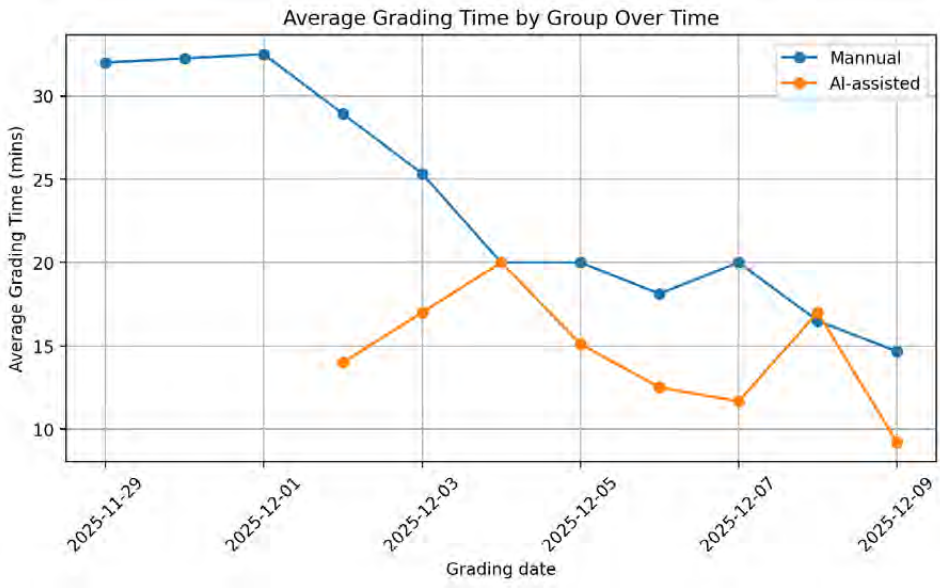


Table 1. Perceived feedback effectiveness by group.

Variable	AI-assisted (= 36)	Instructor-only (= 29)	*t*	*df*	*p-value**
	Mean (SD)	***Mean (SD)***			
Perceived fairness	4.44 (1.21)	4.83 (0.47)	1.75	47.26	0.09
Clarity					
Language/Expression	4.47 (1.18)	4.83 (0.38)	1.69	43.84	0.10
Reference/Specificity	4.28 (1.34)	4.76 (0.51)	1.98	46.80	0.05
Direction for improvement	4.39 (1.15)	4.69 (0.71)	1.29	59.36	0.20
Consistency					
Coverage	4.33 (1.17)	4.66 (0.77)	1.33	60.74	0.19
Tone	4.58 (1.13)	4.83 (0.38)	1.21	44.62	0.23
Scope/Detail	4.53 (1.16)	4.79 (0.62)	1.18	55.53	0.24
Actionability	4.49 (0.98)	4.72 (0.53)	1.24	53.89	0.22
Overall effectiveness	4.53 (0.38)	4.83 (0.38)	1.46	44.19	0.15

** The p-values are not adjusted for multiple testing.*

5 Teacher Professional Development

The Human in the Loop: Empowering Educators to Reclaim Agency in the Age of AI

Boey Chee KIN [1], Paul CHEUNG [1], Tan Lay KHEE [1]
[1] Learning Academy, Temasek Polytechnic, Singapore

Abstract

The rapid rise of generative AI has presented three urgent challenges for educators: accelerating tool development (upskilling), ensuring discipline-specific relevance in training workshops (engagement), and transferring workshop learning into practice (retention). Temasek Polytechnic designed a three-module professional development (PD) programme integrating AI ethics, hands-on exploration, and assessment redesign anchored in authentic practice. Key innovations include co-creating content and co-facilitating with school-based academic mentors to provide contextual training, and designing a post-workshop deliverable requiring educators to report their AI experimentation. Over 97% of educators have completed the programme, reporting increased confidence in harnessing AI to foster deeper learning, heightened ethical awareness, and evidence of assessment redesign to maintain academic integrity. This initiative demonstrates how a human-in-the-loop approach can strengthen educators' agency and ensure that AI adoption remains pedagogically grounded rather than tool-driven.

Keywords: Teacher professional development, generative AI, transfer of learning, AI literacy

1. Introduction

Generative AI has significantly transformed the education landscape since its early releases. As educators at Temasek Polytechnic explore these tools, our primary focus is on navigating their dual impact: leveraging their potential while safeguarding academic integrity and students' ability to develop critical human-centred skills. With the proliferation of GenAI tools, traditional take-home written assessments are increasingly vulnerable to plagiarism, making it challenging for educators to verify genuine learning and uphold academic standards.

Many educators reported uncertainty when redesigning assessments, particularly regarding maintaining assessment integrity and alignment with intended learning outcomes in contexts where student engagement with generative AI is increasingly unavoidable (Boud & Falchikov, 2007; Dawson et al., 2023; UNESCO, 2023). Many grappled with complex decisions during assessment

redesign, such as defining appropriate boundaries for GenAI use, identifying which assessment components require independent human reasoning, and determining how to assess students' ability to work responsibly with AI (Dawson et al., 2023; UNESCO, 2023).

A structured training programme for teaching staff was designed to equip all academic staff with the knowledge and skills to use generative AI effectively and ethically for teaching, learning, and assessment. This programme aims to enhance educators' competence in redesigning assessments that safeguard integrity and ensure validity, as well as harnessing AI tools to augment teaching and learning.

In practice, developing a GenAI course that meets the diverse needs and professional contexts of polytechnic educators is challenging for several reasons:

1. **Fast-Evolving Technology:** GenAI tools and technology evolve rapidly. The fast pace of AI development can compromise training outcomes if content becomes outdated. UNESCO's *Guidance for Generative AI in Education and Research* (2023) warns that the rapid pace of AI development poses risks to educational quality if curricula, materials, and educators' capabilities do not keep pace.
2. **Engaging Diverse Disciplines:** A generic workshop introducing GenAI and AI tools is unlikely to engage participants if examples and use cases applying GenAI in their specific domains are missing (UNESCO, 2023). Without this connection, the audience may feel disconnected, leading to poor training outcomes. This concern was highlighted during a train-the-trainer session debrief.
3. **Continual Application Post-Training:** For an initiative to be relevant and authentic, educators must be able to apply it to their professional practices to effect real changes in the classroom (Boud & Falchikov, 2007). However, translating workshop insights into practice requires time and confidence to experiment (Kirkpatrick & Kirkpatrick, 2006).

2. Methodology

To address the challenges mentioned above, the following approaches were adopted for designing the training programme:

1. **Concurrent Professional Development for Facilitators**
   To maintain the relevance of their skills and training materials/activities, facilitators attended GenAI courses and regularly updated their materials to stay current.

2. **Co-create and Co-facilitate with School Academic Mentors (SAMs)**
   To better engage educators from various disciplines, we co-designed the training materials and activities with academic mentors from the schools or SAMs, including the compilation of cases demonstrating the effective use of GenAI in teaching. Co-facilitation of school-based workshops ensured that discussions remained grounded and contextualised to their domains. Educators worked in small groups to analyse, critique, and redesign their own assessments, enabling the immediate application of theory to practice. Through peer discussion and joint problem-solving, they explored ways to enhance task authenticity, refine assessment criteria, and build AI-resilient assessment designs relevant to their teaching contexts.

3. **Post-Workshop Submission of Learning Application**
   Educators submitted evidence of how they applied workshop insights to their own subjects to demonstrate practice transfer. These submissions showcased redesigned assessments and examples of how they integrated GenAI into teaching and learning, including reflections on their applications.

The GenAI Training Programme

A structured training programme for teaching staff was designed to equip all academic staff with the knowledge and skills to use generative AI effectively and ethically for teaching, learning, and assessment. The learning journey comprises three modules:

Module 1: Introduction to AI in Education (Self-paced online)

Through a self-paced learning format, participants explore what AI is, why it matters in teaching and learning, and how it is reshaping the educational landscape. Educators will gain a baseline in AI literacy and earn a digital badge by achieving at least 80% on the quiz. The quiz assesses academic staff's understanding of the ethical considerations underlying their use of AI for teaching and learning, as well as how students can use AI responsibly in their studies.

Module 2: Using Generative AI to Augment Teaching and Learning (Face-to-face workshop)

This hands-on workshop empowered educators to meaningfully integrate generative AI tools, such as ChatGPT and Gemini, into their teaching practice. Over 3.5 hours, participants explored practical strategies to enhance content creation, feedback, lesson planning, and learner engagement.

Module 3: Assessment in the Age of AI (Face-to-face workshop)

This workshop was designed to help educators redesign assessments for the generative AI era, ensuring alignment with learning outcomes and maintaining assessment integrity. They engaged in small-group discussions to explore strategies for assessment redesign, harnessing the affordances of AI tools while mitigating plagiarism risks.

Modules 2 & 3 involved curating discipline-specific use cases or examples and customising training activities for six schools: Applied Science, Business, Design, Engineering, Humanities & Social Sciences, and Information & IT.

Post-workshop, an additional 2.5 hours were allocated to demonstrate learning, which required educators to apply their learning to actual classroom practices, fostering confident and ethical use of AI in pedagogy. They were given at least six months to apply GenAI to their teaching, such as preparing teaching materials, brainstorming lesson activities, or generating student engagement ideas. After this period, they reported their use cases in a digital form and reflected on the process and outcome.

Programme Evaluation

At the end of each workshop, educators completed an evaluation form regarding the workshop and its facilitators. The items surveyed are available under *Results and Discussion*.

3. Results and Discussion

Programme Completion

Over 97% of academic staff have completed the professional development programme. The workshops, totalling 58 sessions, were conducted over the past year to ensure that knowledge and skills remain current, especially as AI technology advances rapidly. It is also urgent for educators to understand AI's possibilities and risks.

Workshop Evaluation

Overall, 98.9% of educators found the programme effective. Survey results indicate strong evidence that the workshop materials and activities supported a contextual understanding of AI in teaching. High levels of agreement were recorded regarding:

- Discipline-contextualised materials aiding understanding (98.8%)
- Hands-on activities (98.2%)
- Well-sequenced activities (98.8%)

These figures suggest effective cognitive scaffolding. Educators frequently mentioned practical applications, authentic teaching scenarios, and immediate usability. These responses are consistent with learning science research showing that understanding deepens when learning is situated in authentic contexts and supported through experiential engagement (Lave & Wenger, 1991; Kolb, 1984). The combined findings suggest that participants were supported in understanding not only how AI works but also how it can be meaningfully applied within their instructional and disciplinary contexts.

Table 1: Consolidated Findings from Workshop Evaluations

	Dimension Measured	Agree / Strongly Agree	Open-ended Comments (Summarised)
1	Learning outcomes were clear and achieved	98.9%	Objectives were met
2	Hands-on activities were adequate	98.2%	Strong appreciation for hands-on assessment redesign and the use of AI technology
3	Topics and activities were well selected and sequenced *	98.8%	Curriculum flowed well
4	Materials were well-designed and organized *	98.8%	Use cases aided understanding
5	Course duration was appropriate	95.8%	Sentiments slightly more mixed due to varying levels of competencies in educators
6	Facilitator knowledge & practical understanding of AI use*	99.3%	Strong confidence in facilitators; Helped connect new technology to domain knowledge
7	Facilitator communication	99.0%	Clear explanations are valued
8	Facilitator facilitation skills *	99.2%	Good engagement. Exchange of ideas was valued
9	Overall program effectiveness	98.9%	The program was effective as a baseline GenAI training

** Questions 3 & 4 assess the effectiveness of the customised workshop materials and activities co-designed with the SAMs. Questions 6 & 8 assess educators' engagement levels in the workshops.*

Learning Application

The reported use cases by educators were categorised into **content creation** and **student engagement & feedback enhancement**.

- **For Content Creation:** "Question & Quiz Generation" (34.08%) and "Slide and Lesson Creation" (29.55%) accounted for how AI tools

such as ChatGPT and Gamma.app were leveraged to generate quizzes, MCQs, presentation decks, and other teaching materials, saving time and enhancing lesson delivery.

- **For Student Engagement:** Brainstorming lesson activities to better engage students (28.79%) was the third-most-reported example of application.
- **For Feedback:** Embedded AI chatbots (7.58%) helped analyse student responses to support learning, acting as teaching assistants to provide timely answers and feedback.

Table 2: Analysis of GenAI Learning Applications

Category	Applications	Use Cases	Frequency of application (%)
Content creation	Question & Quiz Generation	AI used to generate formative assessments, such as quizzes and MCQs	34.08
	Slides and Lesson Creation	AI used to generate presentation slides, lesson content, including authentic scenarios/case studies	29.55
Enhancing student engagement & supporting student feedback	Lesson Planning and Classroom Activities	AI used to brainstorm for lesson plans, create learning activities, and prepare for class engagement tasks	28.79
	Feedback & Chatbot Integration	AI was used for student feedback and data analysis. AI Chatbots were embedded as interactive tools into learning environments.	7.58

Note: Based on the workshop participants' descriptions in the learning application form, we grouped their use cases into the categories and applications listed above.

4. Implications

The GenAI PD Programme for educators at Temasek Polytechnic was driven by the need to build educators' knowledge of AI tools to an adequate level so they could apply and integrate them into their practice. The workshops were designed to achieve these aims through customised training modules that use a hands-on and dialogic approach. Lecturers from the same disciplines engaged in the training together, thereby encouraging peer discussion using use-

case examples that reflect the realities they face in classrooms. Finally, a learning application component was included to evaluate the transfer of learning to professional practice.

Educators explicitly planned how generative AI would be applied in teaching and assessment contexts and demonstrated this through artefacts and reflective accounts (Kirkpatrick & Kirkpatrick, 2006; Baldwin & Ford, 1988). The initiative has empowered educators at Temasek Polytechnic to integrate AI tools into their pedagogy with greater confidence and thoughtfulness. Through hands-on workshops, collaborative assessment redesign, and ongoing reflection, educators have gained not only technical proficiency but also a deeper understanding of the evolving educational landscape shaped by AI.

As a result, several key impacts have emerged from this initiative:

1. **Ethical Awareness:** Educators gain greater awareness of the ethical implications of AI and recognise their new role in guiding students to use AI responsibly in their studies.
2. **Fostering Critical Thinking:** Educators use AI as a tool to foster critical thinking and problem-solving, ensuring these skills remain central and avoiding over-reliance on AI.
3. **Shift in Perspective:** Educators shifted from merely safeguarding against plagiarism to developing confidence and competence in building students' capacity in using GenAI as a catalyst for deeper learning and critical thinking.

Acknowledgements

We extend our sincere appreciation to the management of Temasek Polytechnic for their steadfast support of educator professional development. Their continued commitment has been instrumental in enabling this initiative and advancing the meaningful integration of artificial intelligence in teaching and learning.

Declaration of Use of AI in the Writing Process

The authors did not use any generative AI or AI-assisted tools in the preparation of this manuscript.

References

Baldwin, T. T., & Ford, J. K. (1988). Transfer of training: A review and directions for future research. *Personnel Psychology*, 41(1), 63–105.

Boud, D., & Falchikov, N. (Eds.). (2007). *Rethinking assessment in higher education: Learning for the longer term*. Routledge. https://doi.org/10.4324/9780203964309

Dawson, P., Bearman, M., Dollinger, M., & Boud, D. (2024). Validity matters more than cheating. *Assessment & Evaluation in Higher Education*, 49(7), 1005–1016. https://doi.org/10.1080/02602938.2024.2386662

Kirkpatrick, D. L., & Kirkpatrick, J. D. (2006). *Evaluating training programs: The four levels* (3rd ed.). Berrett-Koehler Publishers.

Kolb, D. A. (2014). *Experiential learning: Experience as the source of learning and development* (2nd ed.). FT Press.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press.

Miao, F., & Holmes, W. (2023, September 7). *Guidance for generative AI in education and research*. UNESCO. https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research

Evaluating University Teachers' AI Literacy and STEAM Competency: Evidence to Inform Targeted Professional Development

Crystal LUO [1], Gary CHENG [1], Stephanie CHAN [1]
[1] Centre for Learning, Teaching and Technology (LTTC), The Education University of Hong Kong, Hong Kong, China

Abstract

This study developed and piloted two theory-informed questionnaires to evaluate university teachers' AI literacy and STEAM competency in Hong Kong, with the aim of identifying perceived strengths and gaps that may inform professional development. Forty-one teachers completed the AI literacy questionnaire and 38 completed the STEAM competency questionnaire. Grounded in affective, behavioural, and cognitive–metacognitive domains, findings indicate that teachers demonstrate strong intrinsic motivation, positive attitudes, and clear behavioural intention toward AI and STEAM, but report comparatively lower self-efficacy, conceptual understanding, and confidence in hands-on application—particularly in applying, creating, and evaluating AI solutions and in STEAM teaching competency. Descriptive differences between technology and non-technology faculty suggest uneven readiness across disciplines. Taken together, the results provide preliminary evidence that may support the design of tiered, practice-oriented professional development initiatives tailored to diverse disciplinary needs in higher education.

Keywords: AI literacy, STEAM competency, higher education, professional development, evaluation frameworks

1. Introduction

Artificial intelligence (AI), particularly generative AI, is reshaping higher education, making teachers' AI literacy an increasingly important foundation for effective and responsible teaching (Tan et al., 2025). Concurrently, STEAM (Science, Technology, Engineering, Arts, and Mathematics) competencies—spanning critical thinking, creativity, and collaboration—are widely recognised as essential for contemporary educators (Bertrand & Namukasa, 2020), and are closely interconnected with AI literacy. STEAM interdisciplinary thinking supports AI literacy development, while computational thinking from AI strengthens STEAM competency (How & Hung, 2019).

Despite growing recognition, empirical research assessing both capacities among practising university teachers in non-Western contexts appears to remain limited. In addition, existing instruments have been validated primarily with students or pre-service teachers (Hornberger et al., 2023; Long & Magerko, 2020). This study addresses that gap, guided by three research questions:

* **(RQ1)** What are university teachers' current self-reported AI literacy levels across affective, behavioural, and cognitive–metacognitive domains?
* **(RQ2)** What are their current self-reported STEAM competency levels across the same domains?
* **(RQ3)** How do these descriptive patterns differ between technology and non-technology faculty, and what implications might these patterns have for professional development?

Examining these questions is important because professional development provision in higher education often lacks the empirical foundation needed to target resources effectively across disciplinary contexts. By profiling teachers' current strengths and gaps, this study offers preliminary evidence to inform differentiated and scalable professional development in Hong Kong higher education.

2. Methodology

2.1 Development of Instruments

Drawing on Long & Magerko (2020), Hornberger et al. (2023), Ng et al. (2024), and the General Attitudes towards Artificial Intelligence Scale (GAAIS; Schepman & Rodway, 2020), the study developed an AI literacy questionnaire assessing six subcategories: (1) intrinsic motivation for learning AI, (2) self-efficacy in understanding and working with AI, (3) attitudes toward AI, (4) behavioural intention to learn and apply AI, (5) general AI knowledge, and (6) skills in applying, creating, and evaluating AI solutions. A STEAM competency questionnaire adapted from Sandoval-Palomares (2022) assessed five categories: (1) intrinsic motivation, (2) self-efficacy, (3) attitudes, (4) behavioural intention, and (5) teaching competency in STEAM. Both used a 5-point Likert scale and were contextualised for university teachers' professional roles.

Although several source instruments were originally designed for students or pre-service teachers, the items were adapted to reflect the professional roles and instructional responsibilities of in-service university faculty. Figure 1 illustrates the questionnaire components. To strengthen content validity, five experts reviewed the instruments, and their feedback informed item revisions. The adapted questionnaires showed strong internal consistency in this sample, with

Cronbach's α = .91 for AI literacy and α = .94 for STEAM competency. The refined questionnaires were then administered between May and July 2024.

2.2 Participants

Forty-one university teachers completed the AI literacy questionnaire (non-technology: n = 27; technology: n = 14) and 38 completed the STEAM questionnaire (n = 19 per group). All provided informed consent online.

3. Results and Discussion

3.1 AI Literacy

Table 1 presents the AI literacy results across the six subcategories. At the descriptive level, teachers from technology departments reported higher mean scores across all subcategories, with the largest differences appearing in the cognitive–metacognitive domain.

Table 1: Results of the AI Literacy Questionnaire

Category	Subcategory	Non-technology Departments (n=27)		Technology Departments (n=14)	
		M	*SD*	*M*	*SD*
Affective domain	Intrinsic motivation for learning AI	4.51	0.72	4.61	0.55
	Self-efficacy in understanding and working with AI	3.80	0.88	4.17	0.63
	Attitude towards using AI	3.83	1.07	4.18	0.74
Behavioural domain	Behavioural intention towards learning and using AI	3.98	0.92	4.37	0.54
Cognitive and metacognitive domains	General knowledge and understanding of AI	3.82	0.91	4.20	0.65
	Application, creation and evaluation of AI	3.54	1.03	4.01	0.75

Intrinsic motivation was high across both groups (non-technology: M = 4.51; technology: M = 4.61), alongside positive attitudes toward using AI in teaching (M = 3.83 vs. 4.18) and strong behavioural intention to learn and apply AI (M = 3.98 vs. 4.37). These descriptive results suggest that many university teachers perceive AI as relevant and valuable and are willing to invest effort in developing related competencies.

However, self-efficacy in understanding and working with AI was comparatively lower, particularly among non-technology faculty (M = 3.80 vs. 4.17), as were perceived general AI knowledge levels (M = 3.82 vs. 4.20). This pattern

suggests a possible gap between motivation and perceived readiness, consistent with the distinction between affective and cognitive dimensions identified by Ng et al. (2024). In other words, positive attitudes toward AI may not necessarily translate into strong conceptual understanding or confidence in practical application.

The lowest scores appeared in AI application, creation, and evaluation (M = 3.54 vs. 4.01), particularly among non-technology faculty. This pattern may indicate comparatively less hands-on experience with AI tools in teaching contexts and suggests a gap between theoretical familiarity and practical capability. These findings are broadly consistent with prior studies suggesting that practical AI literacy may develop more slowly than motivational dimensions (Hornberger et al., 2023; Tan et al., 2025). This pattern supports the value of structured, practice-oriented professional development that helps teachers move from awareness to application.

3.2 STEAM Competency

Table 2 presents STEAM competency results across the five subcategories. As with AI literacy, teachers from technology departments reported higher mean scores across all categories, with especially notable descriptive differences in the behavioural and cognitive–metacognitive domains.

Table 2: Results of the STEAM Competency Questionnaire

Category	Subcategory	Non-technology Departments (n=19)		Technology Departments (n=19)	
		M	*SD*	*M*	*SD*
Affective domain	Intrinsic motivation for learning STEAM knowledge and skills	4.26	0.87	4.45	0.63
	Self-efficacy in understanding and working with STEAM knowledge and skills	3.68	1.19	4.16	0.49
	Attitude towards STEAM knowledge and skills	3.68	1.12	4.18	0.64
Behavioural domain	Behavioural intention towards learning and using STEAM knowledge and skills	3.52	1.22	4.31	0.59
Cognitive and metacognitive domains	Teaching competency for STEAM	3.50	1.33	4.19	0.66

Teachers from technology departments reported high intrinsic motivation (M = 4.45) and behavioural intention (M = 4.31), while reporting somewhat lower levels of self-efficacy (M = 4.16), attitudes (M = 4.18), and teaching competency (M = 4.19). This pattern suggests that strong motivation may not automatically translate into full confidence in implementation, even among STEAM-aligned faculty. It also highlights the potential value of targeted pedagogical support, such as communities of practice (CoPs), to promote peer learning, collaborative lesson design, and experience sharing across faculties.

Non-technology teachers showed relatively high motivation (M = 4.26) but notably lower self-efficacy (M = 3.68), attitudes (M = 3.68), behavioural intention (M = 3.52), and teaching competency (M = 3.50), reflecting perceived difficulty in applying STEAM in practice. These lower ratings across the behavioural and cognitive–metacognitive domains may indicate that some non-technology faculty perceive STEAM as conceptually distant from their disciplines, which could reduce willingness to engage in related professional learning activities.

Across both datasets, a broadly similar pattern emerges, with motivation appearing stronger than perceived implementation readiness, especially among non-technology faculty. This suggests that an integrated professional development approach addressing both AI literacy and STEAM competency may be beneficial.

4. Implications, Limitations, and Future Research Directions

The convergent pattern across both domains suggests that a staged professional development model may be beneficial. In Phase 1 (foundation), discipline-relevant workshops could introduce core AI and STEAM concepts through worked examples and case studies drawn from participants' own teaching contexts. This may be particularly important for non-technology faculty who may perceive these domains as distant from their existing practice. In Phase 2 (application), teachers could engage in co-design of AI-enhanced lesson materials, structured experimentation with generative AI tools, and peer teaching simulations to translate conceptual knowledge into practical use. In Phase 3 (sustained support), communities of practice could provide opportunities for collaborative reflection, discipline-specific knowledge exchange, and ongoing peer support.

Taken together, the findings suggest that motivational readiness alone may be insufficient to support confident AI and STEAM integration in teaching practice. Instead, professional development may need to be structured, contex-

tualised, and sustained in order to help teachers translate positive attitudes into confident and discipline-relevant practice.

Several limitations should be noted. First, the study involved a small, single-institution sample (AI literacy: N = 41; STEAM: N = 38) drawn from one university in Hong Kong, and the unequal group sizes in the AI literacy dataset (technology: n = 14; non-technology: n = 27) may limit the generalisability of the findings to other institutional and cultural contexts. Second, the study relied on self-report measures, which capture perceived rather than directly observed competence and may therefore be influenced by social desirability or self-assessment bias. Third, because the present analysis focused primarily on descriptive statistics, the observed group differences should be interpreted cautiously.

Future studies should employ larger, multi-institutional, and cross-cultural samples, as well as longitudinal designs tracking competency trajectories before and after professional development interventions, in order to provide stronger evidence regarding programme effectiveness. Mixed-methods approaches incorporating classroom observations, reflective journals, interviews, or student outcome data could help triangulate self-report findings and provide a more comprehensive understanding of AI literacy and STEAM competency in actual teaching practice. Where sample size permits, future studies could further strengthen instrument validation through exploratory or confirmatory factor analysis and examine group differences using inferential statistics, confidence intervals, and effect sizes. Structural equation modelling could also be used to examine how AI literacy and STEAM competency interact, whether growth in one domain predicts or facilitates development in the other, and what individual and institutional factors moderate these relationships. In addition, comparing the effectiveness of different professional development formats—such as structured workshops, mentoring, or CoPs—would further strengthen the evidence base for targeted capacity building in higher education.

Acknowledgments

This study was financially supported by the Fund for Innovative Technology in Education, administered by the University Grants Committee of Hong Kong (Project Numbers: 05400 and 05409).

Ethical Declaration Statement

The project was approved by the Human Research Ethics Committee of the Education University of Hong Kong (Reference Number: 2023-2024-0383).

Generative AI Policy Statement

The authors used ChatGPT during the preparation of this manuscript for language and grammar checking. Following use of the tool, the authors reviewed and revised the manuscript as necessary and take full responsibility for the content of the publication.

References

Bertrand, M. G., & Namukasa, I. K. (2020). STEAM education: student learning and transferable skills. *Journal of Research in Innovative Teaching & Learning, 13*(1), 43–56. https://doi.org/10.1108/JRIT-01-2020-0003

Hornberger, M., Bewersdorff, A., & Nerdel, C. (2023). What do university students know about Artificial Intelligence? Development and validation of an AI literacy test. *Computers and Education: Artificial Intelligence, 5*, 100165. https://doi.org/10.1016/j.caeai.2023.100165

How, M., & Hung, W. L. D. (2019). Educing AI-thinking in science, technology, engineering, arts, and mathematics (STEAM) education. *Education Sciences, 9*(3), 184. https://doi.org/10.3390/educsci9030184

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–16). https://doi.org/10.1145/3313831.3376727

Ng, D. T. K., Wu, W., Leung, J. K. L., Chiu, T. K. F., & Chu, S. K. W. (2024). Design and validation of the AI literacy questionnaire: The affective, behavioural, cognitive and ethical approach. *British Journal of Educational Technology, 55*(3), 1082–1104. https://doi.org/10.1111/bjet.13411

Sandoval-Palomares, J. (2022). STEAM conceptions, competencies and attitudes in higher education: A pilot study. *Journal-Health Education and Welfare, 6*(11), 6–20. https://doi.org/10.35429/JHEW.2022.11.6.6.20

Schepman, A., & Rodway, P. (2020). Initial validation of the general attitudes towards Artificial Intelligence Scale. *Computers in Human Behavior Reports, 1*, 100014. https://doi.org/10.1016/j.chbr.2020.100014

Tan, X., Cheng, G., & Ling, M. H. (2025). Enhancing teachers' AI competency: A professional development intervention study based on intelligent-TPACK framework. *Computers and Education: Artificial Intelligence, 9*, 100521. https://doi.org/10.1016/j.caeai.2025.100521

New Academics' Use of Generative AI in Teaching: Current Adoption and Challenges

Barbara TAM [1], Dick CHAN [1], and Mitesh PATEL [1]
[1] Educational Development Centre, The Hong Kong Polytechnic University, Hong Kong, China

Abstract

This study explores the use of Generative Artificial Intelligence (GenAI) in teaching among new academics at a university in Hong Kong, focusing on their current adaptation and the challenges encountered. Data were collected via a questionnaire survey (n = 78). Findings reveal that nearly one-third of new academics had not used GenAI in their teaching. Among those who did, the majority employed it for content construction, with many expressing doubts about its efficacy. The major challenges identified were concerns over inaccurate information, a lack of skills and knowledge regarding GenAI, associated costs, and its negative impact on student learning. These results suggest that universities need to enhance their tailored professional development offerings and allocate funds to help new academics make the most of GenAI. At the same time, they should implement policies and guidelines to ensure that GenAI supports, rather than undermines, the core goals of education.

Keywords: New Academics, Higher Education, Generative AI, Faculty Development, Technology Adoption

1. Introduction

Generative Artificial Intelligence (GenAI) is significantly reshaping teaching and learning. The use of GenAI tools is becoming increasingly common in higher education (Chan & Hu, 2023), offering innovative opportunities for enhanced educational experiences. New academics, representing a diverse array of backgrounds and cultures, play a crucial role in shaping the future of higher education (Austin et al., 2007). While studies have investigated educators' perceptions and use of GenAI, there remains a gap in understanding new academics' experiences and evolving literacies with GenAI. This research aims to bridge that gap by conducting a study at The Hong Kong Polytechnic University (PolyU), addressing two key questions:

1. What is the current level of adoption of GenAI among new academics in their teaching practices?
2. What challenges do new academics encounter regarding the use of GenAI in teaching?

By gathering insights from new academics, higher education institutions can implement targeted policies and develop guidelines that address the unique training needs of new academics and support their effective use of GenAI to improve student learning outcomes.

2. Methodology

The study emerged from a mandatory training programme for new academics at PolyU designed to equip them with the skills necessary to support student learning effectively. To address the research questions, the teaching team developed a survey to investigate the use of GenAI in teaching practices among new academics. The survey consisted of multiple-choice and short-answer questions to gather insights on both adoption levels and encountered challenges. The questions were informed by relevant literature and previous studies on GenAI in higher education and were structured to capture the nuances of new academics' experiences.

Guided by the Technology Acceptance Model (TAM) (Davis, 1989), which is well-suited for exploring technology adoption, the finalised questionnaire was piloted in January 2025. All participants in the first two offerings of the programme in the 2025–26 academic year were asked to complete the survey. The survey was open for one week, starting in mid-October 2025. Throughout the week, reminders were sent to encourage completion. Eighty-eight responses were collected, of which 78 respondents provided consent (n=78). Upon collecting the data, thematic analysis was employed to identify key concepts and themes.

3. Results and Discussion

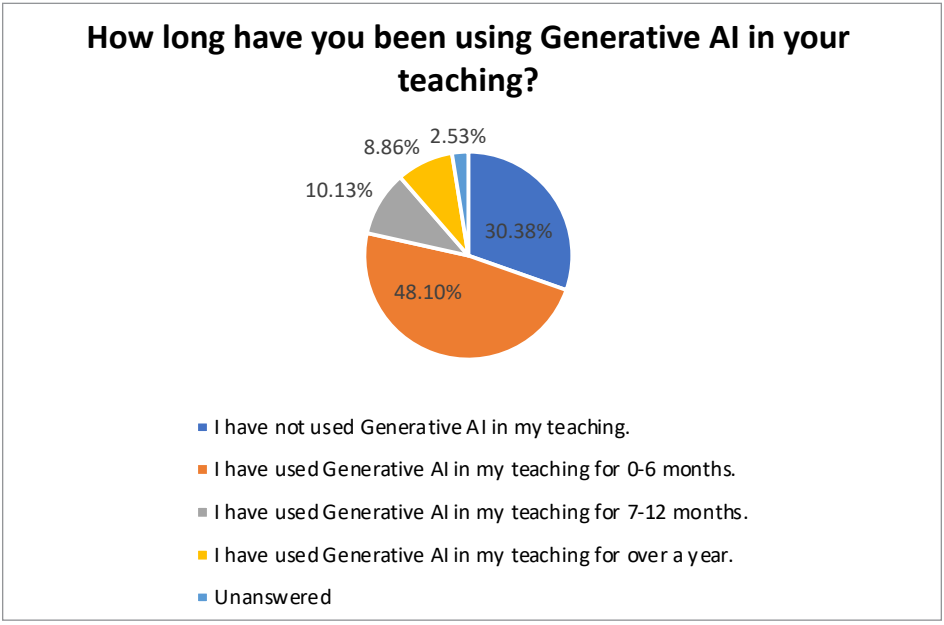
A total of 78 participants consented to use their data from the survey (n=78). In terms of teaching experience, 11 reported that they had never taught, while the others had varying levels of experience, ranging from a few guest lectures to having taught some undergraduate and postgraduate courses. The majority had served as teaching assistants during their postgraduate studies, and a few had taught for several years. One respondent remarked that his teaching "spanned multiple semesters from 2015 to 2023," making him the most experienced among the respondents.

3.1. Current Adoption of GenAI Among New Academics in Teaching

Length of GenAI Adoption in Teaching by New Academics

Results reveal that a significant portion of new academics (30.38%) have not used GenAI in their teaching, implying that a large number of them may

be uncertain about integrating GenAI into their practices. Nearly half of the respondents (48.1%) are relatively new users, having utilised GenAI for less than six months. This suggests that many new academics are still in the early stages of adoption of GenAI in teaching. A small portion of the respondents (8.86%) have used GenAI for over a year. While this percentage is modest, it is still a positive indicator of adoption given that they are new academics. These results highlight the need for targeted support, resources, and tailored professional development to help new academics become more comfortable and proficient in utilising GenAI in their teaching.

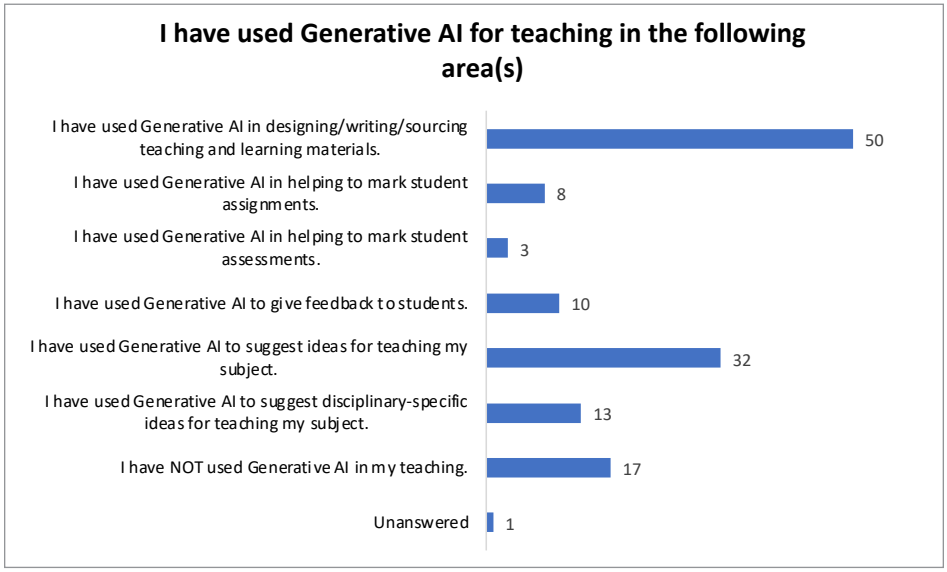


(Figure 1: Length of GenAI Adoption in Teaching by New Academics)

Application of GenAI in Teaching by New Academics

The findings highlight various applications of GenAI among new academics, listed in order of frequency. The most prevalent use, mentioned 50 times, was for "designing, writing, and sourcing teaching materials," indicating a reliance on GenAI to enhance instructional content. The second most common application, with 32 mentions, was "suggesting ideas for teaching my subject," implying that new academics see potential usefulness in leveraging GenAI for creative inspiration. Following this, "suggesting discipline-specific ideas for teaching" was mentioned 13 times, demonstrating a targeted application to meet subject-specific needs. Additional uses included "giving feedback to students" (10 mentions), "helping to mark student assignments" (8 mentions), and "helping

to mark student assessments" (3 mentions). Notably, 17 respondents indicated that they have not used GenAI in their teaching, aligning with previous findings that some new academics have yet to adopt GenAI in their teaching practices.



(Figure 2: Use of GenAI in Teaching by New Academics)

Reasons for Using GenAI in Teaching

Upon thematic analysis, the reasons for new academics' utilisation of GenAI are categorised into three themes: enhancing instructional efficiency, leveraging specific functionalities, and promoting personalised learning experiences.

Many respondents emphasised the time-saving and user-friendly features of GenAI, especially in creating teaching materials. Frequently mentioned terms like "efficient," "quick," and "fast" underscore a strong reliance on GenAI to streamline preparation. Respondents reported that it significantly reduced their preparation time by generating "draft outlines," "examples," and "explanations for lectures and tutorials." Comments like "faster search for various materials compared with traditional search engines" indicate the convenience of GenAI. Many respondents described AI tools as "very easy to use," with specific references to "designing PowerPoint slides" and "drafting engineering teaching materials." These results indicate that new academics view GenAI as essential for saving preparation time. This aligns with the Technology Acceptance Model (TAM), which emphasises that perceived ease of use and usefulness are critical factors in technology adoption.

Respondents also valued the specialised functionalities of GenAI that enhance teaching effectiveness, such as its function to "summarise content and generate varied practice questions." Many benefited from the language improvement feature, with one stating, "its English expression is better than mine." This assists non-native speakers by refining content and communication. Several mentioned interactive features, such as "engineering simulations" and "dynamic quizzes," highlighting how educators use GenAI to enhance learning experiences. These findings show that GenAI's specialised functionalities not only streamline preparation but also foster an engaging learning environment.

Lastly, the findings suggest that new academics recognise GenAI's role in facilitating personalised learning. Many reported using it to create "customised learning paths" and practice problems tailored to individual students, effectively addressing gaps in understanding while promoting advanced challenges. One emphasised generating tailored resources to "support students struggling with core engineering concepts," while another noted the importance of "making content accessible across disciplines." Overall, these findings reveal that GenAI empowers educators to tackle students' unique challenges.

3.2. New Academics' Challenges in Using GenAI for Teaching

This section explores the challenges new academics encounter when using GenAI in teaching. A thematic analysis identified four key themes: lack of relevant knowledge and skills; concerns about the accuracy and trustworthiness of GenAI; cost and accessibility issues; and the impact on student learning.

Lack of Knowledge and Skills in Using GenAI

Many respondents expressed uncertainty about effectively using AI tools, particularly with crafting suitable prompts. Comments like "the biggest challenges might be how to use effective prompts to make the most of it" illustrate that limited familiarity with GenAI may hinder adoption. This suggests the need for tailored training and opportunities to learn from experienced peers, which could facilitate more effective integration of these tools into their teaching practices.

Concerns Regarding Accuracy and Trustworthiness of GenAI

Skepticism about the reliability of AI-generated content is a prominent concern among respondents, particularly in technical fields such as Engineering, where precision is essential. One quote encapsulates the concerns of many: "the biggest obstacle that stops me from using Generative AI for teaching is the accuracy and reliability of the generated content." This reflects that new academics struggle to ensure the quality of content produced by GenAI. This aligns

with the Technology Acceptance Model (TAM), which posits that perceived reliability is vital for the adoption of new technologies (Davies, 1989).

Cost and Accessibility of GenAI

Many respondents noted that GenAI applications often require costly subscriptions, limiting their accessibility. One quote succinctly captures this concern: "the biggest challenge I have about using GenAI for teaching is that many tools require a paid subscription." Although some tools are provided by the university, they often come with limitations, such as token caps that restrict access. These barriers can hinder the effective integration of AI tools and lead to a digital divide in teaching quality.

Impact on Student Learning

The findings highlight concerns regarding the negative impact of GenAI on student learning in several ways. First, new academics struggle to assess genuine student understanding, as GenAI "can produce high-quality essays, code, and answers," making it difficult to differentiate between original student work and AI-generated content. This ambiguity can hinder teachers' ability to provide meaningful feedback, suggesting that a new approach to assessment may be necessary. Second, despite GenAI's capabilities, it does not foster critical thinking. For example, one respondent stated, "it cannot replicate the tactile, problem-solving experience critical to engineering: it can explain machine assembly steps but not guide students through physical adjustments… leaving a key gap in building practical engineering skills that traditional labs prioritise." Lastly, the convenience of GenAI tempts students, creating challenges for educators in preventing overreliance. As one respondent noted, the biggest challenge is "setting clear boundaries so AI supports learning without undermining academic integrity or deep thinking." This perspective aligns with the observations of O'Sullivan et al. (2025): "While AI can scaffold or support some learning tasks, its widespread and uncritical use may erode the very capabilities higher education is intended to cultivate" (p. 19).

4. Implications

This paper investigates new academics' adoption of GenAI for teaching and the challenges they encounter. The insights highlight the need for universities to establish clear policies on the appropriate use of GenAI. New academics need to reflect on the broader implications of GenAI for student learning. Besides technical proficiency, GenAI training for new academics should include reflective components to align their work with the core goals of education, and peer learning to share experiences and strategies. Alternative approaches, such as redesigning assessment, should be explored to help teachers move beyond

merely detecting the use of GenAI. Considering the institutional provision of premium GenAI tools can prevent a 'digital divide' in teaching quality, ensuring equitable access to resources for the effective integration of GenAI. PolyU is already making progress in this direction by implementing tailored professional development programmes and allocating funds to support new academics. Established frameworks and guidelines are in place to help educators use GenAI ethically and effectively. The rapid development of technology and the evolving needs of education, however, indicate that there is always room for improvement. Such a holistic approach will better equip new academics to navigate the complexities of technological integration, ensuring their efforts support the fundamental goals of university education.

Ethical Declaration Statement

The project has been approved by the Institutional Review Board of The Hong Kong Polytechnic University (Reference Number: HSEARS20240204001).

Generative AI Policy

The author(s) used ChatGPT during the preparation of this work to assist with language refinement and correct grammatical errors. 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

Austin, A.E., Sorcinelli, M.D., & McDaniels, M. (2007). Understanding new faculty: Background, aspirations, challenges, and growth. In Perry, R.P., & Smart, J.C. (Eds.) (2007). *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective* (pp. 39-89). Springer. https://link.springer.com/chapter/10.1007/1-4020-5742-3_4

Chan, C.K.Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education, 20*, 43. https://link.springer.com/article/10.1186/s41239-023-00411-8

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319-340. https://doi.org/10.2307/249008

O'Sullivan, J., Lowry, C., Woods, R., Marrinan, B., & Hutchinson, C. (2025). Generative AI in higher education teaching and learning: Sectoral perspectives. Higher Education Authority. https://hea.ie/assets/uploads/2025/09/Gen-AI-in-Higher-Education-Teaching-and-Learning-Sectoral-Perspectives.pdf

GenAI in Higher Education: Disciplinary Epistemologies, Pedagogical Implications, and Recommendations for Professional Development

Verily TAN [1], Yuen Jien SOO [2], Mei Hui LIU [1]
[1] Centre for Teaching, Learning and Technology, National University of Singapore, Singapore
[2] Centre for Development of Teaching & Learning, National University of Singapore, Singapore

Abstract
Generative AI (GenAI) is reshaping pedagogical practices across higher education. It has been observed that faculty respond differently across disciplines. This study seeks empirical evidence through exchanges with both STEM and non-STEM faculty during and after a two-day professional development programme on the integration of AI in the design of teaching and learning activities. Through thematic analysis of interviews, the study illustrates distinct disciplinary orientations: STEM faculty emphasised structure, efficiency, and procedural alignment with GenAI, while non-STEM faculty foregrounded authorship, critical judgement, and epistemic depth. Despite these differences, faculty across disciplines positioned critical evaluation of AI output as a foundational element of responsible AI literacy. The study examines how this is reflected in the assessment or learning activities designed. As a synthesis of the findings, a limited framework for disciplinary professional development for GenAI integration is shared, along with recommendations for professional development.

Keywords: GenAI, STEM, non-STEM, faculty development, activity design

1. Introduction

Generative AI is altering how knowledge is produced, represented, and evaluated in higher education. While enthusiasm for GenAI continues to rise, the patterns of adoption are uneven, shaped by disciplinary epistemologies, teaching traditions, and assessment norms. Research has shown that STEM faculty are more likely to adopt GenAI for tasks such as debugging, solution generation, and procedural support (Garcia et al., 2025). By contrast, non-STEM faculty frequently express concern about voice, authorship, originality, and interpretive depth (Costello et al., 2025). These disciplinary orientations powerfully influence how faculty integrate or resist GenAI in their teaching.

Against this backdrop, universities face increasing pressure to support faculty through discipline-sensitive yet cross-disciplinary professional development (PD). Scholars have highlighted the need for PD that is attuned to disciplinary differences while facilitating collaborative sensemaking across fields. Interdisciplinary learning communities have been shown to broaden pedagogical imagination and challenge entrenched assumptions (University of North Carolina System, 2025). In the literature, there is a lack of rich narratives regarding STEM versus non-STEM orientations, documentation of GenAI use cases, and a guiding framework for professional development.

This study aims to address some of these gaps by studying a two-day professional development programme—an AI-infused Course Enhancement Module. Offered by the university's Centre for Teaching, Learning and Technology, this module was based on voluntary registration. The aim was to offer help to faculty in creating GenAI-integrated teaching and learning activities or assessments which they could implement in their courses.

2. Methodology

Research Questions

1. What are STEM and non-STEM faculty's perspectives on the use of GenAI in teaching and learning?
2. How do STEM and non-STEM faculty use GenAI in the design of assessment and teaching and learning activities?

Participants and Programme Structure

Fifteen educators participated in the PD: six from STEM disciplines (engineering, computing, natural sciences) and nine from non-STEM fields (humanities, social sciences, communications, education). The PD was intentionally designed as interdisciplinary, enabling faculty to discuss their GenAI practices and concerns across epistemic boundaries. Activities included hands-on prompt engineering, the design of assessment and teaching and learning activities in light of AI affordances, and cross-disciplinary peer consultation.

Importantly, the **Human–AI Matrix** was used as a guiding framework (Lloyd & Affias, 2025; Tan et al., 2025). This is a conceptual scaffold that helps instructors distinguish *who* is doing the cognitive work—students or AI. It is represented as four quadrants with the x-axis being AI contribution and the y-axis being human contribution. When these dimensions intersect, four quadrants emerge:

- **Low AI–Low Human activities** are not desirable.
- **High AI–Low Human activities** translate into "Create-for-me" learn-

ing scenarios.

- **Low AI–High Human activities** translate to "Polish, plan and brainstorm with me" learning scenarios.
- **High AI–High Human activities** translate to "Create-with-me" learning scenarios.

Faculty found this tool clarifying because it shifted their focus from AI features to questions of cognitive responsibility, learning value, and ethical boundaries.

Data Collection and Analysis

All participants completed a semi-structured interview of approximately 30 minutes. Interviews explored perceptions of GenAI in teaching and learning, disciplinary considerations, the experience of mixed-disciplinary discussion, and any shifts in perspective following the PD. Twelve participants also submitted designed lesson plans. Interview transcripts were coded inductively following Braun and Clarke's (2014) thematic analysis approach. Codes were clustered into three overarching themes aligned to RQ1. For RQ2, lesson plans were separately analysed by the Principal Investigator and Co-investigator and classified according to categories of GenAI in Higher Education synthesised from literature.

3. Results and Conclusion

Research Question 1: To understand STEM and non-STEM perspectives toward the use of GenAI in teaching and learning

Three themes emerged. The first concerns foundational epistemic commitments; the second, boundaries of appropriate AI use; and the third, the role of interdisciplinary exposure in shaping faculty thinking.

Theme 1: Contrasting Disciplinary Priorities.

Table 1 presents participant quotations supporting this theme.

P1 (Non-STEM)	"I don't think it is stoppable. The world has changed, but we must still keep critical thought and the student's voice central."
P2 (STEM)	"We teach procedures and precision. If the AI gives the right answer for the wrong reasoning, the student learns nothing."
P13 (STEM)	"In my field, the process of argumentation is the learning outcome. The facts matter, but the journey to interpretation is key."

Non-STEM participants such as P1 stressed that the rise of GenAI does not diminish the centrality of critical reasoning and the student's unique voice, particularly in writing. In contrast, the STEM perspective represented by P2 was framed around concerns that students may bypass core cognitive processes tied to procedural mastery and correctness. If GenAI supplies "the right answer

for the wrong reasoning," the foundational aim of STEM education—internalising accepted methods and forms of verification—is compromised. Similarly, P13 noted that in their discipline, "the process of argumentation is the learning outcome," foregrounding the critical nature of knowledge-making. Thus, while non-STEM faculty emphasise voice, interpretation, and argumentation as the core of learning, STEM faculty prioritise methodological rigour, procedural mastery, and disciplined problem-solving.

Theme 2: Positioning of AI in Learning.

Table 2 presents participant quotations supporting this theme.

P8 (STEM)	"The problem is not the final answer, it's the verification and diagnostic thinking. AI can erode the need to debug a problem or check assumptions."
P5 (Non-STEM)	"The main battle for me is authorship. If I can't tell what is their original thought and what is the machine's, how can I grade their learning?"
P2 (STEM)	"Students have to understand why the AI-generated answer is correct or incorrect; otherwise, it's just a copy-paste without true understanding."
P12 (STEM)	"Students have to understand why the AI-generated answer is correct or incorrect."

Participants consistently expected GenAI to function as a supportive tool rather than a substitute for intellectual effort. For non-STEM faculty, this expectation took the form of a defence of authorship and interpretive responsibility. P5's concern that grading becomes untenable "if I can't tell what is their original thought and what is the machine's" captures the anxiety over the erosion of student voice, originality, and ownership in writing-intensive disciplines. STEM participants expressed parallel concerns about student responsibility, but framed around verification and conceptual rigour. P8 noted that "the problem is not the final answer," foregrounding debugging, assumption-checking, and troubleshooting as core disciplinary practices that GenAI might bypass. This was echoed by P2 and P12, who emphasised that students must understand why an AI-generated answer is correct or flawed. For these faculty, AI becomes pedagogically valuable only when its outputs are critically interrogated rather than adopted wholesale.

Theme 3: Interdisciplinary Exposure and Its Effects.

Table 3 presents participant quotations supporting this theme.

P1 (Non-STEM)	"Being with colleagues from other disciplines helped me see that the kind of output you're looking for is so different."
P2 (STEM)	"It's a good springboard for me, hearing how others in humanities approach AI... I saw the poetry example, which gave me an idea to do something similar but technical."
P10 (Non-STEM)	"When we discussed examples from different fields, I realised how varied the expectations are... it helped me question my own assumptions."
P4 (STEM)	"The humanities example made me realize we could use AI to generate incorrect code solutions, forcing the student to debug and apply theory differently."

Interdisciplinary interaction emerged as a powerful catalyst for reflection and pedagogical insight across participants. Non-STEM faculty, such as P1,

noted that engaging with colleagues from other fields illuminated how fundamentally different learning evidence can be across disciplines—whether an interpretive essay or a functional piece of code. STEM participants also described concrete pedagogical gains resulting from this exposure. P2 highlighted the value of hearing how humanities colleagues approach AI, noting that a poetry example inspired them to consider similarly creative yet technically aligned applications in their own course. P4 offered a parallel account, explaining how a humanities-based critical prompt led them to envision AI-generated incorrect code as a productive starting point for debugging and deeper theoretical engagement. These examples show how interdisciplinary settings do not merely foster appreciation of difference but actively stimulate creative transfer, enabling educators to adapt ideas across disciplinary boundaries.

Research Question 2: How Faculty Used GenAI

To aid us in analysing the lesson plans, we synthesised the categories of GenAI use in literature as shown in Table 4 (adapted from Chan, 2023; Gimpel et al., 2025; Mollick & Mollick, 2023; and UNESCO, 2023). The categories were further expanded to include sub-sub-categories/AI roles for a more comprehensive representation of use cases. Table 5 shows the analysis of the lesson plans and the categorisation to demonstrate the role of GenAI in their activity design.

Table 4

Generative AI Use in Higher Education – Categorization Table (Adapted from Chan, 2023; Gimpel et al., 2025; Mollick & Mollick, 2023; and UNESCO, 2023)

	Category	Sub-category
1	AI-assisted content and course design	Lesson planning and preparation
		Content simplification and curation
		Pre-learning support
2	Personalized learning and tutoring support	Adaptive tutoring and study help
		Metacognitive development
		Methodical guidance
		Simulation-based learning
3	Assessment and feedback enhancement	Formative assessment and draft support
		Academic integrity and evaluation
		Bias and quality checks
4	Enhancing student engagement and creativity	Ideation and brainstorming
		Application and integration
		Conceptual linking
		Critical analysis
5	Accessibility and inclusive learning	Professional translation
		Language comprehension support
		Structural and organizational support
		Tone and representation'
		Citation and research compliance
		Ethics and confidentiality
		Career development

Table 5

Use of GenAI by Participants

	Classification	Description
P1 (Non-STEM)	**Ideation & Brainstorming**	***AI-assisted drafting and comparative revision*** **.** In writing, having Draft I without AI. Draft II with the help of AI to gather expressions and examples of phrasing. Reflective comparisons between two drafts. Prevents shortcuts that bypass critical thinking.
P3 (Non-STEM)	**Formative assessment and draft support / AI-assisted content**	***Epistemic challenger & visual prompt generator.*** Using GenAI images to spark dialogue on social geography topics like space, place, dominant landscapes, and alternative landscapes, which had previously drawn stereotypical responses and avoidance of sensitive topics.
P5 (Non-STEM)	**AI-assisted content**	***Content curation / Reasoning support.*** Interdisciplinary groups solving complex ill-structured problems in sociology, making use of ideas from different disciplines. GenAI is used to curate relevant information from the discipline. Reasoning about how ideas from the two disciplines can solve the problem is articulated through a Venn diagram.
P6 (Non-STEM)	**Personalized learning and support**	***AI role-play simulation partner (instructor-mediated).*** Students present their ethical analysis of a case before engaging in a live role play where the LLM adopts a character—the medical patient, family members, etc. There is a reflective debrief on the medical ethics in relation to communication.
P7 (Non-STEM)	**Assessment and feedback enhancement**	***AI authenticity and attribution sentinel.*** After sourcing academic articles, students use GenAI to create an AI-assisted synthesis. They then check for factual errors and unsupported claims as they rephrase and rewrite without GenAI, practising critical judgment. They need to support their work with authentic references.
P9 (Non-STEM)	**Content creation / Enhancing student engagement and**	***Content creation, idea extender.*** Instructor uses GenAI for content creation like slide decks and role-play scenarios. Students are told to incorporate AI analysis in their strategies and negotiations, followed by post-reflections.
P11 (Non-STEM)	**AI-assisted content / Personalized learning and tutoring support**	***Technical scaffold and metacognitive prompt.*** In learning how to code for communications, students work on a group project where AI functions as a technical scaffold by offering code suggestions and visualisation ideas. The process is documented with student justification for why they accept or reject AI iterations, demonstrating conceptual comprehension.
P14 (Non-STEM)	**Personalized learning and tutoring support**	***Reasoning support and diagnostic tutoring.*** In a Psychology module focusing on the quantitative exploration of data, students complete a Python notebook exercise. Scaffolds are given, such as fill-in-the-blank tasks and iterative prompts. AI suggestions are compared to model code. AI acts as a reasoning aid while students practise diagnostic judgement.
P15 (Non-STEM)	**AI-assisted content curation / Enhancing student engagement and creativity**	***AI-assisted content design with critique.*** Students complete a worksheet that structures research prompts, analysis tasks, and presentation scripting. AI generates targeted content in the tutor's presence. Students must digest, critique, and orally present AI-derived materials, demonstrating synthesis and communicative competence. Students document AI use and provide peer feedback.
P2 (STEM)	**AI-assisted content**	***AI-assisted technical scaffold with verification focus.*** Lowers the barrier for students creating code for a Verilog testbench. AI generates code, but this is often prone to subtle errors. Students learn by verification and detecting errors. They develop independent learning and self-help skills.
P12 (STEM)	**Personalized learning and tutoring support**	***AI-assisted debugging with justificatory reflection.*** Students use GenAI to identify and correct intentional errors in code snippets. AI supplies candidate fixes and explanations while students evaluate and explain why changes are correct. Corrected code is submitted together with a reflective explanation The prompts they inpt demonstrate their understanding of the targeted hardware function.
P13 (STEM)	**Personalized learning and tutoring support / Methodological support**	***AI-assisted transcription analysis & reasoning support.*** Otter.ai is used for transcription. Students use GenAI for initial theme extraction, which is then refined and validated through human coding and group discussion. Students need to demonstrate reflection in their methodological awareness.

In reviewing the lesson plans, it became clear that what distinguished the disciplines was not the function of GenAI but the pedagogical emphasis each attached to it. STEM instructors used GenAI primarily as a technical scaffold requiring verification and error detection (P2, P12), while non-STEM instructors applied similar affordances to support idea extension, reasoning articulation, and interpretive judgement (P9, P14). These contrasts reflect discipline-specific conceptions of what counts as meaningful learning evidence. At the same time, a shared emphasis cut across all cases: students were expected to engage critically with GenAI. Lesson designs foregrounded reflective dialogue, peer comparison, and written justification to counter hallucination and bias. Across disciplines, critical evaluation was positioned as central to AI literacy, suggesting that while pedagogical purposes diverge, the core reflective and metacognitive competencies for responsible GenAI use remain common.

Disciplinary Professional Development

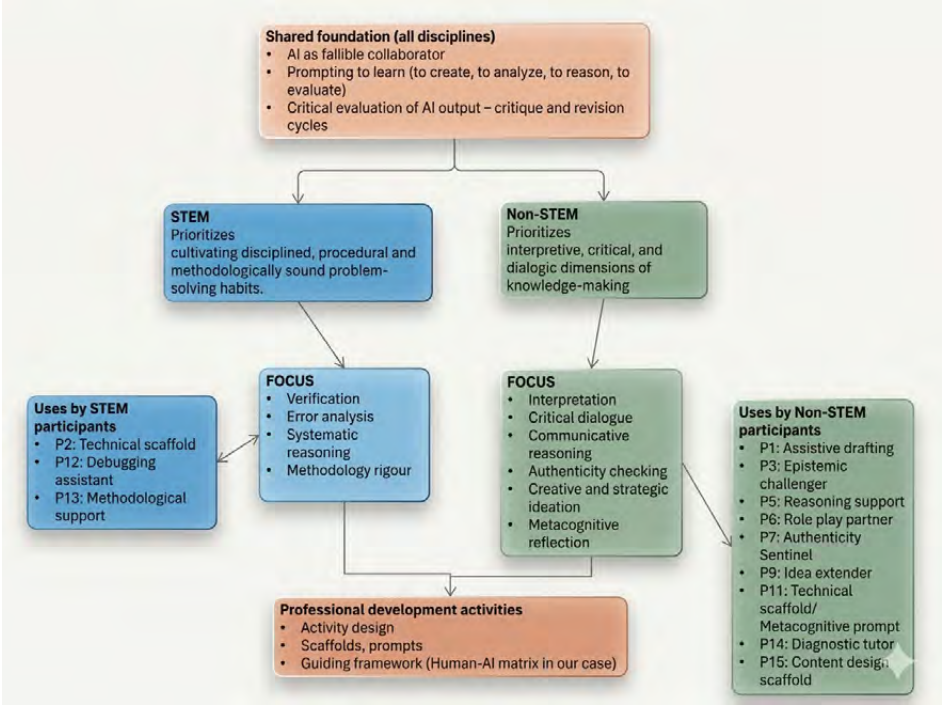
Drawing on these findings, we propose a limited framework for disciplinary professional development centred on three principles (Figure 1). Based on our findings, it is evident that STEM and non-STEM faculty prioritise different aspects of learning when integrating GenAI, and these epistemic commitments are reflected directly in the way they position AI within assessments and learning activities. The practical examples from the 12 use cases help us see alignment as well as possibilities in the disciplines.

How can we have professional development around the focal areas outlined above? Can detailed accounts of how particular assessments and activities were re-designed be one of the ways? Based on literature and our experience, professional development should centre on hands-on design supported by scaffolds and prompting frameworks; in our context, the Human–AI Matrix (Tan et al., 2025) served as a useful organising tool.

Across both STEM and non-STEM settings, however, a shared foundation remains: GenAI is fallible, and instructors must model and design for critical evaluation of its outputs. Developing prompting competence is therefore part of AI literacy, and faculty should understand how iterative, context-rich prompting works.

Figure 1

Disciplinary Professional Development – A Limited Framework



Implications

Professional Development Structures

In light of our findings, a natural question is what effective GenAI-related professional development (PD) should entail. Such PD must be contextual. Disciplinary components work best when there is a critical mass of faculty who already understand the basics of GenAI integration and the guiding framework. This foundation enables more productive disciplinary dialogue. We therefore recommend an initial interdisciplinary session, which—consistent with Theme 3—broadens perspectives for faculty new to GenAI and facilitates meaningful peer exchange. Faculty can then bring these insights back to their own departments. A promising model, currently under discussion with department partners, is a hybrid approach in which discipline-specific workshops are co-led by departmental faculty who share use cases and emergent integration trends. These workshops will be anchored in the Human–AI framework, which can be covered by the teaching and learning arm of the university.

The Human–AI Matrix as a Guiding Framework

In a related study (Tan et al., 2025), faculty found this tool clarifying because it shifted their focus from AI features to questions of cognitive responsibility, learning value, and ethical boundaries. It directly supports the design of tasks that cultivate reflection, critical evaluation of AI output, and metacognitive monitoring. This is a framework that we are continuing to refine based on participant feedback.

Limitations to the Study

The proposed professional development framework is limited by our small sample of 15 participants. While the data provide rich insight into STEM and non-STEM uses of GenAI, the framework is not yet generalisable. We plan to expand the sample to refine the framework for broader application and to explore cross-cultural PD contexts to further test and develop it.

Ethical Declaration Statement

This study was approved by the Institutional Review Board of the authors' university (Reference Number: IRB-2025-27).

Generative AI Policy

The authors used ChatGPT for parts of the analysis (especially in developing themes and cross-checking the quotations). It was also used for drafting support and refinement. All interpretations and final decisions were made by the authors.

References

Bearman, M., Tai, J., Dawson, P., Boud, D., & Ajjawi, R. (2024). Developing evaluative judgement for a time of generative artificial intelligence. *Assessment & Evaluation in Higher Education, 49*(6), 893–905. https://doi.org/10.1080/02602938.2024.2335321

Braun, V., & Clarke, V. (2014). What can "thematic analysis" offer health and wellbeing researchers? *International Journal of Qualitative Studies on Health and Well-being, 9*(1), 26152. https://doi.org/10.3402/qhw.v9.26152

Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education, 20*, 38. https://doi.org/10.1186/s41239-023-00408-3

Costello, J., Doljanin, Z., McAreavey, N., & Walsh, F. (Eds.). (2025). *Teaching humanities and social sciences in the era of generative AI: Case studies from around Ireland*. UCD College of Arts and Humanities. https://www.

ucd.ie/artshumanities/t4media/Case%20Studies%20from%20around%20Ireland.pdf

Garcia, K. F., Ong, A. K. S., Gumasing, M. J. J., & Delos Reyes, C. R. V. (2025). Engineering students' perceptions and actual use of AI-based math tools for solving mathematical problems. *Acta Psychologica, 256*, Article 105004. https://doi.org/10.1016/j.actpsy.105004

Gimpel, H., Hall, K., Decker, S., Eymann, T., Gutheil, N., Lämmermann, L., Braig, N., Maedche, A., Röglinger, M., Ruiner, C., Schoch, M., Schoop, M., Urbach, N., & Vandirk, S. (2025). Using generative AI in higher education: A guide for instructors. *Journal of Information Systems Education, 36*(3), 237–256. https://doi.org/10.62273/QLLG7172

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., Nerowski, M., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., & Kasneci, G. (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.102274

Lloyd, B., & Affias, O. (2025, April 4). The matrix that makes your ai strategy make sense. Rethinking developer productivity in the age of agentic AI. *Dev Interrupted*. https://devinterrupted.substack.com/p/the-matrix-that-makes-your-ai-strategy

Mollick, E. R., & Mollick, L. (2023). Using AI to implement effective teaching strategies in classrooms: Five strategies, including prompts. SSRN. https://doi.org/10.2139/ssrn.4391243

Tan, V., Tan, C. L., Liu, M. H., Soo, Y. J., & He, Y. Y. (2025, October 27–29). *Exploring shifts in faculty attitudes and design practices towards generative AI through the Human–AI matrix* [Conference session]. *Cambridge Generative AI in Education Conference*, Cambridge, United Kingdom. https://cambridge-ai-edu-conference.org.uk/programme

United Nations Educational, Scientific and Cultural Organization. (2023). *Guidance for generative AI in education and research*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000386693

University of North Carolina System. (2025). *Faculty learning communities*. https://www.northcarolina.edu/offices-and-services/academic-affairs/learning-and-technology/faculty-learning-communities/

Re-introducing the Flipped Approach at an Institutional Level

Chi-Ming WONG [1], Julia CHEN [2], Laura ZHOU [2]
[1] Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Hong Kong, China
[2] Educational Development Centre, The Hong Kong Polytechnic University, Hong Kong, China

Abstract

Flipped learning is a pedagogical model that has gained prominence in recent years; it redefines traditional classroom dynamics by shifting direct instruction outside the classroom. This transformation facilitates collaborative and higher-order activities during class time, empowering students to take ownership of their learning journeys and fostering a more interactive educational environment. However, challenges persist in conventional implementations, including student engagement with pre-class materials, high workloads for educators, and inadequate personalised support. This paper presents the approaches an institutional-level initiative has taken to tackle these challenges, highlighting how integrating advanced tools, such as Generative AI (GenAI), contributes to innovative solutions.

Keywords: Flipped Learning, Higher Education, Student Engagement, Educational Innovation, Generative AI

1. Introduction

Flipped learning, a pedagogical model popularised since 2007, redefines the traditional classroom dynamic by shifting direct instruction outside the classroom, thereby facilitating collaborative and higher-order activities during class time. This approach not only empowers students to take ownership of their learning journey but also promotes active engagement and critical thinking—key components of lifelong learning. By enabling students to explore concepts at their own pace before class, flipped learning cultivates a deeper understanding and encourages them to seek knowledge proactively, making them more adept at navigating future challenges.

Although the benefits of flipped learning are well-documented, widespread adoption often depends on the initiative of individual teachers (Baig & Yadegaridehkordi, 2023). Scaling this approach across an institution introduces new

complexities, requiring coordinated strategies and sustained support, especially for Chinese contexts (Raby, 2023). Integrating AI-powered tools into higher education can enhance flipped learning by providing personalised, adaptable, and continuous learning experiences (Ekanayake & Silva, 2025). For example, generative AI (GenAI) can offer tailored feedback and learning pathways that adapt to individual needs, fostering deeper understanding and critical thinking (Wong et al., 2025).

However, insufficient technological support can overwhelm both students and educators. At The Hong Kong Polytechnic University, the Flipped Learning and Teaching Initiative (FLTI) has been launched to address challenges identified in the literature and promote the adoption of the flipped classroom approach. The initiative enhances teaching and learning through innovative methods, such as integrating GenAI-generated videos, and fosters collaboration among stakeholders to overcome challenges and share successes.

To ensure the effective implementation of advanced flipped learning, the initiative is designed around three key objectives:

1. **Establish a Baseline:** Assess the current experiences and perceptions of both students and teachers regarding flipped learning practices.
2. **Respond to Needs:** Develop resources based on gathered data to equip educators with practical strategies for effectively implementing flipped learning in their courses.
3. **Gather Feedback:** Collect feedback from teachers on the effectiveness of these initiatives to evaluate our approach and identify areas for further improvement.

This paper presents our findings from the initiative, outlines our plans for future development, and discusses the limitations encountered during the project.

2. Methodology

This study was conducted at The Hong Kong Polytechnic University and involved a comprehensive, multi-phase data collection and professional development process to evaluate and enhance flipped learning practices across the institution (Figure 1).

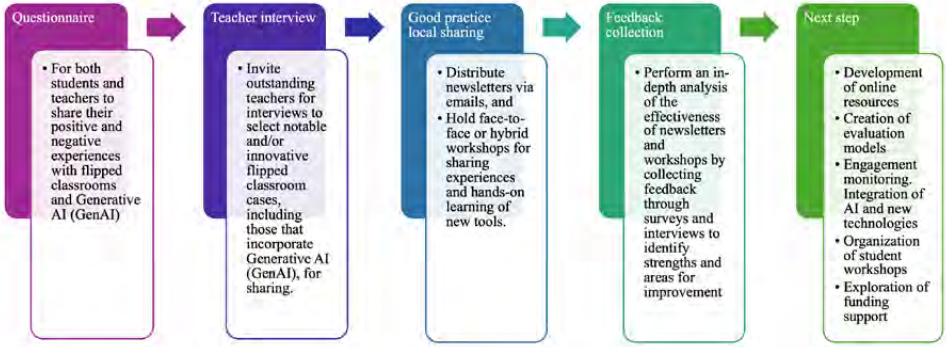


Figure 1. Overview of the five phases of the flipped classroom study.

- **Surveys:** Online surveys for a baseline study were distributed to all students and teachers at the university in November 2023. These surveys collected quantitative data on experiences, perceived barriers, and nominations of effective instructors. A total of 645 students (92 postgraduate and 553 undergraduate) and 89 teachers participated.
- **Teacher Interviews:** Semi-structured interviews were conducted with 24 teachers nominated by students for their exemplary flipped classroom approaches. These interviews explored their design, preparation, delivery strategies, and challenges related to flipped classrooms, providing valuable insights into best practices and areas for improvement.
- **Data Analysis:** Survey responses were analysed using descriptive statistics, while open-ended responses and interview transcripts underwent thematic analysis to identify key themes such as time constraints and technical support needs. Attendance records and survey feedback from workshops were also reviewed to evaluate learning outcomes and inform future programme design.

Table 1. Characteristics of participants at baseline.

Baseline characteristic	UG		PG	
	n	%	*n*	%
Year 1	288	52%	75	82%
Year 2	96	17%	9	10%
Year 3	89	16%	4	4%
Year 4	75	14%	4	4%
Year 5	5	1%	0	0%
Total	553	100%	92	100%

(Note: N = 645. Participants were from different departments or schools.)

3. Results and Discussion

Perceptions and Barriers for Students to Flipped Learning

Our baseline study revealed that 59.6% of students had prior experience with the traditional flipped classroom approach, with a majority understanding its principles. Specifically, 77% of undergraduates (Figure 2) and 74% of post-graduate students (Figure 3) believed that flipped classrooms could enhance their learning outcomes. Many highlighted increased engagement and critical thinking, with 78% of students supporting the idea that flipped classrooms promote active learning.

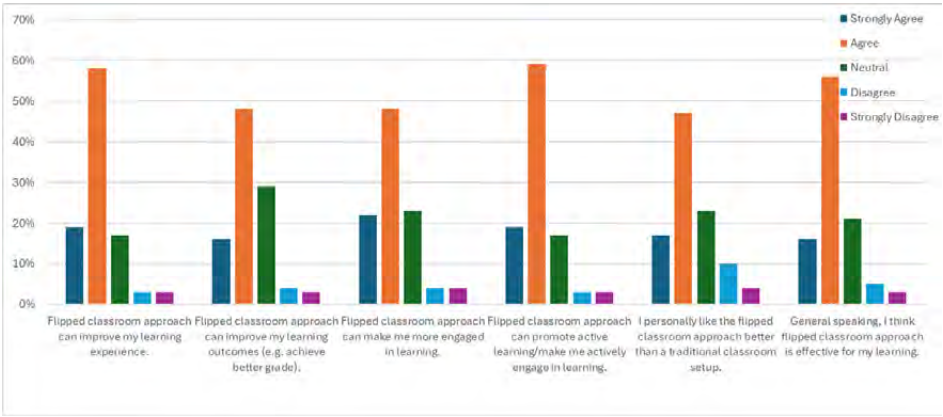


Figure 2. Undergraduate student responses on flipped classroom approach by degree level.

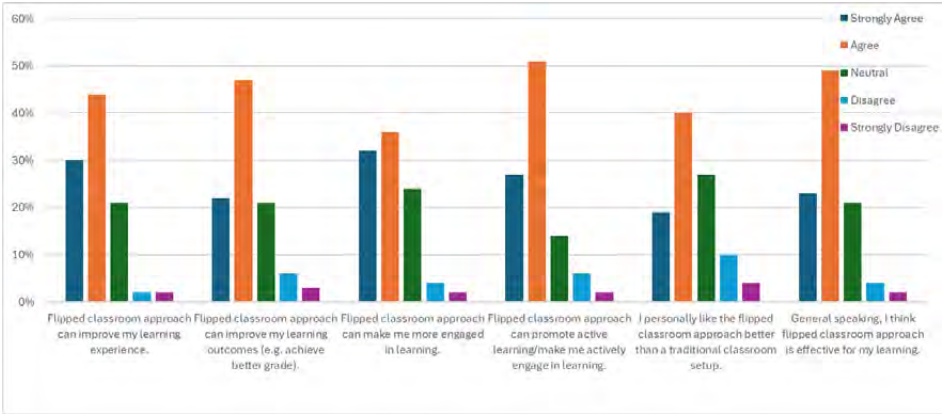


Figure 3. Postgraduate student responses on flipped classroom approach by degree level. *(Note: This bar chart illustrates the responses of students regarding their perceptions of the flipped classroom approach.)*

However, barriers remain that could hinder students from fully embracing this model. Many reported difficulties in completing pre-class assignments due to low motivation, poor time management, or competing academic responsibilities. Notably, 24% of undergraduates and 20% of postgraduates expressed anxiety about independently learning new content (Figures 2 & 3). This lack of immediate support can lead to feelings of unpreparedness during in-class activities. Thus, educators must implement effective strategies to address these challenges, such as providing structured guidance, time management training, and motivational support to enhance student engagement and success.

Challenges Faced by Teachers in Adopting Flipped Teaching and Solutions

The challenges associated with the flipped classroom model extend beyond students; they also significantly affect educators. While this approach has proven to be an effective teaching strategy, its implementation presents several hurdles, including ensuring accessibility, maintaining student engagement, and preparing teachers to integrate new technologies. A key issue frequently cited by educators is that, despite providing teaching materials online well in advance, many students fail to review them before class. Compounding this problem, over 90% of interviewed educators admitted that they rarely verify whether students engage with the materials, further complicating efforts to ensure the model's effectiveness.

To address these barriers, we conducted interviews with student-nominated teachers to gain a deeper understanding of their strategies for overcoming the challenges associated with flipped classrooms. These interviews provided valuable insights into best practices and common obstacles, which were critical for shaping actionable solutions. The findings were shared with all colleagues through a newsletter and made available on our webpage for future reference (Flipped Learning and Teaching Initiative, 2025).

Comprehensive Support for Educators

A crucial first step in scaling flipped learning is providing robust support for educators. We offer various training opportunities, including an introductory teaching course for new staff and hands-on workshops for current faculty. Additionally, flexible online courses allow faculty to learn at their own pace. Recognising that some teachers may doubt the effectiveness of flipped learning in their disciplines, we regularly interview experienced educators and share their best practices in monthly newsletters. This builds confidence and encourages broader adoption.

We also address concerns about the suitability of flipped learning for all

students, particularly those who may struggle academically. We introduce GenAI tools to assist educators in preparing lecture notes and commenting on assignments. We ensure that educators have access to necessary technological resources and financial support for effective implementation. Our goal is to create an empowering environment where teachers feel equipped to embrace flipped learning and leverage innovative tools in their classrooms.

Fostering Collaboration and Community

Teachers learn best from one another through the exchange of ideas and effective practices. To facilitate this, we have established platforms and organised regular sharing sessions to promote collaboration. In our local and international sharing series, educators from various disciplines present their teaching cases, focusing on both pedagogical strategies and practical solutions to common challenges. For instance, some teachers have demonstrated how GenAI tools can efficiently create pre-class videos, reducing preparation workload, while others have shared innovative methods, such as using the metaverse to enhance student engagement. These discussions address real-world issues, ensuring that every educator feels supported.

Creating a Supportive Institutional Culture

Institutional change requires a shared vision and strong support from all university levels. We have collaborated closely with university leadership to secure resources, establish clear guidelines, and foster a culture of innovation and collaboration. The establishment of a Steering Group on Flipped Learning Implementation has been pivotal in driving this initiative. This group includes key stakeholders—such as the Director of Education Development, Associate Heads, and frontline teachers—serving as a vital link between senior management and the teaching community.

The initiative has generated considerable interest, with numerous submissions from educators, leading to increased awareness of the benefits of flipped learning. Through these coordinated efforts, we are laying a strong foundation for sustainable, institution-wide adoption of flipped learning.

Empowering Educators with Advanced Tools

To further support our educators, we organised professional development workshops aimed at addressing technical challenges and boosting confidence in utilising advanced tools. Local and international experts were invited to share their expertise, introducing innovative technologies such as AI-driven video production platforms and immersive VR tools. For instance, in a VR workshop, educators learned how to create interactive anatomy lessons using VR simu-

lations. Participants received laptops, step-by-step instructions, and trial subscriptions to these tools, enabling them to integrate these technologies into their teaching practices effectively. These hands-on experiences not only enhanced technical proficiency but also empowered educators to produce high-quality flipped classroom materials efficiently. Feedback from participant questionnaires is summarised in the following figure.

Figure 4. Participant feedback on workshops organised.

The feedback indicated an overwhelmingly positive reception of our workshops (Figure 4). Many participants credited the success to the speakers, experienced educators who understood the challenges faced by teachers in similar environments. Their practical insights and real-world strategies resonated with attendees, providing an authentic learning experience. This peer-to-peer coaching approach was particularly valued, promoting collaboration and practical applicability, making the workshops impactful and inspiring.

Implications

Flipped learning represents a transformative pedagogical model that, when coupled with Generative AI (GenAI), can serve as a powerful catalyst for educational innovation (López-Villanueva et al., 2024). This report outlines how our initiative focuses on supporting educators in effectively integrating GenAI into their teaching practices (Chen et al., 2025). The encouraging results suggest that this approach not only motivates educators to explore new methodologies but also enhances their overall effectiveness. In the next phase, we will prioritise collecting student feedback on their experiences with the newly implemented flipped learning strategies (Figure 1). This input will provide valuable insights into the study's impact and inform future developments in our approach.

Acknowledgments

We thank our colleagues who generously shared their teaching experiences during the interviews and workshops. This study was supported by the Hong Kong Polytechnic University Teaching Grants and the Fund for Innovative Technology-in-Education (FITE/01).

Ethical Declaration Statement

The project has been approved by the Institutional Review Board of The Hong Kong Polytechnic University (HSEARS20230812001).

Generative AI Policy

The author(s) used ChatGPT during the preparation of this work to assist with content organisation and language refinement. After utilising the tool, the author(s) thoroughly reviewed and edited the content as necessary and assumed full responsibility for the publication's content.

References

Baig, M. I., & Yadegaridehkordi, E. (2023). Flipped classroom in higher education: A systematic literature review and research challenges. *International Journal of Educational Technology in Higher Education, 20*, 61. https://doi.org/10.1186/s41239-023-00430-5

Chen, J., Wong, C., & Zhou, L. (2025). Tips for embedding flipped learning at an institutional level. *Times Higher Education*. https://www.timeshighereducation.com/campus/tips-embedding-flipped-learning-institutional-level

Ekanayake, N. P. K., & Silva, N. K. L. (2025). Reimagining the flipped classroom: Integrating AI, microlearning, and learning analytics to elevate student engagement and critical thinking. *Faculty Focus*. https://www.facultyfocus.com/articles/blended-flipped-learning/reimagining-the-flipped-classroom-integrating-ai-microlearning-and-learning-analytics-to-elevate-student-engagement-and-critical-thinking/

Flipped Learning and Teaching Initiative. (2025). *Flipped Learning and Teaching Initiative*. The Hong Kong Polytechnic University. https://www.polyu.edu.hk/flti/?sc_lang=en

López-Villanueva, D., Santiago, R., & Palau, R. (2024). Flipped learning and artificial intelligence. *Electronics, 13*(17), 3424. https://doi.org/10.3390/electronics13173424

Raby, A. (2023). An exploration of the relationships between Chinese students and their personal tutors: An IPA study (Doctoral dissertation). School of Education, College of Arts, Social Sciences and Humanities.

Wong, C.-M., Zhou, L., Patel, M., Xu, M., & Tang, Y.-M. (2024). Harnessing the power of cutting-edge technologies in flipped learning: A paradigm shifts in university education after the post-COVID-19 era. In *10th International Conference on Higher Education Advances (HEAd'24)* (pp. 1–10). https://doi.org/10.4995/HEAd24.2024.17

Developing the AI Interaction Knowledge Scale for Higher Education Teachers: Pilot Validation

Adit GUPTA [1], Mool RAJ [1], Ankur GUPTA [2]
[1] MIER College of Education (Autonomous), Jammu, India
[2] Model Institute of Engineering and Technology, Jammu, India

Abstract

Despite the rapid adoption of generative artificial intelligence (GenAI) in higher education, teachers consistently struggle to translate awareness of AI's potential into effective pedagogical practice. Existing frameworks extend Technological Pedagogical Content Knowledge (TPACK) to address knowledge about AI but provide no validated instruments to measure the interaction competencies through which teachers operationalise that knowledge. This paper presents the AI Interaction Knowledge (AIK) framework, comprising four interrelated dimensions: Prompt Design Competency, Iterative Refinement Competency, Output Evaluation Competency, and AI Limitation Understanding. Following systematic scale development, a 70-item pool was refined through expert review (S-CVI = .96) to 54 items and then piloted with 35 higher education teachers. Item analysis produced a balanced 40-item instrument demonstrating excellent internal consistency (α range: .923–.955; total α = .967), strong discriminant validity (r = .522–.751), and acceptable criterion validity (r = .393–.728). The AIK Scale addresses a significant measurement gap and offers diagnostic utility for teacher professional development.

Keywords: AI Interaction Knowledge; TPACK; Generative artificial intelligence; Scale development; Teacher professional development

1. Introduction

Generative artificial intelligence (GenAI) has entered higher education at a remarkable speed. Within two years of mainstream accessibility, over 80% of educators worldwide reported using AI tools in their professional practice (Zou et al., 2025), drawn by GenAI's potential to support lesson planning, assessment design, differentiation, and resource creation (Kasneci et al., 2023). Yet widespread adoption has not translated into consistently effective practice. Research across institutional contexts documents a persistent gap: teachers understand what GenAI can do but struggle to realise those outcomes through actual use (Kasneci et al., 2023). This challenge is particularly acute in higher education, where faculty come from highly diverse disciplinary backgrounds with

variable AI experience. Generic professional development programmes cannot adequately address the specific interaction competencies that different academic communities require, nor can they diagnose where individual faculty members need the most targeted support.

Existing frameworks extend TPACK to encompass AI capabilities, ethics, and contextual understanding (Celik, 2023; Mishra et al., 2023). These contributions effectively address what teachers should know about AI. However, none explicitly theorise or provide validated instruments to measure the interaction competencies through which teachers translate that knowledge into effective AI-mediated practice. This paper addresses that gap by presenting the AI Interaction Knowledge (AIK) framework and reporting pilot validation findings for the 40-item AIK Scale.

Two research questions guided this study:

* **RQ1:** What are the psychometric properties of the AIK Scale in terms of internal consistency, discriminant validity, and criterion validity among higher education teachers?
* **RQ2:** Does the four-dimensional structure of the AIK framework hold as conceptually distinct yet interrelated competency domains in pilot testing?

2. Methodology

2.1 The AIK Framework

Effective AI integration in higher education requires more than awareness of GenAI's capabilities. Existing TPACK extensions, including the Intelligent-TPACK (Celik, 2023) and contextualised TPACK (Mishra et al., 2023), comprehensively address knowledge about AI but treat the interaction process as implicit rather than theorising it as a distinct competency domain.

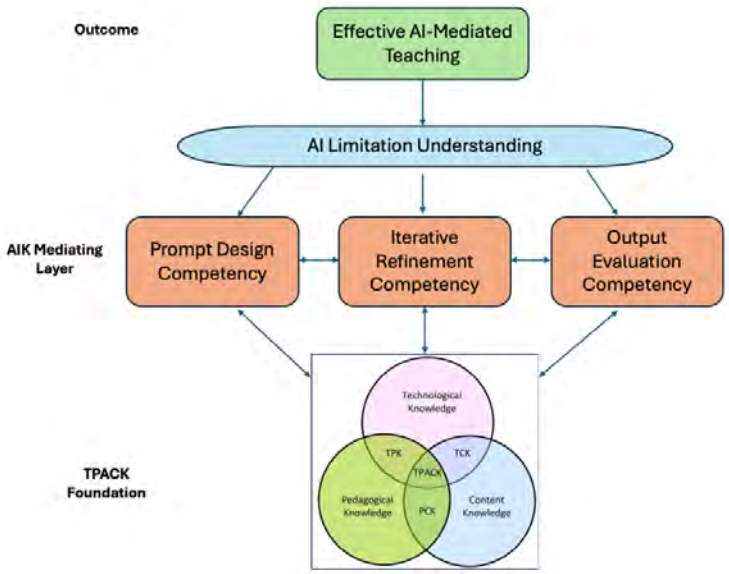
AI Interaction Knowledge (AIK) is proposed as a complementary framework grounded in affordance theory, activity theory, and distributed cognition, comprising four interrelated competency dimensions:

- **Prompt Design Competency (PDC):** Refers to the ability to create clear, contextualised prompts that accurately communicate pedagogical intentions to AI systems (Zamfirescu-Pereira et al., 2023). Without explicit prompt design skills, teachers frequently obtain outputs that are generic, misaligned with curriculum standards, or unsuitable for their learners.
- **Iterative Refinement Competency (IRC):** Encompasses skills re-

quired to progressively improve AI outputs through strategic interaction cycles, recognising that effective human-AI collaboration is inherently iterative rather than transactional (Qian, 2025).

- **Output Evaluation Competency (OEC):** Is the capacity to critically assess AI-generated content for factual accuracy, pedagogical appropriateness, and curriculum alignment, with particular importance given to GenAI's tendency toward hallucination.
- **AI Limitation Understanding (ALU):** Constitutes the meta-cognitive dimension, reflecting awareness of AI's boundaries and the judgement to recognise when AI assistance is inappropriate or insufficient.

As illustrated in Figure 1, the four dimensions are interdependent. PDC initiates effective interaction; IRC sustains and improves it; OEC evaluates outputs; and ALU provides meta-cognitive governance across all three. Together they constitute an integrated competency set complementary to TPACK (Mishra & Koehler, 2006).



(Figure 1: The AI Interaction Knowledge (AIK) Framework)

2.2 Scale Development and Pilot Methodology

Scale development followed established procedures for instrument construction (DeVellis, 2017). Drawing on the AIK theoretical framework and a systematic review of the human-AI interaction literature, 70 initial items were generated across the four dimensions. A five-member expert panel evaluated

all items for relevance and clarity using a four-point Content Validity Index rating protocol. Items with an Item-level Content Validity Index (I-CVI) below .78 were deleted (n = 16), yielding a 54-item pilot instrument. The Scale-level Content Validity Index (S-CVI = .96) substantially exceeded the recommended threshold of .90, confirming strong content validity prior to pilot testing.

A purposive sample of 35 higher education teachers from a teacher education institution in northern India participated in the pilot study. All participants provided informed consent, and institutional ethical approval was obtained. The sample demonstrated characteristics appropriate for instrument validation: 91.4% held postgraduate qualifications (60% doctoral), 94.3% had received formal AI training, and all 35 participants were active GenAI users, with 83% using tools at least several times weekly.

The 54-item instrument employed a five-point Likert scale (1 = Strongly Disagree; 5 = Strongly Agree) across the four AIK dimensions. Ten criterion variables measuring AI integration outcomes were included to establish criterion validity, comprising three items on intention to use AI in teaching, three items on confidence in AI use, and four items on current frequency of AI use. Data were collected during November–December 2025 via an online survey platform and analysed using SPSS Version 28.

3. Results and Discussion

3.1 Descriptive Statistics and Item Refinement

Mean scores across the 54 items ranged from 3.77 to 4.46 on the five-point scale, with standard deviations between 0.55 and 0.95. The elevated means reflected the sample's high AI experience profile. Distribution analysis identified three items with severe kurtosis exceeding 4.0, signalling ceiling effects that would compromise psychometric performance, and these were prioritised for deletion. Following systematic item analysis, 14 items were deleted to produce a balanced 40-item scale with 10 items per dimension. Deletion decisions were guided by corrected item-total correlations below .40, severe kurtosis, improvement to Cronbach's alpha upon deletion, and content redundancy. No dimension required extensive revision.

3.2 Internal Consistency

All four dimensions demonstrated excellent internal consistency, substantially exceeding the .70 threshold recommended for newly developed scales and surpassing the .90 benchmark for excellent reliability. The total scale alpha of .967 is exceptional for a newly developed instrument, with reliability data presented in Table 1.

Table 1: Internal Consistency Reliability for AIK Scale Dimensions

Dimension	Initial Items	Final Items	Cronbach's α
Prompt Design Competency	14	10	.955
Iterative Refinement Competency	12	10	.946
Output Evaluation Competency	16	10	.941
AI Limitation Understanding	12	10	.923
Total AIK Scale	**54**	**40**	**.967**

Note: Final alphas computed after item deletion. All values indicate excellent internal consistency.

3.3 Discriminant Validity

Inter-dimensional correlations ranged from .522 to .751, addressing RQ2 by confirming that the four dimensions, although related, assess conceptually distinct competency domains. No correlation exceeded .80, the conventional threshold above which discriminant validity is considered compromised. The moderate-to-strong correlation range indicates the dimensions are sufficiently related to form a theoretically coherent construct, while remaining distinct enough to justify four separate subscales, consistent with AIK's conceptualisation as a multidimensional rather than unitary competency.

3.4 Criterion Validity

Correlations between AIK dimension scores and the three AI integration outcome variables ranged from .393 to .728, addressing RQ1 by providing evidence that AIK dimensions predict meaningful real-world AI integration behaviours. The strongest association was observed between Output Evaluation Competency and teacher confidence in AI use ($r = .728$). The weakest association, between AI Limitation Understanding and intention to use AI ($r = .393$), nevertheless exceeded the threshold for acceptable criterion validity.

3.5 Discussion

The pilot validation findings provide strong initial evidence for the AIK Scale as a psychometrically sound instrument, with all four dimensions demonstrating excellent reliability, a coherent factor structure, and meaningful criterion validity. The AIK Scale addresses a measurement gap that existing instruments do not fill. Current TPACK-derived scales assess what teachers know about AI, including its capabilities, ethics, and pedagogical applications (Celik, 2023; Mishra et al., 2023), whereas the AIK Scale measures how effectively teachers engage with AI systems to achieve pedagogical goals. Research demonstrates that interaction skill, rather than awareness alone, determines the quality of AI-mediated teaching outcomes (Zamfirescu-Pereira et al., 2023; Zou et al.,

2025), yet no validated instrument previously existed to measure this at scale.

The ceiling effect identified in the pilot sample, attributable to participants' high AI experience profile, indicates the scale will demonstrate stronger discriminatory power when administered to broader faculty populations with more varied AI experience. This finding reinforces the importance of diverse demographic and institutional sampling in the forthcoming main study validation.

4. Implications

The AIK Scale offers direct diagnostic utility for higher education institutions. Administering the scale as a pre-programme instrument enables the identification of dimension-specific competency gaps across an academic workforce. A faculty development programme might find, for instance, that staff demonstrate strong Prompt Design Competency but systematically lower scores on Output Evaluation Competency, enabling targeted rather than generic professional development. This differentiated, evidence-based approach to faculty AI development is particularly relevant in national higher education contexts where AI integration is being actively promoted at the system level (Kasneci et al., 2023).

For research, the validated 40-item scale provides a standardised instrument for comparative studies across institutional types, disciplines, and national contexts. The next phase will undertake confirmatory factor analysis with an internationally diverse sample spanning Australia, China, Indonesia, and Singapore, establishing a cross-culturally validated instrument for the global higher education community.

The AIK Scale makes a distinct contribution to the educational technology measurement landscape. Where existing TPACK-derived instruments assess knowledge about AI, the AIK Scale assesses the interaction competencies through which teachers operationalise that knowledge in practice. This distinction has direct consequences for how institutions diagnose professional development needs and design faculty support programmes for the GenAI era.

Acknowledgements

Not applicable.

Ethical Declaration Statement

This research has received ethics clearance from the institutional Research & Development Cell (RDC) of MIER College of Education (Autonomous), Jammu, India. Ethics Clearance Number: MIER/RDC/2026/001.

Declaration of Use of AI in the Writing Process

The author(s) used Claude (Anthropic) during the preparation of this work to support manuscript organisation, language editing, and improvement of writing clarity. After utilising the tool, the author(s) thoroughly reviewed and edited the content as necessary and assumed full responsibility for the publication's content.

References

Celik, I. (2023). Towards Intelligent-TPACK: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education. *Computers in Human Behavior*, 138, 107468. https://doi.org/10.1016/j.chb.2022.107468

DeVellis, R. F. (2017). *Scale development: Theory and applications* (4th ed.). SAGE.

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., … Kasneci, G. (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

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108(6), 1017–1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x

Mishra, P., Warr, M., & Islam, R. (2023). TPACK in the age of ChatGPT and generative AI. *Journal of Digital Learning in Teacher Education*, 39(4), 235–251. https://doi.org/10.1080/21532974.2023.2247480

Qian, Y. (2025). Prompt engineering in education: A systematic review of approaches and educational applications. *Journal of Educational Computing Research*, 63(7-8), 1782–1818. https://doi.org/10.1177/07356331251365189

Zamfirescu-Pereira, J. D., Wong, R. Y., Hartmann, B., & Yang, Q. (2023). Why Johnny can't prompt: How non-AI experts try (and fail) to design LLM prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1–21). Association for Computing Machinery. https://doi.org/10.1145/3544548.3581388

Zou, D., Chen, X., & Huang, R. (2025). Navigating the future: Establishing a framework for educators' pedagogic artificial intelligence competence. *European Journal of Education*, 60(3), e70117. https://doi.org/10.1111/ejed.70117

6 Student AI Literacy, AI Readiness, and Cognitive Development

Reciprocal Associations of Positive Youth Development and AI Literacy: Longitudinal Evidence from Filipino Adolescents

Jet Uy BUENCONSEJO [1] & Davy Tsz Kit NG [2]

[1] Department of Applied Social Sciences, Faculty of Health and Social Sciences, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

[2] Department of Mathematics and Information Technology, Faculty of Liberal Arts and Social Sciences, The Education University of Hong Kong, Tai Po, New Territories, Hong Kong

Abstract

Amid growing concerns regarding adolescents' readiness to navigate rapidly emerging technologies, scholars have highlighted the need to understand how digital competencies intersect with core developmental assets. Beyond technical skills, non-cognitive psychosocial resources may play a critical role in cultivating artificial intelligence (AI) literacy. Guided by the Positive Youth Development (PYD) framework, this study examined reciprocal longitudinal associations between the 5Cs of PYD (competence, confidence, connection, character, and compassion) and AI literacy dimensions (awareness, usage, evaluation, and ethics). Two waves of survey data from 612 Filipino secondary students were analysed using latent cross-lagged panel modelling with full longitudinal invariance. Results showed strong stability across all PYD dimensions and AI ethics. PYD attributes displayed domain-specific effects on AI literacy, while AI ethics exerted broader influences on socio-emotional Cs (connection and character). A bidirectional relationship emerged between character and AI ethics. Findings underscore the intertwined development of moral, relational, and digital competencies in youth.

Keywords: AI literacy, AI ethics, positive youth development, student thriving

1. Introduction

The rapid expansion of artificial intelligence (AI) technologies across social, economic, and educational sectors has intensified global concerns regarding adolescents' readiness to navigate AI-mediated environments. **AI literacy**, defined as the ability to identify, use, evaluate, and ethically engage with AI technologies, has emerged as a foundational 21st-century competence (Wang et al., 2023). Recent global surveys indicate that more than 80% of students use

AI for academic tasks, with over half engaging with AI tools daily or weekly (DEC, 2024). Despite this widespread adoption, nearly one in two students also feels unprepared to use AI technologies effectively; 58% report lacking adequate AI knowledge and skills, while 49% struggle to critically evaluate the trustworthiness of AI outputs (DEC, 2024). These trends reveal a widening gap between adolescents' high exposure to AI and their limited preparedness to use such tools critically and responsibly.

Current AI literacy initiatives are largely dominated by cognitive-technical approaches that emphasise operational skills, digital competencies, and algorithmic awareness (Wang et al., 2023). However, emerging research proposes that psychosocial factors such as self-efficacy, socio-emotional competencies, and ethical awareness may substantially shape students' engagement with emerging technologies (Martin-Barrado & Gomez-Baya, 2025; Palmquist et al., 2025). Adolescents must not only know how to use AI but also when and why to use it, how to judge the validity of AI-generated outputs, and how to uphold ethical standards in digital contexts. Integrating non-cognitive psychosocial strengths is therefore critical to fostering competent and responsible AI use (Palmquist et al., 2025).

The Positive Youth Development (PYD) framework offers a robust lens for examining how young people's psychosocial assets may support AI literacy development. PYD conceptualises youth thriving through five interrelated dimensions called the 5Cs: competence, confidence, connection, character, and caring (Buenconsejo & Datu, 2022; Lerner et al., 2010). **Competence** reflects perceived ability in academic, social, and task-specific domains; **confidence** refers to self-worth and self-efficacy; **connection** captures positive relationships with family, peers, and institutions; **character** encompasses integrity, moral reasoning, and respect for societal norms; and **caring or compassion** involves empathy and concern for others (Geldhof et al., 2014). Together, these dimensions underscore a holistic perspective on youth development, emphasising personal abilities, relational resources, and moral responsibility.

PYD strengths may influence the development of AI literacy in several ways. Competence and confidence can motivate adolescents to use AI tools to support learning, refine skills, and manage academic tasks more efficiently. Connection may foster collaborative AI use among teachers and peers. Character and compassion may underpin ethical AI practices, integrity in academic work, and sensitivity to issues such as fairness, privacy, and harm. Conversely, AI literacy may also shape PYD attributes by enabling students to engage more confidently with technology, strengthen academic competence, cultivate responsible decision-making, and deepen social connection through collaborative AI-mediated learning environments.

Despite these theoretical linkages, little empirical work has directly examined how adolescents' PYD psychosocial strengths and AI literacy competencies influence one another over time. Moreover, virtually all existing studies have been conducted in high-income Western or East Asian contexts. Investigating these processes in developing countries such as the Philippines is essential, given persistent inequalities in digital access, rapid AI adoption in education, and the heightened risk that students from low- and middle-income countries (LMICs) may be left behind in the accelerating AI-driven global landscape. Understanding how psychosocial strengths and AI competencies co-develop in such contexts can inform more equitable, contextually grounded educational interventions.

This study addresses these gaps by examining the reciprocal longitudinal associations between the 5Cs of PYD and the four dimensions of AI literacy (awareness, usage, evaluation, and ethics) among Filipino secondary school students. By integrating strengths-based developmental science with contemporary AI literacy frameworks, the study provides insight into how psychosocial and digital competencies jointly shape adolescents' readiness to engage with emerging technologies.

2. Methodology

A two-wave longitudinal design was employed to examine the associations between PYD and AI literacy dimensions. Data were collected from 612 Filipino secondary school students (M^{age} = 16.86, SD^{age} = .88; 54% males) enrolled in one private school in Metro Manila. Surveys were administered in August 2024 (Time 1) and April 2025 (Time 2). Parental consent and student assent were obtained at both measurement points, and all procedures adhered to institutional ethical protocols.

PYD was assessed using the 15-item academic version of the 5Cs of PYD measure (Buenconsejo, In Press), which captures competence, confidence, connection, character, and compassion. AI literacy was measured using the 12-item AI Literacy Scale (Wang et al., 2023), assessing awareness, usage, evaluation, and ethics. All items were rated on a 5-point scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The survey was administered through Qualtrics during students' homeroom sessions. Preliminary analyses included descriptive statistics, reliability assessment, bivariate correlations, and confirmatory factor analyses (CFAs). Longitudinal measurement invariance (configural, metric, scalar, and strict) was tested for both PYD and AI literacy constructs to ensure consistency in factor structures over time.

As shown in Figure 1, a latent cross-lagged panel model (LCLPM) was estimated using the `lavaan` package in R (Mackinnon et al., 2022). Gender, age, and socioeconomic status were included as covariates to ensure the incremental validity of the PYD and AI literacy constructs. Goodness-of-fit indices were evaluated using standard criteria (CFI/TLI ≥ .90, RMSEA/SRMR ≤ .08). All analyses were conducted in R.

3. Results and Discussion

All PYD and AI literacy dimensions demonstrated adequate reliability at both time points. CFAs supported the validity of the five-factor PYD model and the four-factor AI literacy model, yielding adequate fit indices. As shown in Table 1, longitudinal measurement invariance was fully established for all PYD and AI literacy factors, permitting meaningful comparison of latent means and relationships across time.

The latent cross-lagged panel model demonstrated strong model fit (SB $\chi^2(1413) = 2252.01$, $p < .001$; CFI = .965; TLI = .960; RMSEA = .031; SRMR = .048). Autoregressive estimates revealed moderate stability for all PYD dimensions and AI ethics ($\beta s \approx .23–.48$), indicating that PYD and moral characteristics remained relatively stable across the eight-month period. Conversely, AI awareness, usage, and evaluation exhibited weaker temporal stability, with non-significant autoregressive paths, suggesting that these dimensions may be more malleable or sensitive to rapid changes in AI tools or technologies. Given the accelerating adoption of AI tools in schools, students' exposure, usage patterns, and evaluative abilities may possibly shift in shorter time frames, contributing to their weak stability.

Several PYD attributes exhibited significant predictive associations with subsequent AI literacy domains. Competence predicted higher AI awareness ($\beta = .16$) and evaluation ($\beta = .07$), supporting the idea that strong academic and cognitive self-perceptions may help students engage more critically with AI tools. Confidence and connection predicted later AI usage ($\beta = .06$; $\beta = .10$), aligning with research suggesting that self-efficacy and positive relational contexts promote exploration and adoption of new digital technologies (Martin-Barrado & Gomez-Baya, 2025). These findings confirm that distinct PYD strengths facilitate domain-specific AI literacy competencies.

AI ethics emerged as the most influential AI literacy dimension, predicting gains in both socio-emotional PYD characteristics and other AI literacy skills. Time 1 AI ethics predicted increases in connection ($\beta = .24$), character ($\beta = .28$), awareness ($\beta = .22$), usage ($\beta = .17$), and evaluation ($\beta = .30$) at Time

2. These results underscore the centrality of ethical sensitivity for navigating AI's risks, responsibilities, and social implications. Ethical orientations may strengthen relational qualities such as trust and respect, reinforce moral behaviour, and enhance students' ability to use AI tools responsibly (Wang et al., 2023).

Notably, character and AI ethics demonstrated a bidirectional relationship: Time 1 character predicted Time 2 AI ethics ($\beta = .12$), while Time 1 AI ethics exerted a stronger reciprocal influence on Time 2 character ($\beta = .28$). This mutually reinforcing pattern reflects the potential co-development of moral character and ethical AI practices, consistent with virtue-oriented perspectives in developmental science that emphasise how moral reasoning and ethical behaviour shape one another over time (Buenconsejo & Datu, 2022; Lerner et al., 2010). This finding also highlights the need for educational interventions that foster both ethical understanding and moral dispositions in AI-mediated learning environments.

Overall, the results indicate that PYD strengths, particularly competence, confidence, connection, and character, may serve as psychosocial resources that can bolster AI literacy development. Conversely, ethical AI literacy may function as a catalyst for both socio-emotional development and domain-specific AI skills. This reciprocal pattern supports a holistic perspective in which digital, moral, and relational competencies interrelate and co-develop, reinforcing calls to broaden AI literacy education beyond technical proficiencies (Palmquist et al., 2025).

4. Implications

The findings highlight the importance of integrating PYD and AI literacy frameworks to support adolescents' development in increasingly AI-mediated learning environments, especially in LMIC educational settings. Educational interventions that cultivate ethical sensitivity alongside self-efficacious and socio-emotional competencies may be particularly effective in promoting competence and responsible engagement with AI. Given the strong predictive role of AI ethics, programmes should incorporate explicit instruction on fairness, privacy, responsible use, and moral reasoning related to AI. In addition to cognitive predictors such as digital literacy, critical thinking, and technological proficiency, educators and policymakers should recognise the value of psychosocial resources in promoting AI awareness, usage, evaluation, and ethical engagement. A holistic educational approach that strengthens both cognitive and non-cognitive developmental assets can enhance students' capacity to navigate AI tools critically and responsibly.

The present study underscores the urgent need for more research on AI literacy in LMIC contexts. Students in developing countries often have limited access to digital infrastructure and AI-related training, increasing the risk of being excluded from the rapidly evolving AI-driven world. Investing in both psychosocial and digital competencies is therefore critical to ensuring that young people from underrepresented and under-resourced communities are not left behind. By revealing how PYD strengths and AI literacy competencies influence one another over time, the study offers a new developmental perspective on preparing the next generation for AI-mediated futures. Strengthening psychosocial resources alongside digital competencies may provide more equitable, developmentally grounded pathways for supporting youth thriving in an AI-driven society.

Acknowledgement

The conduct of this study was partly funded by the Post-Doctoral Fellowship Scheme from the Research Grants Council of Hong Kong SAR (Reference No. PDFS2324-8H02) awarded to the first author. The presentation of this study was financially supported by the Department of Applied Social Sciences of The Hong Kong Polytechnic University (Reference No. P0059996).

Ethical Declaration Statement

The project has been approved by the Human Research Ethics Committee of The Education University of Hong Kong (Reference Number: 2024-2025-0181).

Generative AI Policy

The first author used Grammarly during the preparation of this work to shorten the report and meet word limits. After utilising the tool, the authors thoroughly reviewed and edited the content as necessary and assumed full responsibility for the publication's content.

References

DEC. (2025). *Digital Education Council Global AI Student Survey 2024*. Digital Education Council, Singapore. https://www.digitaleducationcouncil.com/post/what-students-want-key-results-from-dec-global-ai-student-survey-2024

Buenconsejo, J. U. (In Press). Thriving in academic settings: Contextualising and measuring the 5Cs of positive youth development among students. *Journal of Psychoeducational Assessment*, SSCI.

Buenconsejo, J., & Datu, J. A. D. (2022). Towards an integrative paradigm of positive youth development: Implications for research, practice, and policy. *Human Development, 66*(6), 381–396. https://doi.org/10.1159/000527122

Lerner, R. M., von Eye, A., Lerner, J., Lewin-Bizan, S., & Bowers, E. (2010). The meaning and measurement of thriving: A View of the Issues. *Journal of Youth & Adolescence, 39*, 707–719. https://doi.org/10.1007/s10964-010-9531-8

Mackinnon, S., Curtis, R., & O'Connor, R. (2022). Tutorial in longitudinal measurement invariance and cross-lagged panel models using lavaan. *Meta-Psychology, 6*. https://doi.org/10.15626/mp.2020.2595

Martin-Barrado, A. D., & Gomez-Baya, D. (2025). The association between the use of digital technologies and positive youth development: A systematic review. *Frontiers in Psychology, 16*. https://doi.org/10.3389/fpsyg.2025.1552128

Palmquist, A., Sigurdardottir, H. D. I., & Myhre, H. (2025). Exploring interfaces and implications for integrating social-emotional competencies into AI literacy for education: A narrative review. *Journal of Computer in Education*. https://doi.org/10.1007/s40692-025-00354-1

Wang, B., Rau, P. L. P., & Yuan, T. (2023). Measuring user competence in using artificial intelligence: Validity and reliability of artificial intelligence literacy scale. *Behaviour & Information Technology, 42*(9), 1324–1337. https://doi.org/10.1080/0144929X.2022.2072768

Figure 1. Conceptual model of the latent cross-lagged panel model between PYD and AI literacy dimensions.

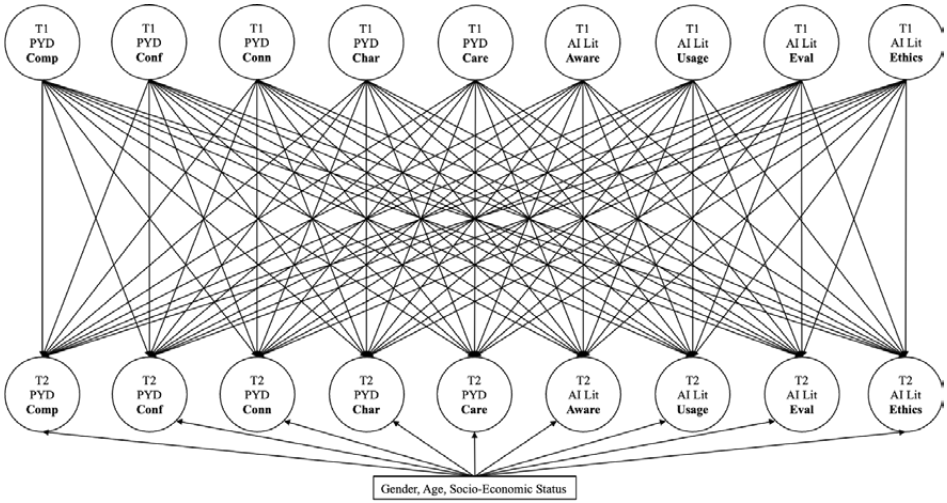


Table 1. Results of the longitudinal measurement invariance on the 5Cs model of PYD and the four-factor model of AI literacy across Time 1 and Time 2 (n = 612).

Model	SB χ^2 (*df*)	CFI	RMSEA	SRMR	ΔCFI	ΔRMSEA	ΔSRMR	Invariant?
Configural	1514.06 (1197)	0.997	0.018	0.028	–	–	–	Baseline
Metric	1576.34 (1224)	0.995	0.021	0.032	-.002	+.003	+.004	Yes
Scalar	1614.05 (1251)	0.995	0.022	0.032	.000	+.001	.000	Yes
Strict	1660.61 (1278)	0.994	0.023	0.035	-.001	+.001	+.003	Yes

Note: SB $\chi 2$ = Satorra–Bentler scaled chi-square test statistic; *df* = degrees of freedom; CFI = Comparative Fit Index; RMSEA = root mean square error of approximation; SRMR, standardized root mean square residual; Δ = change.

AI Interaction and Creative Thinking in Higher Education: The Mediating Role of AI Cognition

Xiaojie FAN [1] & Yan DING [2]
[1] Educational College, Shanghai Normal University
[2] The Research Institute for Higher Education, Center for Faculty Development, Fudan University

Abstract

This study investigates how university students' AI interaction behaviours influence creative thinking through the mediating role of AI cognition. Using survey data collected in 2025 from 396 students at a leading "Double First-Class" university in China, quantitative analyses—including descriptive statistics, hierarchical regression, and structural equation modelling (SEM)—were conducted. Results indicate that students who enrolled in AI-related coursework demonstrated significantly higher levels of AI interaction competence, more positive AI cognition, and stronger creative thinking compared with those without such coursework. Regression analyses confirmed that AI interaction positively predicted creative thinking, and mediation results revealed that positive AI cognition significantly mediated this relationship. Negative AI cognition, however, did not exhibit a significant mediating effect. Moreover, AI course enrolment strengthened the relationship between AI interaction and creative thinking. These findings highlight the cognitive-behavioural mechanism linking AI engagement and creativity, and underscore the importance of structured AI literacy education in fostering students' creative capacities in higher education.

Keywords: AI experience; AI interaction; AI cognition; creative thinking

1. Introduction

With the rapid adoption of generative artificial intelligence (GenAI) tools such as ChatGPT and DeepSeek in higher education, AI-assisted learning has become an essential component of students' daily academic activities. Students increasingly rely on AI for information retrieval, data analysis, academic writing, and programming tasks (Zawacki-Richter et al., 2019; Kasneci et al., 2023). Consequently, AI has evolved from a supplementary instrument into a "cognitive partner" within the learning process, profoundly reshaping how learners construct knowledge and solve problems. As universities continue to integrate AI literacy into curricula and advance AI-enhanced teaching reforms, understanding how AI influences students' cognition and creativity has become

a topic of both theoretical and practical significance.

Creative thinking, regarded as the cognitive foundation of creativity, refers to the mental process through which individuals generate novel ideas by engaging in divergent and integrative thinking (He, 2017). Previous studies have suggested that AI can act as a catalyst for innovative exploration by providing personalised feedback, multiple perspectives, and inspiration (Chang et al., 2023). However, evidence also indicates that the high degree of automation in AI tools may foster excessive dependence on externally generated content, thereby undermining learners' autonomy and critical reflection (Ahmedtelba, 2025). These divergent findings reveal that the mechanisms through which AI affects creative thinking remain ambiguous.

In particular, how students' AI experience translates into cognitive processes that promote or hinder creativity in higher education requires further empirical investigation. In this study, students' engagement with AI tools in learning contexts—conceptualised as *AI interaction*—serves as the primary measure of AI experience. In addition, formal participation in AI-related courses is considered a contextual factor that may moderate the effect of AI interaction on creative thinking. Drawing on the Technology Acceptance Model (Davis, 1989) and the theory of technology-use behaviour (Venkatesh et al., 2003), AI interaction experience is conceptualised as a key factor shaping students' perceptions and attitudes toward AI. Extending this perspective, *AI cognition*—defined as students' understanding, trust, and evaluation of AI's advantages and limitations—may serve as a psychological mediator between AI interaction and creative thinking. Integrating Amabile's (1983) componential model of creativity and Kaufman and Beghetto's (2009) four-level model of creativity, this study posits that positive AI cognition expands students' thinking and fosters innovative ideation, whereas negative cognition may constrain originality and reflective thought. Accordingly, this research, employing descriptive statistics, regression analysis, and SEM, aims to investigate the relationships among AI interaction, AI cognition, and creative thinking using data from undergraduates at a "Double First Class" university in China. The results are expected to contribute theoretically by clarifying the psychological mechanism linking AI learning and creativity, and practically by informing the design of high-quality AI curricula that enhance students' creative capacity in higher education.

2. Methodology

This study adopted a quantitative, cross-sectional design to explore how students' AI interaction behaviours influence their AI-related cognition and multi-level creative thinking under different learning contexts. The overall an-

alytical model followed an "independent–mediator–dependent" logic, in which *AI interaction* served as the independent variable, *AI cognition* as the mediating variable, and *creative thinking* as the outcome variable. AI course enrolment functioned as a contextual factor moderating these relationships.

Questionnaire data were collected in 2025 from 396 students at the aforementioned university; nearly half (49.2%) had taken AI courses. The questionnaire mainly consisted of three sections that measured: (a) AI interaction behaviour, (b) AI cognition, and (c) multi-level creative thinking, using five-point Likert scales.

(1) AI Interaction (Predictor)

This scale captured the extent and sophistication of students' engagement with AI tools in learning and problem-solving contexts. Seven items were developed to assess students' adaptability and reflective strategies when interacting with AI (e.g., "I flexibly adjust my prompts when using AI tools to achieve better solutions"). Higher scores indicated greater interactive and self-regulated use of AI.

(2) AI Cognition (Mediator)

Based on the revised Creativity Support Index (Cherry & Latulipe, 2014) and related creativity assessment instruments, this study developed a localised *Perceived Impact of AI on Creativity* scale containing 12 items. In this context, *AI cognition* refers to students' cognitive appraisal, understanding, and interpretation of how AI technologies influence their learning and creative thought processes. Factor analyses yielded two latent components: *positive AI cognition* and *negative AI cognition*, representing the facilitative and inhibitive aspects of how students view AI's role in supporting creative processes.

(3) Creative Thinking (Outcome)

The outcome variable in this study was *creative thinking*, conceptualised as the cognitive process underlying idea generation, imagination, and problem reconstruction in AI-supported learning contexts. Although Kaufman and Beghetto's (2009) multi-level creativity framework originally describes tiers of creative achievement, it was used here as a referential mapping structure to approximate different levels of students' creative thinking performance. Seven items were designed to evaluate students' creative thinking at different developmental stages:

(a) Mini-C: reflecting individual learning-based creativity and the personal construction of meaning through AI-assisted exploration;

(b) Little-C: representing everyday creative thinking demonstrated in routine academic tasks where AI supports idea refinement and problem-solving; and

(c) Pro-C: indicating more advanced, domain-specific creative thinking applied to professional-level or project-oriented challenges.

Descriptive and correlation analyses summarised basic variable patterns. Hierarchical regression tested the direct effect of AI interaction on creativity and the moderating role of AI course enrolment. SEM further examined the mediating effect of AI cognition, with model fit indices and bootstrapping used for validation. This analytic strategy provided a comprehensive assessment of direct, indirect, and contextual links among AI interaction, AI cognition, and creative thinking.

3. Results and Discussion

3.1 Descriptive Statistics and Mean Differences

Students who had taken AI courses performed significantly better in *AI interaction*, showing greater competence in using AI tools. They also displayed more positive *AI cognition* and less negative *AI cognition*, suggesting that coursework fostered more constructive perceptions of AI's role in creative learning. Furthermore, students with AI course experience demonstrated higher levels of *creative thinking*, and this pattern remained consistent across the Mini-C, Little-C, and Pro-C dimensions. The findings suggest that students exposed to AI education tend to demonstrate stronger *AI interaction* skills, more positive cognitive attitudes toward AI, and higher *creative thinking* across different levels. This provides initial empirical support for the proposed pathway connecting *AI interaction → AI cognition → creative thinking*.

Table 1: Group Differences Between Students With and Without AI Course Enrolment

Variables	No AI Course (n = 201)	AI Course (n = 195)	*p*
AI Interaction	4.154 (0.669)	4.418 (0.717)	< .001
Positive AI Cognition	3.602 (0.980)	3.887 (0.983)	.004
Negative AI Cognition	2.896 (1.193)	2.590 (1.229)	.013
Creative Thinking	3.847 (0.727)	4.155 (0.809)	< .001
Mini-C	3.970 (0.761)	4.267 (0.798)	< .001
Little-C	3.738 (0.818)	4.101 (0.890)	< .001
Pro-C	3.833 (0.814)	4.097 (0.871)	.002

Note: Values are represented as mean (SD). p values are based on independent-samples t tests (two-tailed).

3.2 AI Interaction and Creative Thinking: The Moderating Role of AI Course Enrolment

To investigate the relationship between *AI interaction* and students' *creative thinking*, as well as the potential moderating role of *AI course enrolment*, hierarchical regression analyses were conducted (Table 2). Model 1 confirms a significant positive effect of *AI interaction* on *creative thinking*, indicating that students who engage more actively with AI systems demonstrate higher *creative thinking* performance. After adding positive and negative *AI cognitions* in Model 2, the main effect of *AI interaction* remained significant but slightly attenuated, suggesting a partial explanatory role of AI-related cognitions. Model 3, incorporating demographic and educational controls, shows that these associations are robust and not driven by individual background differences. Importantly, Model 4 reveals a significant interaction between *AI interaction* and *AI course enrolment*. The positive slope indicates that the creativity gains from *AI interaction* are stronger among students with formal *AI course* enrolment. However, *AI course enrolment* alone did not yield a direct effect on *creative thinking*, suggesting that its influence operates by strengthening the link between interaction and *creative thinking* performance. Taken together, these results support the hypothesis that active *AI interaction* promotes *creative thinking*, particularly when accompanied by structured coursework that deepens students' understanding of AI.

Table 2: Hierarchical Regression Analysis Predicting Creative Thinking

Variables	Model 1	Model 2	Model 3	Model 4
AI Interaction	.771*** (.040)	.529*** (.047)	.523*** (.049)	.415*** (.072)
Positive AI Cognition		.309*** (.034)	.314*** (.034)	.307*** (.035)
Negative AI Cognition		−.089*** (.020)	−.084*** (.020)	−.080*** (.020)
Gender (Female)			.009 (.054)	−.015 (.056)
Science & Engineering			.123* (.073)	.133* (.072)
Medical Science			−.019 (.074)	−.018 (.073)
Other Disciplines			.040 (.069)	.048 (.068)
Master's Student			.045 (.066)	.051 (.064)
Doctoral Student			−.146* (.089)	−.110 (.087)
AI Course Enrollment				.057 (.055)
AI Course × AI Interaction				.209*** (.076)
Constant	3.999*** (.028)	3.087*** (.151)	3.024*** (.155)	3.003*** (.158)
N	396	396	389	389
R^2	.481	.614	.623	.632

*Note: Standardised regression coefficients (β) are reported. Standard errors are presented in parentheses. Reference categories: Male (gender), Humanities & Social Sciences (discipline), Undergraduate (education level), and No AI course enrolment. * p < .05; ** p < .01; *** p < .001.*

3.3 AI Interaction and Multilevel Creative Thinking

To examine whether *AI interaction* significantly predicts *creative thinking* across different levels, separate regression models were estimated for Mini-C, Little-C, and Pro-C. Each model included *AI interaction*, *AI course enrolment*, and their interaction, while controlling for *AI cognition*, gender, major category, and education level. Marginal effects were then calculated and visualised in Figure 1.

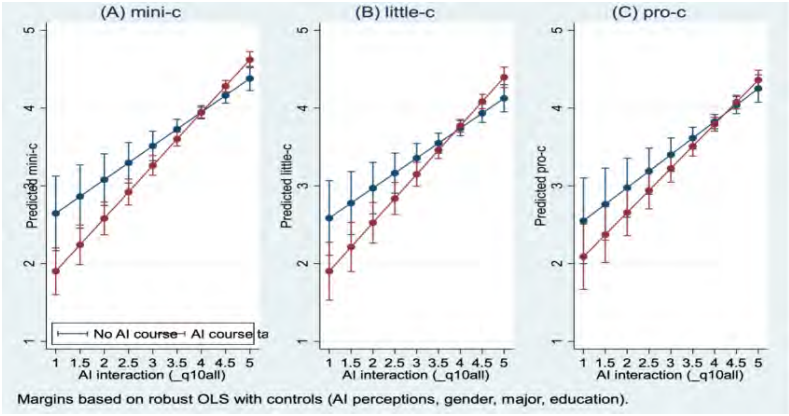
For **Mini-C**, *AI interaction* was positively and significantly associated with Mini-C outcomes ($\beta = .434$, $p < .01$). Moreover, the interaction between *AI interaction* and *AI course enrolment* was positive and statistically significant ($\beta = .247$, $p < .01$), indicating that the positive association between *AI interaction* and Mini-C was stronger among students who had taken AI courses. As shown in Figure 1 (Panel A), predicted Mini-C increased with higher levels of *AI interaction* for both groups, but the increase was steeper among students with *AI course* experience.

Similar patterns were observed for **Little-C**. *AI interaction* exhibited a significant positive relationship with Little-C ($\beta = .385$, $p < .01$), and the interaction term between *AI interaction* and *AI course enrolment* was also significant and positive ($\beta = .238$, $p < .01$). The marginal effects plot in Figure 1 (Panel B) illustrates that differences between students with and without *AI course* experience became more pronounced as *AI interaction* levels increased.

In contrast, for **Pro-C**, *AI interaction* remained a significant positive predictor ($\beta = .425$, $p < .01$), whereas the interaction term with *AI course enrolment* did not reach statistical significance ($\beta = .143$, $p > .10$). The marginal effects shown in Figure 1 (Panel C) suggest that Pro-C increased steadily with *AI interaction* for both groups, with relatively small differences between students with and without *AI course* experience.

In sum, the results above indicate that *AI interaction* is a significant predictor of creativity across all three levels, while the moderating role of *AI course enrolment* is evident for Mini-C and Little-C but not for Pro-C.

Figure 1: The Impact of AI Interaction on Different Levels of Creative Thinking



3.4 Mechanism of the Effect of AI Interaction on Creative Thinking via AI Cognition

Table 3 reports the bootstrap mediation results testing whether *AI interaction* influences students' overall *creative thinking* through *AI cognition*, with comparisons between individuals who have and have not taken AI courses.

For students without *AI course* learning, *AI interaction* exerted a significant direct effect on *creative thinking*, indicating that active engagement with AI tools independently enhances *creative thinking*. A significant indirect effect through positive *AI cognition* further suggests that favourable cognitive beliefs about AI partially mediate this relationship. In contrast, the pathway via negative *AI cognition* remained nonsignificant, implying that anxiety or scepticism toward AI does not substantially weaken *creative thinking* among this group.

Among students who had enrolled in *AI courses*, the direct impact of *AI interaction* on *creative thinking* strengthened considerably, accompanied by a larger indirect contribution from positive *AI cognition*. This pattern indicates that structured *AI* learning magnifies both students' confidence in AI and the creative returns from *AI* engagement. Consistent with the first group, negative *AI cognition* again failed to serve as a meaningful mediator.

Taken together, these findings demonstrate a robust dual pathway: *AI interaction* enhances *creative thinking* both directly and indirectly through positive cognitive mechanisms. Moreover, *AI course enrolment* amplifies this process, suggesting that educational exposure to AI not only improves students' conceptual understanding but also strengthens the cognitive–behavioural foundations that link *AI* use with *creative thinking* outcomes.

Table 3: Mediation of AI Cognition in the Association Between AI Interaction and Overall Creative Thinking by AI Course Enrollment Group

Group	Effect	Mediator	*B*	*SE*	z	p	95% CI
No AI Course (n = 199)	Indirect	Positive AI Cognition	0.194	0.043	4.49	< .001	[0.109, 0.278]
		Negative AI Cognition	0.019	0.017	1.11	.266	[−0.015, 0.053]
		Total Indirect Effect	0.213	0.043	4.99	< .001	[0.129, 0.296]
	Direct	AI Interaction → Overall Creative Thinking	0.407	0.076	5.34	< .001	[0.258, 0.556]
AI Course Enrollment (n = 190)	Indirect	Positive AI Cognition	0.234	0.048	4.89	< .001	[0.140, 0.328]
		Negative AI Cognition	0.010	0.009	1.11	.265	[−0.008, 0.028]
		Total Indirect Effect	0.244	0.047	5.17	< .001	[0.152, 0.337]
	Direct	AI Interaction → Overall Creative Thinking	0.635	0.063	10.15	< .001	[0.512, 0.757]

Note: B = unstandardised regression coefficient; SE = standard error. Estimates are based on 5,000 bootstrap resamples. Indirect effects represent the mediating roles of positive and negative AI cognition in the association between AI interaction and overall creative thinking. Confidence intervals that do not include zero indicate significant mediation effects.

4. Implications

This study highlights how students' interaction with AI tools and their cognitive responses to AI jointly shape *creative thinking* outcomes. The empirical results confirm that active *AI interaction* enhances *creative thinking* both directly and through positive *AI cognition*, and that prior *AI course enrolment* amplifies these effects. These findings contribute to creativity and technology learning theory by demonstrating a cognitive–behavioural pathway through which AI use can foster higher-order creative competencies.

From a practical perspective, the evidence suggests that integrating *AI pedagogy* into higher education can extend beyond technical proficiency to encourage adaptive mindsets and *creative thinking*. Educational policymakers and instructors should consider embedding structured reflective learning activities to cultivate students' positive appraisals of AI and their innovative application of emerging technologies. As AI technologies become increasingly embedded in education, understanding how learners cognitively and behaviourally interact with AI is critical. Encouraging balanced, constructive engagement with AI tools can help prepare future professionals equipped for creative problem-solving in technology-driven environments. Thus, the present findings offer timely insights into fostering creativity in an AI-enabled higher education learning landscape.

Generative AI Policy

GenAI (ChatGPT) was used solely to support language refinement, grammar correction, and stylistic consistency. It was not used for study design, data analysis, interpretation, or the development of theoretical arguments. All intellectual contributions were made by the authors. The authors critically reviewed all AI-assisted outputs and assume full responsibility for the content of this manuscript.

References

Ahmedtelba, E. (2025). Critical Integration of Generative AI in Higher Education: Cognitive, Pedagogical, and Ethical Perspectives. *London Journal of Research in Humanities and Social Sciences, 25*(13), 1-12.

Amabile, T. M. (1983). The social psychology of creativity: A componential conceptualization. *Journal of Personality and Social Psychology, 45*(2), 357.

Chang, D. H., Lin, M. P.-C., Hajian, S., & Wang, Q. Q. (2023). Educational design principles of using AI chatbot that supports self-regulated learning in education: Goal setting, feedback, and personalization. *Sustainability, 15*(17), Article 12921. https://doi.org/10.3390/su151712921

Cherry, E., & Latulipe, C. (2014). Quantifying the creativity support of digital tools through the Creativity Support Index. *ACM Transactions on Computer-Human Interaction, 21*(4), 1–25. https://doi.org/10.1145/2617588

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science, 35*(8), 982–1003. https://doi.org/10.1287/mnsc.35.8.982

Dittmar, E. C. (2025). AI as a catalyst for organizational learning: Moving beyond tool implementation to learning transformation. *Development and Learning in Organizations: An International Journal, 40*(1), 34–37.

He, K. (2017). *A theory of creative thinking: Construction and verification of the dual circulation model*. Springer. https://doi.org/10.1007/978-981-10-5053-4

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274.

Kaufman, J. C., & Beghetto, R. A. (2009). Beyond big and little: The four C model of creativity. *Review of General Psychology, 13*(1), 1–12. https://doi.org/10.1037/a0013688

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425–478. https://doi.org/10.2307/30036540

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education, 16*(1), 39.

Fostering Students' Artificial Intelligence Literacy Through Authentic Learning with Programmable Robots

Chi Fuk SO [1] & Ka Ki LAM [1]
[1] Centre for Learning, Teaching, and Technology, The Education University of Hong Kong, Hong Kong, China

Abstract

This study examined the effectiveness of programmable robots in fostering an authentic learning environment and developing 21st-century skills within that setting. Participants first attended a two-hour structured educational workshop aimed at improving their artificial intelligence (AI) literacy. The workshop introduced students to foundational AI concepts and offered hands-on experience with either AlphAI or Thymio robots. After completing the workshop, participants were asked to complete a survey comprising 35 Likert-type items to evaluate their authentic learning experience and two types of 21st-century skills: ways of thinking and ways of working. The results indicated that the deployment of programmable robots significantly contributed to the creation of an authentic learning environment, which facilitated the acquisition of essential generic skills, including problem-solving, innovation, and collaboration. Students' evaluations also affirmed the effectiveness of programmable robots in improving AI literacy. These findings emphasise the potential of integrating programmable robots into AI education to enhance students' AI literacy and essential competencies for success in higher education.

Keywords: artificial intelligence literacy, authentic learning, educational robotics, 21st-century skills

1. Introduction

AI has become an integral part of contemporary life, influencing diverse domains ranging from business to education. As it can be foreseen that AI will eventually permeate every aspect of our daily lives, there is an increasing necessity to cultivate AI literacy among citizens to facilitate adaptation to the AI-driven era. AI literacy extends beyond merely comprehending AI technologies. Jia et al. (2025) have proposed an AI literacy framework consisting of three dimensions: AI awareness, AI mechanics, and AI impacts, spanning the spectrum from AI definition to responsible practice. This framework included basic to advanced skills, suggesting that the acquisition of AI literacy is a grad-

ual and cumulative process. Promoting AI education is of utmost importance. Tadimalla and Maher (2024) proposed a four-pillar approach, comprising understanding AI-related knowledge, interacting with Gen-AI responsibly, examining the social and ethical issues related to AI, and considering the future development of AI, to deliver AI literacy to individuals with different educational backgrounds. The overarching objective of AI education is to equip individuals with foundational AI knowledge, enable them to utilise AI tools, and encourage them to reflect on the ethical and societal implications of AI applications. AI literacy is essential for comprehending and interacting with AI technologies that are shaping modern society.

Given the transformative role of AI in modern society, fostering AI literacy from an early age is imperative. Consequently, AI education has been incorporated into K–12 curricula to ensure that students are prepared to engage thoughtfully and responsibly with AI technologies. AI education can be implemented with an authentic learning approach. Roach et al. (2018) defined authentic learning as "learning supported by being situated in an environment that aligns learning objectives with real-world tasks, content and context". Authentic learning necessitates the incorporation of real-world elements that manifest in various forms, such as task-based, content-based, and context-based learning. Authentic learning can be implemented in diverse forms to make knowledge practical. It is beneficial for students' learning. A study showed that authentic task-based teaching enhanced students' academic achievement, retention level, attitude, self-confidence, problem-solving skills, and creativity (Pullu & Gömleksiz, 2021). Authentic learning not only equips students with practical knowledge but also with valuable skills, laying a consolidated foundation for their future success.

The deployment of robots in educational settings is prevalent nowadays. Educational robots can construct an authentic learning environment to enhance students' learning efficacy. Incorporating robots into lessons enables students to actively construct knowledge through direct interaction with the robots. Hence, students are provided with the opportunity to transform their practical experiences of operating the robots into a personalised understanding and meaningful learning. Students can also put their knowledge into practice by operating robots. Hands-on experience can enhance students' learning retention rate; they can memorise key learning points for a longer time with the assistance of robots. Besides, learning with robots can raise students' motivation and engagement. The learning motivation and creativity of students in an information technology course were elevated by gamified educational robots (Chen et al., 2023). The integration of robots into educational settings can create an engaging and

enjoyable learning environment, thereby enhancing the overall learning experience for students.

This study aimed to investigate the effectiveness of deploying programmable robots in authentic learning and the influence of authentic learning on the development of 21st-century skills. The research questions of this study were:

1. Can programmable robots create an authentic learning environment?
2. Can authentic learning strengthen students' problem-solving skills and innovation?
3. Can authentic learning strengthen students' collaborative skills?

2. Methodology

Participants in this study included both undergraduate and postgraduate students who voluntarily attended a two-hour structured educational workshop designed to enhance their AI literacy. The workshop began with an introductory session covering foundational AI topics, including its definition, classification, and various practical applications. Following this, students engaged in a hands-on activity to explore different machine learning models, including supervised learning, unsupervised learning, and reinforcement learning. During the workshops, students had the opportunity to operate either an AlphAI or a Thymio robot. Both robots feature a user-friendly graphical interface for operation that is suitable for AI novices. As some of the participants were pre-service teachers, they may consider utilising these robots to deliver AI education to K–12 students in the future.

Participants worked collaboratively in small groups of two to three members, stimulating peer interaction and teamwork throughout the learning process. A survey was developed to evaluate students' learning experience, based on previous research (Herrington & Oliver, 2014; Keinänen et al., 2018; Pu et al., 2016; Tezer et al., 2018). The survey was divided into three sections, focusing on students' authentic learning experience and two types of 21st-century skills: ways of thinking and ways of working. It consisted of 35 Likert-type items, each based on a 5-point scale. Detailed survey items are available in the Appendix.

The "Ways of Thinking" and "Ways of Working" sections evaluated students' development in generic skills, including problem-solving skills, innovation, and collaborative skills. The "Authentic Learning Experience" section assessed the degree of authenticity of the workshop based on the context and activities. In addition to these sections, students rated their perceived improvement in AI literacy using a 4-point Likert scale. SPSS was used to analyse the

data collected. Cronbach's alpha was used to measure internal consistency, while the chi-square goodness-of-fit test was used to examine whether the distributions of responses differed significantly from a uniform distribution across the rating scale, thus providing evidence regarding participants' levels of agreement with statements about authentic learning, 21st-century skills, and AI literacy.

3. Results and Discussion

A total of six identical workshops were conducted between April 2024 and March 2025, involving 98 participants, with a survey completion rate of 97.96%. Following the first workshop, a reliability test was conducted for the survey. It achieved an overall Cronbach's alpha of 0.961 (n = 10) across all 35 items, which is well above the acceptable threshold of 0.7, demonstrating excellent reliability and internal consistency. Hence, this survey was adopted throughout the study.

Regarding the "Ways of Thinking" section, the mean rating was 4.33 (SD = 0.65). The chi-square goodness-of-fit test results showed a significant difference in the distribution of responses, with χ^2 values ranging from 17.06 to 89.08 and all $p < .001$, indicating that participants' responses were not evenly distributed among the scale options but were instead skewed towards agreement. The chi-square goodness-of-fit test results provided statistical evidence that students significantly agreed that they had developed problem-solving and innovation skills within the authentic learning environment. Given the close relevance of AI to everyday life, students were highly motivated to engage with topics related to AI. Enhanced knowledge of AI empowered students to devise novel solutions to real-world problems. These findings were consistent with prior research (Güven & Nas, 2022), which has shown that activities designed using an authentic learning approach positively impact students' creative problem-solving abilities.

Table 1: Descriptive Statistics of Survey Results

Section	N	Mean	Standard Deviation	Maximum	Minimum
Ways of Thinking	96	4.33	0.65	5	2
Ways of Doing	96	4.45	0.56	5	2
Authentic Learning Experience	96	4.39	0.57	5	2

Table 2: Chi-Square Goodness-of-Fit Test Results for Ways of Thinking Section

Item	df	χ^2	p
1	2	17.06	<.001
2	3	51.17	<.001
3	2	40.19	<.001
4	2	31.69	<.001
5	3	66.08	<.001
6	2	28.31	<.001
7	2	31.19	<.001
8	2	30.06	<.001
9	3	61.25	<.001
10	3	89.08	<.001
11	2	34.19	<.001
12	3	70.58	<.001
13	2	22.94	<.001
14	3	66.92	<.001
15	3	61.50	<.001
16	3	51.92	<.001

Note: χ^2 = Chi-square value rounded to two decimal places; $p < .001$ for all tests; all expected cell frequencies were adequate.

Note: N = number of respondents.

The "Ways of Working" section recorded a mean rating of 4.45 (SD = 0.56). The chi-square goodness-of-fit test results showed a significant difference in the distribution of responses in most questions, with χ^2 values ranging from 1.04 to 45.06, and all $p < .001$ except item 4. The strong skew towards agreement confirmed that students perceived substantial enhancement of their collaborative skills through the authentic learning experience. The group setting facilitated communication among students, while the hands-on activity encouraged collaboration as they worked collectively towards a shared objective. It could be concluded that an authentic learning environment enabled students to strengthen their 21st-century skills, including problem-solving, innovation, and collaborative skills.

Table 3: Chi-Square Goodness-of-Fit Test Results for Ways of Working Section

Item	df	χ^2	p
1	2	36.75	<.001
2	2	42.75	<.001
3	2	42.25	<.001
4	1	1.04	.307
5	2	45.06	<.001
6	2	31.94	<.001
7	2	28.00	<.001
8	2	37.00	<.001
9	2	39.44	<.001

Note: χ^2 = Chi-square value rounded to two decimal places; p-values reported as < .001 when below .001.

Regarding the "Authentic Learning Experience" section, the mean rating was 4.39 (SD = 0.57). The chi-square goodness-of-fit test results showed a significant difference in the distribution of responses, with χ^2 values ranging from 29.25 to 91.17 and all $p < .001$, indicating that students' responses were concentrated towards agreement. These results showed that participants recognised a high level of authenticity in the learning experience. Students were able to engage directly with the AI training process by programming the robots to compete in the robot race. Overall, these results have proved that the integration of programmable robots is effective in creating a genuinely authentic learning environment that promotes the development of key generic skills.

Table 4: Chi-Square Goodness-of-Fit Test Results for Authentic Learning Experience Section

Item	df	χ^2	p
1	2	45.44	<.001
2	2	44.44	<.001
3	2	34.94	<.001
4	3	79.17	<.001
5	2	29.25	<.001
6	2	37.00	<.001
7	3	82.83	<.001
8	2	41.31	<.001
9	3	91.17	<.001
10	2	40.19	<.001

Note: χ^2 = Chi-square value rounded to two decimal places; $p < .001$ for all tests; all expected cell frequencies were adequate.

Regarding AI literacy, the mean rating of the AI literacy-related item was 3.57 out of 4 (SD = 0.52). The chi-square goodness-of-fit test yielded a χ^2 value of 49.56 with $p < .001$, indicating that responses were again skewed towards agreement rather than evenly distributed across the scale. Students strongly agreed that they had gained a better understanding of machine learning and related concepts through the hands-on activities. The deployment of programmable robots made abstract AI concepts more concrete and observable, thereby facilitating students' comprehension and retention of these ideas.

Table 5: Descriptive Statistics of Students' Evaluation on AI Literacy

Item	N	Mean	Standard Deviation	Maximum	Minimum
The hands-on activities help me better grasp the idea of machine learning and related concepts	96	3.57	0.52	4	2

Note: N = number of respondents.

Table 6: Chi-Square Goodness-of-Fit Test Results for Students' Evaluation on AI Literacy

Item	df	χ^2	p
The hands-on activities help me better grasp the idea of machine learning and related concepts	2	49.56	<.001

Note: χ^2 = Chi-square value rounded to two decimal places.

4. Implications

In conclusion, this study demonstrated that the deployment of programmable robots is effective in establishing an authentic learning environment and enhancing AI literacy. Students can experience a simplified AI training process by operating the programmable robots. As AI literacy becomes an increasingly strategic priority in society, it is essential to equip individuals with the necessary skills to navigate an AI-driven future. AI education is a requisite for K–12 education and even higher education. Authentic learning is an efficacious pedagogical approach for delivering AI education. By adopting an authentic learning framework, abstract AI concepts can be transformed into tangible, real-world applications, enabling students to connect theoretical knowledge with practical contexts. Furthermore, this approach facilitates the development of essential 21st-century skills, including problem-solving, innovation, and collaboration, and prepares individuals for their future academic and career endeavours. It is recommended that educators incorporate authentic learning methodologies into their AI curricula to enhance student engagement and skill acquisition. In

the future, interviews will be conducted to gain more insights into students' authentic learning experience. It is suggested that future research can explore the effectiveness of authentic learning environments in K–12 education, thereby promoting the authentic learning pedagogy for AI education.

Acknowledgments

The project was supported by the Fund for Innovative Technology-in-Education provided by the University Grants Committee, Hong Kong.

Ethical Declaration Statement

The project has been approved by the Human Research Ethics Committee of The Education University of Hong Kong (Reference Number: 2023-2024-0139).

Generative AI Policy Statement

The author(s) used Perplexity during the preparation of this work to edit languages. 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

Chen, T.-I., Lin, S.-K., & Chung, H.-C. (2023). Gamified educational robots lead an increase in motivation and creativity in STEM education. *Journal of Baltic Science Education, 22*(3), 427–438. https://doi.org/10.33225/jbse/23.22.427

Güven, İ., & Nas, C. (2022). The effectiveness of inquiry based activities based on authentic learning approach on 5th grade students' academic achievement and creative problem solving skills. *Academy Journal of Educational Sciences, 6*(1), 52-58. http://dx.doi.org/10.31805/acjes.1091917

Herrington, J., Reeves, T. C., & Oliver, R. (2014). Authentic learning environments. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of Research on Educational Communications and Technology* (pp. 401–412). Springer, New York, NY. https://doi.org/10.1007/978-1-4614-3185-5_32

Jia, K., Leung, T. H. M., Cheung, N. Y. I., Li, Y., & Yu, J. (2025). Developing a holistic AI literacy framework for children. *ACM Transactions on Computing Education, 25*(2), Article 21. https://doi.org/10.1145/3727986

Keinänen, M., Ursin, J., & Nissinen, K. (2018). How to measure students' innovation competences in higher education: Evaluation of an assessment tool in authentic learning environments. *Studies in Educational Evaluation, 58*, 30–36. https://doi.org/10.1016/j.stueduc.2018.05.007

Pu, Y.-H., Wu, T.-T., Chiu, P.-S., & Huang, Y.-M. (2016). The design and implementation of authentic learning with mobile technology in vocational nursing practice course. *British Journal of Educational Technology, 47*(3), 494–509. https://doi.org/10.1111/bjet.12443

Pullu, E. K., & Gömleksiz, M. N. (2021). Authentic teaching tasks for academic success, attitude, problem solving, and creative thinking skills. *Discourse and Communication for Sustainable Education, 12*(2), 108–123. https://doi.org/10.2478/dcse-2021-0020

Roach, K., Tilley, E., & Mitchell, J. (2018). How authentic does authentic learning have to be? *Higher Education Pedagogies, 3*(1), 495–509. https://doi.org/10.1080/23752696.2018.1462099

Tadimalla, S. Y., & Maher, M. L. (2024). AI literacy for all: Adjustable interdisciplinary socio-technical curriculum. *2024 IEEE Frontiers in Education Conference (FIE)*, 1–9. https://doi.org/10.48550/arXiv.2409.10552

Tezer, M., Yildiz, E. P., & Uzunboylu, H. (2018). Online authentic learning self-efficacy: A scale development. *Quality & Quantity, 52*(1), 639–649. https://doi.org/10.1007/s11135-017-0641-1

Appendix: Survey Items

Ways of Thinking Section

1. I can produce alternative solutions to real problems.
2. I can present real-life problems.
3. I can improve my problem-solving skills.
4. I can be motivated.
5. I can relate the knowledge I acquired to real life.
6. I can find opportunities to express myself to others.
7. I can apply my knowledge to different subject fields.
8. I can relate new information to my old knowledge.
9. I can take different roles in learning environments.
10. I can easily and openly attend to learning activities.
11. I suggest ideas for others to explore how the job should be done.
12. I suggest new ideas for solving problems.
13. I suggest new practical solutions to reach a goal.
14. With my competence, I help the group achieve its goals.
15. I offer new ideas to make them available to others.
16. I can steer the group towards the goal.

Ways of Working Section

1. I can construct information by cooperating with other learners.
2. I can create social bonds with people around me (friend, teacher, specialist, etc.).
3. I can define myself as a member of the group I am in.
4. I can work in harmony with my fellow group members.
5. I can acquire advantages like brainstorming and discussing ideas in collaborative environments.
6. I can keep up with new technological advancements by using social networks (YouTube, Wikipedia, Facebook, etc.).
7. I can keep up with new technological advancements by using online communication tools (WhatsApp, Zoom, Skype, Gmail, etc.).
8. I can solve the problems more easily when I work with other learners.
9. I can reflect on what I have learned both individually and collaboratively with activities.

Authentic Learning Experience Section

1. I can learn better by interacting in real life.
2. I can understand how the information is used in real life.

3. I can apply the skills and knowledge I have acquired in real life.
4. I can face complex tasks.
5. I can come up with solutions to complex tasks.
6. I can acquire rich learning experiences.
7. Through the activities, I am able to analyse and critically consider what I have learned, which helps me understand and organise my knowledge.
8. Through the activities, outstanding students receive recognition for their performance, which helps motivate me to perform better.
9. I feel that through each activity, I am able to ascertain areas in which I need further personal development and learning.
10. Each activity allows everyone to participate together, which allows students to provide each other with recommendations.

Bridging Learning and Impact: OpenImpactLab for Developing AI Readiness, Literacy and Cognitive Skills

Liang QIAN[1], Li-Ta HSU[1], Max Jwo Lem LEE[1], Ruoyu CHEN[1], Xiaoyu LUO[2], Yu Jen LIN[3]

[1] Department of Aeronautical and Aviation Engineering, The Hong Kong Polytechnic University, Hong Kong, China

[2] Educational Development Centre, The Hong Kong Polytechnic University, Hong Kong, China

[3] CS Equity AI, The United States of America

Abstract

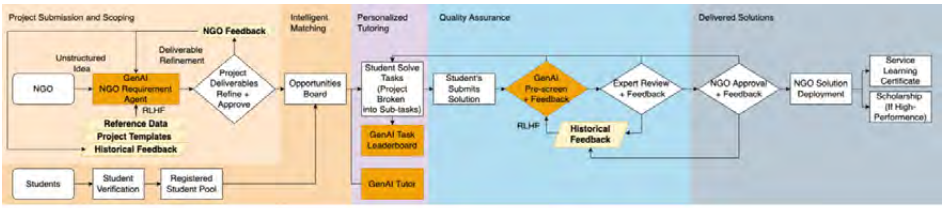
As Generative AI (GenAI) reshapes higher education, there is a pressing need for pedagogies that move beyond tool usage to foster AI literacy and cognitive development. This paper introduces *OpenImpactLab*, an AI-powered platform that connects students with NGOs to co-create social impact projects. We present a specialised GenAI curriculum embedded in the platform, "From Code to Cognition." The course follows a three-phase structure: (1) building AI readiness through Python foundations; (2) enhancing AI literacy via agent construction; and (3) fostering cognitive development through a capstone "AI Data Scientist" project. We report baseline data from 40 upper-secondary students (Pre-test) and outline the mixed-methods research design. Baseline results reveal a gap between high affective motivation and moderate technical readiness. We discuss expected outcomes and show how this ecosystem transforms learning from theoretical instruction to applied, experiential problem-solving.

Keywords: AI Literacy, AI Readiness, OpenImpactLab, Project-Based Learning, Cognitive Development

1. Introduction

The future of higher education demands a shift from theoretical instruction to applied, experiential problem-solving. In response, the *OpenImpactLab* project proposes an AI-powered platform aligned with Industry 4.0 readiness that connects students, educators, and Non-Governmental Organisations (NGOs) to co-create real-world social impact projects. Using Generative AI, the platform transforms complex NGO challenges into structured project missions and supports guidance, collaboration, and assessment throughout the learning process. Figure 1 illustrates this ecosystem.

Figure 1. *OpenImpactLab Platform Design: Key Components Highlighting the Cyclical Feedback Loop Between Students, NGOs, and AI Tutors*



However, successful participation in such projects requires specific competencies. AI literacy is increasingly defined as a set of competencies enabling effective and critical interaction with AI (Long & Magerko, 2020; Ng, Su, Leung, & Chu, 2024). In parallel, "AI readiness" refers to the prerequisite knowledge and dispositions for AI-rich work, including technical skills and psychological preparedness (Dai et al., 2020; Karaca, Çalışkan, & Demir, 2021). Project-Based Learning (PBL) has been identified as an effective approach for developing these competencies in AI education (Chiu et al., 2022). This paper describes the preparatory curriculum, "From Code to Cognition," embedded in *OpenImpactLab*, and reports the initial phase of an embedded mixed-methods study evaluating its impact on student development.

2. Methodology

This study utilises an Embedded Mixed-Methods Design to evaluate the educational intervention. The methodology comprises the pedagogical design of the curriculum and the research framework used to assess it.

2.1 Pedagogical Intervention: "From Code to Cognition" Curriculum Design

The course is structured into three phases, each mapping to a specific developmental goal (Table 1).

Table 1. *Phases of the Curriculum and Construct Alignment*

Phase	Project Form	Example Tasks	Main Construct Focus	Future OpenImpactLab Project Type
1	Structured micro-projects	Playlist analyser; Spam message classifier	AI Readiness (Python, data wrangling, basic ML)	Data cleaning; descriptive analytics
2	Medium-scale AI agent tasks	CSV QA agent; API-based data helper	AI Literacy (ABCE; critical GenAI use)	FAQ bots; internal data assistants
3	Capstone "AI Data Scientist"	End-to-end agent for a data problem	Cognitive Development (Decomposition, systems thinking)	NGO dashboards; decision-support agents

- **Phase 1: The Foundation (Building AI Readiness)**
  Rationale: This phase builds AI readiness through targeted, auto-gradable tasks. By mastering core code and data operations, students establish the technical floor for later complexity.
- **Phase 2: The AI Agent Lab (Enhancing AI Literacy)**
  Rationale: Through open-ended exploration and component-based construction, students deepen AI literacy by moving beyond treating AI as a "magic box" to understanding inputs, context windows, and tool calls.
- **Phase 3: The AI Data Scientist (Fostering Cognitive Development)**
  Key milestones include model retrieval, code generation, and refinement/debugging. This capstone promotes cognitive development by requiring problem decomposition and systems thinking as students integrate LLMs with Python runtimes.

2.2 Research Design

Research Questions:

- **RQ1 (AI Readiness):** How does the course affect students' technical skills and AI readiness?
- **RQ2 (AI Literacy):** How does the curriculum influence students' affective, behavioural, cognitive, and ethical AI literacy?
- **RQ3 (Cognitive Development):** What evidence of problem decomposition, systems thinking, and metacognition emerges from the capstone projects?

Participants: The study involves 40 upper-secondary students (M_age = 15.78), serving as a pilot cohort for university-level implementation.

2.3 Operationalisation and Assessment Framework

To ensure the rigour and replicability of the study, the core components of the pedagogical intervention are operationalised as follows:

- **Capstone Performance Assessment:** The Phase 3 "AI Data Scientist" project is evaluated using a multi-dimensional rubric. This assesses: (1) Technical Execution (code accuracy and API integration); (2) Systems Thinking (the integration of LLMs with Python runtimes); and (3) Requirement Alignment (the efficacy of the solution in addressing the specific social impact needs of the partner NGO).
- **Process Data Analysis:** We analyse student development through Git metadata. Specifically, we track Commit Granularity to evaluate problem-solving decomposition skills and Debugging Loops (the temporal interval between system errors and corrective commits) to assess metacognitive persistence and self-correction strategies.

2.4 Data Collection

Data is collected across a timeline:

- **Pre-Course:** Demographic survey, AI Readiness Test (10 items), and the AI Literacy Questionnaire (AILQ, 31 items) adapted from Ng, Wu, Leung, Chiu, and Chu (2024).
- **In-Course:** Project process data (Git commits), code documentation, and reflective journals.
- **Post-Course:** Post-test (Readiness), Post-survey (AILQ), semi-structured interviews, and focus groups.

2.5 Data Analysis

- **Quantitative:** Paired sample t-tests will compare pre- and post-test scores.
- **Qualitative:** Thematic analysis will be applied to interview transcripts and project artefacts.
- **Triangulation:** Quantitative scores will be cross-referenced with qualitative evidence to validate whether self-reported literacy increases match demonstrated cognitive depth.

Quantitative data are analysed using descriptive statistics and paired-sample t-tests to measure growth. For qualitative data, a thematic analysis is conducted on reflective papers. The study employs a data transformation strategy for mixed-methods integration: qualitative themes regarding coding behaviours are cross-tabulated with quantitative literacy scores. This triangulation validates whether self-reported literacy gains align with the technical and cognitive depth demonstrated in the students' final artefacts.

3. Results and Discussion

3.1 Baseline Findings (Pre-Test Data)

Currently, baseline data has been collected for 40 students. Table 2 summarises the descriptive statistics.

Table 2. *Baseline AI Readiness and AI Literacy Scores (N = 39–40)*

Measure	N	Mean (M)	SD	Range
AI Readiness (0–10)	40	6.50	2.29	1–10
AI Literacy (1–5)				
Affective	39	4.03	0.58	2.00–5.00
Behavioural	39	4.03	0.60	2.12–5.00
Cognitive	39	3.92	0.61	2.17–5.00
Ethical	39	4.69	0.39	3.75–5.00

- **AI Readiness:** The mean score was 6.50/10. Students showed competence in basic logic but struggled with API concepts and data structuring—skills critical for the *OpenImpactLab* context.
- **AI Literacy (ABCE):** Students reported high Ethical awareness (M = 4.69) and Affective motivation (M = 4.03). However, correlation analysis revealed a weak relationship between Readiness and Literacy (r ranges from .13 to .29), indicating an "illusion of competence" where motivation exceeds technical ability. This necessitates the structured approach found in our curriculum to bridge the gap.

3.2 Expected Outcomes and Discussion

Based on the research design and pilot observations, we anticipate the following outcomes upon study completion:

- **Addressing RQ1 (Readiness):** We expect quantitative post-test data to show significant improvements in technical scores, validating Phase 1's efficacy in establishing the necessary technical "floor."
- **Addressing RQ2 (Literacy):** While self-reported scores are expected to rise, the qualitative data (interviews/demos) will likely reveal a qualitative shift in how students understand AI—moving from a vague, abstract perception to a concrete, structured mental model.
- **Addressing RQ3 (Cognitive Development):** We anticipate that the qualitative analysis of the "AI Data Scientist" capstone will yield rich evidence of cognitive growth. Specifically, we expect to see students actively applying systems thinking (integrating components), problem decomposition (following milestones), and metacognition (iterative debugging loops).

These expected findings will demonstrate that the *OpenImpactLab* model successfully bridges the gap between novice enthusiasm and the rigorous demands of real-world social impact projects.

3.3 Limitations and Future Work

Whilst this study provides essential baseline insights, it is currently more programmatic than empirical. As a "Work in Progress" submission, the conclusions regarding student development and the effectiveness of the "From Code to Cognition" curriculum remain preliminary. To validate the proposed impacts, the next phase of research will include:

- **Post-Course Comparative Analysis:** Utilising paired-sample t-tests to quantify growth from baseline AI readiness and literacy scores to final outcomes.

- **Empirical Artefact Analysis:** Applying thematic coding to Git metadata and capstone project artefacts to provide objective evidence of cognitive growth and systems thinking.

4. Implications

The integration of the "From Code to Cognition" curriculum with the *OpenImpactLab* platform offers significant implications for higher education institutions. Firstly, it validates the strategic focus on applied learning found in many modern educational initiatives. By using GenAI to break down complex NGO challenges into manageable missions, the platform provides the necessary scaffolding for students to contribute meaningfully to society.

Secondly, the study highlights the importance of diagnostic assessment. The baseline data revealed a gap between student interest and readiness. Higher education institutions must not assume "digital natives" are automatically "AI ready" (Dai et al., 2020). Curricula must explicitly design for the "Foundation" phase before expecting students to orchestrate complex AI agents.

Finally, this project demonstrates a scalable pathway. The curriculum acts as a gateway, equipping students with the specific literacies and cognitive skills needed to utilise the *OpenImpactLab*'s AI Tutor and Collaborative Dashboard effectively. This ensures that when students eventually engage with NGOs, they do so not just as learners, but as capable co-creators of social value.

Beyond the local pilot, these findings address critical international challenges in AI education. The identified "readiness-literacy gap" suggests that global institutions must move beyond treating AI as a standalone digital tool, instead adopting structured pedagogical models that bridge novice enthusiasm with rigorous computational foundations. The *OpenImpactLab* ecosystem serves as a scalable blueprint for the "AI for Social Good" (AI4SG) agenda, demonstrating how universities worldwide can foster civic engagement whilst developing Industry 4.0 competencies. By aligning structured scaffolding with real-world NGO needs, this model provides a pathway for global higher education to standardise AI literacy frameworks and democratise AI creation in diverse geographical contexts.

Ethical Declaration Statement

The ethics approval process for this project is currently in progress.

Generative AI Policy: Declaration of Use of AI in the Writing Process

Statement: The author(s) used ChatGPT/Gemini during the preparation of this work to assist in structuring the argument and refining the language. After utilising the tool, the author(s) thoroughly reviewed and edited the content as necessary and assumed full responsibility for the publication's content.

References

Chiu, T. K. F., Meng, H., Chai, C.-S., King, I., Wong, S., & Yam, Y. (2022). Creation and evaluation of a pretertiary artificial intelligence (AI) curriculum. *IEEE Transactions on Education, 65*(1), 30–39. https://doi.org/10.1109/TE.2021.3085878

Dai, Y., Chai, C. S., Lin, P. Y., Jong, M. S. Y., Guo, Y., & Qin, J. (2020). Promoting students' well-being by developing their readiness for the artificial intelligence age. *Sustainability, 12*(16), Article 6597. https://doi.org/10.3390/su12166597

Karaca, O., Çalışkan, S. A., & Demir, K. (2021). Medical artificial intelligence readiness scale for medical students (MAIRS-MS) – Development, validity and reliability study. *BMC Medical Education, 21*, Article 112. https://doi.org/10.1186/s12909-021-02546-6

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–16). Association for Computing Machinery. https://doi.org/10.1145/3313831.3376727

Ng, D. T. K., Su, J., Leung, J. K. L., & Chu, S. K. W. (2024). Artificial intelligence (AI) literacy education in secondary schools: A review. *Interactive Learning Environments, 32*(10), 6204–6224. https://doi.org/10.1080/10494820.2023.2255228

Ng, D. T. K., Wu, W., Leung, J. K. L., Chiu, T. K. F., & Chu, S. K. W. (2024). Design and validation of the AI literacy questionnaire: The affective, behavioural, cognitive and ethical approach. *British Journal of Educational Technology, 55*(3), 1082–1104. https://doi.org/10.1111/bjet.13411

Critical Reading in the AI Era: Do Summarisation Tools Hinder Deep Reading for EFL Learners?

SUN Zhuo [1] & ZHU Dianyong [1]
[1] School of Foreign Studies, Harbin Engineering University, China

Abstract

As GenAI integrates into higher education, its efficiency in summarising texts raises concerns about cognitive offloading and the atrophy of EFL learners' deep reading skills. This study empirically investigates how AI-generated summaries affect surface-level factual recall versus deep-level critical inference. Fifty TEM-4 undergraduate English majors from Harbin Engineering University were randomly assigned to a ChatGPT-assisted or a traditional reading control group. While the AI group completed the reading task in half the time and excelled at surface fact retrieval, they performed significantly worse on deep critical analysis tasks requiring metaphor interpretation and logical evaluation. Qualitative findings reveal that over-reliance on AI summarisation fosters an "illusion of competence," reducing the necessary cognitive effort for high-order thinking. To counter this, we propose a three-step post-reading comparison strategy— independent reading, structured AI assistance, and critical comparative evaluation—to preserve and develop critical reading skills in the AI era.

Keywords: generative AI, deep reading, EFL learners, cognitive offloading, critical thinking

1. Introduction and Literature Review

In the contemporary academic context, university students face immense pressure to process vast volumes of complex literature. Traditional academic reading requires not only solid vocabulary and syntactic decoding skills but also the ability to engage in higher-order critical thinking. Recent guidelines from UNESCO (2023) explicitly state that in the digital age, educational systems must staunchly defend and cultivate students' critical thinking capacities while integrating digital literacy. Historically, higher education's pursuit of high-level text comprehension fostered the systematic cultivation of deep reading skills. Renowned reading neuroscientist Maryanne Wolf (2018) defines deep reading as a slow, immersive cognitive process involving not just basic text decoding, but complex inferential reasoning, critical analysis, self-reflection, and empathy.

However, the proliferation of GenAI tools, such as ChatGPT and specialised reading assistants, is fundamentally reshaping this cognitive landscape. These tools empower students to instantly generate long-text summaries, extract core arguments, and simplify complex academic syntax, ostensibly maximising learning efficiency. While the leap in efficiency is evident, the underlying cognitive risks have sparked widespread academic debate. The theory of cognitive offloading in cognitive psychology posits that when humans become accustomed to relying on external tools to process, store, or organise information, they unconsciously reduce their own internal cognitive engagement (Risko & Gilbert, 2016). In the specific context of academic reading, this raises a critical question: does the extensive use of AI summarisation tools downgrade academic reading into superficial information scanning? Researchers worry that if students habitually bypass the necessary intellectual effort of decoding complex texts, they may lose the opportunity to develop higher-order critical thinking skills.

To fill the empirical research gap regarding the impact of GenAI in authentic reading scenarios, this study aims to answer two core research questions through a comparative experiment. First, compared to traditional close reading, does the use of AI-generated summaries differentially affect learners' accuracy in surface factual recall versus deep logical inference and critique? Second, to what extent does the use of summarisation tools alter the time efficiency of the reading process, and does this increased efficiency come at the expense of contemplative depth?

2. Methodology

2.1 Participants and Design

This study selected 50 undergraduate English majors from Harbin Engineering University. All participants had passed the Test for English Majors Band 4 (TEM-4), ensuring an upper-intermediate to advanced English foundation and minimising the confounding influence of basic linguistic competence on observed comprehension differences. The 50 participants were randomly and equally assigned to two homogeneous conditions: the AI-assisted group and the traditional control group (n = 25 per group). Prior to the main experiment, a 15-minute critical reading pre-test was administered to establish baseline equivalence. The pre-test material consisted of a 300-word argumentative excerpt from Johann Hari's article "Your attention didn't collapse. It was stolen" (*The Guardian*), a text comparable in difficulty and structural complexity to the main experimental passage. Participants answered two short-answer questions (5 points total) requiring them to infer the author's implied attitude and analyse a core metaphor, without access to any digital assistive tools. Two independent

raters scored the pre-test responses, yielding high inter-rater reliability (Pearson $r = .869$, $p < .001$). An independent-samples t-test on the pre-test scores revealed no statistically significant difference between the prospective control group ($M = 3.21$, $SD = 0.74$) and the AI-assisted group ($M = 3.07$, $SD = 0.73$): $t(48) = 0.674$, $p = .504$. This baseline equivalence supports the interpretation that subsequent comprehension differences reflect the effect of the intervention rather than pre-existing ability differences.

2.2 Materials Selection and Development

The main reading passage was Nicholas Carr's seminal argumentative essay "Is Google Making Us Stupid?" (*The Atlantic*, 2008). The article is rhetorically rich, and its focus on how technology reshapes cognition forms a meta-textual resonance with the experiment's purpose. To fit the 40-minute session, the original text was truncated to approximately 1,150 words, retaining the key rhetorical structures including the opening metaphor, the neuroplasticity argument, and the concluding philosophical reflection. A reading comprehension post-test (10 points total) was designed across two dimensions. Section A (Surface Comprehension, 5 points) comprised five multiple-choice questions assessing students' ability to extract explicitly stated information. Section B (Deep Comprehension, 5 points) comprised two short-answer questions requiring critical inference beyond literal meaning, including metaphor analysis and logical evaluation of the author's central argument.

2.3 Experimental Procedure

To maximise ecological validity and accurately simulate authentic digital reading behaviours, the experiment was conducted via a professional online psychometric platform with strict backend time-tracking. All participants logged in concurrently during a pre-scheduled 40-minute session. The control group received a secure link to the original English text. The platform's anti-cheating mechanism disabled copy-pasting functions, and participants were instructed to perform traditional close reading without any digital assistive tools. The AI-assisted group received the same text but were provided with a standardised set of three authorised prompts on their landing page: (1) "Summarise the core arguments in 200 words," (2) "Extract the paragraph outlines," and (3) "Explain the literal meaning of [complex sentence]." Participants were mandated to use only these predefined prompts in a parallel ChatGPT window. This controlled intervention design mitigated variance arising from heterogeneous prompt-engineering proficiencies, whilst maintaining consistency of AI support across the experimental group. Direct input of post-test question stems was strictly prohibited.

2.4 Data Processing and Analysis

The platform automatically recorded each participant's time on task. Quantitative data were imported into IBM SPSS Statistics and analysed using independent-samples *t*-tests. For the short-answer items in Section B, two independent raters employed thematic analysis to qualitatively code students' written responses, enabling characterisation of differences in depth of critical engagement.

3. Results and Discussion

3.1 Reading Efficiency: Algorithmic Acceleration at the Expense of Contemplation

Quantitative analysis revealed a substantial divide in reading efficiency. The AI-assisted group completed the reading task in a mean of 14.45 minutes (SD = 1.71), approximately half the time of the control group (M = 26.11 min, SD = 2.74). This difference was highly statistically significant ($t(48) = 18.07$, $p < .001$). These data confirm that GenAI functions as a powerful efficiency amplifier: equipped with ChatGPT, students no longer needed to spend extended time parsing complex syntax, allowing rapid acquisition of the text's structural skeleton. However, while this acceleration saved time, it simultaneously reduced the space for pausing, re-reading, and contemplative engagement during the reading process.

Table 1: Independent Samples t-Test Results for Reading Efficiency and Comprehension Variables

Variables	Control Group (n=25)	AI-Assisted Group (n=25)		t	p
	M	SD	M	SD	
Time on Task (mins)	26.11	2.74	14.45	1.71	18.07	< .001
Surface Comprehension (Score A, /5)	4.08	0.70	4.64	0.49	−3.27	.002
Deep Comprehension (Score B, /5)	3.98	0.60	2.16	0.53	11.29	< .001

Note: Two-tailed independent-samples t-tests, df = 48. Negative t-values for Surface Comprehension indicate higher scores in the AI-assisted group.

3.2 Surface Comprehension: AI as Cognitive Support for Fact Extraction

On the five-item multiple-choice surface comprehension test (Section A), the AI-assisted group scored significantly higher (M = 4.64, SD = 0.49) than the control group (M = 4.08, SD = 0.70): $t(48) = -3.27$, $p = .002$. This finding indicates that when processing explicitly stated information, ChatGPT functioned as a highly effective cognitive scaffold. AI-generated summaries accurately mapped key concepts and historical details, effectively compensating for the attentional lapses that commonly occur during long-text reading. For learning tasks centred on information retrieval, AI is thus a demonstrably effective assistive tool.

3.3 Deep Comprehension: Overconfidence and the Absence of Critical Thinking

In the deep comprehension section (Section B), however, the data exhibited a sharp reversal. The control group significantly outperformed the AI-assisted group on tasks requiring metaphor analysis and logical critique (Control: M = 3.98, SD = 0.60; AI-assisted: M = 2.16, SD = 0.53; $t(48) = 11.29$, $p < .001$). Qualitative thematic analysis of students' written responses illuminated the cognitive mechanism underlying this gap. Control group responses exhibited a quality of productive intellectual effort: participants worked through interpretive ambiguity, capturing not only the literal contrast in the author's metaphors but also the implicit emotional and rhetorical tensions. Conversely, while AI-assisted group submissions were grammatically fluent and logically coherent, they were highly homogenised and lacked interpretive depth; the AI had filtered out the rhetorical nuances and emotional texture of the original text. These findings are consistent with the hypothesis that over-reliance on AI summaries fosters unwarranted overconfidence in text comprehension—an "illusion of competence" (cf. Koriat & Bjork, 2005) in which students mistake the fluency of AI-generated paraphrase for genuine understanding. By delegating the demanding work of decoding and inference to the AI, students are deprived of the cognitive friction that constitutes the developmental substrate of critical reading skills (Wolf, 2018).

Table 2: Comparison of Representative Student Responses to the "Scuba Diver" Metaphor Dimension

Dimension	Control Group (Traditional Reading)	AI-Assisted Group (ChatGPT)
Cognitive Focus	Engaged with the emotional tone and implied conflict of the metaphor	Focused on extracting the literal meaning and generating a structural summary
Sample Response	"The scuba diver represents someone who disconnects from the noise to explore hidden meanings, symbolizing the patience required for deep thinking."	"Carr compares deep reading to scuba diving and skimming to jet skiing. The internet makes us act like jet skiers because we move too fast."
Criticality Level	High: identified the author's tragic stance on the loss of mental solitude and concern for irreversible neural change	Low: factually correct but superficial; failed to capture the author's implicit warning or emotional register

Note: Responses are lightly edited for readability and selected as representative of the dominant qualitative profile within each group.

4. Implications, Limitations, and Conclusion

4.1 Pedagogical Implications

Summarising the primary conclusion of this study, while GenAI significantly accelerates reading efficiency, its unstructured use among EFL learners inadvertently outsources cognitive processing, silently eroding the deep reading habits required for critical inference. As Baron (2021) has argued, digital mediums inherently tend to downgrade reading into information consumption—a tendency that GenAI undeniably exacerbates. In a global context, this poses a profound challenge. EFL learners are not merely acquiring a language; they are developing the cross-cultural competencies and critical thinking skills necessary for global citizenship and international discourse. If students uncritically defer to the often homogenised and culturally flattened summaries generated by AI, they risk losing the ability to critically engage with diverse global perspectives. Attempting a blanket ban on AI in higher education is, however, both unrealistic and contrary to the trajectory of technological development. The more productive response is to design instructional approaches that harness AI's efficiency advantages while preserving the cognitive engagement necessary for deep global literacy development.

4.2 Limitations and Future Research

Several limitations of this study should be acknowledged. First, the sam-

ple is restricted to a single institution and a specific English proficiency group (TEM-4 level); future research should incorporate larger, more diverse samples across different academic backgrounds and language proficiency levels to assess the generalisability of the findings. Second, the experiment captured only the short-term impact of AI tools on immediate comprehension outcomes. Delayed post-tests spanning weeks or months would be valuable in assessing whether extended reliance on AI produces lasting changes to students' critical thinking development. Third, while this study employed a standardised prompt set to control for heterogeneous prompt-engineering proficiency, future research should investigate how different types of AI prompting strategies—including analytical and Socratic questioning prompts—might differentially affect comprehension outcomes, and whether structured prompt training can mitigate the negative effects of cognitive offloading.

4.3 Conclusion

This study provides empirical evidence that AI summarisation tools, while demonstrably effective at enhancing reading efficiency and surface fact retrieval, are associated with significantly reduced performance on deep comprehension tasks requiring metaphorical interpretation and critical reasoning. The qualitative findings suggest that this deficit is mediated by an illusion of competence: AI-generated summaries create the impression of understanding whilst bypassing the cognitive engagement through which genuine critical literacy is developed. The proposed post-reading comparison strategy offers a provisional pedagogical response, though its efficacy requires validation through future intervention research. In an era in which AI can outperform humans at information extraction, the indispensable competency of academic reading is the capacity to think critically about what AI cannot fully represent: the rhetorical, affective, and argumentative texture of human writing.

Acknowledgments

This research was supported by the School of Foreign Studies, Harbin Engineering University. The authors thank the participants and the two independent raters for their contribution to this study.

Ethical Declaration Statement

This study received ethics approval from the School of Foreign Studies Research Ethics Committee, Harbin Engineering University. All participants provided informed consent prior to participation.

Generative AI Policy

The authors used ChatGPT-4o to assist in refining the grammatical accuracy of sections of the manuscript. All AI-assisted content was reviewed and edited by the authors, who assume full responsibility for the published work.

References

Baron, N. S. (2021). *How we read now: Strategic choices for print, screen, and audio*. Oxford University Press. https://doi.org/10.1093/oso/9780190084097.001.0001

Carr, N. (2008). Is Google making us stupid? *The Atlantic*, 302(1), 56–63. https://www.theatlantic.com/magazine/archive/2008/07/is-google-making-us-stupid/306868/

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

UNESCO. (2023). *Guidance for generative AI in education and research*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000386693

Wolf, M. (2018). *Reader, come home: The reading brain in a digital world*. Harper. https://www.harpercollins.com/products/reader-come-home-maryanne-wolf

Exploring the Benefits and Hesitations of Using GenAI for Students in Media Design

Jae-Eun OH [1], Qianhui SHEN [1], Tsz Ning TSANG [1], Yuet Kai CHAN [1]
[1] School of Design, The Hong Kong Polytechnic University, Hong Kong, China

Abstract

The rapid development of AI is well-known within the creative industry; however, the approach to using AI tools in innovative processes and design education remains untested. This study explores the benefits and hesitations of using GenAI for media design students. Focus groups were conducted and analysed with 33 Media Design students, respectively, after one lecture and seven lessons on GenAI. Most students had experience with GenAI before the workshops. After the workshops, students see AI as an assistant during the creative process: a brainstorming tool, an idea reorganiser, and a troubleshooter, resulting in a tailor-made study partner, while showing hesitations toward AI and obstacles to the creative process. Findings offer insights for educators and suggest that future research investigates the teaching approach for AI tools in the media design curriculum.

Keywords: Generative AI, Media Design, Design Education

1. Introduction

AI-generated visual content has seen widespread use in recent years for various purposes (Chaisatitkul et al., 2023). The creative industry has found that GenAI can be adapted to improve the creative process with fewer barriers (Caramiaux et al., 2019; Chaisatitkul et al., 2023). In the educational context, most studies focus on the benefits of GenAI. For instance, it is said to enhance project efficiency and provide an ideal learning path for individual students that suits their learning style and aesthetic (Jin et al., 2024; Lu & Wang, 2024).

Beyond the benefits of using GenAI tools, some researchers note that concerns arise when exploring the role of artificial intelligence in the creative field. These concerns revolve around ethical issues regarding copyright and authorship of AI-generated work (Ivchyk, 2024), as well as workers' displacement through AI automation tasks (Bender, 2024; Jin et al., 2024). Nevertheless, some scholars express concern about craftsmanship, noting a loss of originality (Enache et al., 2024). However, the authors of this study observed that students in Media Design rarely use GenAI tools in the creative process, and there are very few discussions about GenAI in the classroom. This study aims to bridge the knowledge gap regarding future users' concerns about AI tools by exploring

GenAI's role, benefits, and hesitations for students in Media Design. The findings aspire to provide educators with insights to enhance students' adaptability to advanced technology in the future of media design education.

To achieve these aims, this study was conducted to answer the following research questions:

- **RQ1:** What role do GenAI tools play in the creative process for Media Design students?
- **RQ2:** What are the benefits and hesitations of using GenAI tools in the creative process for Media Design students?

2. Methodology

2.1 Research Design and Participants

This study adopted a mixed-methods design, combining a quantitative pre-lesson survey with qualitative focus groups, to examine how Media Design students integrate GenAI tools into their creative process and conceptual thinking. A purposive sample of thirty-three Year 3 and Year 4 undergraduate students was recruited to ensure disciplinary consistency and familiarity with production workflows.

2.2 Intervention

To establish a shared foundation of GenAI literacy, the research team coordinated one lecture and seven hands-on sessions across the 2024/2025 academic year. Sessions were led by industry professionals who actively incorporate GenAI into media production, covering applications in 3D character animation, image and video generation, music generation, video editing (DaVinci Resolve 19), and industry-level AI use at Filmart (see Table 1).

Table 1. The delivery method and tools of GenAI in the media design lesson

No.	Lesson's Topic	Method	AI Tools Introduced
1	Introduction of AI in the media industry	Lecture	Explain the type of AI involved in each process of production.
2	AI in 3D Character Animation (Motion capture, 3D modelling)	Workshop *2	Tripo 3D, Meshy AI, Deep Motion, Wonder Dynamics, Rokoko
3	Image generation and video generation	Workshop *2	Leonardo.ai, Fooocus, Stable Diffusion (Comfy.ui), Kling
4	Music generation with AI	Workshop *2	Suno, Deepseek
5	DaVinci Resolve 19: AI-integrated video editing tools	Talk	AI-integrated video editing tools in DaVinci Resolve 19
6	AI in film (AI Hub in Filmart)	Site visit	cineDESK, Deepfake

Note: All workshops provided hands-on tasks. Sessions 5 and 6 were elective events open to voluntary attendance.

2.3 Data Collection

Before and after the sessions, participants completed a structured questionnaire (see Appendix 1) capturing their prior experience, frequency of use, tools employed, and barriers to adoption. Seven focus groups were then conducted, each guided by a semi-structured protocol (see Appendix 2) covering three domains: (1) students' reflections following GenAI instruction, (2) their independent AI use in media design production, and (3) their critical views on AI's potential role in film and animation. All focus group sessions were audio-recorded and transcribed verbatim.

2.4 Thematic Analysis Procedure

Transcripts were analysed thematically using MAXQDA 2020 following a six-phase inductive process. In Phase 1 (familiarisation), two members of the research team read all transcripts independently and recorded initial observations. In Phase 2 (initial coding), both researchers coded the data independently, generating descriptive codes directly from participants' language—for example, codes such as "speeds up research," "AI as idea generator," "aesthetic inconsistency," and "loss of creative ownership." In Phase 3 (code comparison and discrepancy resolution), the two coders convened to compare their code sets. Discrepancies—instances where the same utterance received different codes—were discussed and resolved through consensus, with reference back to the original transcript to confirm the intended meaning. In Phase 4 (theme development), related codes were grouped into candidate themes using MAXQDA's visualisation tools; codes sharing an underlying conceptual concern were merged into sub-themes (e.g., brainstorming, organising ideas, and technical assistance under the overarching theme of Benefits). In Phase 5 (theme review), themes were reviewed against the full dataset to confirm they were coherent, internally consistent, and grounded in the data. In Phase 6 (theme definition and naming), final themes and sub-themes were defined and named to accurately reflect their analytical content (see Table 2). To further ensure reliability and validity, researcher triangulation was employed throughout: coding decisions were cross-checked between the two independent coders before finalisation. Member-checking was additionally conducted by sharing a summary of preliminary themes with a subset of focus group participants, who confirmed that the themes accurately reflected their expressed views.

3. Results

3.1 Media Design Students' Usage of AI Tools Before the GenAI Tools Lessons

Out of 33 participants, 31 (94%) had prior experience with GenAI tools, while two did not. Most experienced users had used AI tools for between 6

months and 2 years; fewer than half reported frequent use (see Figures 1 and 2). The most common tool category was conversational AI, such as ChatGPT, used by the majority. Fourteen students (42%) had experience with AI art-generation tools like Midjourney and Adobe Firefly, whereas only two participants each had used AI video-generation tools or the AI research platform Perplexity.

Figure 1. Media Students' experience of AI tools

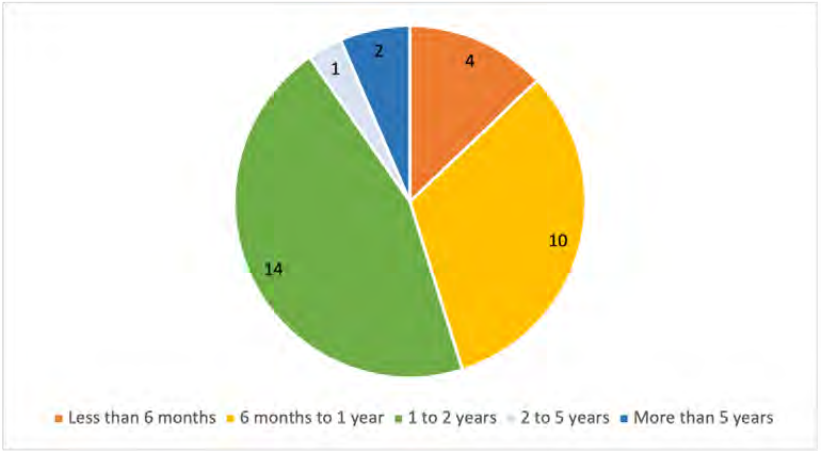
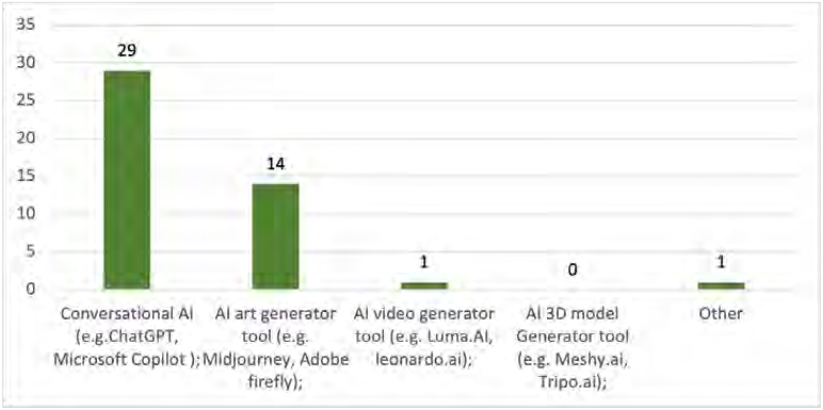


Figure 2. Media Students' usage of AI tools



These trends are intentional. The widespread use of conversational AI—despite the minimal adoption of AI for video and image creation—indicates a specific limitation within the domain: students see the value of language-focused AI for tasks like scripting and research but have not yet developed effective workflows for incorporating generative visual or video AI into their workflows. As noted in Section 3.2, this foundational understanding influences how students later assess the tools during the workshops.

The primary barriers to AI adoption among non-users were unfamiliarity with available tools and the perception that existing methods were sufficient. One participant also cited complexity and concerns about data privacy and security—a concern that, as we will see, resonated thematically among students who later used the tools.

3.2 Students' Impressions of GenAI Tools After Lessons

The focus group analysis revealed two main themes—Benefits and Hesitations—each with sub-themes based on participants' firsthand experiences with GenAI during production tasks. Importantly, these themes were interconnected and often contrasted, showing that students' relationship with these tools is complex and varies by discipline, rather than simply being acceptance or rejection.

3.2.1 Benefits: GenAI as a Creative Assistant in Pre-Production

Students consistently viewed GenAI as most helpful during the pre-production stage of media design, when they focus on gathering information, brainstorming ideas, and planning the structure before diving into the hands-on work. Three sub-themes characterise this role.

Brainstorming. Students described GenAI as a rapid, low-barrier entry point for gathering background information and perspectives at the start of a project. One participant explained: "I tried to use AI in the Character Design Assignment, and I asked AI about the views on housewives, and I received the result from AI's big data." Another noted: "AI tools can help when we need ideas and information to start a project quickly." This function—using GenAI to surface large bodies of contextual knowledge—was valued specifically because it compressed the early research phase, freeing up time for design iterations.

Organising Ideas. Beyond raw information, students found GenAI useful for structuring and expanding their own thinking. One participant articulated this dynamic particularly clearly: "Sometimes I have messy thoughts, but AI can help me contemplate things I might not have considered before, and it helped me organise my previously chaotic thoughts. Although the results were sometimes too far-fetched, I can adapt them to my own way of thinking. It is a kind of inspiration." Another described using AI as a collaborative scenario planner: "I would ask AI to role-play a character and interact with it to see which scenarios could be adapted to the script." These quotes show that students are not simply accepting AI output passively; instead, they actively compare AI suggestions with their own creative ideas. This interaction is more like a back-and-forth dialogue than just consuming output.

Technical Assistance. Students also appreciated GenAI as readily accessible technical assistance for operating production software. One participant described: "I took a screenshot and asked AI what the next step should be when I want to achieve the style I want. AI provided me with some steps in speed." Another shared: "I directly asked ChatGPT, and I copied AI's answer and shared it with my groupmates to test it. The problem was resolved." In a field where software proficiency is integral to professional identity, students' use of GenAI to accelerate technical problem-solving reflects a pragmatic and self-directed learning strategy.

3.2.2 Hesitations: Obstacles Rooted in Discipline-Specific Values

While students acknowledged GenAI's utility in pre-production, their hesitations intensified when they tried to use AI-generated outputs as finished or near-finished creative material. These hesitations are not generic concerns about AI; they are grounded in what it means to practice media design.

Obstacle to the Creative Process. More than 80% of our focus group members considered AI-generated visuals aesthetically insufficient to meet their expectations. This was not just a matter of quality but also highlighted the gap between AI's uncertain output and the exactness needed in animation and film creation. One student noted: "The image quality that AI generated was fair, but not thrilling." Another was more pointed: "I do not think Visual GenAI is helpful for 2D animation; in fact, it might increase the time spent revising compared to drawing on my own." The consistency problem was widely cited: "From our experience using AI, it was inherently inconsistent." Students highlighted the labour paradox of working with GenAI in animation: "I think the key relies on prompting skill. You might spend much time resulting in useless images, but occasionally, some of the generated images would be ok," and "If I must go through the details again, I could just draw it myself. When it comes to animation, we probably need to generate hundreds of iterations to achieve the style we want." These concerns directly account for the low adoption of AI video and image tools seen in the pre-lesson survey: students did not avoid these tools due to ignorance; many had experimented with them and found the output-to-effort ratio insufficient for the precise requirements of animation and film production.

Ethics. Students raised principled concerns about the provenance of AI training data and the ethical implications for working artists. One participant articulated the structural issue: "Most artists have a negative attitude toward AI, as no one would like their secrets exposed and used. After undergoing AI data training, the concept of theft becomes irrelevant." Another observed: "I believe that people dislike AI primarily due to the actions of its users. Someone would

steal art from a famous artist trained and generate an image to post on social media but did not claim that the image was generated by AI." A third raised the unresolved legal dimension: "The biggest controversy still revolves around the issue of intellectual property rights—where exactly are they?" These concerns align with broader scholarly debates on copyright and authorship (Ivchyk, 2024), but students articulated them with a personal stake: as emerging designers who will produce original work and enter a labour market affected by AI displacement (Bender, 2024), these questions are not abstract.

Craftsmanship and Originality. The primary concern—particularly unique to media design education—centred on how GenAI might threaten craftsmanship and creative authorship. Students viewed art as a process-dependent and emotionally driven activity. "Art is about personal feelings and the process. If AI is used for everything and the process is overlooked, it raises the question of art's purpose." One student foregrounded the irreducible humanity of creative practice: "Even modern art these days, no matter how simple it may seem, is created by humans and involves human emotions." Another identified the core tension directly: "AI has taken the original creative process, which determines that the work is not based on humanity." These statements show that students aren't just concerned about quality or legality; they see GenAI's involvement in creation as a challenge to what it means to be a designer. In a field where mastering skills like drawing, compositing, and sound design is closely tied to professional identity, this concern is especially significant.

Table 2. Selective data & themes from focus group results

Theme	Sub-theme	Coding unit
Benefit	Brainstorming	"I tried to use AI in the Character Design Assignment, and I asked AI about the views on housewives, and I received the result from AI's big data."
Benefit	Brainstorming	"AI tools can help when we need ideas and information to start a project quickly."
Benefit	Organise ideas	"Sometimes I have messy thoughts, but AI can help me contemplate things I might not have considered before, and it helped me organise my previously chaotic thoughts. Although the results were too far-fetched sometimes, I can adapt them to my own way of thinking. It is a kind of inspiration."
Benefit	Organise ideas	"I would ask AI to role-play a character and interact with it to see which scenarios could be adapted to the script."
Benefit	Technical assistance	"I took a screenshot and asked AI what the next step should be when I want to achieve the style I want. AI provided me with some steps in speed."

Theme	Sub-theme	Coding unit
Benefit	Technical assistance	“I directly asked ChatGPT, and I copied AI’s answer and shared it with my groupmates to test it. The problem was resolved...”
Benefit	Technical assistance	“How to easily collect information about arson chemicals that can be found in everyday life, as we found asking others would inevitably lead to problems.”
Hesitations	Obstacle to the creative process	“The image quality that AI generated was fair, but not thrilling.”
Hesitations	Obstacle to the creative process	“I do not think Visual GenAI is helpful for 2D animation; in fact, it might increase the time spent revising compared to drawing on my own.”
Hesitations	Obstacle to the creative process	“From our experience with using AI, it was inherently inconsistent.”
Hesitations	Obstacle to the creative process	“I think the key relies on prompting skill. You might spend much time resulting in useless images, but occasionally, some of the generated images would be ok.”
Hesitations	Obstacle to the creative process	“If I must go through the details again, I could just draw it myself. ……. When it comes to animation, we probably need to generate hundreds of iterations to achieve the style we want.”
Hesitations	Ethics	“Most artists have a negative attitude to AI, as no one would like their secrets to be exposed and used. After undergoing AI data training, the concept of theft becomes irrelevant.
Hesitations	Ethics	“I believe that people dislike AI primarily due to the actions of its users.” Someone would steal art from a famous artist to train and generate an image to post on social media but did not claim that the image was generated by AI.”
Hesitations	Ethics	“The biggest controversy still revolves around the issue of intellectual property rights—where exactly are they?”
Hesitations	Craftsmanship	“Art is about personal feelings and the process. “If AI is used for everything and the process is overlooked, it raises the question of art’s purpose.”
Hesitations	Craftsmanship	“Even modern art these days, no matter how simple it may seem, is created by humans and involves human emotions.”
Hesitations	Craftsmanship	“AI has taken the original creative process, which determines that the work is not based on humanity.”

4. Discussion / Conclusion

This study makes a specific contribution to the field of Media Design education by surfacing the discipline-grounded tensions that undergraduate students experience when engaging with GenAI—tensions that generic accounts of AI in higher education do not capture.

The most significant finding is not that students find GenAI useful or that they have reservations about it: both are well-documented in adjacent fields. The novel contribution lies in the structure of that ambivalence and its roots in the epistemology of media design practice. Students in this study did not resist GenAI due to unfamiliarity; the pre-lesson survey confirmed that 94% had prior experience with AI tools before the intervention. Rather, their hesitations crystallised precisely through hands-on engagement with generative visual and motion AI—tools they had not previously used—and were articulated in terms specific to animation and film production: inconsistency across frames, the iterative cost of prompt refinement, the loss of the embodied making process, and the erasure of authorial identity in the final output.

This discipline-specific dimension advances the existing literature in two respects. First, it extends Enache et al.'s (2024) concern about originality by showing that media design students conceptualise craftsmanship not simply as a quality standard but as a process-constitutive value: the act of drawing, compositing, or sound designing is itself understood as the site of meaning-making, not merely a means to an end. Students' insistence that "art is about personal feelings and the process" and that "even modern art involves human emotions" indicates that their resistance to AI is not conservatism but a principled defence of what defines creative practice in their field. This represents a meaningful departure from findings in architectural education (Jin et al., 2024) and environmental design (Lu & Wang, 2024), where AI-assisted efficiency is more readily embraced, suggesting that the acceptability of GenAI in design education is modulated by the degree to which hands-on craft is central to disciplinary identity.

Second, this study reveals an asymmetric workflow: students are highly receptive to GenAI at the pre-production stage—for research, idea structuring, and technical problem-solving—but reject it as a generator of production-ready assets. This asymmetry has direct pedagogical implications. It suggests that media design curricula should explicitly distinguish between GenAI's role in the conceptual phase and in the craft execution phase and should provide students with frameworks for making that distinction consciously rather than defaulting to either wholesale adoption or wholesale rejection. The ethical concerns

students raised—particularly around data provenance, intellectual property, and the labour implications for practising artists—further underscore the need for a critically mediated pedagogical approach. Students entering the industry will encounter these questions professionally; equipping them to navigate them requires dedicated curriculum time for critical AI literacy alongside technical workshops.

Taken together, these findings argue that media design education should position GenAI as a phase-appropriate, critically mediated tool: valuable as a pre-production thinking partner, but requiring explicit pedagogical scaffolding around its limitations and its relationship to creative authorship. Future research should investigate specific instructional strategies—such as structured reflection tasks, portfolio documentation of AI-assisted versus hand-crafted work, or curriculum frameworks that embed AI ethics alongside production practice—that support students in developing a grounded and critical relationship with GenAI in media design.

Ethical Declaration Statement

The project has been approved by the PolyU Institutional Review Board (PolyU IRB) (or its Delegate) of The Hong Kong Polytechnic University. (Reference Number: HSEARS20240903006).

Generative AI Policy

The author(s) used Google NotebookLM during the preparation of this work to do general research and assist with focus group results. 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

Bender, S. (2024). Generative-AI, the media industries, and the disappearance of human creative labour. *Media Practice and Education*, 26(2), 200–217. https://doi.org/10.1080/25741136.2024.2355597

Chaisatitkul, A., Luangngamkhum, K., Noulpum, K. et al. The power of AI in marketing: enhancing efficiency and improving customer perception through AI-generated storyboards. *Int. j. inf. tecnol.* 16, 137–144 (2024). https://doi.org/10.1007/s41870-023-01661-5

Enache, I. C., Valter, N. E., Răduică, F. F., & Chivu, O. R. (2024). Integrating Artificial Intelligence into the Photographic Design Process. *Journal of Industrial Design and Engineering Graphics*, 19(1), 143–148. http://sorging.ro/jideg/

Ivchyk, V. (2024). Overcoming Barriers To Artificial Intelligence Adoption. *Three Seas Economic Journal*, 5(4), 14-20. https://doi.org/10.30525/2661-5150/2024-4-3

Jin, S., Tu, H., Li, J., Fang, Y., Qu, Z., Xu, F., Liu, K., & Lin, Y. (2024). Enhancing Architectural Education through Artificial Intelligence: A Case Study of an AI-Assisted Architectural Programming and Design Course. *Buildings*, 14(6), 1613. https://doi.org/10.3390/buildings14061613

Lu, M., & Wang, H. (2024). The practice and reflection of generative AI in the cultivation of aesthetic education in colleges and universities: centred on environmental design major. *MATEC Web of Conferences*, 395, 01021. https://doi.org/10.1051/matecconf/202439501021

OpenAI. (2019, November 5). GPT 2: 1.5B release. https://openai.com/index/gpt-2-1-5b-release/

Yadav, N., Sharma, L., & Dhake, U. (2023). Artificial Intelligence: The Future. *International Journal Of Scientific Research In Engineering And Management*, 07, 1–10. DOI: 10.55041/IJSREM27796.

Appendix (1) Pre-lesson AI Usage Research

Name of participant: _____

I have read and understand the above consent form. By clicking the "Next" button to enter the survey, I indicate my willingness to voluntarily take part in this study.

☐ Agree ☐ Disagree

Gender: ☐ Male ☐ Female

Which year are you in? ☐ Year 3 ☐ Year 4

Have you ever used any AI tools in your studies? ☐ Yes ☐ No

How often do you use AI tools?

☐ Rarely ☐ Sometimes ☐ Often ☐ Always ☐ Daily

What kind of AI tools did you use? (Multiple Choice)

☐ Conversational AI (e.g., ChatGPT, Microsoft Copilot)
☐ AI art generator tool (e.g., Midjourney, Adobe Firefly)
☐ AI video generator tool (e.g., Luma.AI, leonardo.ai)
☐ AI 3D model Generator tool (e.g., Meshy.ai, Tripo.ai)
☐ Other: _____

How long have you been using AI tools?

☐ Less than 6 months
☐ 6 months to 1 year
☐ 1 to 2 years
☐ 2 to 5 years
☐ More than 5 years

Why haven't you used AI before? (Multiple Choice)

☐ I haven't had the opportunity to try AI tools.
☐ I find AI tools are hard to access.
☐ I'm not familiar with how it works or its applications.
☐ I haven't felt the need to use AI tools as my current methods get the job done effectively.
☐ I haven't felt confident enough to start using them as AI tools are too complex for me.
☐ I find AI tools are not affordable to me.
☐ I have concerns about data privacy and security when using AI tools.
☐ I don't know which AI tools fit my production style.
☐ I never hear about AI tools.
☐ Other: _____

- End of questionnaire -

Appendix 2 Focus Group Questions

General: What is the AI using habit in this year? Any changes?

Section 1: Student's thoughts after attending lessons integrated with AI

1.1 How did the AI-integrated lessons change your understanding of AI's role in production?

1.2 Did AI enhance or hinder your learning experience? Why? In which way?

1.3 Would you like future lessons to incorporate AI? What specific AI features would you want to explore?

1.4 Additional thoughts on AI-integrated lessons.

Section 2: Student's off-lesson usage of AI in film / animation production

2.1 Apart from experiencing AI tools on lessons, do you explore/use AI tools in your creative process/production process?

2.2 If yes, what AI have you been using so far?

- **2.2.1** At which stages of your workflow do you apply AI?
- **2.2.2** What percentage of your creative process involved AI? Can you walk us through your creative process?
- **2.2.3** How does AI affect your creative process?
- **2.2.4** Does AI reply with unexpected answers? How can they inspire you? (Example, please.)
- **2.2.5** What motivates you to use AI more frequently? (Pros)
- **2.2.6** What hinders your experience on using AI? (Cons)

2.3 If no, what is the reason(s) affecting you not to use AI?

- **2.3.1** What would motivate you to start using AI tools?

Section 3: Critical thinking

3.1 How would you rate the accuracy of AI-generated content in your field? What factors determine this? (e.g., relevance, factual correctness, contextual understanding)

3.2 What criteria do you use to evaluate whether AI-generated content is truly useful? (e.g., time saved, creativity boost, error reduction.)

3.3 How do you assess the practical value of AI-generated content in your workflow? Does it align with your initial expectations?

3.4 **Verification Practices**

- Text: Do you routinely fact-check or edit AI-generated text (e.g., scripts, reports)? If so, how?
- Visuals: How do you validate the accuracy of AI-generated visuals (e.g., reference comparisons, expert review)?

-End of Focus Group Questions-

From Classroom to Career: Understanding AI Dependency and AI Literacy of Hong Kong University Students

Chun Yin LAM [1], Jonathan Man Lung KWOK [2], Macy Mei Chi WONG [1]
[1] College of Professional and Continuing Education, The Hong Kong Polytechnic University, Hong Kong, China
[2] Division of Business and Hospitality Management, College of Professional and Continuing Education, The Hong Kong Polytechnic University, Hong Kong, China

Abstract

This study examines perceptions of artificial intelligence (AI) in the workplace, students' dependence on AI, and how such dependence mediates the usefulness of AI in their academic studies and perceived work skills in future workplaces, as well as the moderating role of AI literacy. Indeed, global higher education and future workplaces are undergoing significant technological disruption, especially regarding AI, which has a widespread impact across industry sectors. This study fills an essential gap in existing literature by exploring AI literacy. Grounded in the Technology Acceptance Model (TAM) and Self-Efficacy Theory, this study adopts a cross-sectional design and collects data from a sample of 353 Hong Kong undergraduate students through online surveys. Results reveal positive relationships among perceptions of AI, dependence on AI, AI usefulness, and future work skills. Moreover, AI literacy significantly moderates these relationships. These findings highlight AI literacy as an essential factor and underscore its importance in curriculum design and employability.

Keywords: AI literacy, AI dependence, AI usefulness, Future work skills, Higher education

1. Introduction

The AI era is akin to a new form of electricity, becoming the central driver of change in higher education and across different industry sectors. According to Feuerriegel et al. (2024), generative AI is gaining prominence for generating new materials such as words and images, and, in turn, reshaping future student learning, career preparation, and workplaces. It is reshaping the expectations of graduating students regarding these AI skills, preparing them for future workplaces. Due to this transformation, many higher education institutions have launched AI courses as a compulsory requirement for graduates. It is believed

that AI integration can benefit students in their academic studies and future employability. Therefore, this research aims to examine how students perceive and interact with AI in educational and workplace contexts, and to consider whether students' AI literacy is an essential competency for the future workforce.

In an AI-driven future, this research proposes a mediated-moderated model for AI literacy research, as it remains in its infancy. Grounded in the Technology Acceptance Model (TAM) and Self-Efficacy Theory, this study aims to address these gaps by investigating how perceptions of AI in the workplace relate to dependence on AI and how this dependence influences two outcomes: the usefulness of AI in study and work skills in the future workplace. This study will further explore whether AI literacy moderates these relationships.

Therefore, this study addresses two overarching research questions:

1. How does dependence on AI mediate the relationship between perceptions of AI in the workplace and the usefulness of AI in study and the development of future work skills?
2. To what extent does AI literacy moderate the effects of AI dependence on the usefulness of AI in study and work skills in the future workplace?

While the Technology Acceptance Model (TAM) utilises both perceived usefulness and perceived ease of use to predict user behaviour regarding initial technology adoption, this study focuses on assessing functional utility, as universities have already provided various AI subscription services to students who widely use them. By situating the research in the Hong Kong context, this study advances the literature on higher education, AI readiness, and future workforce preparation.

2. Methodology

A quantitative survey method was employed, and participants were recruited using a convenience sampling approach via mass email invitations. A total of 353 valid responses were collected from students at a higher education institution in Hong Kong between 2024 and 2025. Participation was voluntary and anonymous. The overall response rate was approximately 2.26% based on the total number of students contacted. The institution has integrated AI coursework across all academic programmes so as to ensure all respondents have baseline exposure to AI literacy and application. This was done to enhance the relevancy and internal consistency of the sample.

All measures were drawn from previous research literature. Respondents were asked to rate items on a 7-point Likert scale (1 = strongly disagree to 7 = strongly agree).

- **Perception of AI in the workplace** (8 items), **usefulness of AI in study** (7 items), and **work skills in the future workplace** (7 items) were adapted from Rožman et al. (2023). Sample items for perception of AI in the workplace include: *"AI makes faster business decisions based on outputs from cognitive technologies."*
- Sample items for usefulness of AI in study include: *"I think that using AI would save me a lot of time in studying."*
- Sample items for work skills in the future workplace include: *"Analytical thinking and innovation are important skills to have in the workforce,"* and *"Active learning and learning strategies are important skills to have in the workforce."*

Dependence on AI was drawn from Morales-García et al.'s (2024) 5-item scale. Sample items include: *"I constantly need validation or feedback from AI systems to feel confident in my decisions*."

Finally, the **AILQ scale** (32 items) was adapted from Ng et al. (2024). The Ng et al. (2024) AILQ scale is designed to assess students' AI literacy development through the affective, behavioural, cognitive, and ethical (ABCE) framework.

- The **affective dimension** measures intrinsic motivation and self-efficacy. Sample items include: *"Learning AI is interesting,"* and *"I believe I can master AI knowledge and skills."*
- The **behavioural dimension** evaluates commitment and collaboration. Sample items include: *"I will continue to use AI in the future,"* and *"I try to work with my classmates to complete AI learning tasks and projects."*
- The **cognitive dimension** gauges lower- to higher-order skills. Sample items include: *"I know what AI is and recall the definitions of AI,"* and *"I can evaluate AI applications and concepts for different situations."*
- The **ethical dimension** examines students' critical understanding of responsible AI use. Sample items include: *"I understand how misuse of AI could result in substantial risk to humans,"* and *"I think that people should be accountable for using AI systems."*

Data Analysis

Data were analysed using Partial Least Squares Structural Equation Modelling (PLS-SEM) because this method is more appropriate than others for analysing an overall model with multiple latent variables and indicators simultaneously, where normal distribution is not required (Hair et al., 2017).

Measurement and Structural Models

We followed the guidelines suggested by Hair et al. (2017) to assess the measurement model. First, factor loadings, composite reliability (CR), and Cronbach's alphas were used to evaluate internal consistency. Table 1 shows that all items demonstrated satisfactory loadings, which were all above the recommended threshold of 0.70. Moreover, all constructs met the 0.70 criteria for CR and Cronbach's alpha, reflecting acceptable internal consistency.

Table 1 also indicates satisfactory results for convergent validity, as indicated by average variance extracted (AVE) values. All constructs exceeded the 0.50 benchmark for AVE values, thus indicating adequacy of item convergence within each construct. To assess discriminant validity, suggested by Henseler et al. (2015), we used the heterotrait-monotrait (HTMT) ratio of correlations. All constructs showed sufficient discriminant validity as all HTMT values were below the threshold of 0.90 (see Table 2).

After confirming the measurement model, we tested the structural model using PLS-SEM with 5,000 bootstrap resamples (Figure 1).

3. Results and Discussion

Overview of Hypotheses

Table 3 shows all path coefficients and significance levels, and Figure 1 visualises the results with all hypotheses. Indeed, perception of AI in the workplace was positively related to dependence on AI ($\beta = 0.450$, $p < 0.001$), supporting H1. Moreover, dependence on AI was positively related to the usefulness of AI in study ($\beta = 0.494$, $p < 0.001$) and work skills in the future workplace ($\beta = 0.242$, $p < 0.001$), supporting H2a and H2b. Regarding the indirect relationships, results also show that there were indirect relationships between perception of AI in the workplace and usefulness of AI in study ($\beta = 0.223$, $p < 0.001$) and work skills in the future workplace ($\beta = 0.109$, $p < 0.001$) via dependence on AI, offering support to H3a and H3b.

Finally, regarding the moderating effects of AI literacy, results show that AI literacy moderated the direct relationships between dependence on AI and usefulness of AI in study ($\beta = -0.098$, $p < 0.01$), and between dependence on AI and work skills in the future workplace ($\beta = -0.12$, $p < 0.01$), such that when AI literacy was high, the positive direct relationships were dampened, thereby offering support to H4a and H4b. Finally, for the mediated moderating effects of H5a and H5b, Table 4 shows that results also offer support for the mediated moderating relationships at different levels; specifically, AI literacy moderated

the indirect relationships between perception of AI in the workplace and usefulness of AI in study via dependence on AI, and between perception of AI in the workplace and work skills in the future workplace via dependence on AI.

Perception of AI in the Workplace, Dependence on AI, and Usefulness of AI in Study

Students who perceived AI as beneficial in workplace settings usually display stronger dependence on AI for personal and academic tasks (Rožman et al., 2023). These findings align with the TAM, which posits that repeated interactions with and task completion using AI will enhance perceived usefulness of AI in study and future work skills. Moreover, Self-Efficacy Theory can explain the reciprocal process of using AI that leads to positive outcomes, which can enhance the perceived usefulness of AI and self-efficacy.

Bandura's (1997) self-efficacy theory posits that one's self-efficacy is developed through mastery experience, vicarious experience, verbal persuasion, and physiological states. When individuals have confidence in their ability to attain a task, they are more likely to engage with it, persist through difficulties, and achieve better outcomes. For instance, students accumulate mastery experience using AI tools for academic tasks. These interactions boost their confidence in their technological capabilities. Vicarious experiences, such as observing peers using AI in other areas, such as integrating AI to assist their study routine, and receiving encouragement from the institution, further reinforce their self-efficacy. If students achieve a successful outcome, such as improved efficiency in completing the task with less effort while producing a consistent level of output, or resulting in better grades, this increases their self-efficacy, and in turn promotes further and more frequent usage of AI. They perceive it as a useful tool for achieving their goals. Eventually, this creates a positive reciprocal and self-reinforcing cycle. Indeed, successful AI use enhances self-efficacy; stronger self-efficacy promotes deeper engagement with AI; and repeated positive outcomes amplify perceived usefulness.

However, to support healthy reliance rather than overdependence on AI, higher education institutions should build students' self-efficacy in handling core academic work before AI use.

Dependence on AI, Usefulness of AI in Study, and Work Skills in the Future Workplace

Apart from the positive relationship between dependence on AI and the usefulness of AI in study, dependence on AI also predicted confidence in future work skills. Using AI, students tend to improve their data-handling, communi-

cation, and problem-solving skills, which are valuable in contemporary workplaces and better prepare them for future employment. Indeed, Santos-Jaén et al. (2025) state that exposure to AI increases perceived competence even when actual skill development may vary.

Moderating Role of AI Literacy

Our study further explores the moderating role of AI literacy, and it moderates the relationship between dependence on AI and both outcomes, namely academic and career-related outcomes. Drawing on Bandura's (1997) self-efficacy theory to explain an interesting finding—that is, high AI literacy reduced the strength of these relationships—a certain level of AI literacy does not simply increase the positive efficacy cycle but promotes balanced engagement, prevents over-reliance, and supports professional autonomy and competency.

As defined by its four aspects (to know and understand, to use and apply, to evaluate and create, and ethical use), students with high AI literacy possess stronger skills in evaluating the results created for them, leading to a clearer understanding of their capabilities and limitations. For example, they may use AI tools to generate points for an assignment, but then critically review, modify, refine, and connect the ideas using their own judgment. Therefore, even though they use AI often, their perceived usefulness stems from a rational evaluation of practical benefits rather than emotional dependence or insecurity.

In contrast, students with lower literacy doubt their own abilities and view AI as a useful tool to fill their skills gap. They may constantly seek validation from AI, ask the system to generate the whole answer for them to adopt, or feel anxious about not using it to complete academic tasks. Eventually, they rely on it for support, which reinforces perceived usefulness. These findings help to explain that AI literacy serves as a protective factor.

4. Implications

Despite the benefits of AI, increasing excessive dependence on AI requires careful institutional attention and proactive monitoring in global education and across industry sectors. Productive use occurs when AI supports learning by providing analysis or suggestions, while students still critique the result using their own judgment and make the final decisions. In this case, AI strengthens their abilities and professional skills rather than replacing them. The boundary into harmful dependence is crossed when its application transitions from augmentation to substitution, resulting in decreased self-confidence. In this situation, AI is no longer a tool to support learning and leads to weakened independent thinking skills and reduced personal autonomy.

Given the inevitable trend, global universities are advised to continue to establish guidelines for students to strengthen the ethical usage of AI and embed such training into their learning modules. A mandatory disclosure is required for all coursework to state whether they have used such tools. The key is to promote students' self-efficacy while preventing over-reliance. A formative assessment, such as a written critique towards an AI-generated essay, is a balanced way for students to evaluate the weaknesses of AI while applying the knowledge they have learnt throughout the course. Providing their own justification points for each idea can strengthen students' critical thinking and their behavioural competency.

Nonetheless, the targeted approach should view AI as a complement, rather than a substitute for human abilities (Morales-García et al., 2024). In an AI-dominated environment, universities must help to enhance students' AI skills, as well as emotional and psychological competencies; these work skills are needed in the future workplace.

Acknowledgements

N/A

Ethical Declaration Statement

The project has been approved by the CPCE Research Committee of the College of Professional and Continuing Education, The Hong Kong Polytechnic University (Reference Number: RC_ETH_H_439).

Generative AI Policy

The author(s) used Grammarly during the preparation of this work for grammar checking. 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

Bandura, A. (1997). *Self-efficacy: The exercise of control*. WH Freeman/Times Books/ Henry Holt & Co.

Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2024). Generative AI. *Business & Information Systems Engineering, 66*(1), 111-126. https://doi.org/10.1007/s12599-023-00834-7

Hair, J., Sarstedt, M., Ringle, C., & Gudergan, S. (2017). *Advanced Issues in Partial Least Squares Structural Equation Modeling*. Sage Publications. https://doi.org/10.1007/978-3-319-05542-8_15-1

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science, 43*(1), 115-135. https://doi.org/10.1007/s11747-014-0403-8

Morales-García, W. C., Sairitupa-Sanchez, L. Z., Morales-García, S. B., & Morales-García, M. (2024). Development and validation of a scale for dependence on artificial intelligence in university students. *Frontiers in Education, 9*, 1-7. https://doi.org/10.3389/feduc.2024.1323898

Ng, D. T. K., Wu, W., Leung, J. K. L., Chiu, T. K. F., & Chu, S. K. W. (2024). Design and validation of the AI literacy questionnaire: The affective, behavioural, cognitive and ethical approach. *British Journal of Educational Technology, 55*(3), 1082-1104. https://doi.org/10.1111/bjet.13411

Rožman, M., Tominc, P., & Vrečko, I. (2023). Building skills for the future of work: Students' perspectives on emerging jobs in the Data and AI Cluster through artificial intelligence in education. *Environment and Social Psychology, 8*(2). https://doi.org/10.54517/esp.v8i2.1670

Santos-Jaén, J. M., León-Gómez, A., Valls Martínez, M. d. C., & Llorente Muñoz, V. (2025). University students' perceptions of the impact of artificial intelligence in the business sector on their educational and professional development. *Education and Information Technologies*. https://doi.org/10.1007/s10639-025-13712-4

Appendices

Table 1: Results of the measurement model

Scale and items	Cronbach's Alpha	Composite reliability	AVE
Perception of AI in workplace (PAIW)	0.985	0.986	0.908
Dependence on AI (DAI)	0.942	0.944	0.812
Usefulness of AI in Study (UAIS)	0.952	0.953	0.776
Work Skills in Future Workplace (WFW)	0.982	0.982	0.903
AI Literacy (AIL)	0.993	0.993	0.819

Table 2: HTMT of the constructs

Constructs	AI Literacy	Dependence on AI	Perception of AI in Workplace	Usefulness of AI in Study
AI Literacy				
Dependence on AI	0.488			
Perception of AI in Workplace	0.827	0.465		
Usefulness of AI in Study	0.674	0.698	0.656	
Work Skills in Future Workplace	0.697	0.491	0.690	0.689

Note: Diagonal values represent correlations within constructs; off-diagonal values represent HTMT ratios between constructs.

Table 3: Results summary of the hypotheses

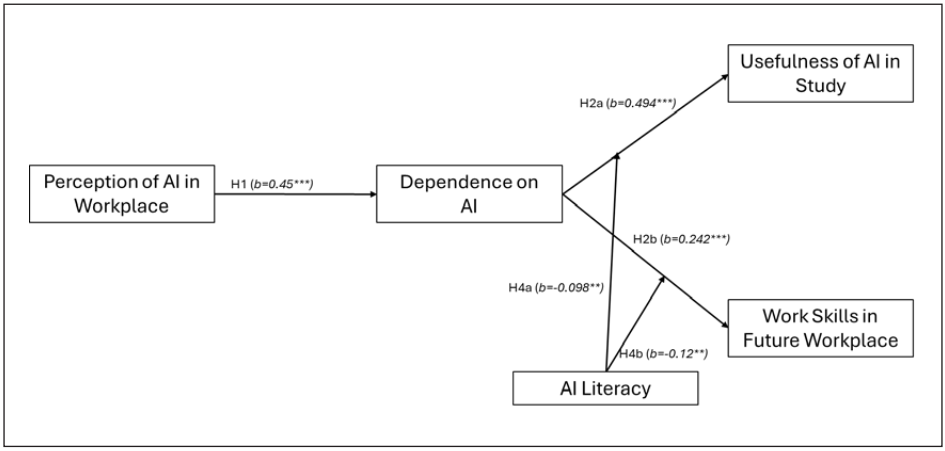
Path	Hypothesis	Path coefficients	t-statistics	p-values	Conclusion
PAIW → DAI	H1	0.450	9.340	0.000	Support
DAI → UAIS	H2a	0.494	9.953	0.000	Support
DAI → WFW	H2b	0.242	4.418	0.000	Support
PAIW → DAI → UAIS	H3a	0.223	6.248	0.000	Support
PAIW → DAI → WFW	H3b	0.109	3.769	0.000	Support
DAIxAIL → UAIS	H4a	-0.098	2.788	0.005	Support
DAIxAIL → WFW	H4b	-0.120	3.107	0.002	Support

Note: PAIW = Perception of AI in workplace; DAI = Dependence on AI; UAIS = Usefulness of AI in study; WFW = Work Skills in Future Workplace; AIL = AI Literacy.

Table 4: Results of the mediated moderating model

PAIW → DAI → UAIS			
Levels of moderator	**Coefficient**	**t-statistics**	**p-value**
High AI Literacy	0.178	4.998	0.000
Low AI Literacy	0.267	6.359	0.000
PAIW → DAI → WFW			
High AI Literacy	0.055	2.403	0.016
Low AI Literacy	0.163	3.889	0.000

Figure 1: Results of the model and hypothesis



Scaffolding Metacognitive Engagement with Generative AI: Design and Classroom Use of the 5P Framework

Chris HARWOOD [1]
[1] Faculty of Liberal Arts, Sophia University, Japan

Abstract

This paper reports on the design and classroom implementation of a structured critical AI interaction framework—comprising **Plan, Probe, Pause, Prove,** and **Ponder (the 5P)**—developed to scaffold metacognitive engagement with generative AI. The framework was embedded within a critical thinking course comprising 25 undergraduate students at C1 level (CEFR) at a private university in Japan. Adopting a design-based classroom action research approach, the study draws upon reflective journals and instructor observations to examine how students engaged with the framework in authentic academic contexts. Analysis suggests that the 5P model increased students' awareness of AI influence on reasoning and argumentation; however, engagement with the reflective stages was uneven and context-sensitive. During periods of high academic workload, students appeared to prioritise efficiency, selectively bypassing deeper metacognitive reflection. Rather than reporting learning outcomes, this study contributes pedagogical design insight into the tensions between AI literacy scaffolding, cognitive offloading, and self-regulated learning.

Keywords: Critical AI literacy; pedagogical design; self-regulated learning; metacognition; generative AI

1. Introduction

Generative AI tools are reshaping how students plan, draft, and evaluate academic work. While such tools offer efficiency and expanded access to linguistic and organisational support, concerns persist regarding cognitive offloading, surface-level learning, and reduced critical engagement. These concerns are particularly salient in courses that foreground reasoning, argumentation, and evaluative judgement. Prior research indicates that students often increase their reliance on AI under time pressure, prioritising task completion over reflective engagement (Abbas et al., 2024). From a self-regulated learning perspective, this presents a pedagogical challenge. Self-regulated learning involves learners' active control over planning, monitoring, and evaluation processes, which are fragile under cognitive load (Zimmerman, 2008). Digital learning environ-

ments can amplify this fragility, particularly when AI systems perform forms of cognitive work such as idea generation, structuring, and phrasing. Without explicit scaffolding, AI use risks shifting students from active regulation to strategic delegation. In response, this study presents the design and classroom implementation of a five-stage critical AI interaction framework: **Plan, Probe, Pause, Prove,** and **Ponder (the 5P)**. Rather than positioning the framework as a solution that guarantees critical engagement, the study examines how such a scaffold functions in practice and where its limits become visible under authentic academic conditions.

2. Pedagogical Design and Context

The 5P framework was designed as a structured sequence of metacognitive prompts guiding students before, during, and after AI use. Each stage targets a distinct regulatory function: goal setting and task framing (**Plan**), analytical questioning of prompts and outputs (**Probe**), perspective-taking and interruption of automation (**Pause**), verification and evidentiary checking (**Prove**), and reflective evaluation of learning and authorship (**Ponder**). The framework was integrated into a critical thinking course at a private university in Japan with 25 undergraduate students operating at C1 level (CEFR). Students were explicitly introduced to the framework and encouraged to apply it across AI-supported tasks. The study adopted a classroom-based, design-oriented action research approach. Data sources included weekly student reflective journals and systematic instructor observations (Burns, 2010). This design-based study prioritised qualitative evidence from journals and observations to examine engagement with the framework, rather than collecting quantitative outcome data or conducting inferential analysis. The aim was to examine how students adopted, adapted, or resisted the scaffold under real academic constraints such as workload, deadlines, and assessment pressures.

3. Results and Discussion

Student reflections indicate that the 5P framework functioned primarily as an awareness-raising scaffold. Students frequently articulated increased sensitivity to how AI shaped their reasoning, structure, and assumptions, often noting that such influence became visible only when prompted to reflect on their interaction process. In this sense, the framework supported **metacognitive noticing** rather than behavioural compliance. However, engagement across the framework was uneven. Early stages, such as *Plan* and *Probe*, were more readily adopted, particularly when aligned with task completion and productivity goals. In contrast, reflective stages, especially *Pause* and *Ponder*, were more likely to be abbreviated or bypassed. Students often described these stages as desirable but difficult to sustain under time pressure. Importantly, this selective

engagement was framed by students as strategic rather than uncritical. During periods of high workload, reflections emphasised efficiency and pragmatism, even while acknowledging the value of deeper reflection. This pattern aligns with research suggesting that metacognitive regulation is often the first component of self-regulated learning to deteriorate under pressure (Zimmerman, 2008). Rather than indicating a failure of the framework, these patterns suggest that the 5P model made visible the trade-offs students negotiate between efficiency and reflective engagement when working with AI.

4. Implications and Limitations

The findings suggest that structured AI literacy frameworks can increase metacognitive awareness but cannot override institutional and temporal pressures shaping student behaviour. The 5P framework offers a pedagogical design tool that surfaces these tensions explicitly, enabling instructors to address AI use as strategic decision-making rather than rule compliance. This design-based perspective may be particularly valuable in contexts where efficiency pressures must be critically negotiated rather than denied. As a design-based classroom study relying on reflective and observational data from a single course, the findings are context-specific and intended to inform pedagogical design rather than support generalisable claims.

Ethical Declaration Statement

The project adhered to the institutional research ethics of the participating university.

Generative AI Policy

The author used ChatGPT to support language refinement and structural clarity during manuscript preparation; all content was critically reviewed and revised, and full responsibility for the final text is assumed.

References

Abbas, M., Jam, F.A. & Khan, T.I. (2024). Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students. *International Journal of Educational Technology in Higher Education*, 21, 10. https://doi.org/10.1186/s41239-024-00444-7

Burns, A. (2010). *Doing Action Research in English Language Teaching: A Guide for Practitioners*. Routledge.

Zimmerman, B.J. (2008). Investigating Self-Regulation and Motivation: Historical Background, Methodological Developments, and Future Prospects. *American Educational Research Journal*, 45(1), 166–183. https://doi.org/10.3102/0002831207312909

7 Hybrid (Human AI) Intelligence

Discerning Learners, Human Intelligence, Artificial Tools: Learning History with Hybrid Resources in the Age of Generative AI

Lik Hang TSUI
Associate Professor, Department of Chinese and History
City University of Hong Kong

Abstract

This presentation provides a pedagogical approach that repositions generative AI as both a catalyst for historical reasoning and an object of critical analysis in history education. Drawing on a teaching development project implemented at undergraduate and postgraduate levels, it demonstrates how assignments can train students to verify, critique, and reverse-engineer AI outputs rather than passively consume them. This approach adopts a hybrid model combining digital tools—text repositories, collaborative annotation platforms, and LLM-powered chatbots—with material-rich resources including libraries, archives, and museum collections. Concrete teaching designs include AI-mediated reflective writing, historiography exercises treating chatbot outputs as secondary sources, and student-led podcast projects. The presentation will offer reusable resources—assignment templates, video guides, and workshop materials—that support teachers in integrating historical source criticism as a foundation for information literacy education. Implications for history pedagogy in an AI-saturated environment are also discussed.

Keywords: digital humanities, information literacy, generative AI, history education

1. Introduction

In an era increasingly shaped by generative AI technology, the fundamental value of humanities education is being discussed and re-examined. This presentation addresses the challenge directly, proposing a pedagogical paradigm for history education that leans into—rather than resists or ignores—the digital landscape. As AI increasingly masters content delivery, the focus must pivot from recall to inquiry in historical learning. At the core of this is the re-centering of historical reasoning and source criticism as a practical framework for analyzing claims, reflecting on evidence, and creating original work amid the deluge of algorithmically generated content that is impacting everyone, especially our university students.

This paper contributes a pedagogical framework that repositions generative AI as both a catalyst for historical reasoning and an object of critical analysis, demonstrating through undergraduate and postgraduate teaching designs how assignments can train students to verify, critique, and reverse-engineer AI outputs, rather than passively consume them. The project that this study is based on further offers reusable, adaptable resources—including AI-collaboration assignment templates, video guides, and training for podcast projects—that bridge classroom practice with public-facing dissemination and support teachers in integrating historical source criticism as the foundation for information literacy education in an AI-saturated environment.

2. Pedagogical Design and Classroom Practice

Drawing on outcomes from my teaching development project, "Truth be Told: Information Literacy for All with History Learning Aids," [1] I describe and implement a pedagogical approach that puts historical reasoning at the center of AImediated learning. I adopt a hybrid model that combines digital tools—digital text repositories and LLMpowered chatbots—with "lowtech" but materialrich resources such as libraries, physical books (including threadbound editions), museum exhibits and objects, and field trips. Within the augmentation of this blended learning environment for students, the project aims to move from a teacherled to a studentled approach: instead of merely receiving curated content, students are tasked with asking their own questions, designing ways to test AIgenerated claims, and choosing appropriate combinations of digital and physical sources to support their historical reasoning and arguments.

The project redesigns assignments and class activities so that AI becomes a catalyst for historical thinking rather than a substitute for it. Students are asked not only to place AIgenerated outputs alongside primary and secondary sources, but also to critique AI responses, iteratively refine their questions in dialogue with AI, and reverseengineer the likely evidence, proof, or references behind an AIgenerated text. Historical reasoning then becomes the organizing principle for these informationliteracy practices: I ask students to identify who or what is "speaking," what kind of data the response appears to draw on, what is omitted, and how claims could be tested or verified with other (historical) sources. Within this framework, LLMs (and especially chatbots) function as dialogic partners under explicit guardrails—such as prompts that require source tracing, routines for spotting bias and hallucination, and systematic citationverification—rather than as invisible ghostwriters.

---

[1] "Truth be Told: Information Literacy for All with History Learning Aids," Hong Kong UGC Teaching Award project grant, 2024-2026.

These principles are put into practice in the project in concrete teaching designs at both undergraduate and postgraduate levels. In an undergraduate course on cultural exchanges in a global context, I use an individual reflective assignment on the cultural biographies of things that deliberately builds AI into the workflow: each student first prompts a chatbot for a short, source-cited answer to a historical question about a chosen object, then treats that answer as raw material to be verified, reorganized, expanded, and rewritten in their own voice. Designed in this way, the task trains critical thinking, academic verification, and integrative abilities by asking students to turn an initially plausible but unreliable AI draft into a historically grounded reflection that draws on course readings, primary sources, and peer feedback. The grading criteria focus not on what AI generates, but on how students use the AI-generated responses to reconstruct argumentative frameworks, deepen and reorganize content, correct errors, integrate course readings and peer feedback, and perform rigorous scholarly checking—such as identifying and replacing fabricated references produced by AI.

In a postgraduate historical methods course, I ask students to understand AI historically and to evaluate it as both a source and a tool. To make this concrete, the assignments center on explicitly AI-co-created materials: students must treat chatbot outputs as if they were unfamiliar secondary texts, asking who is "speaking", what kinds of sources are implied, how training data and design choices shape the narrative, and how far such material can count as evidence. In this setting, they are required to answer, in very literal terms, "on what evidence can this claim be based?", and hence at the same time tracing, testing, and, where necessary, challenging the assertions embedded in AI-generated text or images. In practice, they begin from chatbot-generated overviews or draft paragraphs—such as ChatGPT's treatment of changing assessments of China's first emperor Qin Shihuang—and then test these responses against primary and secondary sources, reworking, correcting, and restructuring the text through close reading, comparison, and source checking. In both cases, I require students to make their AI use explicit and to justify how they move from chatbot outputs to historically grounded analysis, thereby demonstrating the project's paradigm in action.

As humanists, we live by stories, and within the project, I therefore designed a student-led group podcast as a way for students to practice historical storytelling. The assignment grows out of my courses and functions as a capstone that moves students from reading history to transmitting historically informed views to others. Working in small groups, students select themes and sources, define target audiences, and decide how to present their interpretations in accessible yet historically responsible ways, while I provide scaffolding in

the form of guiding questions and materials. They must think about how historical content "travels through sound": what kind of show to create, how much background a non-specialist listener will need, and how to weave documentary material into an engaging format. In one such podcast focusing on the history of the Silk Road, for example, a group used translated documentary sources from Hansen's book to script a short narrative following a merchant family's journey across Central Asia and China. Throughout, students draw on information-literacy habits cultivated elsewhere in the course—checking claims, locating and comparing sources, and situating their stories within broader historiographical conversations—while also practicing how to communicate these insights through narrative. The finished episodes, published in digital formats such as WeChat public accounts, stand as public-facing outcomes of the project that require students' careful reflection on how humans receive these cultural products, even in an age of AI-saturated content.

Alongside these narrative projects, I make systematic use of collaborative annotation tools to bring historical texts to life. Using platforms such as Hypothes.is on digital editions of canonical Chinese history sources such as the Shiji [Records of the Grand Historian], I guided students to jointly highlight, annotate, and discuss passages in the text in situ. This shared, layerbylayer close reading surfaces interpretive processes that are often invisible, giving quieter students greater agency to contribute, and turns online texts and web pages into spaces for slow, collective reasoning rather than rapid consumption. These practices sit together with other datadriven exercises in my courses—mapping locations, drawing timelines, or exploring digital biographies—to demonstrate how digital tools can expand students' access to curated historical resources while deepening their engagement with them. (Tsui, 2022)

Ultimately, across these diverse assignments, student work reveals a distinct evolution in historical reasoning. Moving beyond passive consumption, students consistently demonstrate three analytical shifts: (1) reverse-engineering algorithmic narratives to expose errors; (2) applying "traditional" source criticism to cross-verify claims against primary documents; and (3) treating AI outputs as raw baselines for evidentiary judgment rather than authoritative answers. A brief example from an undergraduate course exercise on co-writing and critiquing AI-generated narratives of cultural histories of everyday biographies captures this perfectly: tasked with tracing the history of tobacco, a student detected and dismantled a chatbot's Eurocentric, hallucination-filled draft, replacing fabricated citations with verified and actual scholarship on the history of transatlantic labor (see 5. Appendix). Such outcomes demonstrate that generative AI and digital tools, when properly constrained, actively sharpen students' source-critical skills.

3. Dissemination and Adaptation

The project further generates reusable resources that support immediate classroom adoption. One set of outputs is video learning guides, including a short popular-science video produced for the author's departmental "Chinese Culture Goes Virtual (雲上中華)" series, with videos titled "What makes a good digital humanities resource?" and so on. Rather than simply showcasing tools, the video series gives viewers clear methods for recognizing what counts as a good online resource, including how to distinguish between more and less reliable primary and secondary materials and so on. Students can use it independently as a step-by-step guide to evaluating digital resources for history learning and research in an AI-saturated environment.

In parallel, I have distilled the undergraduate and postgraduate AI-collaboration assignments described above into a set of shareable, ready-to-use teaching assignment designs. These illustrate how to embed AI explicitly into history coursework, require students to document and reflect on their AI use, and align assessment with source-checking, evidence-based argument, and critical evaluation of AI-mediated texts.

Building on this foundation, as mentioned, a sequence of student-led history podcast assignments on topics such as Dunhuang and the Silk Road, or otherwise on topics such as Eileen Chang's modern Chinese fiction. This allows students to "do" history into a more participatory and less textbook-bound form. By asking students to turn historically grounded analysis into short audio programmes for non-specialist audiences, the podcasts extend information-literacy and historical-reasoning practices into collaborative storytelling that makes historical knowledge vivid and accessible beyond the classroom, even when AI-generated content is so prevalent.

Beyond my own courses, I extend the project's methods and materials through workshops for teachers in Hong Kong and abroad. In Hong Kong, I collaborate with the Education Bureau, local secondary schools, and universities to run hands-on sessions for both secondary-schoolteachers and university instructors, including a joint "digital history" workshop with the Education Bureau on helping high school teachers learn Chinese history with AI and online resources. I also conduct similar workshops in universities overseas. Across these settings, participants work through examples that integrate digital text repositories, collaborative online annotation, mapping and timeline exercises, and carefully framed interactions with generative-AI tools. Rather than redesigning entire syllabi, we focus on the level at which change is most immediately possible: individual reading-intensive or materials-based tasks. Teachers

experiment, from the standpoint of their own disciplinary expertise, with adapting sample prompts, classroom activities and marking criteria to their specific institutional constraints and curricula and reflect on how to teach in ways that make students' handling of sources and AI outputs more transparent and more assessable. These workshops give teachers practical experience with digital tools, provide materials they can adapt for their own classrooms, and introduce strategies for treating AI outputs and online resources as "sources" that must be contextualized, questioned and verified.

I also use invited talks and panel presentations in Hong Kong and internationally to extend the teaching framework developed in the project. In research-oriented and teaching-and-learning venues, my "History and AI" lectures initially focus on the conceptual side of this framework: they argue for historical reasoning and source criticism as a practical backbone for information-literacy education in a GenAI-saturated environment, and show how assignments from my Hong Kong classrooms—such as AI-mediated reflective writing, historiography exercises that treat chatbot outputs as a new types of "secondary source" for historical research, as well as public-facing podcast and annotation projects—can be used to redesign course tasks and assessment towards inquiry, verification, and explicit reflection on AI use. Across these activities, I keep the emphasis on concrete practice: how to reconfigure assignments, structure AI-assisted reflection, and support information literacy through historically grounded tasks, rather than on speculation about technology or resistance to their use in history assignments.

4. Implications

The examples discussed here point towards a shift in how history assignments can be conceived in a GenAI-saturated environment. Instead of treating AI mainly as a threat to traditional essays, it is more productive to design tasks that make AI a study partner and an object of critical analysis for students. When students are asked to critique AI-generated responses, refine questions in dialogue with AI, and reverse-engineer the evidence behind AI-produced text, generative tools become a sharpening stone for historical reasoning and information literacy rather than a shortcut around them.

At the same time, collaborative digital tools and creative projects make history learning more participatory and outward-facing. Shared annotation platforms, digital maps, timelines and other visual tools open up different ways of encountering curated historical resources, while student-led podcasts and similar projects invite students to turn historically grounded analysis into an accessible narrative for public audiences. Assignments that require students to reflect explicitly on their use of AI and other digital tools encourage careful thinking about accuracy, attribution and voice.

Taken together, these practices show how historical reasoning can anchor information-literacy education: students learn to work with and against digital tools rather than simply accepting their outputs, to question what they see, and to use multiple media to communicate historically informed insights to wider audiences. However, these developments of course do not mean abandoning "low-tech" material practices. Physical books, libraries, archives, museum collections and field visits remain central to historical education and at the core of the skillsets of many history scholars and teachers; digital and AI-mediated work sits alongside them rather than replacing them. For students, historically grounded information-literacy training equips them to navigate AI-mediated, data-saturated environments with critical awareness, helping them judge online content, resist misinformation and make informed decisions in academic work and everyday life.

For the wider community, a project using historical methodologies for information literacy, such as "Truth be Told: Information Literacy for All with History Learning Aids", contributes reusable teaching resources, public-facing videos and podcasts, and school–university collaborations that connect history education with emerging digital practices. This integrated approach to AI use and information-literacy education fosters deep engagement and contemporary relevance while building core critical skills for an AI-augmented era. It preserves the rigour of humanistic academic practice and offers a resilient, adaptive path forward for historical inquiry that can harness generative AI.

5. Appendix—Sample Undergraduate Student Work on Co-writing and Critiquing AI Narratives: Individual Reflection Piece on Cultural Biographies of Everyday Objects

CAH2510 Cultural Exchange in Global Context
2023/24 Semester B
Individual Reflection Piece

Box A

Tobacco, a plant native to the Americas, has played a significant role in global cultural exchanges and has become integral to everyday life and culture. Its importance lies in its multifaceted uses across different cultures, its economic value, and its social implications.

1. Cultural Significance: In many indigenous cultures, tobacco was used for ceremonial and medicinal purposes long before it was commodified (Wilbert, 1987). With the arrival of Europeans in the Americas, tobacco was introduced to the Old World, where it was initially used as a medicinal herb before its recreational use became widespread (Gately, 2003).

2. Global Commodity: Tobacco's transformation into a global commodity was facilitated by the expansion of European colonial powers. As settlers established plantations in the Americas, tobacco became a significant export crop. The growth of the tobacco trade led to its widespread use across different cultures (Klein, 2015).
3. Economic Value: The economic value of tobacco significantly contributed to its importance. It became a key driver of the colonial economy, particularly in regions like Virginia and the Caribbean. The demand for tobacco also played a role in the development of the transatlantic slave trade (Cox, 2000).
4. Social Implications: Over time, smoking tobacco became a social activity, associated with leisure and sociability. It also became a symbol of status and identity in various cultures (Gilman & Zhou, 2004).
5. Health Impact: In the 20th century, the recognition of tobacco's adverse health effects led to changes in societal attitudes and government policies, highlighting the complex relationship between culture, commerce, and public health (Brandt, 2007).

In conclusion, tobacco's journey from a native American plant to a global commodity illustrates the profound impact of cultural exchange, economic forces, and social trends on everyday life and culture.

References:

1. Wilbert, J. (1987). Tobacco and Shamanism in South America.
2. Gately, I. (2003). Tobacco: A Cultural History of How an Exotic Plant Seduced Civilization.
3. Klein, J. S. (2015). The Empire of Tobacco.
4. Cox, N. (2000). The Complete History of Tobacco: Its Use, Trade, and Secret Formulas.
5. Gilman, S. L., & Zhou, X. (2004). Smoke: A Global History of Smoking.
6. Brandt, A. M. (2007). The Cigarette Century: The Rise, Fall, and Deadly Persistence of the Product That Defined America

Box B

Tobacco is one of the significant and representative Points that marked the success of the European Age of Discovery. It originated in the American continent, and its first cultivation was arguably in South America. Although Americans used it for different purposes centuries ago, it has become a more prevalent Point in global exchange since the expansion of European colonial power. The use of tobacco then varied and became a commodity integrated into people's daily lives. With different uses, it has also brought distinct implications such as social status or wealth. In the modern era, with advanced medical technology and knowledge, people have found that tobacco could harm our health. People's perceptions towards tobacco have changed from time to time. Therefore, this essay investigates tobacco's different uses in distinct cultures, its economic values under globalisation, its social implications, and the derived health concerns. It helps explore how the role of tobacco in society altered in different eras.

First of all, tobacco was initially a significant Point for the native Americans, mainly for ceremonial and medicinal uses. Before it was introduced to Europe and other continents, it had a tight correlation with ceremonies and rites in South America. According to Wilbert (1987)[2], tobacco Shaman was prevalent before it became a commodity in the 16th century. Tribes and religious practitioners would use tobacco for religious activities. Tobacco associated with Shaman was because the story of the mythical lord and the growth of tobacco were coincidently similar. Therefore, Shaman people used tobacco as an offering to communicate with the superficial spirits in the rites. Besides ceremonial uses, Americans also utilised tobacco as a medicine to heal minor and serious illnesses by chewing, eating, drinking, applying on wounds, or passive smoking (Gately, 2002)[3]. For instance, indigenous people would put tobacco leaves or juice on minor illnesses such as toothache or wounds bitten by snakes to relieve pain and cure wounds. As for major diseases, such as fever, people believed they were caused by evil spirits (Gately, 2002). Shamans would diagnose the illness by using tobacco smoke. It was also reckoned that passive smoking could protect and bless patients with a good supernatural force. The above examples illustrated that tobacco played an important role in American aboriginal culture, particularly in religious and medicinal aspects, before the arrival of Europeans.

Since the Age of Discovery, tobacco has been no longer a tribal Point but a commodity with economic values circulating in the global market. As European sovereigns wanted to expand their international influence and gain economic benefits from other undeveloped countries, colonisation occurred on the American, African, and Asian Continents, which was the New World. Tobacco was arguably a colonial product since European colonists found its medicinal value and the pleasure brought by smoking (Gately, 2002). With a rising demand for tobacco in Europe, British colonists started tobacco plantations in their American colonies, for instance, in Virginia, and sold them in Europe (Hahn, 2007). According to Morgan (1971), in the early 17th century, English people sold Virginian tobacco at a high price in their homeland, encouraging the colonised Americans to massively plant it. As it had a mounting economic importance, it even became a legal currency in North America. For instance, men could 'buy' their spouses by paying 'one hundred and twenty pounds of the best leaf tobacco' (Gately, 2002). Like other plantations in European colonies, slavery issues also existed. Native enslaved people were exploited and were forced to work for European landlords until Abraham Lincoln's presidency (Hahn, 2007; Pomeranz & Topik, 2006). Hahn (2007) stated that there were 31% enslaved Black people to the total population in Virginia in 1860. It illustrated that tobacco was a typical colonial commodity associated with slavery trading in the international market, and it brought considerable profit to European powers.

After the introduction of tobacco in Europe, it was predominantly used for leisure activities and entertainment. Tobacco use or ownership could correlate to one's social status and wealth. Taking England, the largest tobacco market in Europe in the 17th century, as an illustration, tobacco was sold in coffeehouses, where male guests were only permitted, mostly bourgeoises (Pomeranz & Topik, 2006). Not only did it offer a space for art and recreation, but also a 'democratic masculine

[2] Wilbert, J. (1987). Tobacco and shamanism in South America. Yale University Press.

[3] Gately, I. (2002). Tobacco: A Cultural History of How an Exotic Plant Seduced Civilization. New York: Grove Press.

sociability'. It allowed the middle class to exchange ideas concerning knowledge and politics without government control (Pomeranz & Topik, 2006). This example demonstrated that tobacco encouraged gatherings, conversations, and social connections (Gilman & Zhou, 2004)[4]. Besides, smoking in the Elizabethan era could imply wealth and economic power due to its exoticness. Since England had not yet had tobacco plantations in the Indies, they relied on capture, foreign trade, and smuggling (Gately, 2002). This circumstance accelerated its value with the increasing popularity of tobacco, and only people with extra income could buy it at that time. Hence, the consumption of tobacco revealed one's social class and wealth.

In contemporary times, although tobacco is still a prevalent commodity, social attitudes and perceptions have altered because of rising health awareness. In the mid-20th century, scientists and scholars discovered the negative health impacts of smoking tobacco (Gately, 2002). It was verified that heavy and long-term smokers have a ten times higher chance of having lung cancer than non-smokers, and the cancerous cells could spread to other organs, adversely affecting their functions. Apart from lung cancer, they also have a higher risk of suffering from various illnesses, not limited to heart disease, stroke, and diabetes (Gupta, 2001). Also, passive smokers, such as children, could be impacted by having asthma and other respiratory issues. (World Health Organisation, 2023). Since the public has learnt about the repercussions of smoking, there have been growing public negative attitudes and descending acceptability (Brandt, 2007)[5]. Meanwhile, institutions implemented different policies to curb the smoking population by having mandatory warnings on tobacco packages and imposing heavy tobacco tax (Wipfli et al., 2020). Dai, Gakidou and Lopez pinpointed that the global smoking population has shrunk since 1990, and more developed countries and regions have a particularly significant decline (2022). It exhibited the changes in social acceptance and popularity.

In conclusion, tobacco is a vital cultural Point and has had different roles in different eras. It first served for medicinal and ceremonial purposes in South American tribes. After the discovery of the Americas, tobacco was brought to Europe and other continents through transnational trading and colonial power expansion. Since then, it has become a popular commodity for entertainment and has brought distinct social implications. In recent decades, scientists found tobacco and its by-products have adverse impacts on the health of smokers and passive smokers, which discouraged the smoking population. People's attitudes in different eras towards tobacco have been highly correlated to the historical and social contexts, which helps create different cultural meanings.

[4] Gilman, S. L., & Zhou, X. (2004). *Smoke: a global history of smoking.* Reaktion Books.

[5] Brandt, A. M. (2007). *The cigarette century : the rise, fall, and deadly persistence of the product that defined America*. Basic Books.

Reference

American Historical Association. (2025, August 5). *Guiding principles for artificial intelligence in history education*. https://www.historians.org/resource/guiding-principles-for-artificial-intelligence-in-history-education/

Artificial intelligence and the practice of history: A forum. (2023). *American Historical Review, 128*(3), 1345–1389. https://doi.org/10.1093/ahr/rhad337

Brandt, A. M. (2007). *The cigarette century : the rise, fall, and deadly persistence of the product that defined America*. Basic Books.

Dai, X., Gakidou, E., & Lopez, A. D. (2022). Evolution of the global smoking epidemic over the past half century: strengthening the evidence base for policy action. *Tobacco control*, *31*(2), 129-137.

Fickers, A. (2012). Towards a new digital historicism? Doing history in the age of abundance. *Journal of European Television History and Culture, 1*(1), 19–26. https://doi.org/10.18146/2213-0969.2012.jethc004

Gately, I. (2002). *Tobacco: A Cultural History of How an Exotic Plant Seduced Civilization.* New York: Grove Press.

Gilman, S. L., & Zhou, X. (2004). *Smoke: a global history of smoking*. Reaktion Books.

Gupta, P. C. (2001). The public health impact of tobacco. *Current Science*, 475-481.

Hahn, B. (2007). Making Tobacco Bright: Institutions, Information, and Industrialization in the Creation of an Agricultural Commodity, 1617-1937. *Enterprise & Society*, *8*(4), 790–798. https://doi.org/10.1093/es/khm080

Jones, J. S. (2023, September 7). Students critique a ChatGPT essay: A classroom experiment. *Perspectives on History*. https://www.historians.org/perspectives-article/students-critique-a-chatgpt-essay-a-classroom-experiment-september-2023/

Morgan, E. S. (1971). The Labor Problem at Jamestown, 1607-18. *The American Historical Review*, *76*(3), 595-. https://doi.org/10.2307/1851619

Mollick, E. R., & Mollick, L. (2023). Using AI to implement effective teaching strategies in classrooms: Five strategies, including prompts. *Wharton School Research Paper*. https://doi.org/10.2139/ssrn.4391243

Pomeranz, Kenneth., & Topik, Steven. (2006). *The world that trade created : society, culture, and the world economy, 1400 to the present* (2nd ed.). M.E. Sharpe, Inc.

Tsui, L. H., Zhu, B., & Chen, J. (2023). Finding flexibility to teach the 'next big thing': Digital humanities pedagogy in China. In B. Croxall & D. Jakacki (Eds.), *What we teach when we teach DH: Digital humanities in the classroom* (pp. 274–291). University of Minnesota Press.

Tsui, L. H. (2022). Using digital humanities in the Chinese history classroom and online [邁向道器並重的努力：數碼人文和中國歷史教學]. In *History teaching in the twenty-first-century: Challenges and opportunities* [21世紀中國歷史教育的挑戰與機遇] (pp. 82–98). Chung Hwa Book Company.

Wilbert, J. (1987). *Tobacco and shamanism in South America*. Yale University Press.

Wineburg, S. (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology, 83*(1), 73–87. https://doi.org/10.1037/0022-0663.83.1.73

Wineburg, S. (2018). *Why learn history (when it's already on your phone)*. University of Chicago Press.

Wipfli, H., Bhuiyan, M. R., Qin, X., Gainullina, Y., Palaganas, E., Jimba, M., Saito, J., Ernstrom, K., Raman, R., & Withers, M. (2020). Tobacco use and E-cigarette regulation: Perspectives of University Students in the Asia-Pacific. *Addictive Behaviors*, *107*, 106420–106420. https://doi.org/10.1016/j.addbeh.2020.106420

World Health Organisation. (2023). *Tobacco.* https://www.who.int/news-room/fact-sheets/detail/tobacco

Acknowledgments: The research for this study is supported by Hong Kong's University Grants Committee Teaching Award project grant "Truth be Told: Information Literacy for All with History Learning Aids" (2024-2026).

Generative AI Policy

Statement: *The author(s) used ChatGPT and Claude during the preparation of this work to proofread the text and rephrase some expressions. After utilising the tool/service, the author thoroughly reviewed and edited the content as necessary and assumed full responsibility for the publication's content.*

Human–AI Collaboration in Writing Feedback: Exploring Hybrid Intelligence in Students' Writing Performance

Daniel FUNG [1] & Yu Hsuan WANG [1]
[1] Faculty of Education, The University of Hong Kong, Hong Kong, China

Abstract

Given that AI-generated feedback has become increasingly popular in education, this study examines whether it benefits all learners equally. While AI provides instant feedback, its impacts on students are not even. This study investigates how AI-assisted writing feedback affects students at different proficiency levels and how they perceive the role and nature of AI and teachers. Seventy-two secondary school students (aged 14–17) participated in an eight-week writing training, completed pre- and post-writing tests, and a questionnaire. Overall, stronger students improved more than weaker students, but they showed little gain in content development in their writing. Stronger students used AI in a diagnostic, strategic way, but weaker students tended to over-rely on it. Theoretically, this study refines the Hybrid Intelligence Framework by showing that engagement with AI feedback is heterogeneous and shaped by proficiency and feedback literacy. Practically, it calls for differentiated hybrid pedagogy with teacher training and curriculum redesign.

Keywords: Hybrid Intelligence, AI-Assisted Feedback, AI Literacy, Second Language Writing, Teacher Feedback

1. Introduction

Integrating Artificial Intelligence (AI) into education is no longer a futuristic concept; it is a dynamic reality actively reshaping teaching and learning. From diagnosing textual errors to identifying writing problems and offering suggestions (Alharbi, 2023), AI has moved from science fiction speculation to a central actor in classrooms worldwide. Despite increasing enthusiasm for AI in education, fully automating feedback remains controversial, as many AI platforms struggle with emotional intelligence and making human-like judgments to accommodate learner diversity (Weber et al., 2025). While the initial goals of AI in education were to make learning more effective, the key values of human teaching—human communication, interaction, and contextualisation—are often missing in AI tools (Alharbi, 2023). This stops AI from providing down-to-earth and contextualised feedback. When these essential values are missing, AI

feedback alone could be insufficient without human mediation.

As a result, the academic focus has moved towards hybrid intelligence (HI), which emphasises integrating human expertise with AI capabilities (Weber et al., 2025). Figure 1 presents the conceptual framework of hybrid intelligence, adapted from Weber et al. (2025). Within this hybrid intelligence approach, complementary roles are established: AI offers immediate and objective evaluation on language elements, while teachers contextualise, explain, and extend the feedback to cultivate higher-order thinking.

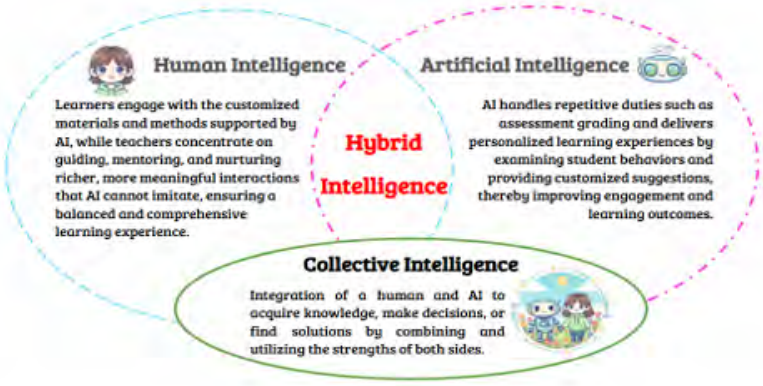


Figure 1. Conceptualisation of Hybrid Intelligence adapted from Weber et al. (2025)

While the theory of human–AI collaboration is becoming more popular, empirical evidence on how AI-assisted feedback affects writing performance at various skill levels and students' perception of their roles in this hybrid-intelligence learning environment is rather limited. The study addresses these gaps through an investigation of students' writing performances and their views on AI-assisted writing feedback. Specifically, this study aims to answer two research questions (RQs):

- **RQ1:** How does hybrid intelligence involving both teacher and AI feedback influence students' writing performance?
- **RQ2:** What perceptions do students hold about the roles of AI and teachers in hybrid-intelligence writing feedback?

2. Methodology

2.1 Research Design

This study employs a mixed-methods quasi-experimental design grounded in the theory of HI—a collaborative paradigm where human teachers and AI feedback dynamically interact to advance students' writing development. This

eight-week phase is part of a one-year school-based research project evaluating AI-assisted feedback in English classrooms. The investigation involved 72 secondary-school students aged between 14 and 17 who were divided into two intact English classes: Class A (36 higher-proficiency students) and Class B (36 lower-proficiency students). Participants were grouped based on their prior English exam performance. To ensure consistency in teaching, both classes were instructed and actively managed by the same English teacher.

2.2 Learning Context and Platform

Students received feedback on Lingo Task, an AI-powered English learning platform aiming to enhance writing performance. The main features include automatic feedback, instant grading, and progress tracking. Under the hybrid-intelligence framework, AI gives immediate formative feedback, and teachers explain and contextualise the comments, making the feedback more relatable to students' learning progress.

2.3 Procedure

The eight-week intervention involves three phases: Phase 1 (preparation and pre-test), Phase 2 (hybrid feedback intervention), and Phase 3 (post-test and evaluation). The procedure is summarised below in Table 1.

Table 1. Overview of the Study Design and Timeline

Phase	Description	Duration
Phase 1 – Preparation and Pre-test	Students completed Writing Task 1 (300 words, argumentative topic) to establish baseline writing quality.	Weeks 1–2
Phase 2 – Hybrid-Feedback Intervention	Orientation and training using LingoTask. Over six weeks, students completed four AI-assisted writing tasks (one every 1–1.5 weeks) via LingoTask. AI provided instant feedback, and the teacher conducted follow-up mini-sessions to explain, contextualize, and model revisions. Students maintained brief reflection logs documenting how they interpreted and used AI feedback.	Weeks 3–8
Phase 3 – Post-test and Evaluation	A post-test writing task on a comparable topic. Scripts were rated using a standardized analytic rubric. Student perceptions were collected through questionnaires.	Week 8

Total intervention duration: 8 weeks (as part of a 1-year research project).

2.4 Instruments and Data Sources

Ensuring authenticity and alignment with local standards, the pre-test and post-test writing tasks are adapted from the Hong Kong Diploma of Secondary Education (HKDSE) past papers, making them relevant to the students'

study environment. HKDSE writing rubrics were used to assess students' performance: content, language, and organisation. Then, students completed a questionnaire consisting of closed-ended rating-scale items and open-ended questions. Given the limited scope of this paper, the analysis focuses on the open-ended responses.

3. Results and Discussion

3.1 RQ1: Effects of AI-Assisted Feedback on Students' Writing Performance

The quantitative findings showed significant improvement in students' overall writing performance following AI-assisted feedback with a moderate effect size ($d = 0.50$, $p < .001$). In the three domains in focus, students showed the greatest improvement in language ($d = 0.51$, $p < .001$), followed by organisation ($d = 0.45$, $p < .001$) and content ($d = 0.34$, $p = .0047$). The findings could indicate that AI-assisted feedback is particularly useful in enhancing linguistic accuracy and lexical precision, but less so in idea elaboration and development.

Table 2. Pre-test and Post-test Writing Performance Across Groups

Domain	Group	Pre-test M (SD)	Post-test M (SD)	t(df) / p-value	Cohen's d	Interpretation
Content	Overall	4.54 (0.80)	4.88 (0.86)	t(72) = 2.93, p = .0047	0.41	Moderate
	Stronger	4.87 (0.86)	5.20 (0.76)	p = .00766	0.41	Moderate
	Weaker	4.32 (0.67)	4.66 (0.86)	p = .00304	0.44	Moderate
Language	Overall	3.85 (0.82)	4.26 (0.74)	t(72) = 4.30, p < .001	0.53	Moderate
	Stronger	4.30 (0.79)	4.73 (0.69)	p = .0046	0.58	Moderate
	Weaker	3.55 (0.70)	3.93 (0.59)	p = .00295	0.58	Moderate
Organization	Overall	4.24 (0.77)	4.62 (0.72)	t(72) = 3.75, p < .001	0.51	Moderate
	Stronger	4.57 (0.63)	4.97 (0.72)	p = .0159	0.59	Moderate
	Weaker	4.02 (0.79)	4.39 (0.62)	p = .0062	0.52	Moderate
Total Mark	Overall	12.64 (1.95)	13.76 (2.16)	t(72) = 4.30, p < .001	0.54	Moderate
	Stronger	13.73 (2.00)	14.90 (1.84)	p = .0077	0.61	Moderate
	Weaker	11.89 (1.96)	12.98 (1.62)	p = .0028	0.61	Moderate

Note: The maximum score of Content, Language, and Organisation is 7.

Although both groups of students received benefits from the AI-assisted feedback, their gains were different. Weaker students had statistically signifi-

cant improvement in all areas, especially in content and language. This implies that the integration of AI and teachers contributed to a more effective way of enhancing writing performances. Stronger students showed comparatively less improvement in content. While they achieved moderate gains in content (d = 0.41), their effect sizes were notably larger in language (d = 0.58), organisation (d = 0.59), and total mark (d = 0.61). A possible reason for this distinction might be that stronger students were already equipped with the ability to construct ideas with sufficient development and elaboration. AI-assisted feedback, therefore, might have less impact on the content domain since it might primarily focus on surface-level language correction.

To analyse the outcome difference between the two groups, an ANCOVA was performed, controlling for initial pre-test scores. Results showed a significant main effect of proficiency, $F(1, 71) = 5.92$, $p = .018$, partial $\eta^2 = .08$. After adjusting for pre-test performance, higher achievers (adjusted $M = 14.95$) improved more in the post-test than their weaker counterparts (adjusted $M = 13.24$). The finding showed that although AI-assisted feedback was beneficial to students with all abilities, stronger students were able to gain more improvements.

The consistent, moderate effect shows that AI feedback can offer differentiated yet equitable gains in learners' writing. While stronger students seemed to improve more overall than the weaker students, in terms of the three domains, stronger ones did not improve much in content.

3.2 RQ2: Students' Perceptions of AI and Teacher Roles in Hybrid-Intelligence Writing Feedback

Students' Perception

Given the limited scope of the current paper, only the responses from an open-ended question were reported. The responses were analysed and categorised into different themes using thematic analysis. Tables 3 and 4 summarise the similarities and differences of strong and weak students' perception on AI-assisted writing feedback, respectively.

The responses were grouped into two main categories: role and nature. For similarities, students perceived teachers having a more interpretive role by explaining "why mistakes occur" and "how to correct" them (n=33), and they tended to focus more on content, organisation and higher-order thinking (n=24). AI primarily emphasises surface-level accuracy (e.g., grammar and lexis) (n=28) and provides alternative phrasings and model sentences for students to follow (n=19). This finding echoes the results from previous research that automated

tools may mainly target language accuracy instead of content arguments (Alharbi, 2023).

More importantly, students perceive teachers and AI as complementary, having different contributions in enhancing students' writing performance. For the nature of feedback, while AI was often viewed as "fast, detailed, strict, and systematic" (n=26), teacher feedback was described as "humane, personal, and trustworthy" (n=27), which was easier to understand. These perceptions resonate with the hybrid intelligence framework (Weber et al., 2025), where AI targets mechanical precision and the teacher provides socio-emotional and contextual understanding. This aligns with Vygotsky's (1978) concept of the "more knowledgeable other" (MKO) and the Zone of Proximal Development (ZPD)—the gap between what a learner can do independently and what they can achieve with guidance from a more capable person (e.g., the teacher). In other words, even though AI helps learners in their ZPD by giving instant corrections, it is the human teacher who acts like the primary MKO and guides the learning process.

In short, while students value AI's precision, they value the important role of teachers. This balance demonstrates calibrated trust, where AI offers mechanical correction while teachers give the relatable, contextualised guidance needed for meaningful learning.

Table 3. Similarities in Strong and Weak Students' Perception on the Roles of AI and Teachers

Category	Theme	Description	Representative Quotes	Freq.
ROLE	A1. Teachers explain reasons and guide improvement	Teachers help students understand *why* errors occur and how to fix them.	"Teacher explains why I'm wrong."	33
	A2. AI identifies grammar errors and fixes sentences	AI spots mistakes and improves surface-level accuracy.	"AI corrects my grammatical mistakes."	28
	A3. Teachers improve content, organisation, and higher-order thinking	Teachers focus on ideas development, structure, and coherence.	"Teachers focus on content and organisation."	24
	A4. AI provides examples, advanced phrasing, and rewording	AI generates alternative sentences and better vocabulary.	"AI gives advanced phrases."	19
	A5. AI and teachers complement each other	Students see each as useful in different ways.	"AI focuses on language; teacher on content."	17
NATURE	B1. AI is fast, detailed, strict, systematic	AI is efficient and thorough with language problems.	"AI feedback is fast and detailed."	26
	B2. Teacher feedback is more humane, personal, trustworthy	Teachers are more reliable and human in their guidance.	"Teachers' feedback is more personal" and "more humane"	27

For differences, stronger learners adopt a diagnostic approach to AI feedback for error detection (n=17) and use AI strategically and critically (n=11). Weaker learners tended to rely on AI, especially for model writing and grammar checkers (n=21). They tended to accept AI feedback without questioning it (n=10). These findings are consistent with previous literature (Rong et al., 2025) research, in which strong students have more cognitive interactions (e.g., critical evaluation and strategic use) with AI compared to their weaker counterparts.

Stronger students could identify the different roles of teachers and AI more clearly than their weaker counterparts. As for nature, stronger students could explain that teacher feedback was personal and trustworthy (n=20), while weaker students showed their preferences but did not explain them clearly. Stronger students recognised the trade-offs between AI's speed and teachers' depth, but weaker ones tended to focus on the benefits of AI rather than its limitations.

In short, higher achievers demonstrated higher AI literacy and metacognitive awareness. This echoes findings in automated writing evaluation (AWE) literature, where effective use of automated tools depends on learners' ability to critically check, evaluate, and engage with the automated feedback, rather than overly relying on its automated function without questioning its limitations (Link et al., 2014).

4. Implications and Conclusion

This study highlights the need to perceive AI as a complementary tool to human expertise, rather than a simple replacement. While AI's automated feedback gives instant correction to linguistic errors, teachers' interpretation, guidance, and clarification are also crucial for students. Theoretically, the clear division of labour of AI and teachers adds nuance to the hybrid intelligence frameworks (e.g., Weber et al., 2025). The current study shows that "hybrid" is not a uniform experience. Higher-achievers describe it as a strategic, diagnostic, sequenced model, while lower achievers show a tendency of over-reliance on AI as a language model provider. This extends the AWE research (Link et al., 2014) that usually compares the usefulness of AWE and teacher feedback. The current study shows that engagement with AI feedback is not homogeneous: both learner proficiency and feedback literacy can contribute to different learning strategies and outcomes of hybrid intelligence.

This research also adds a new element, trust, that has been previously explored by scholars such as Shin (2021). In our study, we found that students exhibited different levels of trust in teachers and AI based on their knowledge and

interpretations. In the classroom, stronger students tend to strike a balance by critically examining AI feedback and articulating that teacher feedback is more trustworthy. When using AI, students need to determine the extent to which they trust AI feedback based on their knowledge and interpretation.

Practically, differentiated hybrid pedagogy is essential in L2 writing classrooms. While stronger students are more capable of managing AI and teacher feedback, weaker students need more explicit scaffolding. There is a need for teacher training programmes and curriculum frameworks to undergo revision to formally integrate AI-informed pedagogical strategies. Moorhouse et al. (2024), for example, have exemplified how an intervention programme could enhance pre-service teachers' professional GenAI competence. In relation to our study, in particular, teachers should be trained not only to interpret AI automated feedback but also, more critically, to teach students how to evaluate and question algorithmic suggestions into thoughtful revisions. With the hybrid intelligence pedagogical model, automated platforms could manage linguistic enhancement, thereby freeing up teachers to dedicate their expertise to cultivating students' cognition and content development.

Despite aforementioned contributions, some limitations should be noted. First, our analysis focused only on stronger and weaker students without delving into the nuanced differences exhibited by intermediate learners. However, given that the stronger and weaker students came from intact classes, we believed that it was more ecologically valid for this study to contrast the stronger class of students with the weaker class. Furthermore, the relatively small sample size (N=72) and the short-paper format also make it difficult for us to include a more nuanced analysis of students with different proficiency levels. We call for future research to work with larger samples to further investigate the proficiency variable. Second, we mainly reported findings from questionnaires and writing tasks in this paper. Future research can consider diversifying the methods used to collect data through interviews or think-aloud protocols to yield a more comprehensive and in-depth analysis.

Generative AI Policy Statement

The author(s) used ChatGPT and Gemini during the preparation of this work to proofread. 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

Alharbi, W. (2023). AI in the foreign language classroom: A pedagogical overview of automated writing assistance tools. *Education Research International*, 2023(1), 4253331. https://doi.org/10.1155/2023/4253331

Link, S., Dursun, A., Karakaya, K., & Hegelheimer, V. (2014). Towards better ESL practices for implementing automated writing evaluation. *Calico Journal*, 31(3), 323-344.

Moorhouse, B. L., Wan, Y., Wu, C., Kohnke, L., Ho, T. Y., & Kwong, T. (2024). Developing language teachers' professional generative AI competence: An intervention study in an initial language teacher education course. *System*, 125, 103399. https://doi.org/10.1016/j.system.2024.103399

Rong, M., Yao, Y., Li, Q., & Chen, X. (Winnie). (2025). Exploring student engagement with artificial intelligence-guided chatbot feedback in EFL writing: Interactions and revisions. *Computer Assisted Language Learning*, 1–30. https://doi.org/10.1080/09588221.2025.2539979

Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies*, 146, 102551. https://doi.org/10.1016/j.ijhcs.2020.102551

Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.

Weber, F., Wambsganss, T., & Söllner, M. (2025). Enhancing legal writing skills: The impact of formative feedback in a hybrid intelligence learning environment. *British Journal of Educational Technology*, 56(2), 650-677. https://doi.org/10.1111/bjet.13529

Rethinking Hybrid Human–AI Intelligence in Learning: The AI-Augmented Creative Resilience Framework

Henry MA [1] & Ka Yan CHUNG [1]
[1] School of Design, The Hong Kong Polytechnic University, Hong Kong, China

Abstract

The increasing adoption of Generative Artificial Intelligence (GenAI) in creative learning environments is reshaping how intelligence, creativity, and agency are understood within higher education. This integration introduces significant psychological stressors, particularly Creative Displacement Anxiety (CDA). This conceptual paper proposes the AI-Augmented Creative Resilience Framework (AICRF) to conceptualise hybrid human-AI intelligence as a cyclical process of recognition, reframing, co-creation, and regeneration. The framework demonstrates how learners transition from fears of obsolescence toward sustained agency. By positioning AI as a collaborative partner, the AICRF provides actionable pedagogical strategies for design education. The model highlights that AI literacy requires both technical mastery and the emotional regulation necessary for effective human-AI collaboration.

Keywords: Creative resilience; Generative artificial intelligence (GenAI); Creative Displacement Anxiety (CDA); Human–AI collaboration; Design education

1. Introduction

While GenAI expands ideational capacity in higher education, it introduces significant psychological challenges regarding authorship and creative identity. Creative Displacement Anxiety (CDA), characterised by fears of skill obsolescence and diminished creative value (Caporusso, 2023), represents a significant challenge in creative education where originality and personal expression are central objectives. This phenomenon signals the emergence of hybrid human–AI intelligence, in which cognition and creativity are distributed across human learners and intelligent systems. The construct is empirically operationalised via the validated Creative Displacement Anxiety Scale (CDAS), which identifies seven psychosocial stressors, including skills atrophy and loss of creative identity, relevant to AI-augmented environments (Chung et al., 2025).

To address these challenges, this conceptual paper proposes the AI-Augmented Creative Resilience Framework (AICRF) to explain how learners adapt through reflective and co-creative engagement with AI to sustain creative agency.

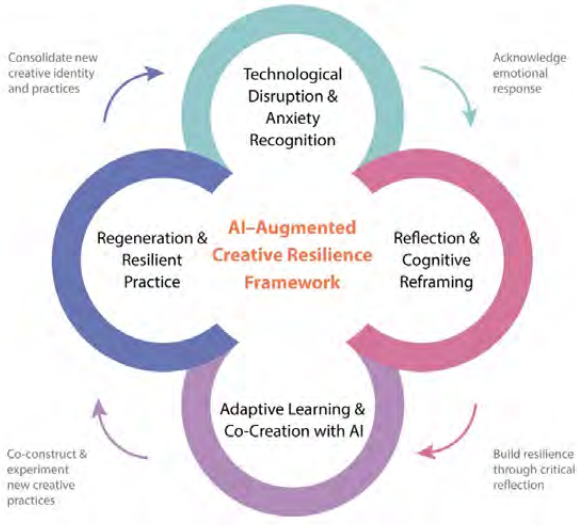
2. Methodology

This study utilises narrative synthesis to develop a conceptual framework by integrating research across three primary domains: generative AI and creativity (Wingström et al., 2024), creative resilience (Lin & Chen, 2024), and human-AI co-creativity (McGuire et al., 2024). Through iterative thematic analysis, recurring constructs concerning emotional response, cognitive adaptation, and collaborative intelligence were synthesised into a theoretically grounded structure. This methodology provides a coherent lens for interpreting how hybrid human–AI intelligence develops within creative learning environments supported by technology.

3. Results and Discussion

The AICRF (Figure 1) conceptualises hybrid human–AI intelligence as a cyclical process of emotional and cognitive adaptation. The cycle initiates with **Anxiety Recognition**, where encountering sophisticated GenAI outputs triggers CDA related to perceived threats to creative identity and skills. This affective response serves as the critical catalyst for the transition to **Cognitive Reframing**, which is facilitated by reflective engagement rather than resistance (Hwang & Wu, 2025).

Figure 1: The Proposed AI-Augmented Creative Resilience Framework



By reinterpreting AI as a collaborative partner, learners align with resilience theory's focus on meaning-making under disruption (Masten et al., 2021), enabling the shift into **Adaptive Co-creation**. In this phase, cognition becomes distributed, merging human intentionality with machine-generated variation (Rezwana & Maher, 2023). Finally, the iterative success and mastery achieved during collaboration drive the transition to **Regenerative Learning**, where creative confidence and agency are restored. This continuous loop reinforces hybrid intelligence as a developmental capacity sustained through ongoing human-AI interaction.

4. Implications and Conclusion

The AICRF provides a conceptual blueprint for integrating GenAI into curricula by framing hybrid intelligence as a socio-emotional process rather than a purely technical one. Educators can operationalise the framework through structured pedagogical interventions that build observable resilience. For instance, in a visual identity design studio, students often encounter CDA when comparing manual ideation speed to rapid AI outputs. Instructors facilitate Cognitive Reframing by positioning AI as a lateral thinking partner rather than a replacement. During Adaptive Co-creation, students curate and refine diverse AI-generated permutations to sustain creative agency, while the Regenerative Learning phase requires students to document their specific aesthetic interventions. This process makes creative resilience observable through the learner's ability to articulate their agency over the machine.

Ultimately, the AICRF suggests that AI literacy must transition from technical mastery toward the emotional and cognitive regulation required for human-AI collaboration. By emphasising reflection and co-creation, the framework supports pedagogical approaches that prepare learners to navigate evolving technological landscapes while sustaining creative well-being across diverse educational contexts.

Generative AI Policy

The author utilised ChatGPT-5 for language refinement and proofreading to ensure clarity and grammatical accuracy. No other AI tools were employed. All conceptual content, including ideas and conclusions, is original. The author assumes full responsibility for the final manuscript.

References

Caporusso, N. (2023). Generative artificial intelligence and the emergence of creative displacement anxiety. *Research in Psychology and Behavior*, 3(1). https://doi.org/10.53520/rdpb2023.10795

Chung, K.Y., Ma, H., & Chan, Y.K. (2025). Measuring Creative Displacement Anxiety in the Age of Generative AI: Scale Development and Validation. In C.-Y. Chang, C.-H. Chen, & Y. Hsu (Eds.), *IASDR 2025: Design Next*, 02–05 December, Taiwan. https://doi.org/10.21606/iasdr.2025.46

Hwang, Y., & Wu, Y. (2025). The influence of generative artificial intelligence on creative cognition of design students: a chain mediation model of self-efficacy and anxiety. *Frontiers in Psychology*, 15, 1455015. https://doi.org/10.3389/fpsyg.2024.1455015

Lin, H., & Chen, Q. (2024). Artificial Intelligence (AI)-integrated educational applications and college students' creativity and academic emotions: students and teachers' perceptions and attitudes. *BMC Psychology*, 12, 487. https://doi.org/10.1186/s40359-024-01979-0

Masten, A.S., Lucke, C.M., Nelson, K.M., & Stallworthy, I.C. (2021). Resilience in development and psychopathology: Multisystem perspectives. *Annual Review of Clinical Psychology*, 17(1), 521–549. https://doi.org/10.1146/annurev-clinpsy-081219-120307

McGuire, J., De Cremer, D., & Van de Cruys, T. (2024). Establishing the importance of co-creation and self-efficacy in creative collaboration with artificial intelligence. *Scientific Reports*, 14(1), 18525. https://doi.org/10.1038/s41598-024-69423-2

Rezwana, J., & Maher, M.L. (2023). Designing creative AI partners with COFI: A framework for modelling interaction in human-AI co-creative systems. *ACM Transactions on Computer-Human Interaction*, 30(5), 1–28. https://doi.org/10.1145/3519026

Wingström, R., Hautala, J., & Lundman, R. (2024). Redefining creativity in the era of AI? Perspectives of computer scientists and new media artists. *Creativity Research Journal*, 36(2), 177–193. https://doi.org/10.1080/10400419.2022.2107850

8 Technology, Curriculum Innovation and Sustainability in Higher

Enhancing Conceptual Understanding of Crystal Structures through Holographic Projection and Team-Based Learning

Fatwa Firdaus ABDI[1] & Hoi Ying CHUNG[1]
[1] School of Energy and Environment, City University of Hong Kong, Hong Kong, China

Abstract

This study investigates the integration of holographic projection with a team-based learning (TBL) framework to improve student understanding of crystallographic structures. Students often struggle to visualise three-dimensional atomic arrangements, and conventional diagrams or physical models provide limited spatial immersion. As generative artificial intelligence (GenAI) increasingly supports content creation and instructional design in higher education, shared volumetric holographic displays offer a complementary pathway for enhancing spatial cognition and collaborative learning. When combined with a structured TBL sequence, the holographic visualisation enables teams to explore symmetry, packing, and coordination in common crystal structures. Students reported high levels of engagement, improved spatial comprehension, and a strong preference for holographic visualisation compared with 2D images, physical models, and AR/VR tools. Performance metrics showed substantial gains from individual to team readiness assessments and higher subject-relevant mid-term scores relative to a previous cohort. The results demonstrate that holographic projection, when paired with collaborative learning, effectively enhances conceptual understanding and supports both cognitive and social dimensions of learning. This approach offers a scalable, low-barrier innovation that complements emerging GenAI-driven educational technologies for teaching spatially complex topics.

Keywords: 3D holographic projection, immersive visualisation, team-based learning, crystallographic structures, interactive STEM pedagogy

1. Introduction

In typical materials-related science and engineering education, students explore fundamental topics such as atomic bonding, crystal structures, defects, phase transformations, mechanical properties, and characterisation techniques. Of these, crystal structures and defects are critical, as they determine material properties ranging from strength and hardness to conductivity and failure

behaviour. However, these concepts are abstract; students frequently struggle to visualise atomic arrangements, identify symmetry elements, and connect structures to properties (Komaro et al., 2017; Krause et al., 2003). Traditional tools—two-dimensional diagrams, static images, and physical ball-and-stick models—often fail to convey the full spatial complexity of crystallography. Although 3D computer visualisation tools exist (Beekman, 2022; Li, 2020), they may still lack the interactivity and immersion needed to engage learners.

Recent pedagogical innovations, including augmented-reality (AR) or virtual-reality (VR) visualisation and active-learning frameworks, have shown promise in improving spatial cognition and engagement. Immersive, interactive representations support mental model formation, while team-based learning (TBL) promotes deeper understanding through structured collaboration. TBL is a well-established instructional strategy grounded in constructivist learning theory, in which students first develop individual understanding and then refine their knowledge through team interaction, feedback, and application-focused tasks (Hrynchak & Batty, 2012; Roossien et al., 2022). This structure is particularly effective for conceptually demanding topics, as it encourages articulation of reasoning, peer discussion, and clarification of misconceptions.

However, many AR/VR approaches require individual headsets or specialised hardware, limiting scalability in classrooms. In contrast, emerging research indicates that pseudo-hologram devices, such as holographic pyramids and LED fans, offer notable benefits over traditional aids. Students using holograms show higher motivation, improved performance, and greater learning enjoyment (Prado Ortega et al., 2020; Yu et al., 2024). These devices create immersive, three-dimensional visuals without requiring wearable equipment and allow multiple learners to view and interact with models simultaneously, making them well-suited for collaborative classroom settings. Reviews and meta-analyses confirm that holographic projection not only increases engagement and motivation but also fosters a deeper understanding of abstract concepts (Flores-Cáceres et al., 2024; Yoo et al., 2022). As such, 3D holograms represent a promising technology for improving visualisation and comprehension across disciplines.

From a pedagogical perspective, integrating holographic visualisation with TBL offers a potentially complementary approach. A shared volumetric display provides a common visual reference for team discussion, supporting joint attention, spatial referencing, and collaborative reasoning—processes central to team-based learning. In this context, the hologram is not merely a passive visualisation tool but can act as an active discussion anchor around which teams test interpretations and apply concepts. This alignment between visualisation tech-

nology and instructional design provides a plausible basis for supporting both individual spatial understanding and group-level reasoning.

This study investigates the integration of holographic projection devices, capable of displaying shared volumetric 3D models, with a TBL instructional sequence to address challenges in teaching crystallographic structures, including body-centred cubic (BCC), face-centred cubic (FCC), and hexagonal close-packed (HCP) arrangements. The combined approach was designed to strengthen spatial reasoning, collaborative engagement, and conceptual understanding. The goal of the study is to evaluate the effectiveness of holographic visualisation paired with TBL using student perceptions, engagement metrics, and performance data. Through a mixed-methods design, the study contributes empirical evidence on how shared 3D visualisation embedded within structured collaboration can advance engineering education. While beyond the scope of this study, complementary technologies, such as generative artificial intelligence (GenAI), could further enhance shared visualisation-based learning approaches.

2. Methodology

2.1. Instructional Intervention Design

The intervention took place during a lesson on crystal structures. Before class, students watched a short video in which the instructor introduced basic structures using animations. In class, a TBL sequence was implemented:

- **Individual Readiness Assurance Test (iRAT):** Students independently completed a short quiz assessing core concepts related to lattices, packing arrangements, and atomic coordination in common crystal structures.
- **Introduction to Holographic Models:** The instructor revisited key concepts (unit cells, symmetry, packing, coordination) from the pre-class video using holographic projections. A large holographic fan projector (Fig. 1a) was used by the instructor in the middle of the classroom, providing dynamic control to pause, zoom, and rotate models in real time for detailed examination. At the same time, smaller tabletop holographic pyramids (Fig. 1b) were distributed on each group's table, enabling students to explore the models from all angles by physically rotating the pyramids for a full 360° perspective. Models included BCC, FCC, HCP, and simple cubic, with clearly rendered atoms and coordination environments. This provided a shared spatial visualisation experience accessible to all group members.
- **Team Readiness Assurance Test (tRAT):** Students retook the quiz in teams, using holograms to support discussion.

- **Team Application Activities:** Groups completed two problem-solving activities involving coordination numbers, distinguishing crystal structures, and determining atomic packing factors:
  - **Crystal Structure Bingo:** Students received a 5 × 5 grid containing structure names, coordination numbers, atoms per unit cell, atomic packing factor, and other descriptions/properties. For each hologram shown, students circled the matching entries, continuing until a vertical/horizontal/diagonal line was completed. A short time limit was imposed for each hologram to keep students engaged and promote quick identification of structural features.
  - **Crystal Structure Mission Match:** Students were given energy-related applications (e.g., aircraft fuselage, turbine blade, battery electrode, wind turbine magnet) and selected suitable crystal structures based on observable features like symmetry, packing density, and layering. Visual clues and holographic pyramids of the available crystal structures were provided to guide their reasoning.
- **Class Discussion:** Teams shared insights, clarified misconceptions, and reflected on which visual cues were most helpful. The instructor aided the discussion by providing suggested answers to consolidate learning.

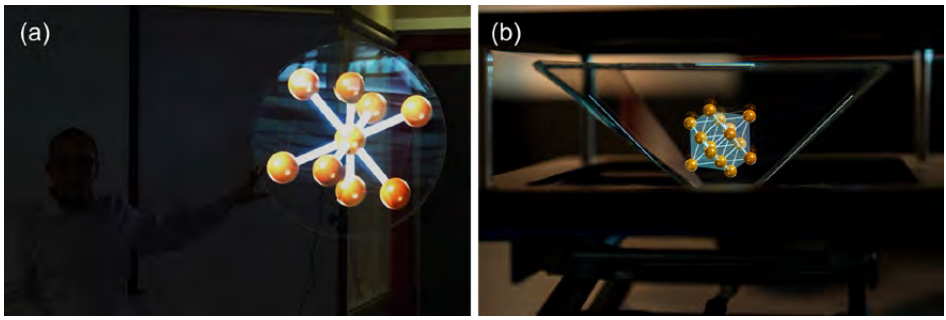


Figure 1. (a) Holographic fan projector and (b) holographic pyramid projector used in this study.

2.2. Participants

Participants were undergraduate students in an introductory materials engineering course. The intervention cohort included ~18 students, and a prior cohort (n ≈ 27) served as a performance benchmark. No identifying or demographic data were collected to preserve anonymity.

2.3. Data Collection Instruments

Five categories of data were gathered. A Likert-scale survey assessed student

perceptions of the holographic tool, including ease of viewing, understanding of symmetry, atomic packing, coordination number, and overall usefulness. The survey also assessed perceptions of the TBL format, covering preparatory quizzes, engagement in group tasks, discussion quality, and perceived learning gains. Comparative items asked students to evaluate the holographic device against 3D computer models, AR/VR headsets, physical models, and textbook images. Performance data were collected from individual/team quizzes, group activities, and mid-term assessments. Finally, open-ended questions solicited qualitative feedback on effective aspects, limitations, and suggested improvements.

2.4. Data Analysis

Quantitative data were summarised using descriptive statistics, and performance improvements were analysed by comparing individual and team test scores as well as mid-term results across cohorts. Qualitative responses were thematically coded to identify common patterns. This mixed-methods approach allowed triangulation of perception, performance, and qualitative insights for a comprehensive evaluation.

3. Results and Discussion

3.1. Student Perceptions of Holographic Visualisation

Student perceptions of the holographic device were overwhelmingly positive (Fig. 2). Most students found it easy to view from multiple angles (94% rated 4–5), facilitated understanding of symmetry (81% rated 4–5), and effective for improving comprehension of atomic packing (100% rated 4–5) and coordination numbers (94% rated 4–5). All participants rated the tool as highly useful, indicating that it successfully addressed challenges in visualising complex three-dimensional arrangements and enabled more intuitive structural interpretation.

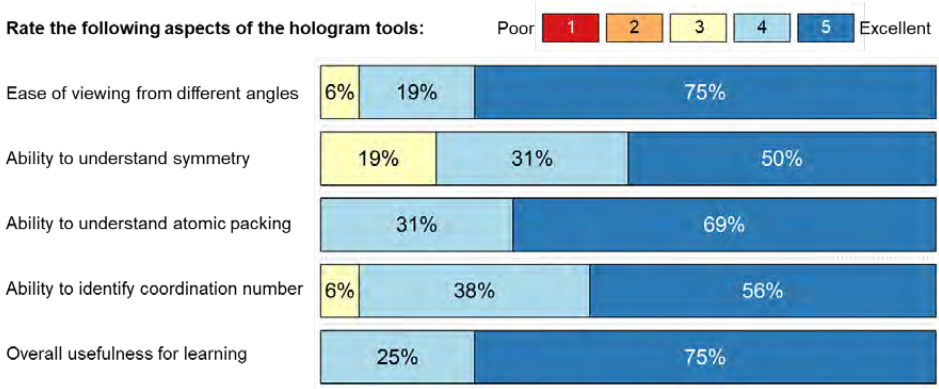


Figure 2. Students' survey results on their perception of the hologram tools used in this study.

Qualitative feedback echoed these findings, emphasising clearer symmetry perception, shared group viewing, and improved recall of atomic arrangements. Students noted that the device supported group discussion and made abstract concepts more tangible. Reported limitations—brightness variation, small graphical details, and occasional interface issues—were minor, with some students requesting complementary notes. Overall, these concerns did not diminish the tool's pedagogical value.

3.2. Student Perceptions of the TBL Format

The TBL format was also positively received (Fig. 3). Students indicated that preparatory quizzes supported readiness, and that group activities were engaging and discussion-driven. Most reported that teamwork enhanced conceptual understanding and that post-activity discussions helped clarify ideas. Overall, students viewed TBL as an effective way to learn crystallographic structures, consistent with previous research showing that structured peer collaboration paired with shared visualisations strengthens conceptual development.

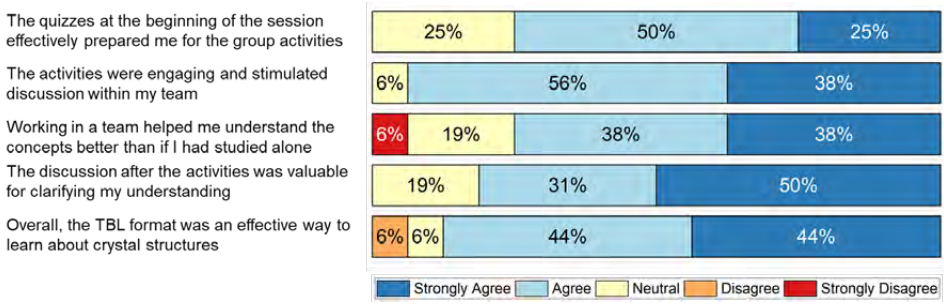


Figure 3. Students' survey results on their perception of the team-based learning (TBL) elements incorporated in this study.

3.3. Comparison with Other Visualisation Modalities

Comparisons with other tools further illustrate the value of holographic projection. Students reported that the holographic models were more interactive than conventional 3D computer models and enabled better collaboration than AR/VR headsets, which require individual devices. Relative to physical ball-and-stick models, the holograms were easier to manipulate, view, and share within teams. Nearly all students agreed that, relative to textbook images, the holographic visualisation offered clearer three-dimensional understanding and improved their ability to distinguish BCC/FCC/HCP structures. These comparisons underscore the advantage of a shared volumetric display for promoting spatial understanding and collaborative learning.

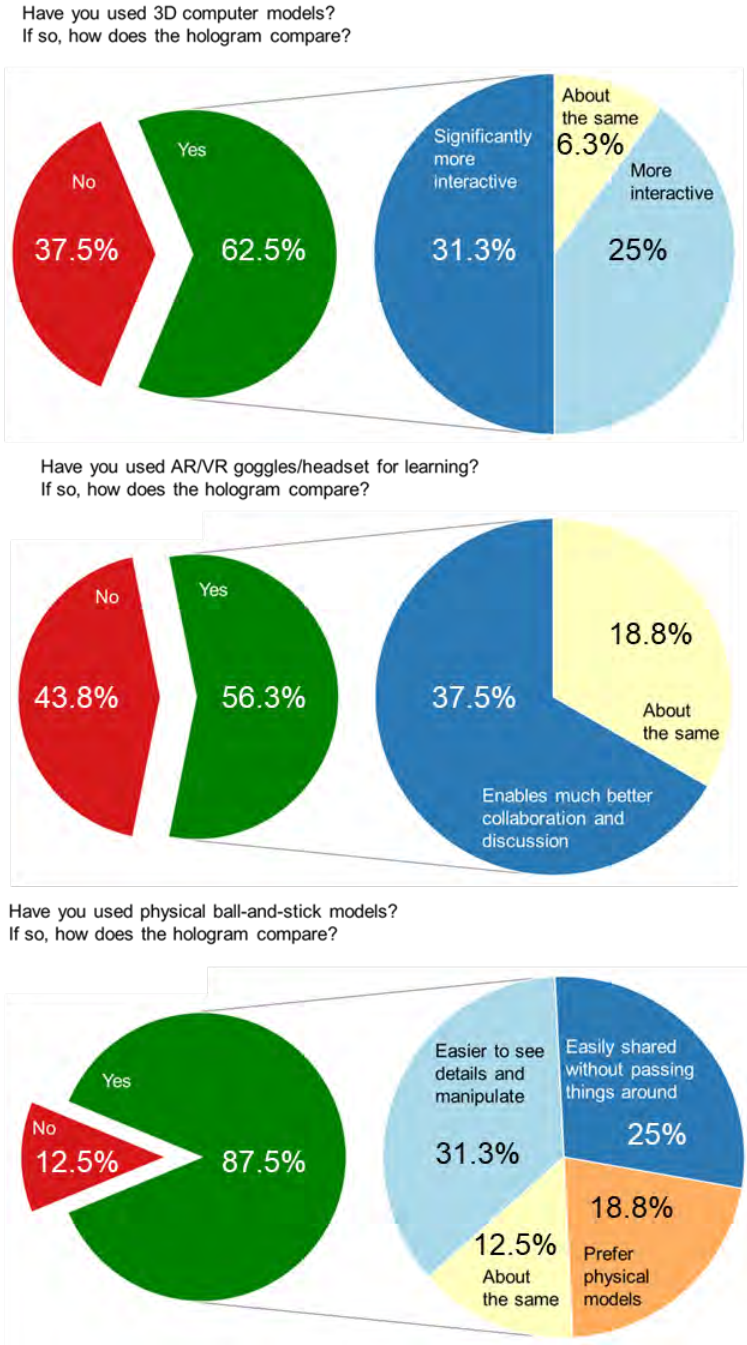


Figure 4. Students' survey results on the comparison between the hologram tools and other crystal structure visualisation methods.

3.4. Learning Performance and Conceptual Gains

Performance metrics corroborated the positive perceptions. iRAT scores averaged 80.0±17.3, while tRAT following exposure to the holographic visualisation increased to an average of 98.1±5.0. The substantial increase suggests that teams effectively resolved misconceptions. Scores on team activities were also high (92.2 for Crystal Structure Bingo; 91.7 for Crystal Structure Mission Match), indicating active engagement and effective application of knowledge. Mid-term results on problems related to crystal structures showed a marked improvement relative to the historical cohort: the pre-intervention cohort (2023) scored 63.3±24.2, while the intervention cohort (2025) achieved 75.5±13.8. Although not a controlled experiment, the higher mean and reduced variance suggest meaningful conceptual gains associated with combining holographic visualisation and TBL.

4. Implications

In conclusion, holographic visualisation enhanced students' spatial understanding of crystal structures, particularly symmetry and coordination. Integrated with TBL, the shared visualisations facilitated discussion, collective reasoning, and correction of misconceptions. Students reported strong engagement and perceived learning gains, and performance data indicated tangible improvement across readiness tests and formal assessments. While larger studies would strengthen generalisability, these results provide promising evidence for this pedagogical approach. Practically, the holographic device is easy to deploy, requires minimal training, and supports medium-to-large classes without individual headsets or screens. Its affordability and simplicity make it accessible to institutions with limited resources (Ravikumar et al., 2025; Salloum et al., 2024). The findings also align with research on spatial cognition and collaborative learning, suggesting that shared volumetric visualisation enhances joint attention and collective reasoning (Kang et al., 2024; Lee et al., 2021). Future integration with GenAI could further enhance this approach by enabling automated creation of customised visualisations tailored to specific learning needs or conceptual difficulties. Finally, this approach is broadly transferable to disciplines involving spatial complexity—including engineering, chemistry, medicine, and architecture—offering a low-barrier alternative to AR/VR environments. By combining immersive visualisation with structured collaboration, the intervention addresses both cognitive and social dimensions of learning and offers a replicable model for pedagogical innovation in higher education.

Acknowledgments

The work described in this paper was fully supported by the University Grants Committee's Teaching Development Grant (Project No. 6000913). The

authors acknowledge the support of Mr Alphonso Lei, Mr Jimmy Chong Hin Fei, Mr Marco Leung Yiu Lam, Dr William Cheung Man Yin and Ms Lan Wen during the preparation and delivery of the TBL session.

Ethical Declaration Statement

The project has been approved by the Human Subjects Ethics Sub-Committee of the City University of Hong Kong (Reference Number: HU-STA-00001172).

Generative AI Policy

The author(s) used ChatGPT-5.1 during the preparation of this work to organise ideas based on research findings, improve clarity and conciseness throughout the writing process. After utilising the tool, the author(s) thoroughly reviewed and edited the content as necessary and assumed full responsibility for the publication's content.

References

Beekman, M. (2022). Teaching basic crystallography and diffraction using open access structure visualisation software. *MRS Advances, 7*(22), 482-487. https://doi.org/10.1557/s43580-022-00238-x

Flores-Cáceres, R., Olivares-Zegarra, S., Flores-Velásquez, C. H., Arévalo-Tuesta, J. A., Aldana-Trejo, F., Alvarado-Bravo, N., Durán-Herrera, V., Tarazona-Bernal, J., Torres-Quiroz, A., & Aranda-Dextre, C. C. (2024). Bibliometric Review on the Application of Holograms in Immersive Environments in the Context of University Teaching, 2013-2023. *International Journal of Learning, Teaching and Educational Research, 23*(3), 169-185. https://doi.org/10.26803/ijlter.23.3.9

Hrynchak, P., & Batty, H. (2012). The educational theory basis of team-based learning. *Medical Teacher, 34*(10), 796-801. https://doi.org/10.3109/0142159X.2012.687120

Kang, J., Zhou, Y., Rajarathinam, R. J., Tan, Y., & Shaffer, D. W. (2024). Unveiling joint attention dynamics: Examining multimodal engagement in an immersive collaborative astronomy simulation. *Computers & Education, 213*, 105002. https://doi.org/10.1016/j.compedu.2024.105002

Komaro, M., Djohar, A., Setiawan, A., Hasan, B., & Wibi, A. (2017). The implementation of multimedia animation to improve the mastery of material about crystal defects in engineering courses. In *Regionalization and Harmonization in TVET* (pp. 207-210). Routledge. https://doi.org/10.1201/9781315166568

Krause, S., Decker, J. C., Niska, J., Alford, T., & Griffin, R. (2003). Identifying student misconceptions in introductory materials engineering classes.

Proceedings of the ASEE Annual Conference. Washington: ASEE Annual Conference. https://doi.org/10.18260/1-2--12519

Lee, B., Hu, X., Cordeil, M., Prouzeau, A., Jenny, B., & Dwyer, T. (2021). Shared Surfaces and Spaces: Collaborative Data Visualisation in a Co-located Immersive Environment. *IEEE Transactions on Visualisation and Computer Graphics, 27*(2), 1171-1181. https://doi.org/10.1109/TVCG.2020.3030450

Li, X.-Z. (2020). SVAT4: a computer program for visualisation and analysis of crystal structures. *Journal of Applied Crystallography, 53*(3), 848-853. https://doi.org/10.1107/S1600576720004586

Prado Ortega, M. X., Delgado Ramírez, J. C., Valarezo Castro, J. W., Armijos Carrión, J. L., Ávila Carvajal, A. A., & González Segarra, A. N. (2020). Application of the technical - pedagogical resource 3D holographic LED-fan display in the classroom. *Smart Learning Environments, 7*(1), 32. https://doi.org/10.1186/s40561-020-00136-5

Ravikumar, B., Gopinathan, S., Ajitha Angelian, J., Shivamalar, M., & Pavithra, A. (2025, 28-30 Oct. 2025). Enhancing Education with Holographic Interaction. *2025 10th International Conference on Communication and Electronics Systems (ICCES)*. https://doi.org/10.1109/ICCES67310.2025.11337191

Roossien, L., Boerboom, T. B. B., Spaai, G. W. G., & de Vos, R. (2022). Team-based learning (TBL): Each phase matters! An empirical study to explore the importance of each phase of TBL. *Medical Teacher, 44*(10), 1125-1132. https://doi.org/10.1080/0142159X.2022.2064736

Salloum, S. A., Alhumaid, K., Alfaisal, A. M., Aljanada, R. A., & Alfaisal, R. (2024). Adoption of 3D Holograms in Science Education: Transforming Learning Environments. *IEEE Access, 12*, 70984-70998. https://doi.org/10.1109/ACCESS.2024.3402549

Yoo, H., Jang, J., Oh, H., & Park, I. (2022). The potentials and trends of holography in education: A scoping review. *Computers & Education, 186*, 104533. https://doi.org/10.1016/j.compedu.2022.104533

Yu, Q., Li, B.-m., & Wang, Q.-y. (2024). The effectiveness of 3D holographic technology on students' learning performance: a meta-analysis. *Interactive Learning Environments, 32*(5), 1629-1641. https://doi.org/10.1080/10494820.2022.2124424

Promoting Institutional GenAI-in-Assessment Reform through the Lens of Kotter's Change Model

Dawn LO [1], Julia CHEN [1], Wing Yuk Leo CHON [1]
[1] Educational Development Centre, The Hong Kong Polytechnic University, Hong Kong, China

Abstract

The rapid emergence of generative artificial intelligence (GenAI) tools, such as ChatGPT, presents both opportunities and challenges for higher education, particularly regarding assessment practices. This paper examines a university's initiative in Hong Kong to redesign assessment in response to GenAI, utilising Kotter's eight-step Change Model as an analytical framework. The study reviews the university's actions in establishing urgency, forming a guiding coalition, creating a strategic vision, communicating that vision, empowering others, achieving short-term wins, consolidating improvements, and institutionalising new approaches. It identifies both strengths and areas for improvement in the implementation of GenAI integration. The analysis reveals the importance of clear communication and iterative improvement strategies, offering practical implications for other institutions navigating similar technological shifts in education.

Keywords: Generative AI; Assessment Redesign; Kotter's Model; Institutional Change

Introduction

The emergence of generative artificial intelligence (GenAI) tools, such as ChatGPT, has significantly impacted higher education (HE), challenging traditional teaching and assessment methods. While GenAI may enhance learning, it raises plagiarism concerns, prompting educators to reconsider assessment purposes and evidence gathering (Chen, 2023). This paper applies Kotter's eight-step change model (1996) to evaluate an assessment redesign initiative at a Hong Kong university, identifying strengths and offering insights for similar institutions.

Literature Review

Kotter's Change Model (1996) is a prominent framework for managing change in complex organisations (Wentworth et al., 2018). It comprises eight steps—spanning urgency, coalition-building, vision, communication, empow-

erment, short-term wins, consolidation, and institutionalisation—and has successfully guided higher education changes, including assessment systems, organisational culture, and faculty development (Guzman et al., 2011; Springer et al., 2012; Steinert et al., 2007). Appelbaum et al. (2012), in particular, reviewed fifteen years of research to determine the effectiveness of Kotter's model in organisational change efforts. Although lacking quantitative empirical data, Appelbaum et al. (2012) recommend the model over alternatives. Wentworth et al. (2018) also found Kotter's model well-suited for practical use in higher education. The studies of Kang et al. (2022) and Potter et al. (2023) demonstrate that Kotter's change model can be used for the analysis of change efforts at a faculty/department level and at a university level, respectively.

Methodology

This paper adopts a case study approach to examine the institutional change process of a university in embracing GenAI through assessment redesign. Multiple data sources were drawn upon, including staff surveys related to GenAI use in teaching and assessment, institutional documents (such as GenAI guidelines, committee minutes, policy drafts, funding calls, and assessment exemplars), and participant observations. Data were analysed through thematic mapping against Kotter's change model. Evidence from different sources was iteratively compared to identify convergence, divergence, and gaps across stages of the change process. Rather than aiming for statistical generalisability, the analysis focused on examining how and why particular change strategies were enacted, their limitations, and generating insights that may be transferrable to similar institutional contexts.

All three authors were directly involved in initiating and supporting the change and therefore adopted a participant-observer stance. While this positioning provided rich, insider insight into institutional decision-making and implementation processes, it also required careful attention to potential bias. To mitigate potential bias, methodological triangulation was employed by substantiating observations with evidence from multiple independent data sources and by grounding interpretations in documented institutional records rather than the authors' personal opinions. In addition, interpretations of key change milestones were discussed and cross-checked among the authors to reduce individual subjectivity and to ensure that claims were grounded in verifiable evidence.

Institutional Context

The case study university is one of Hong Kong's eight government-funded institutions. It enrolls approximately 28,000 students with over 1,200 faculty members. The University aspires to be an innovative world-class university that

pursues excellence in education, research, and knowledge transfer. With ten faculties/schools of diverse disciplines, the University ranks among the top 100 universities worldwide.

Results

This section presents the findings from the case study, using Kotter's eight-step Change Model as an analytical framework. Each subsection corresponds to one stage of the model and explains how the University undertook the GenAI-in-assessment reform.

Step 1: Establishing a Sense of Urgency

The November 2022 emergence of ChatGPT and the subsequent proliferation of GenAI tools precipitated a significant sense of urgency within the educational sector. It sparked intense academic discussions, particularly concerning their potential implications for learning processes, assessment methodologies, and academic integrity. To raise staff awareness of the need to change, the Educational Development Centre (EDC) at the case study university organised Hong Kong's first university staff workshop on ChatGPT in February 2023, followed by student sharing sessions, to highlight impacts on academic integrity and assessment validity.

Step 2: Forming a Powerful Guiding Coalition

The second step in Kotter's model involves assembling a strategic transformation team with the necessary knowledge, skills, and influence to lead the change effort effectively. The Director of EDC formed a dedicated team and, recognising the need for senior support, secured backing from the President and senior management to embrace GenAI-in-education via discussions and meetings held by university committees, including the University Senate and the Learning and Teaching Committees (LTC).

Step 3: Forming a Strategic Vision

Kotter's third step involves creating a clear picture of the desired future state and establishing a plan for achieving that vision (Kotter, 1996). The University established its vision through drafting ethical guidelines, followed by consultation and final approval by the LTC. The process refined the vision and strategy, established a common goal, and ensured alignment across the University.

Step 4: Communicating the Vision

In Kotter's fourth step, organisations must communicate the vision and

plan efficiently and effectively. Major change is successful only when people buy-in and drive in the same direction (Besterfield-Sacre et al., 2014). To ensure clear and effective communication of the University's stance on the ethical use of GenAI, GenAI guidelines for all teachers and students were disseminated via email by the LTC on 28 April 2023. The EDC followed up with faculty outreach and departmental consultations to clarify the vision and strategies. In addition, the Director of EDC penned two articles (in July 2023 and April 2024) and shared content on a podcast (in August 2024) on *THE Campus* regarding assessment redesign in the age of GenAI (Chen, 2023), the future of higher education (Chen, 2024), and the University's attempts at managing AI respectively.

Step 5: Empowering Others to Act on the Vision

Kotter advocates expanding change implementation by delegating responsibilities more broadly to increase organisational buy-in and commitment to the transformation. The case study university encountered three key challenges: ensuring equitable access to GenAI tools, developing AI literacy and competence among teachers, and securing funding and resources for innovative practices.

To address the first challenge, the EDC established an 'AI Playground' in March 2023, providing students and teachers with free access to a curated selection of GenAI tools. To foster AI literacy and competence among teachers, the EDC offered professional development opportunities through workshops, consultations, and sharing sessions, covering introductory GenAI concepts and their potential educational impact, as well as more advanced sessions focused on developing teachers' technical skills and pedagogical approaches for using GenAI.

As another key initiative, in July and December 2023, the EDC conducted two rounds of 'Collection of and Consultation on Assessment Tasks to Accommodate the Use of GenAI'. All academic departments and teaching centres were invited to submit two or three redesigned, GenAI-ready assessment tasks. An assessment consultant reviewed the submitted tasks and provided feedback to teachers. Following each collection round, the EDC identified and disseminated good practices of effective assessment redesign through faculty outreach meetings, sharing sessions, and workshops (Chen et al., 2024).

To address the final challenge, the University established Seed Funds for Promoting AI Literacy and Use of GenAI for Authentic Learning and Assessment. Departments could use the fund to develop initiatives, create supporting resources, and/or evaluate the redesign of learning, teaching, and assessment (LTA) strategies and the development of AI literacy among their students and staff.

Step 6: Creating Short-Term Wins

Creating and celebrating short-term wins provides emotional rewards and builds momentum for change (Appelbaum et al., 2012; Kotter, 1996). In the University, short-term wins were embedded within the assessment redesign timeline as departments were expected to complete assessment redesigns before the 2023/24 academic year. The EDC reviewed the collected redesigned assessment tasks and identified teachers who had effectively integrated GenAI. These teachers were invited to share their insights and experiences through various channels. Showcasing good practices and successful case studies not only acknowledges teachers' innovative work in LTA, providing them with a sense of accomplishment, but also motivates other teachers to adopt similar changes.

Step 7: Consolidating Improvements

Kotter (1996) argues that success fosters further success. He also emphasised the importance of reinforcing changes with new systems and policies. The EDC updated the Honour Declaration Form for Submitting Assignments in June 2023 to facilitate GenAI usage declaration. The form was further enhanced in June 2024 to reinforce academic integrity. Another example is the regular review of teacher guidelines to reflect rapid technological advancements, such as the use of GenAI in providing feedback.

Step 8: Institutionalising New Approaches

Anchoring changes involves normalising the new culture and approaches (Kotter, 1996). GenAI became the 2023/24 university's strategic pedagogic development focus. The LTC revised departmental guidelines to include teaching innovation as a KPI. Updating and expanding the student declaration form on the use of GenAI in assessment (mentioned in Step 7) is another example. The University's new strategic plan for 2025–31, approved in June 2024, further demonstrates its commitment to these changes. One strategic goal specifically aims to "Develop the technical acumen of both students and staff and promote the use of generative AI and other advanced technologies to support student learning". These measures aim to ensure the successful integration of GenAI into education and assessment over the next five years.

Discussion and Implications

Through the lens of Kotter's change model, the University has effectively created urgency, formed a vision, and empowered staff to take actions. This effectiveness could be attributed to the timely institutional responses following the emergence of ChatGPT in late 2022, including the establishment of the university stance and guidelines, early professional development, seed funding

schemes, and assessment support structures, which helped lower barriers to staff participation in assessment redesign.

Retrospective analysis reveals the need for an earlier, interdisciplinary working group. While the EDC responded promptly, a broader coalition of an interdisciplinary working group could have provided more timely and effective support across the institution. A second area for improvement lies in communicating the vision. While the need for assessment redesign in the age of GenAI was conveyed to teachers through teacher GenAI guidelines and promoted through various channels, adoption was inconsistent across the university. For example, despite availability since June 2023, some teachers remained unaware of the updated Honour Declaration Form in late 2024, highlighting critical communication gaps. This suggests that future efforts could build on existing communication approaches by complementing dissemination with more iterative and dialogic engagement, so as to support ongoing sense-making and foster more consistent uptake across the institution.

Finally, regarding anchoring changes in the university culture, while GenAI integration into assessment has begun, it is not yet clear that these changes are fully embedded across the university. A shift in focus from initial implementation to long-term impact is needed. This entails actively tracking student outcomes and the effectiveness of assessment changes to demonstrate the benefits and positive impact of GenAI-enabled assessment reforms on both the institution and student learning.

Conclusion

As a single-institution case study, this paper provides a rich and context-specific account of institutional change processes, offering practical insights for other universities navigating GenAI-related assessment reform. Kotter's model proved effective for analysing the University's GenAI assessment reform, highlighting how initiatives established urgency, coalitions, and vision while revealing areas for sustaining long-term change. Our analysis revealed the strengths and areas for further development to sustain assessment reform. Kotter's change model can serve as a common analytical framework for comparative studies of different universities' efforts to address GenAI challenges. Results from such studies can shed light on best practices, common challenges, and effective strategies for addressing GenAI challenges, which in turn can inform and enhance GenAI integration in diverse educational contexts.

Generative AI Policy

The author(s) used Gemini 3 Pro and Microsoft Copilot during the prepa-

ration of this work to refine and enhance the language quality. After utilising the tool, the author(s) thoroughly reviewed and edited the content as necessary and assumed full responsibility for the publication's content.

References

Appelbaum, S. H., Habashy, S., Malo, J.-L., & Shafiq, H. (2012). Back to the Future: Revisiting Kotter's 1996 Change Model. *Journal of Management Development, 31*(8), 764–782. https://doi.org/10.1108/02621711211253231

Besterfield-Sacre, M., Cox, M. F., Borrego, M., Beddoes, K., & Zhu, J. (2014). Changing engineering education: Views of US faculty, chairs, and deans. *Journal of Engineering Education, 103*(2), 193-219. https://doi.org/10.1002/jee.20043

Chen, J. (2023, July 21). Four directions for assessment redesign in the age of generative AI. *THE Campus*. https://www.timeshighereducation.com/campus/four-directions-assessment-redesign-age-generative-ai

Chen, J. (2024, April 18). The three questions that will define the future of higher education. *THE Campus*. https://www.timeshighereducation.com/campus/three-questions-will-define-future-higher-education

Chen, J., Lo, F.Y., & Chon, L. (2024). Redesigning and implementing GenAI-expected assessment: Strategies developed, features implemented, good practices and teacher development. Presentation at the Digital Learning Week 2024 of United Nations Educational, Scientific and Cultural Organization (UNESCO), Paris, France.

Guzman, W. Z., Gely, M. I., Crespo, K., Matos, J. R., Sanchez, N., & Guerrero, L. M. (2011). Transformation of a Dental School's Clinical Assessment System through Kotter's Eight-Step Change Process. *Journal of Dental Education, 7*(4), 485–495.

Kang, S. P., Chen, Y., Svihla, V., Gallup, A., Ferris, K., & Datye, A. K. (2022). Guiding change in higher education: An emergent, iterative application of Kotter's change model. *Studies in Higher Education, 47*(2), 270-289.

Kotter, J. P. (1996). *Leading change*. Harvard Business School Press.

Potter, J., Welsh, K., & Milne, L. (2023). Evaluating an institutional response to generative artificial intelligence (GenAI): Applying Kotter's change model and sharing lessons learned for educational development. *Journal of Perspectives in Applied Academic Practice, 11*(3). https://doi.org/10.56433/jpaap.v11i3.582

Springer, P. J., Clark, C. M., Strohfus, P., & Belcheir, M. (2012). Using Transformational Change to Improve Organizational Culture and Climate in a School of Nursing. *Journal of Nursing Education, 51*(2), 81–88. https://doi.org/10.3928/01484834-20111230-02

Steinert, Y., Cruess, R. L., Cruess, S. R., Boudreau, J. D., & Fuks, A. (2007). Faculty Development as an Instrument of Change: A Case Study on Teaching Professionalism. *Academic Medicine, 82*(11), 1057–1064. https://doi.org/10.1097/01

Wentworth, D., Behson, S. J., & Kelley, C. L. (2018). Implementing a new student evaluation of teaching system using the Kotter change model. *Studies in Higher Education, 45*(3), 511-523. https://doi.org/10.1080/03075079.2018.1542434

Investigation on Learning Enhancement through AI-empowered Immersive Environment

Ping GENG [1]
[1] Department of Food Science and Nutrition, The Hong Kong Polytechnic University, Hong Kong, China

Abstract

Traditional lecture-based education exhibits well-documented limitations, including relatively short knowledge retention times, low student engagement, and limited personalisation. This study presents a supplementary approach that enhances knowledge retention through engaging and immersive experiences. Piloted in the Department of Food Science and Nutrition, a metaverse platform integrating AI-driven virtual agents was developed to enhance students' understanding and motivation. Through surveys and focus group interviews with two cohorts of Year-3 students (N = 92) and analysis of 171 AI avatar responses in a representative food production scene, we find that students value the platform's interactivity, hands-on engagement, and novelty, reporting improved comprehension and enthusiasm over traditional methods. Thematic analysis highlights the AI avatars' supportive and explanatory communication style, which contrasts with students' short and fragmented inputs. Results also indicate students' preference for different immersion levels and provide valuable insights for the development of more engaging, personalised, and effective learning environments in higher education.

Keywords: AI agents, Immersive learning, Personalised learning pathway

1. Introduction

In the learning of science and engineering-related subjects, several challenges exist that need to be addressed. First, most subjects are lecture-based and exam-oriented, resulting in low knowledge retention after exams (Maceiras et al., 2025). Second, students crave more interactive and individualised approaches to learning (Walkington, 2013). Third, traditional learning materials may not sufficiently explain key concepts in vivid and interesting ways. For example, in the food technology domain, it is difficult for students to understand industrial food processing practices using only flowcharts, pictures, or videos. An immersive learning experience would be much more helpful for their knowledge construction process when the students behave as food engineers in food manufacturing factories (Gorman et al., 2022).

To address these challenges and help students achieve a higher level of learning beyond simple "remembering and understanding", this project aims to create a unique, engaging, interactive, and immersive metaverse learning environment through a VR-based metaverse empowered by AI. The metaverse design is grounded in the theoretical foundations of: (1) gaming elements in immersive learning to encourage problem-solving (Mochizuki et al., 2021); (2) narratives in stories to promote sustained attention and deeper cognitive engagement (Dahlstrom, 2014); and (3) AI-powered virtual agents to simulate complex human interactions and provide timely and customised feedback (Zhou et al., 2020).

This study investigates three key research questions:

1. Can immersive technologies like the metaverse significantly enhance students' learning engagement compared with traditional educational methods?
2. How do students interact with AI in an immersive environment?
3. How do students' acceptance levels differ between VR and desktop/mobile devices for the exploration of a metaverse environment, given their different levels of immersion?

Through exploring these questions with empirical evidence, we can gain a deeper understanding of students' perceptions in immersive, AI-powered learning environments and optimise platform design to enhance students' learning experience. The findings from this study could provide insights for the development of more engaging, personalised, and effective educational platforms for a wide range of science and engineering disciplines.

2. Methodology

2.1 Participants

Participants (N = 92) in this study comprised two cohorts (23/24 and 24/25) of Year-3 students in the Department of Food Science and Nutrition taking the subject *Food Engineering and Processing II*. The gender distribution was balanced overall (48 Female / 44 Male), with an age range of 18–25 years.

2.2 Platform Design

This study has developed two metaverse scenes (a corn starch manufacturing plant and a beer brewery plant), each featuring different food production processes with over 100 smart non-player characters (NPCs). Students interacted with NPCs in the metaverse platform by tapping interactable icons on desktop computers or VR controllers (Figure 1a–d), which triggered dialogues and allowed for decision-making. The NPCs' responses and guidance were dy-

namically tailored to the students' choices and actions, leading them through a sequence of approximately 30 tasks within each scene, enabling a personalised learning experience. A representative scene design is shown in Figure 1e–f.

Meanwhile, AI-powered avatars were integrated into the metaverse scenes, allowing students to interact with the avatars using text or voice and receive answers to any inquiries during their exploration (Figure 1h–i). The design of these GPT avatars utilises proprietary prompt templates to ensure that avatars delivered content relevant to the learning objectives. Instructors could easily modify the questions or define the AI avatars through the authoring tool in the platform (Figure 1g).

2.3 Implementation

During the implementation, students experienced both scenes: one with VR headsets (Meta Quest 3) and the other with less immersive desktop computers, for around 40 minutes respectively. After experiencing two metaverse scenes in two different modes, students were invited to complete a semi-structured survey designed to evaluate their user experience and acceptance of the metaverse platform. The survey incorporated both 5-point Likert scale questions and open-ended questions based on the modified Technology Acceptance Model (TAM) for immersive design (Fussell & Truong, 2022). A focus group interview was also conducted with 20 randomly selected participants.

2.4 Data Analysis

Descriptive analysis and T-tests were employed to identify distribution differences in the constructs of TAM and preference in different immersion levels (VR and desktop computers). Meanwhile, thematic analysis informed by word frequency counts was used to analyse students' feedback in focus group interviews and identify key concepts and patterns in the AI avatar's replies.

3. Results and Discussion

3.1 Qualitative Feedback on the Platform's Strengths and Weaknesses

The qualitative results came from both the open-ended qualitative responses in the survey and subsequent focus group interviews through thematic analysis. The representative positive feedback could be divided into the following three categories:

(1) Interactivity and hands-on engagement: Students appreciated the active, task-based nature of the immersive learning environment. They enjoyed opportunities to physically interact with content rather than passively absorbing information. Representative comments include:

- "Fun and more interactive."
- "I like interactive tasks, like grabs and selection, better than MCQ and text. I think it can have more interactive tasks and animations."

(2) Enhanced learning and comprehension: The participants noted that the immersive learning experience enhanced their understanding of information compared to traditional methods. The ability to visualise and handle simulated equipment was highlighted as a key benefit. Representative comments include:

- "Memorising the information was easier because you see the actual equipment."
- "I like the grabs and while I was doing it I felt I was really there. It's better than reading the text and it helps me to learn the process."
- "VR simulation... helped me to see the whole process that I can't see in real life and it enhances my learning."

(3) Novelty and engagement in the learning experience: Many student participants expressed enthusiasm for the fresh, innovative approach of the AI-empowered metaverse platform. Representative comments include:

- "I think it's really fresh and I prefer VR to the traditional way."
- "Walking in a factory doing this thing and that thing... it's interesting and better than reading the text."

Meanwhile, we also received students' feedback indicating challenges for further improvement. The weaknesses mainly involve dizziness issues and technical difficulties when using VR headsets, as well as relatively slow response times and fast speaking speeds of AI avatars.

3.2 Gen-AI Usage in the Metaverse Platform

We analysed 171 replies generated by the Gen-AI avatar in the scene of the beer brewery plant after cleaning (lowercasing, removing punctuation, URLs, and stop words). It was discovered that on the avatar side, all 171 replies were analyzable and were substantially longer (median 12 content words, M = 12.85). The vocabulary was strongly dominated by explanatory brewing language (*fermentation, process, purpose, mashing, primary, secondary, convert, sugar, enzyme, alcohol*), reflected enthusiastic introductions ("Great! Let's dive into…"), and supportive context management ("I understand you're stepping away…").

Through comparing the input from students and replies from Gen-AI, results show that students bring short, sometimes fragmented or emotional input ("What is the difference between …", "Uh primary fermentation is …, right",

"Pretty good"), whereas the Gen-AI avatar responds with longer, more uniformly structured, and generally positive explanations. The replies not only mirror the brewing content but also add layers of encouragement, clarification, and repair, especially when students appear confused or frustrated.

3.3 Comparison Between Different Immersion Modes

The evaluation survey was developed on the basis of the modified Technology Acceptance Model (TAM) (Fussell & Truong, 2022) to identify students' preference on different immersion levels (VR and desktop computers). Key aspects include the functionality of interactive elements and user interface (TECH), attitudes towards learning in the metaverse (ATT), ease of use (EASE), enjoyment (ENJ), perceived productivity (PERF), perceived usefulness (USE), intent to use (INT), and usage behaviour (BEH).

As shown in Table 1, notable device-based differences could be observed. Desktop users reported significantly higher ratings across multiple constructs, including behavioural intention (BEH) (4.023 vs. 3.375) and perceived usefulness (USE) (4.227 vs. 3.438), highlighting a preference for desktop interactions. Additionally, ease of use (EASE) was rated higher for desktop users (4.182) compared to VR headset users (3.500), indicating that desktop-based interactions were perceived as more intuitive and accessible. T-test results further exhibit the significant differences, with desktop users reporting higher usage behaviour (BEH) ($p = 0.027$), enjoyment (ENJ) ($p = 0.014$), ease of use (EASE) ($p = 0.014$), and perceived usefulness (USE) ($p = 0.008$) than VR headset users. The comparison suggests that students using desktop computers experience greater independence and enjoyment on the metaverse platform, feel more confident in using it, and perceive it as more useful compared to those using VR headsets.

4. Implications

The application of an interactive metaverse with AI-empowered avatars offers a promising enhancement to traditional lecture-based teaching, providing students with an engaging and enjoyable learning experience that reinforces their knowledge. Positive student feedback, along with the increase in eSFQ scores from 4.3 in 2022/23 to 4.6 in 2023/24 and 4.7 in 2024/25, suggests that this approach can be further leveraged to maximise its benefits, particularly in supporting knowledge construction through ongoing reinforcement.

Here are some recommendations based on the findings from this study:

- While the current design incorporates multiple knowledge delivery methods (MCQs, videos, drag-and-drop interactions), students ex-

pressed a preference for interactive tasks. Therefore, future metaverse design should prioritise user-friendly interactive elements to enhance learning effectiveness (Ali, 2025).

- Statistical analysis revealed that desktop users reported significantly higher behavioural intention, enjoyment, and ease of use compared with VR users. Although students acknowledged VR's novelty, discomfort and dizziness remained concerns (Rebenitsch & Owen, 2016). A desktop or mobile version of the metaverse could offer greater accessibility, reduce motion sickness, allow flipped learning before class, and self-tutoring after class.
- In addition, AI avatars increase the interactivity of the metaverse platform. Due to students' varied requirements for AI avatars, efforts could be made to further customise the styles of AI avatars based on students' individual needs, creating a more personalised learning experience (Kim & Baylor, 2016).

Acknowledgments

This work was supported by PolyU Teaching Development Grant (Grant No. 49SZ) and Fund for Innovative Technology-in-Education-Inter-Institutional Collaborative Activities (FITE-IICA) (Grant No. 89VR). The authors gratefully acknowledge the financial support provided by both projects.

Ethical Declaration Statement

The project has been approved by the Institutional Review Board of The Hong Kong Polytechnic University (Reference Number: HSEARS20230808006).

Declaration of Use of AI in the Writing Process

The author(s) used PolyU GenAI during the preparation of this work to assist with grammar polishing. After utilising the tool, the author thoroughly reviewed and edited the content as necessary and assumed full responsibility for the publication's content.

References

Ali, M. (2025). Fostering Engagement and Motivation: Evaluating Gamified Learning in the Metaverse for Architectural History Education. *International Journal of Interactive Mobile Technologies*, 19(6). https://doi.org/10.3991/ijim.v19i06.53841

Dahlstrom, M. F. (2014). Using narratives and storytelling to communicate science with nonexpert audiences. *Proceedings of the National Academy*

of Sciences, 111(supplement_4), 13614–13620. https://doi.org/10.1073/pnas.1320645111

Fussell, S. G., & Truong, D. (2023). Accepting virtual reality for dynamic learning: An extension of the technology acceptance model. *Interactive Learning Environments*, 31(9), 5442–5459. https://doi.org/10.1080/10494820.2021.2009880

Gorman, D., Hoermann, S., Lindeman, R. W., & Shahri, B. (2022). Using virtual reality to enhance food technology education. *International Journal of Technology and Design Education*, 32(3), 1659–1677. https://doi.org/10.1007/s10798-021-09669-3

Kim, Y., & Baylor, A. L. (2016). Research-Based Design of Pedagogical Agent Roles: a Review, Progress, and Recommendations. *International Journal of Artificial Intelligence in Education*, 26, 160–169. https://doi.org/10.1007/s40593-015-0055-y

Maceiras, R., Feijoo, J., Alfonsin, V., & Perez-Rial, L. (2025). Effectiveness of active learning techniques in knowledge retention among engineering students. *Education for Chemical Engineers*, 51, 1–8. https://doi.org/10.1016/j.ece.2025.01.003

Mochizuki, J., Magnuszewski, P., Pajak, M., Krolikowska, K., Jarzabek, L., & Kulakowska, M. (2021). Simulation games as a catalyst for social learning: The case of the water-food-energy nexus game. *Global Environmental Change*, 66, 102204. https://doi.org/10.1016/j.gloenvcha.2020.102204

Rebenitsch, L., & Owen, C. (2016). Review on cybersickness in applications and visual displays. *Virtual Reality*, 20, 101–125. https://doi.org/10.1007/s10055-016-0285-9

Walkington, C. A. (2013). Using adaptive learning technologies to personalize instruction to student interests: The impact of relevant contexts on performance and learning outcomes. *Journal of Educational Psychology*, 105(4), 932. https://doi.org/10.1037/a0031882

Zhou, L., Gao, J., Li, D., & Shum, H. Y. (2020). The design and implementation of xiaoice, an empathetic social chatbot. *Computational Linguistics*, 46(1), 53–93. https://doi.org/10.1162/coli_a_00368

Table 1: Descriptive ratings and T-test result of VR and desktop/mobile modes of metaverse learning environments

Constructs	Desktop/mobile devices		VR headsets		p
	Mean	SD	Mean	SD	
BEH	4.023	0.794	3.375	1.182	0.027
ENJ	4.136	1.049	3.438	1.116	0.014
ATT	4.250	0.856	3.604	1.170	0.030
INT	4.136	0.743	3.417	1.204	0.019
EASE	4.182	0.839	3.500	1.063	0.014
USE	4.227	0.736	3.438	1.126	0.008
TECH	3.886	0.912	3.563	1.097	0.177
PERF	4.114	0.925	3.438	1.155	0.019

Note: N = 92. Participants were on average 21 years old with 52% female participants.

Abbreviations: BEH: usage behaviour; ENJ: enjoyment; ATT: attitudes towards learning in the metaverse; INT: intent to use; EASE: ease of use; USE: perceived usefulness; TECH: interactive elements and user interface; PERF: perceived productivity.

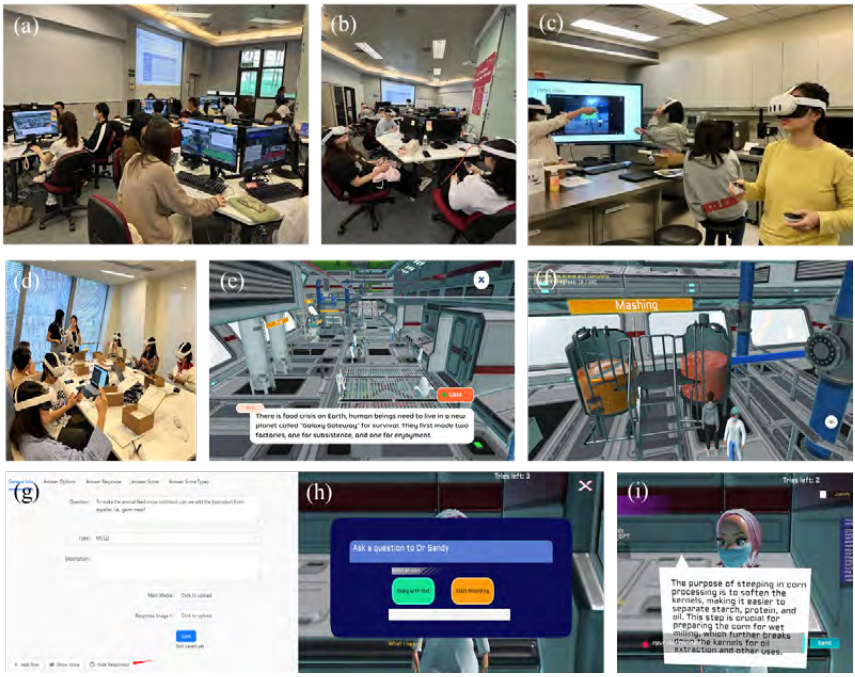


Figure 1: AI-powered metaverse environment.

(a)–(d) Implementation of metaverse platform for immersive learning.

(e)–(f) Part of the immersive learning environment design for food science education.

(g) Authoring tool interface allowing the modification of questions by instructors.

(h)–(i) AI-powered avatars in the metaverse scene for personalised learning.

Exploration and Practice of Teaching Methods Assisted by Artificial Intelligence

Jing LEI [1] & Ying HUANG [1]
[1] College of Electronic Science and Technology, National University of Defense Technology, China

Abstract

This paper focuses on the theory and methods of smart teaching, refining and forming a "Six-in-One" teaching methodology. It proposes a practical approach system for smart teaching, which includes: AI-assisted autonomous learning; personalised learning path planning guided by agents; dynamic classrooms and deep interaction; problem-guided integration of teaching and research; real-time monitoring under a task-driven framework; and accurate assessment guided by learning situation analysis. This system effectively integrates "knowledge graphs and large models" with various teaching methods (such as problem-based, task-based, and seminar-based learning), thereby bridging the gap between theoretical methodologies and practical applications. We carried out reform practices over three semesters among students majoring in electronic information. A survey was conducted among over 100 students using a questionnaire encompassing four core competencies and ten questions. The statistical results indicated that the aforementioned method can effectively enhance students' competency development.

Keywords: Smart Teaching; Knowledge Graphs; Large Models; Problem-Based Learning; Task-Based Learning

1. Introduction

With the strategic deployment outlined in the *Outline of the Plan for Building a Strong Education Country (2024–2035)* and the advancement of digital transformation, the ternary interactive teaching model based on "Teacher-Student-AI" has prompted a shift in the role of teachers, from mere knowledge transmitters to intelligent guides in the learning process. As an "intelligent node", AI can interact with students in real time, dynamically adjust learning strategies, and usher in a new era of personalised and immersive learning (Li, Z., 2025). Adhering to the "student-centred" educational philosophy, constructing a smart teaching system and method remains a challenge for higher education.

Fudan University has formed three circles in AI-empowered discipline in-

tegration: the core circle, the advanced circle, and the cultivation circle (Jiang, Y., 2025). In AI-empowered educational and teaching practice, Jiang, X. et al. (2025) deeply analysed the basic characteristics of smart education in universities and explored ways to integrate digital twin technology into teaching. Xu, X. and Zhang, C. (2025) explored how Generative AI empowers higher engineering education under the goal of capability cultivation, proposing specific measures for constructing a capability cultivation system and an ability level evaluation and certification system. Bao, N. (2022) proposed the M-PERT (Management-Pedagogic-Environment Resources-Target) practice model of smart teaching. Cui, Y. (2023) explored the TPDCK (Teacher Professional Development Content Knowledge) knowledge framework for teacher transformation under smart teaching from the perspective of teachers. Measures and implementation methods for the reform of smart teaching models in higher education have been explored by Zhao, D. (2024), Gui, X., and He, Q. (2024).

From current research results, the "last mile" of smart teaching—specifically its implementation in course teaching—has not yet been fully achieved. Our research focuses on addressing issues regarding method construction and strategy implementation for smart teaching. The main innovations and contributions of this paper are as follows:

1. We have formulated a "Six-in-One" teaching method for smart teaching, promoting achievements presented in literature (Bao, N., 2022).
2. We have established a practical approach system for smart teaching that breaks the barriers between theoretical methods and practical approaches.
3. The evaluation of reform effects was conducted through questionnaire surveys.

2. Methodology

2.1 Statement of the Research Question

Our research question can be characterised as exploring how to build the relationship between teaching methods and practical approaches in smart teaching, so as to enhance the applied value of smart teaching.

2.2 Research and Exploration of Solutions

Relying on the development of artificial intelligence technology, smart teaching has been endowed with new vitality. A teaching method of "Six-in-One" has been summarised and formed (see Table 1).

Table 1: Characteristics of Teaching Methods

No.	Method name	Method features	AI empowerment
1	The diversification of learning	Through diverse forms including projects, tasks, and problems, we optimize the learning process.	Leverage AI to access high-quality learning resources and experience a comprehensive learning process.
2	Innovation in teaching mode	Effectively and deeply integrate technology with teaching	Knowledge graphs and large models can be used.
3	Task-based teaching activities	Through task - based teaching activities, students gain a sense of achievement, and realize a learning - centered and active - construction - oriented approach.	With AI assistance, students' high-order cognitive abilities can be cultivated.
4	Deepening teacher-student interaction	Implement the "Teacher-Student-AI" ternary collaborative model.	AI-enabled tools can be utilized to deepen different interaction.
5	Implement personalized teaching	We should meet students' individual differences and highlight their individuality during the teaching process.	AI can assist in formulating personalized teaching plans and assistance.
6	Precise teaching evaluation	Conduct multi-dimensional and precise evaluations.	Under the drive of big data, precise assessment is carried out to construct student profiles.

Our implementation integrates the 'Teacher-Student-AI' collaborative model to create a structured approach system (Figure 1) featuring six practical methods organised into three hierarchical tiers: student-AI engagement, teacher-student-AI collaboration, and teacher-AI interaction.

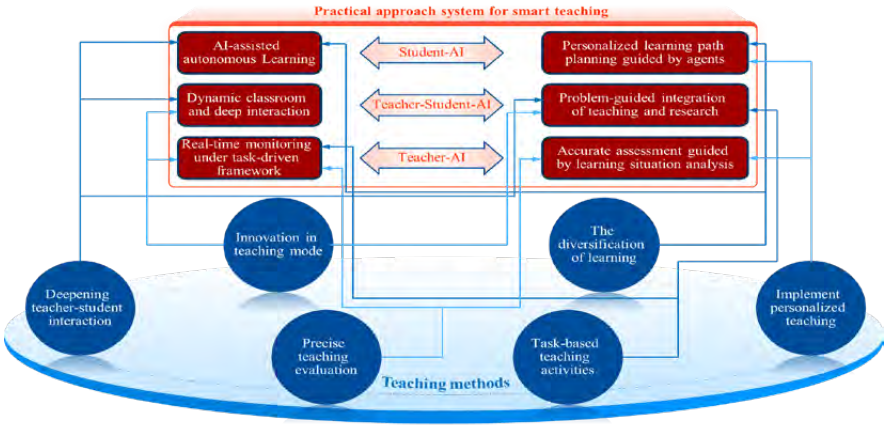


Figure 1: The practical approach system for smart teaching supported by teaching methods.

- **AI-assisted autonomous learning:** Based on the knowledge graph, students can engage in autonomous learning centred around knowledge points.
- **Personalised learning path planning guided by agents:** Utilise intelligent agents as teaching assistants to guide students in self-exploration and self-construction, and facilitate learning and reflection based on individually planned paths.
- **Dynamic classroom and deep interaction:** Teachers utilise AI-powered classroom interaction tools to design instructional activities, encouraging students to think critically and provide feedback.
- **Problem-guided integration of teaching and research:** By leveraging interactions among teachers, students, and AI, we create a research-oriented and exploratory classroom atmosphere, with AI facilitating the implementation of case-based teaching and the advancement of problem exploration.
- **Real-time monitoring under task-driven framework:** By monitoring students' task progress, teachers gain insights into each student's learning status. Special attention and reinforcement are given to key individuals or challenging topics.
- **Accurate assessment guided by learning situation analysis:** By leveraging artificial intelligence, we conduct in-depth analysis of students' learning conditions to optimise instructional design. Precision evaluations are carried out based on students' attainment of learning abilities, with the aim of refining learning paths and teaching tasks.

2.3 Experimental Design

Among undergraduate students majoring in electronic information at our university in the 2022 cohort, 100 students were selected as the subjects of analysis. We conducted questionnaire surveys among the same group of students after the completion of the courses *Principles of Communications* (Fall 2024), *Fundamentals of Information Theory and Coding* (Spring 2025), and *Fundamentals of Wireless Communications* (Fall 2025), respectively.

The specific questions along with their corresponding value labels (APi_Qj) are presented in Table 2.

Table 2: Survey Questionnaire Content

No.	Ability point	Corresponding problem	Answer options
1	Integration of knowledge (*AP*1)	I can clearly trace the logical connections between various knowledge points in this course, and no longer understand and memorize them in isolation. (*AP*1_*Q*1)	Yes/No (1/0)
2		I am able to effectively correlate the knowledge points learned in this course with those from other related courses. (*AP*1_*Q*2)	Yes/No (1/0)
3		When facing new and complex problems, I am able to flexibly apply knowledge from multiple chapters or disciplines to conduct reasonable analysis. (*AP*1_*Q*3)	Yes/No (1/0)
4	Cultivation of scientific thinking (*AP*2)	Facing unknown problems, I am able to conduct critical thinking with the assistance of AI, objectively evaluating the source of information, the sufficiency of evidence, and the reasonableness of conclusions. (*AP*2_*Q*1)	Yes/No (1/0)
5		When analyzing data or phenomena, I am able to grasp the essential laws of things and reduce subjective assumptions based on my knowledge system and reasoning logic. (*AP*2_*Q*2)	Yes/No (1/0)
6		After undergoing relevant training, I have developed the habit of questioning, verifying, and reflecting in scientific research, and have applied this habit to relevant scientific research practices. (*AP*2_*Q*3)	Yes/No (1/0)
7	Communication and interaction skills (*AP*3)	In team discussions or group collaborations, I am able to articulate my viewpoints clearly and accurately, listen effectively to others' opinions, and provide constructive feedback. (*AP*3_*Q*1)	Yes/No (1/0)
8		I am capable of effective communication, utilizing various forms such as charts, language, and body language, based on the characteristics of the communication partner, to better describe complex issues, knowledge, or solutions. (*AP*3_*Q*2)	Yes/No (1/0)
9	Applied innovation competence (*AP*4)	I am capable of applying the theoretical knowledge I have learned to solving specific problems in real life or work, formulating feasible solutions, and possessing the awareness and ability to continuously optimize. (*AP*4_*Q*1)	Yes/No (1/0)
10		In project practice or case analysis, I am able to identify problems in a timely manner and proactively seek solutions, possessing the ability to transform creativity into specific and executable plans. (*AP*4_*Q*2)	Yes/No (1/0)

2.4 Method of Result Analysis

The questionnaire was distributed via Wenjuanxing. A student's response of "Yes" is recorded as 1, and a response of "No" is recorded as 0. Normalised mean competency scores for each ability point can be calculated by Equation (1):

$$APi = \frac{\sum_{s=1}^{Ns}\sum_{j=1}^{Ni} APi_Qj(s)}{Ns \times Ni}$$

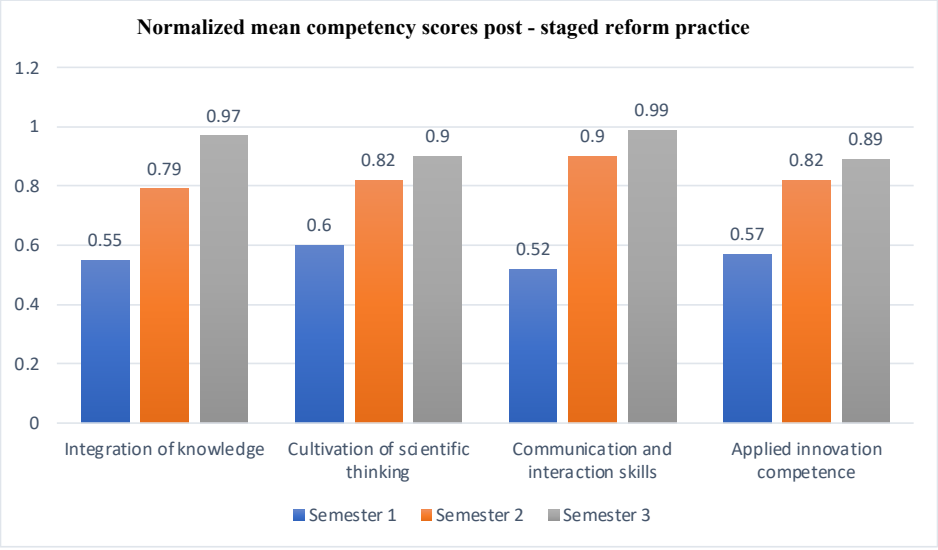
Here, APi_Qj(s) represents the answer value given by the *s*-th student for the *j*-th question regarding the *i*-th ability point. *Ni* represents the total number of questions included in the *i*-th ability point, and *Ns* denotes the total number of students who participated in the questionnaire.

3. Results and Discussion

3.1 Statement of Key Results

Figure 2 presents the evaluation results of core competencies. The results of the first semester showed a low level of achievement, with a notable increase in the second semester and a continued rise in the third semester. When comparing the results of the first semester with those of the third semester, the largest increase was 47% (for communication and interaction skills), while the smallest was 30% (for the cultivation of scientific thinking).

Figure 2: Normalised mean competency scores post-staged reform practice.



3.2 Discussion

The main reasons for the low level of achievement in the first semester were as follows:

1. The hardware platform for smart teaching was not sufficiently developed to fully support the rollout of teaching reforms.
2. The proportion of teaching reforms implemented in the assessment course for Fall 2024 was unsatisfactory.

From the second semester onward, both of these issues were effectively resolved. Meanwhile, teachers' AI literacy also demonstrated a marked improvement. Therefore, a comparison of the results from the subsequent two semesters with those of the first semester reveals that the practical approaches to smart teaching proposed in this study can be effectively implemented in real-world teaching scenarios.

3.3 Critical Evaluation

However, we must also objectively point out the following limitations:

1. There was no control group in our study. Nevertheless, by conducting the same questionnaire survey on the same group of students over three consecutive semesters, we were still capable of revealing the changing trend of the results.
2. The rationality of the evaluation indicators and the hierarchy of the issues currently require further improvement; these will be investigated as specific issues in future research.
3. Due to the inadequate number of survey respondents and time constraints, our team will conduct further follow-up on effect evaluation and correction.

4. Implications

With the continuous advancement of digital and intelligent transformation in education, this paper primarily explores the practical implementation of smart teaching, culminating in the formulation of a "Six-in-One" smart teaching method. By integrating intelligent teaching tools such as knowledge graphs and AI large language models, we propose a systematic framework for smart teaching implementation. This framework bridges the gap between theoretical methodologies and practical pathways. After three semesters of reform practice, the questionnaire results demonstrate that the smart teaching reform pathways proposed in this paper are effective.

Generative AI Policy

AI tools were exclusively employed for language refinement and polishing during the English writing and revision stages of this manuscript. No AI-generated content was used to replace critical intellectual contributions, including methodological research, teaching design, data analysis, or interpretation of results.

References

Bao, N. (2022). *Study on the construction of the practice model of smart teaching in the universities* [Doctoral dissertation, Northeast Normal University]. CNKI. https://link.cnki.net/doi/10.27011/d.cnki.gdbsu.2022.000106

Cui, Y. (2023). *A research on construction and application of TPDCK knowledge framework under smart educational environment* [Doctoral dissertation, Northeast Normal University]. CNKI. https://link.cnki.net/doi/10.27011/d.cnki.gdbsu.2023.000167

Gui, X., & He, Q. (2024). Exploration of systematic reform of university computer general education empowered by AI. *China University Teaching, (4)*, 2, 4–11. https://doi.org/10.3969/j.issn.1005-0450.2024.04.002

Jiang, X., Shao, M., & Shi, J. (2025). Empowering smart education in universities with "AI+Digital Twins": Iterative integration, application scenarios, and practical approaches. *Journal of Heilongjiang Institute of Technology, 39*(1), 66–70, 80. https://doi.org/10.19352/j.cnki.issn1671-4679.2025.01.011

Jiang, Y. (2025). Artificial intelligence empowers high-quality development of higher education. *China Education Network, (1)*, 1–3. https://qikan.cqvip.com/Qikan/Article/Detail?id=7200858472

Li, Z. (2025). Artificial intelligence and higher education: The symbiosis of technical rationality and humanistic value. *China Education Network, (Z1)*, 5–7. https://qikan.cqvip.com/Qikan/Article/Detail?id=7201149484

Xu, X., & Zhang, C. (2025). Generative AI empowers engineering education and student ability cultivation, assessment and certification systems. *Research in Higher Education of Engineering, (4)*, 1–9. https://link.cnki.net/urlid/42.1026.G4.20250624.1120.044

Zhao, D. (2024). Research on the digital intelligent teaching mode of higher education from the perspective of standardization. *China Standardization, (22)*, 241–243. https://doi.org/10.3969/j.issn.1002-5944.2024.22.044

Boosting Creative Learning with Gen-AI-Enabled Semi-Flipped Classes and Project-Based Design Thinking Method

Nan Nathan XU [1] & Arianna BIONDA [2]
[1] Department of Mechanical Engineering, The Hong Kong Polytechnic University, HKSAR
[2] Department of Design, Politecnico di Milano, Italy, Europe

Abstract

This study centres on Generative AI (GenAI) as a core enabling pedagogical tool to examine how integrating GenAI-enabled semi-flipped classes and the Project-Based Design Thinking Method fosters students' creative learning, interdisciplinary collaboration, and real-world problem-solving. Conducted at a local university, this study analyses GenAI-integrated course methodology, student performance, feedback, and iterative improvements across the 2023 (non-GenAI) and 2025 (GenAI-enabled) cohorts. GenAI is embedded as a foundational enabler across the entire pedagogical lifecycle, applied to three core and interconnected pedagogical links with explicit design logic and measurable outcomes: AI-aided interdisciplinary team formation (via ChatGPT + MBTI personality analysis tools), AI-enhanced pre-class knowledge scaffolding (via domain-specific GenAI platforms), and AI-supported project iteration (via design prototype feedback tools). This work demonstrates that GenAI-enabled hybrid pedagogy is a highly effective method in nurturing core competencies for interdisciplinary youth aiming to address global challenges, and provides a replicable GenAI integration framework for higher education course design.

Keywords: Generative AI (GenAI), GenAI-Enabled Method, Design Thinking, Flipped Classroom, Interdisciplinary Learning, Higher Education Pedagogy, Creative Problem-Solving

1. Introduction

The current era faces global challenges such as climate change, energy crises, and the societal impact of AI (Yim et al., 2024). However, traditional lecture-based teaching is often fragmented and fails to engage students in active knowledge application, leaving them inadequately prepared for complex real-life issues. *"What is Design and Why Design"* has been iterated to address this gap, framing design as "purposeful creative problem-solving" (Lakey, 2009).

The course adopts a GenAI-enabled hybrid method combining a semi-flipped class (Barrows, 1996) with three integrated components, with GenAI serving as the central enabling layer connecting and optimising the other two:

- **Semi-flipped class** (Barrows, 1996), which delivers pre-class theoretical/practical content with in-class guidance (Bishop & Verleger, 2020), augmented by GenAI for personalised pre-class scaffolding and in-class interactive support;
- **Project-Based Design Thinking** (Bionda & Incitti, 2024), involving collaborative in-class work with teacher facilitation, optimised by GenAI for team formation, ideation, and prototype iteration;
- **GenAI Integration** as the core enabling framework, embedded to systematically resolve 2023 cohort pain points (homogeneous teams, insufficient personalised pre-class support, low project iteration efficiency) with explicit, measurable model design and outcomes.

Focused on the 2025 Spring semester cohort, this study addresses two core questions: How can we structure a GenAI-enabled hybrid model for interdisciplinary problem-solving? How can we enhance the model's effectiveness and replicability via iteration? By comparing it with the 2023 non-GenAI cohort, this study validates a scalable GenAI-enabled hybrid model and provides a replicable GenAI integration blueprint for engineering and design education, with all student performance and feedback data explicitly linked to GenAI's pedagogical impact.

2. Methodology

2.1 Course Methodology Preparation

Figure 1: The 14-week Teaching and Learning Journey of the Subject in the 2025 Spring Semester.

Weeks	1	2	3	4	5	6	7
Lecture	Openning	Design Disciplines	Design History	Survey & Questionnaire	What is a Good Design	Human-Centered Design	Sustainability
GenAI Enaled Labs	NA	**GenAI-Aided** Initial Grouping	Design Analysis	Save a Failure Design	DT-I Empathy	DT-II Define	DT-III Ideate
Milestones	0	5%	10%	15%	20%	25%	30%

Weeks	8	9	10	11	12	13	14
Lecture	Color	Study Break	Design for Water Sports	Q&A	Q&A	NA (Public Holiday)	NA
GenAI Enaled Labs	Midterm (DT-IV Prototyping & Test, **GenAI Feedback Support**)		Color Lab	**GenAI-Aided** Project Time		NA (Public Holiday)	Final Presentation + Individual Report
Milestones	50%		55%	60%	65%	65%	100%

As illustrated in Figure 1, the course was structured as a 14-week GenAI-enabled instructional framework, aligned with the university's standard semester timeline, to balance theory and hands-on practice. Core preparatory measures were all designed with GenAI as a central consideration, including:

- **Universal Learning Outcomes:** Four core transferable competencies (creative problem-solving, interdisciplinary collaboration, systems thinking, prototype development) plus a mandatory GenAI literacy competency (the ability to critically use GenAI tools for design/problem-solving, transferable to global challenge scenarios).
- **Flexible Content Modules:** Design fundamentals, theme-specific topics (nautical ecosystems, water sports; swappable for climate action), and GenAI literacy core modules (ChatGPT/Midjourney instruction, ethical GenAI use in design, validating GenAI outputs against user needs, cross-disciplinary knowledge translation via GenAI).
- **Interdisciplinary Teaching Team:** Three core human faculty members (architectural design, IoT, yacht/marine tech) overseeing GenAI tool selection, integration logic, and faculty/student training, replacing vague "GenAI-encouraged policies" with a formal role for systematic GenAI integration.
- **GenAI-Infused Assessment Rubric:** 35% lab performance, 20% midterm showcase, 40% final showcase, 5% individual report. GenAI-specific criteria were applied to the quality of GenAI-aided team collaboration, critical use of GenAI for prototype iteration, and validation of GenAI outputs against user needs. GenAI served as a cross-cutting evaluation dimension for all project themes.

2.2 Course Running (Pre-Class, In-Class, Post-Class)

The implementation followed a structured yet adaptable three-phase approach, designed to accommodate different real-life pain-point scenarios while maintaining core teaching logic.

2.2.1 Pre-Class: AI-Assisted Team Formation and Background Learning

Students completed a survey (majors, skills, interests) with pre-class learning supported by GenAI-augmented resources (lecture notes + domain-specific GenAI platforms). For team formation, anonymised survey/self-intro data were input into a hybrid GenAI system (ChatGPT + MBTI collaboration tool) (Usahanunth, 2020) that analysed three critical dimensions for interdisciplinary balance:

- **Major Diversity:** ≥2 academic domains per 3–5 student team.
- **Skill Complementarity:** Matching technical skills (prototyping) with soft skills (user research).
- **Personality Compatibility:** Pairing "organiser-type" (ESTJ) and "creative-type" (ENFP) students with GenAI-suggested team roles.

Figure 2: Combined Initial (left) and Updated (right) Grouping of the Subject in the 2025 Cohort.

	Major	MBTI Type	Gen-AI Recommended Role
Group 1	Business	INFP	Designer
	Engineering	ENTP	Project Manager
	Business	ENFJ	Researcher
	Business	INTJ	Developer
	Engineering	ISFJ	Designer
Group 2	Engineering	ESTJ	Project Manager
	Social Science	ESTP	Developer
	Social Science	ISFP	Designer
	Business	ESFP	Researcher
Group 3	Engineering	INFP	Developer
	Business	ESFJ	Researcher
	Engineering	INTJ	Designer
	Business	INTP	Project Manager
	Social Science	INFP	Researcher
Group 4	Engineering	ESFP	Developer
	Business	INTJ	Researcher
	Science	ISTJ	Designer
	Business	ISFJ	Project Manager
	Science	ESTP	Researcher
Group 5	Science	ISFP	Researcher
	Business	INTJ	Designer
	Business	ENFP	Project Manager
	Engineering	ESFJ	Developer
Group 6	Business	ESTP	Designer
	Science	ISFP	Researcher
	Business	ESFJ	Project Manager
	Business	INFP	Designer
	Engineering	INTJ	Developer
Group 7	Business	ESTP	Researcher
	Engineering	ISFP	Developer
	Business	ESTJ	Designer
	Business	INTP	Project Manager
	Business	ENTJ	Designer
Group 8	Business	INFP	Project Manager
	Engineering	INTJ	Developer
	Business	ESFP	Researcher
	Engineering	ISTP	Designer
	Science	ISFJ	Researcher

	Major	MBTI Type	Actual Role for Final Project
Group 1	Business	RESFP	Designer
	Engineering	ENTP	Developer
	Business	ENFJ	Researcher
	Business	INTJ	Project Manager
	Business	ISFJ	Designer
Group 2	Engineering	ESTJ	Project Manager / Developer
	Social Science	ESTP	Withdrawal
	Engineering	ISTP	Designer
	Science	ISFJ	Researcher
Group 3	Engineering	INFP	Developer
	Business	ESFJ	Researcher
	Engineering	INTJ	Designer
	Business	INTP	Project Manager
	Social Science	INFP	Researcher
Group 4	Engineering	ESFP	Developer
	Business	INTJ	Researcher
	Science	ISTJ	Designer
	Business	ISFJ	Project Manager
	Science	ESTP	Researcher
Group 5	Science	ISFP	Designer
	Business	INTJ	Designer
	Business	ENFP	Project Manager
	Engineering	ESFJ	Developer
	Engineering	ISFP	Researcher (Newly joined)
Group 6	Business	ESTP	Designer
	Science	ISFP	Researcher
	Business	ESFJ	Project Manager
	Business	INFP	Designer
	Engineering	INTJ	Developer
Group 7	Business	ESTP	Researcher
	Engineering	ISFP	Developer
	Business	ESTJ	Designer
	Business	INTP	Project Manager
	Business	ENTJ	Designer
Group 8	Business	INFP	Project Manager
	Engineering	INTJ	Developer
	Business	ESFP	Researcher
	Business	INFP	Developer
	Social Science	ISFP	Designer

Business	20
Engineering	11
Social Science	3
Science	5

Figure 2 above shows the initial and final grouping of the 2025 cohort students. Students reviewed GenAI-generated recommendations, adjusted by preference, and finalised with instructors. Few changes occurred (late joins/withdrawals, project-driven role shifts). This GenAI-enabled practice reduced

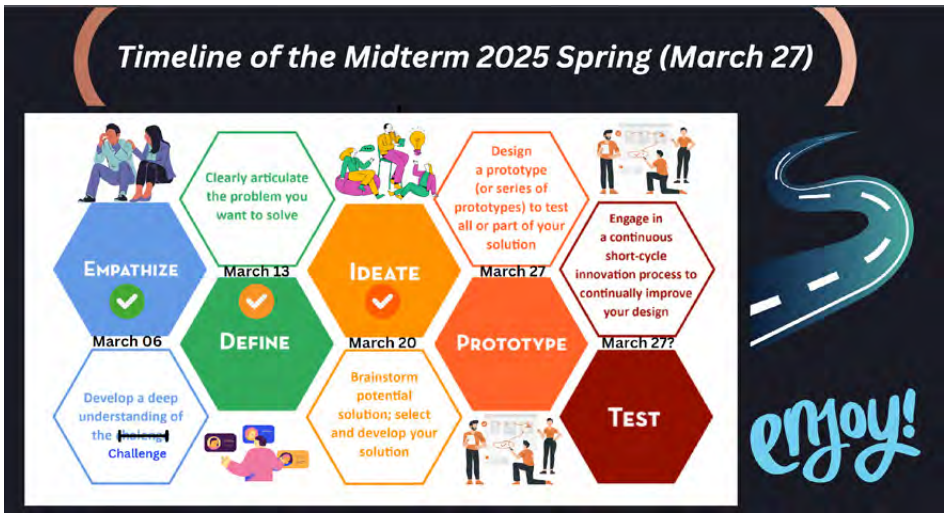
instructor workload by 40% (vs. 2023 manual grouping) and boosted team satisfaction by 28% (SFQ data).

2.2.2 In-Class: Design Thinking Application and Thematic Exploration

In-class activities (Figure 3 on page 5) nurtured three core capabilities with phase-specific GenAI support for the Design Thinking cycle, culminating in a low-fidelity midterm prototype:

- **Foundational Skill Building:** Labs (*Design Analysis, Save a Failure Design*) helped students identify team roles (Project Manager, Designer, Researcher, Developer) and build collaborative critical thinking.
- **GenAI-Embedded Design Thinking Iteration:** Four sequential labs guiding Empathy → Define → Ideate → Prototype → Testing, with GenAI supporting each phase: analysing user interview data (Empathy), refining "How Might We" statements (Define), suggesting cross-disciplinary ideation (Ideate), and providing real-time prototype feasibility feedback (Prototype/Testing).
- **GenAI-Augmented Thematic Expansion:** A post-midterm *Nautical Ecosystem* guest lecture and trend lab expanded systems thinking; optional team reorganisation (aided by GenAI's updated skill-matching algorithm) aligned teams with thematic interests.

Figure 3: The Timeline of the Design Thinking Labs till the Midterm Evaluation.



2.2.3 Post-Class: GenAI-Optimised Consultation and Societal Engagement

Post-class activities focused on GenAI-driven efficiency in project development, resolving the 2023 cohort's low project iteration efficiency pain point:

- **GenAI-Optimised Guided Project Development:** Three weeks of dedicated project time with rotating instructor consultations. Students submitted draft progress reports, first reviewed by GenAI project management tools (Notion AI, design-focused GenAI editors), which identified gaps, generated revision suggestions, and standardised the structure. This reduced instructor feedback time by 35% (vs. 2023) and enabled targeted, high-value instructor input.
- **GenAI-Enhanced Societal Input Integration:** Teams tested Minimum Viable Products (MVPs) (Camacho, 2016) with external stakeholders. GenAI feedback analysis tools categorised, synthesised, and prioritised stakeholder comments (functionality, aesthetics, feasibility) and generated data visualisations for final presentations, enhancing prototype iteration rigor vs. 2023's manual feedback analysis.

2.3 Course Evaluation

Evaluation combined quantitative and qualitative data to assess pedagogical effectiveness across iterations as follows:

- **Student Performance:** Graded via a standardised rubric; the final showcase was judged by the teaching team plus two senior PhD students (Design/Technical backgrounds) for objectivity.
- **Student Feedback Questionnaire (SFQ):** Eight structured questions (satisfaction with materials, support, feedback, overall experience) benchmarked against university averages.
- **Lecturer Reflections:** Documented challenges, engagement patterns, and iterations (2023–2025: team formation adjustments, module refinements).

3. Results and Discussion

3.1 Students' Performance

The 2025 Spring GenAI-enabled cohort achieved an overall course average of 82.99/100 (final showcase: 17.52/25, SD=0.89), marking a statistically significant 5.24% improvement over the 2023 non-GenAI cohort (78.86/100). This improvement is directly attributed to GenAI integration across the teaching lifecycle. Gains were largest in GenAI-aided phases:

- **Final Showcase:** 4.8% higher scores (16.72→17.52/25), with the GenAI-related criteria (15% of total score) achieving an average of 13.8/15 (92% success rate) for all teams.
- **Lab Performance:** 5.6% higher scores (76.3→80.6/100), with GenAI-aided team formation and Design Thinking labs scoring 8–10% higher than non-GenAI lab activities in the 2025 cohort.

Final projects spanned diverse themes (sustainable campus, nautical ecosystems), with 70% of presentations using multi-media deliverables (animations, user testing videos) vs. 35% in 2023. A direct result of GenAI tools reducing technical barriers to content creation (e.g., AI-generated animation drafts, data visualisations). This reflects enhanced engagement and creative output, all enabled by GenAI's ability to let students focus on design thinking rather than technical production.

3.2 Student Feedback Survey Data (SFQ)

Table 1: SFQ Results in Both 2023 and 2025 Cohorts vs. University Averages (The full mark is 5).

Question No.	Statements	Weighted Marks for this subject (2023)	Weighted Marks for this subject (2025)	Average Weighted Marks for the University (2025)
1	Learning activities support outcomes	4.42	4.71	4.01
2	Teaching Team explanation clarity	4.40	4.86	3.99
3	Teaching Team encouragement of participation	4.56	4.86	3.94
4	Informative answers to questions	4.42	4.71	4.04
5	Teaching Team availability outside class	4.58	4.86	4.03
6	Helpful/timely feedback on assignments	4.33	4.86	3.97
7	Satisfaction with teaching support	4.42	4.71	3.96
8	Overall satisfaction with tutorials/labs	4.38	4.57	4.02

As shown in Table 1 above, SFQ results indicate that 2025 scores outperformed the 2023 cohort (6/8 items improved) and 2025 university averages (14–21% higher). Standout items include:

- "Helpful/timely feedback" (4.86/5, +21% vs university).
- "Learning activities support" (4.71/5, +17.46% vs university).

Key 2023–2025 gains included "Teaching Team explanation clarity" (+10.24%) and "participation encouragement" (+6.38%), with consistent high scores in instructional support.

3.3. Challenges Encountered

A key challenge was GenAI over-reliance: many students initially used GenAI to generate entire prototype drafts without critical revision. This was addressed by reinforced "GenAI output validation" during individual group coaching, requiring students to compare GenAI suggestions with user needs and justify adjustments. Another key challenge was supporting neurodiverse learners: one student with ADHD excelled in individual tasks but struggled with group collaboration, leading to withdrawal. GenAI-based individual task scaffolds (e.g., AI-generated step-by-step project checklists) or tailor-made AI tutors could be expanded for future cohorts.

4. Implications

4.1. Implications

Theoretically, this study validates a hybrid model that embeds GenAI to resolve two core tensions in engineering design education (Azuka et al., 2024). Practically, the 14-week modular framework offers a replicable blueprint for global challenges, adaptable to diverse themes via interdisciplinary teaching and societal engagement. GenAI should target pedagogical pain points (e.g., team formation, feedback delays) rather than replacing human interaction.

4.2. Limitations and Future Work

Limitations include a small cohort size (30–40 students) at a single institution and limited long-term outcome tracking. Future research should test scalability across universities and disciplines, track alumni application of skills in careers, and strengthen inclusive support for neurodiverse learners (e.g., flexible team structures, individual task options).

4.3. Conclusion

This study demonstrates GenAI as the central, transformative enabler of a powerful hybrid teaching approach for engineering design education, which integrates Semi-Flipped-Class, Project-Based Design Thinking, and a core GenAI-enabling framework. This approach is far more effective than non-GenAI hybrid teaching in nurturing creative learning, interdisciplinary collaboration,

and real-world problem-solving skills. GenAI's role across the entire teaching lifecycle (team formation, pre-class support, Design Thinking iteration, project development) is the key to the 2025 cohort's 5.24% performance improvement, 28% higher team satisfaction, and 70% multi-media deliverable adoption. The model's modularity ensures replicability across global challenges, with GenAI as the scalable, personalised core that adapts to diverse themes and learner backgrounds. By balancing structured pedagogy, interdisciplinary collaboration, and critical GenAI augmentation, this GenAI-enabled hybrid model empowers cross-disciplinary students to become GenAI-literate, innovative, collaborative problem-solvers—the exact learners needed to address the world's most pressing global challenges in the GenAI era. This study provides a foundational, replicable GenAI integration blueprint for higher education institutions and educators aligned with advanced GenAI and pedagogical innovations in higher education.

Acknowledgments

We thank the three instructors (architectural design, electronics/IoT, yacht design/marine tech) for their collaboration, the two senior PhD students for evaluation support, and the GenAI resources for guiding tool integration. We also acknowledge the department and university for budgetary and institutional backing.

Ethical Declaration Statement

Notably, the subject does not require an Ethical Declaration Statement. This is because the course's project work focuses on publicly observable real-world problem exploration (e.g., brainstorming solutions for general social interaction needs) rather than research involving human subjects, sensitive data, or ethical risk. Thus, eliminating the need for an ethical declaration.

Generative AI Policy Statement

Authors used ChatGPT 5.0 during the preparation to polish the grammar and logistics flow. After utilising the tool, authors thoroughly reviewed and edited the content as necessary and assumed full responsibility for the publication's content.

References

Azuka, C. V., Wei, C. R., Ikechukwu, U. L., & Nwachukwu, E. L. (2024). Inclusive instructional design for neurodiverse learners. *Current Perspectives in Educational Research, 7*(1), 56–67. https://doi.org/10.46303/cuper.2024.4

Barrows, H. S. (1996). Problem-based learning in medicine and beyond: A brief overview. *New Directions for Teaching and Learning*, 1996(68), 3–12. https://doi.org/10.1002/tl.37219966804

Bionda, A., & Incitti, G. (2024). Modelling Interior Yacht Design Concepts crossing multiple AI tools: Teaching in an uncertain and flexible framework (pp. 181–186). *26th International Conference on Engineering and Product Design Education*. https://doi.org/10.35199/epde.2024.31

Bishop, J., & Verleger, M. (2020). The Flipped Classroom: A Survey of the Research (p. 23.1200.1-23.1200.18). *2013 ASEE Annual Conference & Exposition*. https://doi.org/10.18260/1-2--22585

Camacho, M. (2016). David Kelley: From design to design thinking at Stanford and IDEO. *She Ji, 2*(1), 88–101. https://doi.org/10.1016/j.sheji.2016.01.009

Lakey, J. (2009). Purposeful, creative problem solving. *Gifted Education International, 25*(1), 60–70. https://doi.org/10.1177/026142940902500109

Usahanunth, N. (2020). The Application of MBTI to Develop A Learning Program: A Case Study for the Personality Type of the Occupational Health and Safety Students. *International Journal of Engineering Research & Technology, 9*(4). https://doi.org/10.17577/ijertv9is040493

Yim, T., Leung, C. Y., & Woo, K. T. (2024). Enhancing Integrated STEM Education through Underwater Robotics Competitions: A Project-based Learning Approach Incorporating Engineering Design Processes and Bloom's Taxonomy (pp. 1–5). *2024 IEEE Global Engineering Education Conference (EDUCON)*. https://doi.org/10.1109/educon60312.2024.10578695

Integrating Technology-Enhanced and Traditional Pre-Service Learning: Preparing University Students for Service-Learning in Nursing Homes

Elaine Shuk-ping CHEUNG [1], Ka Yan MA [1], Ka Yee LUI [1], Kin CHEUNG [1], Si Jian LI [1]

[1] School of Nursing, The Hong Kong Polytechnic University, Hong Kong, China

Abstract

Service-learning (SL) (Bringle & Hatcher, 1995) integrates credit-bearing coursework with community service addressing real-world needs. This study examines technology-enhanced pre-service learning—specifically the Hybrid Immersive Virtual Environment (HiVE), metaverse platforms, and virtual reality (VR)—compared with traditional approaches in preparing university students for nursing home service delivery. To our knowledge, this is the first study to examine the role of technology-enhanced pre-service learning in fostering student readiness and engagement, as well as its comparative effectiveness against traditional pre-learning approaches regarding SL outcomes. A pre/post-survey assessed four cohorts between 2023 and 2025. HiVE and mobile-based metaverse platforms provided high-quality learning experiences and effective familiarisation with nursing home settings, whereas VR caused discomfort. Technology-enhanced modes improved health-promoting behaviours (health responsibility, physical activity, nutrition, spiritual growth, and interpersonal relations), while traditional pre-learning excelled in intellectual, cognitive, and civic development. Integrating both approaches maximises outcomes; virtual tools should be used to deliver foundational knowledge and cultural context prior to fieldwork.

Keywords: HiVE, Metaverse, Service-Learning

1. Introduction

Service-learning (SL) (Bringle & Hatcher, 1995) is a credit-bearing, course-integrated educational approach in which students engage in structured service activities that address authentic community needs. The COVID-19 pandemic disrupted traditional in-person SL, particularly within health promotion contexts, due to social distancing requirements and institutional closures. This necessitated a rapid transition toward electronic and hybrid SL models. This shift underscored the importance of incorporating hybrid pre-service learning sessions alongside face-to-face service delivery.

Emerging technologies, such as the Hybrid Immersive Virtual Environment (HiVE) and gamified metaverse platforms, provide immersive and interactive pre-learning experiences by simulating real-world environments and practice scenarios. These tools support students in developing foundational knowledge, cultural competence, and needs assessment skills virtually, thereby overcoming access barriers while effectively preparing them for direct community engagement.

To our knowledge, this is the first study to examine the role of technology-enhanced pre-service learning in fostering student readiness and engagement, as well as its comparative effectiveness against traditional pre-learning approaches regarding SL outcomes.

2. Methodology

This study employed a quasi-experimental design. A pre- and post-survey was conducted with university students enrolled in the subject "Engaging in Workforce Health for Health Care Workers at Nursing Homes." This course required 30 hours of group service within nursing homes in Hong Kong. Data were collected before (T0) and after (T1) the service-learning course from four cohorts between 2023 and 2025.

Students attended pre-service lectures, conducted needs assessments with nursing assistants, and designed two tailored health promotion activities. Among the 2024/25 cohorts, 82 students participated in technology-enhanced pre-service learning using HiVE and gamified metaverse platforms. These tools enabled students to virtually explore nursing home settings, understand on-site services, and practice needs assessment skills through filmed simulations prior to service delivery, followed by a post-experience evaluation of these technologies.

As the data were not normally distributed, the Kruskal-Wallis test was used to compare the Technology Acceptance Model across teaching modes. The Wilcoxon signed-rank test compared SL learning outcomes between pre- and post-SL delivery. The Mann-Whitney U test compared SL learning outcomes between groups with and without metaverse learning. The significance level was set at $p < 0.05$.

3. Results and Discussion

Table 1 presents the demographic characteristics of the participants. A total of 82 students (57.3% female, mean age 21.6 years) completed the post-experience questionnaire on technology-enhanced learning. Most lacked prior experience with HiVE (54.9%) or the metaverse (68.3%), although 54.9% had

previously used VR in an educational context.

Significant differences were observed across teaching modes regarding perceived learning productivity, intent to use, usability, and health impacts ($p < .05$). Mobile-based metaverse was preferred over other modalities ($p < .05$), aligning with previous research (Choi et al., 2025). HiVE and mobile metaverse scored higher on production quality and nursing home familiarisation ($p < .05$), whereas VR-based metaverse induced greater discomfort.

However, no prior studies have directly compared HiVE, metaverse across different devices, and VR in this specific context; previous studies have confirmed the environment familiarisation capabilities of the metaverse (Kye et al., 2021).

All students demonstrated improvement in all assessed aspects post-SL compared to pre-SL ($p < .01$), consistent with previous findings regarding improvements in civic and personal development following SL (Nuriyeva et al., 2025).

Among students who received technology-enhanced pre-service learning, significant improvements were observed in health responsibility ($p < .01$), physical activity ($p < .01$), nutrition ($p < .01$), spiritual growth ($p < .05$), and interpersonal relations ($p < .01$) of the Health-Promoting Lifestyle Profile II (HPLP-II) after service-learning delivery. In contrast, students in the traditional pre-service learning group showed significant improvements only in civic development ($p < .01$) and health responsibility ($p < .05$) among common learning outcomes.

There were no significant differences between groups in any variables prior to SL delivery. After SL delivery, students in the traditional pre-service learning group demonstrated greater improvements in intellectual and cognitive development ($p < .01$) and civic development ($p < .01$) compared to those in the technology-enhanced group. Conversely, students in the technology-enhanced group showed greater improvement in health responsibility ($p < .05$) than those in the traditional group.

The findings indicate that both technology-enhanced and traditional groups improved across all assessed aspects, with each approach offering distinct advantages in specific domains. Therefore, these results suggest the development of a new SL framework that strategically integrates both approaches to maximise student outcomes.

4. Implications

This study suggests that technology-enhanced pre-service learning modes, such as HiVE and mobile-based metaverse environments, effectively prepare university students for service delivery in nursing homes by offering high-quality learning experiences and familiarisation with the setting. However, VR-based metaverse teaching is associated with greater discomfort. Technology-enhanced pre-service learning significantly improved students' health-promoting behaviours. In contrast, traditional pre-service learning was more effective for intellectual, cognitive, and civic development.

These results suggest that integrating technology-enhanced and traditional approaches is essential for maximising student learning outcomes. A hybrid model can utilise HiVE and mobile-based metaverse technologies to deliver foundational knowledge and cultural context—such as for service locations in Thailand and mainland China—before students begin fieldwork. This virtual exposure can enhance cultural understanding and better prepare students for on-site service delivery.

Ethical Declaration Statement

The project has been approved by the Departmental Research Committee (on behalf of the PolyU Institutional Review Board) of The Hong Kong Polytechnic University (Reference Number: HSEARS20240906008).

Generative AI Policy Statement

The author(s) used PolyU Gen AI during the preparation of this work for grammar checking. 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

Bringle, R. G., & Hatcher, J. A. (1995). A Service-Learning Curriculum for Faculty. *Michigan Journal of Community Service Learning*, 2, 112-122.

Choi, H., Wong, P., Shen, J., Francisco, M., & Nurgissayeva, A. (2025). Uncovering the drivers of intent to use the metaverse: diverse experiences in sustainability education. *Discover Sustainability*, 6. https://doi.org/10.1007/s43621-025-00903-9

Kye, B., Han, N., Kim, E., Park, Y., & Jo, S. (2021). Educational applications of metaverse: possibilities and limitations. *Journal of Educational Evaluation for Health Professions*, 18. https://doi.org/10.3352/jeehp.2021.18.32

Nuriyeva, A., Mammedov, O., & Zeynalov, S. (2025). Integrating Social-Emotional Learning into Hybrid Platforms: A Model for Student Digital Well-being and Engagement. *Journal Emerging Technologies in Education*. https://doi.org/10.70177/jete.v2i6.2248

Table 1: Demographic Characteristics of Participants

Baseline characteristic	Technology-enhanced learning approaches		Traditional learning approaches		Full sample	
	n	%	*n*	%	*n*	%
Gender						
Male	34	54.0	24	47.1	58	50.9
Female	29	46.0	27	52.9	56	49.1
Current year of study						
Year 1	0	0	0	0	0	0.0
Year 2	10	15.9	0	0	10	8.8
Year 3	7	11.1	4	7.8	11	9.6
Year 4	46	73.0	47	92.2	93	81.6
SYA Students	45	71.4	32	62.7	77	67.5
Program						
Health related	19	30.2	7	13.7	26	22.8
Engineering and Building	19	30.2	16	31.4	35	30.7
Business	12	19.0	8	15.7	20	17.5
Design	1	1.6	5	9.8	6	5.3
BA Art	7	11.1	4	7.8	11	9.6
Science	4	6.3	7	13.7	11	9.6
Hotel	1	1.6	4	7.8	5	4.4

Note: N = 114 (n = 63 for Technology-enhanced learning approaches and n = 51 for Traditional learning approaches). Participants were on average 21.2 years old (SD = 1.78), and all characteristics did not differ by condition.

9 Smart Learning Environments

Math City: A Futuristic Math Classroom for STEM Skills
(A Virtual City for Real Knowledge)

Manpreet SINGH [1]
[1] Department of Mathematics and Information Technology, The Education University of Hong Kong, Hong Kong, China

Abstract

In today's technology-driven world, mathematics is crucial for developing essential skills, particularly in STEM education. However, many students find abstract concepts challenging due to conventional teaching methods. To address this issue, **Math City**, a 3D gamified virtual platform based on the 5E inquiry model, was developed. This platform integrates GenAI tools and authentic assessments to make abstract math topics engaging and interactive. Unlike traditional textbooks, Math City enables students to visualise 3D shapes within an immersive environment, thereby enhancing comprehension and retention. Built using the Unity game engine and hosted on the metaverse website spatial.io, Math City seamlessly integrates into online and hybrid learning settings, promoting inquiry-based learning and 21st-century skills, including creativity, critical thinking, collaboration, and communication. The prototype was pilot-tested with around 1,400 users from diverse backgrounds, and the findings were published in the peer-reviewed journal *Computers and Education: Artificial Intelligence* (Singh et al., 2025).

Outcomes-based Learning Objectives

- Experience Math City's immersive environment first-hand and discover how virtual quests and digital avatars promote inquiry-based learning in mathematics.
- Assess the pedagogical effectiveness of Math City by engaging with its 3D spaces, game-based learning features, and instructional design supported by GenAI tools.
- Participate in a mini-lesson focusing on authentic assessments within the virtual setting of Math City.

Target Audience

This workshop is designed for higher education faculty, instructional designers, edtech researchers, and postgraduate students who are interested in innovative pedagogies, AI integration, and immersive learning environments.

Technical Support

Technical support will be provided to participants during the workshop as a guided walkthrough in Activity 1 to help them log in to the spatial.io platform, create digital avatars, and practice basic navigation.

Technical Requirements

The workshop runs on standard laptops or desktop computers (Windows or macOS) with an updated web browser and a stable internet connection.

Topic Relevance

Future of Learning / Smart Learning Environments

Workshop Design

Workshop Format and Activities

1. **Introduction (5 minutes)**

   Welcome remarks and discussion questions.

2. **Main Content Delivery (10 minutes)**

   * **Pedagogical Framework:** Discussion of the 5E Inquiry Model and Game-based learning (GBL).
   * **Platform Features:** Introduction to virtual spaces designed in Math City and their alignment with the 5E inquiry model. Introduction to gamification elements, such as NPCs, quests, and badges embedded in Math City.
   * **Usability Insights:** Sharing findings from usability trials of Math City.

3. **Interactive Activities (30 minutes)**

   Activity 1: Guided Walkthrough (10 minutes)

Participants will create digital avatars to enter Math City and divide themselves into different teams for collaborative activities within Math City.

Activity 2: Exploring Math City (15 minutes)

Participants will enter Math City by clicking the link provided during the workshop and interact with NPCs to receive further instructions for navigating through Math City and completing game quests.

The first stop is the **Knowledge Academy**, which activates prior knowledge of 3D shapes. Next, participants will visit the **Exploration Lab**, where

they will collaborate to answer inquiry questions. In this virtual space, two GenAI portals, Google AI and Poe.com, are embedded for participants to conduct the inquiry and submit a joint report via Google Sheets.

Finally, in the **Exhibition Hall**, they will have an immersive experience before the evaluation stage of the 5E inquiry model.

Activity 3: Authentic Assessment (5 minutes)

In the authentic assessment area, "Amusement Park," in Math City—aligned with the evaluation stage of the 5E inquiry model—participants will interact with real-life 3D objects and answer MCQs that reflect the knowledge acquired in Math City.

4. **Discussion and Q&A (8 minutes)**
   - **Math City Badge:** Participants will earn a digital Math City badge by completing all game quests and authentic assessments within Math City.
   - **Future Implications:** Discussion about the future version of Math City.

5. **Closing Remarks (2 minutes)**
   - Invitation to join future pilot studies or collaborative research on Math City.

Acknowledgments

This workshop is based on research published in *Computers and Education: Artificial Intelligence* and supported by The Education University of Hong Kong.

Ethical Declaration Statement

For the usability trials, prior approval was obtained from the participants, and their involvement was completely voluntary. They understood that the information gathered from them could be used in future research and may be published without revealing their personal details.

Generative AI Policy

The author(s) used Grammarly tools during the preparation of this work to polish the language and grammar. 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

Singh, M., Sun, D., & Yang, Y. (2025). Design and implementation of Math City: An inquiry-based metaverse platform for mathematics education. *Computers and Education: Artificial Intelligence*, 9, 100462. https://doi.org/10.1016/j.caeai.2025.100462

C

Contributors

1	ABDI	Fatwa Firdaus ABDI	Fatwa Firdaus ABDI, Fatwa Abdi is an Associate Professor and Major Leader (Energy Science & Engineering) at School of Energy and Environment, City University of Hong Kong. He obtained his PhD (Chem. Eng.) from TU Delft, Netherlands. He teaches courses on energy applications and is the recipient of SEE Teaching Excellence Award 2025.
2	ANDERS	Abram ANDERS	Abram ANDERS, Dr Abram Anders is the Jonathan Wickert Professor of Innovation and Associate Director of the Student Innovation Center at Iowa State University, where he leads the AI Innovation Studio. He created a pioneering Artificial Intelligence and Writing course and conducts research on AI literacies in education. His recent Computers & Education: Artificial Intelligence article demonstrates how integrating comprehensive AI literacies—functional, critical, ethical, and creative—with self-regulated learning can promote student agency and human-in-the-loop practices that augment rather than replace disciplinary expertise. His work on technological innovation, communication, and education appears in leading journals including College English and the International Journal of Business Communication. The Association for Business Communication has recognized his scholarship with multiple awards including the 2022 Outstanding Article Award. Learn more at abramanders.com.
3	ANDERSON	Esme ANDERSON	Esme ANDERSON
4	BEN-SAID	Selim BEN-SAID	Selim BEN-SAID is an Associate Professor at National Sun Yat-Sen University, Taiwan. His research focuses on sociolinguistics, linguistic landscapes, urban art, and technology-enhanced language learning. He has published extensively on multilingualism, identity, and innovative approaches in EFL and higher education across Taiwan, Tunisia, Hong Kong, Malaysia, and Singapore.
5	BETTONI	Marco BETTONI	Marco BETTONI, Senior Engineer at the Industrial Centre, The Hong Kong Polytechnic University, and PhD student in the Department of Health Technology and Informatics. Holds MSc and BSc in Computer Engineering from Politecnico di Torino, Italy. Focuses on AI and robotics applications, with research on digital twin solutions for healthcare and biomedical data integration
6	BIONDA	Arianna BIONDA	Arianna BIONDA, PhD in Design, is a researcher and professor at the Design Department, Politecnico di Milano. Architect, sailor, and yacht designer. Her

			research activities are mainly focused on innovative technologies and materials' role in a yacht design project, their current dynamic, and practices of interaction.
7	BUENCONSEJO	Jet UY BUENCONSEJO	Jet UY BUENCONSEJO is a Research Assistant Professor whose work focuses on Positive Youth Development, including its measurement, developmental correlates, and cross-cultural applicability in East and Southeast Asia. His broader interests span youth well-being, positive education, and AI literacy. He is also a registered psychologist and psychometrician.
8	CAI	Louis CAI	Louis CAI, PhD, focuses on the intersection of technology, pedagogy, and educational data for the betterment of tertiary education. With a multidisciplinary background, he is passionate about implementing technology, including Generative AI, to augment learning across various disciplines. He also draws on learning analytics for evidence-based pedagogical decisions.
9	CHAN	Chi Ho CHAN	Chi Ho CHAN is a Research Associate in the Department of Electrical and Electronic Engineering at The Hong Kong Polytechnic University. He has participated in several research projects related to game-based learning and the use of AI in helping students develop writing and speaking skills.
10	CHAN	Christy CHAN	Christy CHAN, Ms Christy Chan is a Senior Tutor in the Language Centre at City University of Hong Kong. Her interests are genre analysis, English across the curriculum, and Edtech. She has received the Hong Kong UGC Teaching Excellence Award (Team) in 2022 and QS Reimagine Education Silver Award in 2024.
11	CHAN	Henry CHAN	Henry CHAN
12	CHAN	Ivan CHAN	Ivan CHAN
13	CHAN	Ka Yee CHAN	Ka Yee CHAN, Ms. Shirley Chan is a Lecturer at the Centre for Language in Education at The Education University of Hong Kong (EdUHK) . She actively participates in projects related to technologies in education, including the creation of an English learning mobile app, "Everyday English in Hong Kong", etc. Her research interests include technology-mediated language learning and gamification in ELT.
14	CHAN	Kitty CHAN	Kitty CHAN , Dr. Kitty Chan, Associate Professor of Practice at The Hong Kong Polytechnic University, leads innovative and interprofessional healthcare education projects, including IVR and online game platforms. Her award-winning initiatives, such as the

			Virtual Hospital and vCare, have enhanced learning, attracted funding, and earned international recognition for advancing technology-integrated pedagogies.
15	CHAN	Stephanie CHAN	Stephanie CHAN, Miss Stephanie Chan is an Executive Officer of LTTC at EdUHK. She has extensive experience in government-funded research projects focused on student development and performance in AI literacy and STEAM competency.
16	CHAN	Yuet Kai CHAN	Yuet Kai CHAN
17	CHARITONOS	Koula CHARITONOS	Koula CHARITONOS
18	CHEN	Hsuan-Yu CHEN	Hsuan-Yu CHEN
19	CHEN	Hsueh Chu CHEN	Hsueh Chu CHEN
20	CHEN	Julia CHEN	Julia CHEN (PhD, PFHEA) is Director of Educational Development at HKPolyU, and Chair of UGC-initiated HK Teaching Excellence Alliance. Her research interests include leveraging technology for education, learning analytics, and English Across the Curriculum (EAC). The inter-institutional EAC team she leads has received multiple awards, including a QS Reimagine Award.
21	CHEN	Nuo CHEN	Nuo CHEN is an undergraduate student in the Qian Xuesen Honors Program (40 selectees), NUDT, specializing in Atmospheric Sciences. Inspired by the advancements in AI, she is interested in the outcomes of student-AI collaboration. Leveraging her science background, she focuses on data processing and analysis in the research.
22	CHEN	Ruoyu CHEN	Ruoyu CHEN
23	CHEN	Wu CHEN	Wu CHEN
24	CHEN	Xuchuan (Winnie) CHEN	Xuchuan (Winnie) CHEN is Associate Professor at the National University of Defense Technology. She obtained her PhD in Education from The Chinese University of Hong Kong. Her research interests include L2 writing and technology-enhanced pedagogy, with publications in Journal of Second Language Writing, Computer Assisted Language Learning, and Language Teaching Research.
25	CHENG	Shuk Ling CHENG	Shuk Ling CHENG is currently a lecturer at City University of Hong Kong, and also a doctoral student at the Hong Kong Polytechnic University. Her research interests include social identity, AI in

			education, sociolinguistics, and second language acquisition.
26	CHEUNG	Elaine Shuk-ping CHEUNG	Elaine Shuk-ping CHEUNG, Dr. Elaine CHEUNG is an Assistant Professor of Practice and specializing in innovative teaching strategies. Her expertise includes flipped classrooms, simulation-based learning, and virtual reality in nursing education. Passionate about enhancing student engagement and learning outcomes, Dr. Cheung actively develops and implements cutting-edge educational methodologies. With a Doctorate in Health Science, she is dedicated to fostering an interactive learning environment that prepares students for real-world challenges. In addition to her teaching, she is an advocate for using technology to advance nursing education and improve healthcare quality.
27	CHEUNG	Paul CHEUNG	Paul CHEUNG is an experienced and skilled educational developer at Temasek Polytechnic's Learning Academy. He is well-versed in integrating edtech into the classroom and in online teaching and has extensive experience with eLearning. Committed to helping others grow, Paul is passionate about developing individuals' skills and knowledge.
28	CHIEN	Chi-Wen CHIEN	Chi-Wen CHIEN, Chi-Wen Chien (Will) is an Associate Professor committed to enhancing learning through innovative approaches, including problem-based learning, experiential activities, and serious games. Supported by multiple teaching development grants, he develops teaching-related database, pedagogy, and game, and has published these initiatives in educational and professional journals.
29	CHIYANIKA	Chileka CHIYANIKA	Chileka CHIYANIKA, Dr Chileka Chiyanika, FHEA, is a Research Assistant Professor at The Hong Kong Polytechnic University with over 20 years' experience in radiography, CT, MRI, and ultrasound. He specialises in imaging research, innovative anatomy education using VR and AI, and has received multiple awards for his contributions to radiological science and teaching.
30	CHON	Wing Yuk Leo CHON	Wing Yuk Leo CHON, Mr Leo Chon is an Assistant Educational Development Manager at the Educational Development Centre, The Hong Kong Polytechnic University. He supports academics in designing and implementing innovative assessment strategies to incorporate the use of GenAI. Leo also works on projects that enhance grade integrity and consistency, and foster student-staff partnership.
31	CHU	Beatrice CHU	Beatrice CHU

32	CHU	Maggie CHU	Maggie CHU
33	CHUNG	Hoi Ying CHUNG	Hoi Ying CHUNG
34	CHUNG	Ka Yan CHUNG	Ka Yan CHUNG, Fafa Chung Ka Yan is a design researcher with a background in advertising design and STEM education. Her research focuses on design education, creativity, the psychological and educational implications of AI integration in creative industries.
35	DING	Lin Qian DING	Lin Qian DING, Ms. Linqian Ding is a doctoral student in English Language Education at EdUHK. Her research focuses on AI-assisted language learning. She has published in journals such as Education and Information Technologies, and has five years of experience teaching English listening and writing.
36	DING	Yan DING	Yan DING, Dr. Yan Ding earned her doctoral degree in Educational Development from Nagoya University, Japan. She is currently an Associate Professor at the Research Institute for Higher Education and serves as Associate Director of the Center for Faculty Development at Fudan University. Dr. Ding is among the scholars actively engaged in faculty development research and practice in China. Since 2019, she has served as Deputy Director of the Academic Committee of the China Higher Education Development Network (CHED), the largest national alliance for faculty development in China. Her research interests include faculty development, the evaluation of teaching and learning, and human resource management in higher education. She has authored and translated four books and published more than thirty papers in Chinese, Japanese, and English academic journals.
37	DURÁN DEL FIERRO	Francisco DURÁN DEL FIERRO	Francisco DURÁN DEL FIERRO, Dr Francisco Durán del Fierro is Senior Research Fellow at the UCL Knowledge Lab, Department of Culture, Communication and Media - IOE, UCL's Faculty of Education and Society. His research examines the social and cultural dimensions of the digital transformation of science, understanding Digital Research Infrastructures (DRIs) as a site of epistemic transformation where objects, knowledge, and meaning intersect. This subfield of sociology is interdisciplinary, combining theoretical frameworks from philosophy of science, social epistemology, cultural sociology, the history of science,

			poststructuralism, and innovative qualitative methodologies.
38	FAN	Xiaojie FAN	Xiaojie FAN is a Lecturer at Shanghai Normal University, specializing in educational policy and teacher development. Her work examines higher education reform and professional learning.
39	FEAST	Luke Edward Eric FEAST	Luke Edward Eric FEAST, Luke Feast is an Assistant Professor in Product Design at PolyU Design, specializing in design entrepreneurship and higher education innovation. With a PhD in Design and over 18 years of teaching experience, my research advances sustainability, interdisciplinary collaboration, and culture-based innovation, and has been published in leading journals and international conferences.
40	FENG	Hai Wei FENG	Hai Wei FENG received a master's degree from Hebei University of Technology, Tianjin, China, in 2025. He joined the Algorithm Engineer Department of the System Engineering College of National University of Defense Technology on July 2025. His research field is AI large model algorithms and application development.
41	FRIGINAL	Eric FRIGINAL	Eric FRIGINAL is Professor and Head of the Department of English and Communication at The Hong Kong Polytechnic University. He specializes in applied corpus linguistics, technology-enhanced language teaching, and the analysis of spoken professional discourse in the workplace. He is the founding Co-Editor-in-Chief of the Applied Corpus Linguistics (ACORP) journal.
42	FU	Jo-Shan FU	Jo-Shan FU is a research assistant professor at Hong Kong Polytechnic University. With a Ph.D. in ESL from Southern Illinois University, she specializes in AI-integrated language learning, inventing tools like DocsDocs and WriteRight, and leads projects to enhance EFL writing through technology.
43	FUNG	Daniel FUNG	Daniel FUNG is an Assistant Professor at The University of Hong Kong. His primary research interests include learning strategies, second language listening, vocabulary acquisition, and English Medium Instruction. His recent research papers have been published in Language Teaching Research, ELT Journal, Language and Education, and System.
44	FUNG	Sui Leung Walter FUNG	Sui Leung Walter FUNG
45	GAO	Ruiya GAO	Ruiya GAO is an undergraduate student at the National University of Defense Technology, affiliated with the Qian Xuesen Honors Class. She is interested in the intersection of artificial intelligence and

			education, with a particular focus on how AI technologies can support and enhance English language learning.
46	GENG	Hongyan GENG	Hongyan GENG is an Assistant Professor of Teaching at Science Unit, Lingnan University. She has extensive experience in teaching general education science courses to non-science majors. Her areas of specialization include environmental geochemistry and interdisciplinary studies on sustainability. Her recent interest is science general education in the digital era.
47	GENG	Ping GENG	Ping GENG, Dr. Geng is dedicated to teaching innovation as PI/Co-PI of six teaching grants, focusing on active learning technologies, such as immersive and AI-empowered learning. Her efforts have been recognized with the Faculty Outstanding Young Teacher Award, Bronze Award in E-learning Forum Asia and shortlist nomination for QS Reimagine Award 2024.
48	HARWOOD	Chris HARWOOD	Chris HARWOOD is an associate professor in the Faculty of Liberal Arts at Sophia University, Japan. His research examines academic literacy, critical thinking, and technology mediated learning in higher education, with particular attention to self-regulated learning, ethical technology use, and pedagogical design in English medium instruction contexts.
49	HAU	Sam Chun Kit HAU	Sam Chun Kit HAU, Dr. Sam Hau earned his B.Sc. (First Class Honors) in Chemistry from HKBU (2005) and his Ph.D. in Organic Synthesis from CUHK (2010). After postdoctoral research in X-ray Crystallography at CUHK, he became a CUHK lecturer in 2019. His interests include innovative teaching, AI in chemistry education, and student development.
50	HO	Jenifer HO	Jenifer HO is Associate Professor at the Department of English and Communication at The Hong Kong Polytechnic University. Her research interest is in language teaching and intercultural communication in digitally mediated professional and educational contexts, with a specific focus on video-based communication.
51	HOU	Junyao HOU	Junyao HOU
52	HUANG	Jinmei HUANG	Jinmei HUANG is a lecturer at Guangzhou College of Commerce. Her research focuses on design education and AI. She leads 3 related projects, has won national and provincial teaching awards, and is a DDES candidate at PolyU Design.

53	HUANG	Xiaofeng HUANG	Xiaofeng HUANG
54	HUANG	Ying HUANG	Ying HUANG received the M.S. and Ph.D. degrees in Information and Communication Engineering from National University of Defense Technology, in 2002, and 2014, respectively. She has published in excess of 40 journal and international conference papers. Her research interests include teaching and learning research, covert communication, physical layer security, channel codes, etc.
55	HUI	Vivian HUI	Vivian HUI
56	JI	Rui JI	Rui JI recently completed a Master's degree in TESOL at the University of Nottingham Ningbo China. Her research focuses on AI-supported IELTS speaking preparation and the strategies learners employ to enhance performance. She is particularly interested in the intersection of educational technology and language assessment.
57	KANAMARU	Toshiyuki KANAMARU	Toshiyuki KANAMARU, Ph.D., is an Associate Professor at Kyoto University. His research interests focus on teaching English for Academic Purposes through the application of natural language processing, including generative AI. He has published articles in journals such as Review of Cognitive Linguistics and Asia Pacific Journal of Corpus Research.
58	KIDANEMARIAM	Yonathan Aklilu KIDANEMARIAM	Yonathan Aklilu KIDANEMARIAM, Yonathan is a second year Global Business student at HKUST. He works in RAG and agentic AI systems and is diving deep into how these systems could better pedagogy. He has partnered with faculty to bring his platform to courses in the School of Business and Management.
59	KUO	Yi-Chun KUO	Yi-Chun KUO
60	KWOK	Jonathan Man Lung KWOK	Jonathan Man Lung KWOK, Dr Jonathan Kwok is the Senior Lecturer at the College of Professional and Continuing Education, the Hong Kong Polytechnic University. His research interests include leadership, self-regulatory focus, and innovation and creativity. His work has been published in the Journal of Enterprise Information Management, Informatics, Asia Pacific Education Review, Sustainability, etc.
61	KWONG	Theresa KWONG	Theresa KWONG, Ph.D., SFHEA, is the Director of the Centre for Holistic Teaching and Learning at Hong Kong Baptist University. She leads the promotion of quality teaching, provides expertise on pedagogy and e-learning, and advises on innovative assessment methods. Her research interests include

			academic integrity and technology-enhanced learning.
62	LAI	Pauli LAI	Pauli LAI, Dr Pauli Lai is a Senior Lecturer in the Department of Electrical and Electronic Engineering at The Hong Kong Polytechnic University. She is keen on leveraging AI and EdTech to enhance teaching methodologies. She has received the Faculty Outstanding Teacher Award in 2024, and the 2024 QS Reimagine Education Awards.
63	LAI	Zhongyu LAI	Zhongyu LAI
64	LAM	Chun Yin LAM	Chun Yin LAM, Mr Jack Lam is a Research Assistant at the College of Professional and Continuing Education, the Hong Kong Polytechnic University. His research interests include Chinese Medicine Nephrology, employability, and international business. He has served as a peer tutor at the University of Hong Kong, teaching life science and nursing courses.
65	LAM	Ka Ki LAM	Ka Ki LAM, Lam Ka Ki is an Assistant Educational Development Officer at the Centre for Learning, Teaching and Technology, The Education University of Hong Kong. Her research interests include AI education and STEM education.
66	LAM	Kim Hung LAM	Kim Hung LAM, Dr. Joe Lam is Senior Fellow of the Advance Higher Education (FHEA), Lecturer, Programme Leader, Academic Advisor at Department of Applied Biology and Chemical Technology (ABCT) of Faculty of Science (FS), as well as Fellow of PolyU Institute for Higher Education Research and Development (IHERD), The Hong Kong Polytechnic University.
67	LAU	Peter LAU	Peter LAU, FHEA, EdD, is an advocate for SaP as an innovative pedagogy. He engages in curriculum innovation and interdisciplinary collaboration alongside faculty and students. His projects created platforms supporting SaP development in HK and received international awards and research grants. He recently joined the Co-editorial Team (HK) for IJSaP.
68	LEE	Jonathan LEE	Jonathan LEE
69	LEE	Max Jwo Lem LEE	Max Jwo Lem LEE
70	LEI	Jing LEI	Jing LEI received the B.Sc., M.Sc., and Ph.D. degrees from the National University of Defense Technology, in 1990, 1994, and 2009, respectively. She has published many papers in various journals and conference proceedings and five books. Her research interests include teaching and learning research,

			information theory, and wireless communication technology, etc.
71	LEI	Yue LEI	Yue LEI, Ms Lei Yue graduated from the Hong Kong Polytechnic University in 2024. She then got a Master degree in Business Analytics from University of Illinois Urbana-Champaign, 2025. She was on Dean's list at Polyu, and presented her capstone project in America Conference of Information Systems, 2024.
72	LEUNG	Man Kin LEUNG	Man Kin LEUNG, Mr. Man Kin Leung has been teaching various subjects in tertiary mathematics for over a decade. He has been actively involved in educational development and innovation. He currently leads the "Enhancing Teaching and Learning Experience with AI Technology" project aiming to promote integration of AI technologies into teaching and learning.
73	LI	Hung Wing LI	Hung Wing LI
74	LI	Kristen Yuanxi LI	Kristen Yuanxi LI
75	LIAO	Weihang LIAO	Weihang LIAO, Dr. Weihang LIAO joined the department of Data Science and Artificial Intelligence at the Hong Kong Polytechnic University in 2025. Before that, he was a postdoctoral fellow in the department of Health Technology and Informatics at the same university from 2024.
76	LIAO	Yuanyi LIAO	Yuanyi LIAO, the Project Associate at the School of Design, The Hong Kong Polytechnic University. Her research focuses on design education, examining how reflective, process-driven creative practice generates knowledge and informs the development of design teaching and learning in higher education.
77	LIM	Steven LIM	Steven LIM is experienced lecturer specializing in pedagogy, technology, law, and ethics. Plays a leading role in shaping educational policy for digital learning and generative AI across the Polytechnic & ITE sector. At Ngee Ann Polytechnic, develops colleagues in digital flipped learning design, facilitation, and Gen-AI-integrated assessment.
78	LIN	Yu Jen LIN	Yu Jen LIN
79	LITTLEJOHN	Allison LITTLEJOHN	Allison LITTLEJOHN, "Professor Allison Littlejohn is Pro-Vice Provost for UCL's Grand Challenge Data Empowered Societies and is Professor of Learning and Technology in the UCL Knowledge Lab at University College London. Her research expands our understanding of the socio-technical effects of digital transformation at work and how it impacts work and learning. Professor Littlejohn was previously

			Director of the UCL Knowledge Lab, University College London (2020-2025); Dean (Learning & Teaching) at the University of Glasgow (2019-20); Academic Director for Digital Innovation at the Open University, UK (2015-2019); Professor and Director of the Caledonian Academy at Glasgow Caledonian University (2006-2015). Allison holds the Ruth Wong Visiting Professorship at the National Institute of Education, Singapore."
80	LIU	Xintao LIU	Xintao LIU is an Associate Professor in the Department of Land Surveying and Geo-Informatics at The Hong Kong Polytechnic University (PolyU). I received my PhD degree in Geo-Informatics from the Royal Institute of Technology KTH, Sweden in 2012. Prior to joining PolyU in 2016, I worked as a postdoc fellow and sessional lecturer at the GeoCollaboration and GIS Lab at Toronto Metropolitan University, Toronto, Canada.
81	LO	Dawn LO	Dawn LO, Dr Dawn Lo is a Senior Educational Development Manager at the Educational Development Centre, The Hong Kong Polytechnic University. She works in the areas of teacher learning and professional development, institutional quality enhancement, and teacher assessment and grading practices, and supports the University's strategic initiatives on teaching and learning.
82	LODGE	Jason LODGE	Jason LODGE, Professor Jason M. Lodge, PFHEA, is a Professor of Educational Psychology and Director of the Learning, Instruction, and Technology Lab in the School of Education at The University of Queensland, Australia. His work explores the cognitive and emotional mechanisms of learning with digital technologies, addressing critical questions of how technology, particularly AI, is shaping learning and education. Jason's research informs educational policy and practice across Australia and internationally. He serves as an expert advisor for the Australian Government and OECD, applying his work to enhance equitable learning for all students.
83	LOPEZ-OZIEBLO	Renia LOPEZ-OZIEBLO	Renia LOPEZ-OZIEBLO
84	LUI	Richard Wing Cheung LUI	Richard Wing Cheung LUI, Dr. Richard Lui is a Senior Lecturer in the Department of Computing at The Hong Kong Polytechnic University. He has extensive experience in adopting innovative pedagogies and developing educational platforms. His projects received the Silver Award in the "AI in

			Education" category at the QS Reimagine Education Awards 2025.
85	LUO	Crystal LUO	Crystal LUO, Miss Crystal Luo is a Teaching and Learning Manager of LTTC at EdUHK, with extensive experience in instructional design, professional development training, and teaching in higher education. Her research interests include AI in education and learning analytics. She has published articles in conferences and journals, and has previously received a Best Paper Award.
86	LUO	Xiaoyu LUO	Xiaoyu LUO is a Project Assistant at the Educational Development Centre, The Hong Kong Polytechnic University. She holds an MSocSc in Applied Social Science from City University of Hong Kong. Her work focuses on a project that delved into exploring effective strategies for flipped learning and teaching in higher education.
87	MA	Henry MA	Henry MA, Professor Henry Ma is the leader of the Creativity and Design Education Lab and the Doctor of Design Programme at The Hong Kong Polytechnic University. He has secured funding for various research in technology and education development. Recently, his focus has shifted to nurturing creativity and advancing design thinking research.ntegration in creative industries.
88	MA	Ka Yan MA	Ka Yan MA, Ms Ka Yan MA is a Project Associate. Her expertise spans mental health well-being, art-based interventions to enhance mental health, and extensive experience in university service learning. Holding a Master's degree in Applied Psychology, she brings a strong foundation in psychological principles to her educational practice. She is deeply committed to promoting positive learning environments and raising awareness of mental health issues among students, inspiring them to embrace well-being and holistic personal development throughout their academic journey.
89	MENG	Fanfei MENG	Fanfei MENG, Ms. FanFei Meng is an MBA graduate from the faculty of business at City University of Macau.
90	MENGONI	Paolo MENGONI	Paolo MENGONI, Dr. Paolo Mengoni is a Senior Lecturer in Interactive Media at Hong Kong Baptist University and Associate Programme Director for the MSc in AI and Digital Media. He earned a PhD in Computer Science from the University of Florence. His research spans behaviour analysis, complex networks, NLP, evolutionary computation, and recommender systems, with publications in leading journals.

91	MOK	Betty MOK	Betty MOK, Dr Betty Mok is an Instructor in the Department of Applied Social Sciences at The Hong Kong Polytechnic University. She received the Faculty Team Award for Outstanding Achievement in Teaching in 2024. Her research interests involve positive youth development and service-learning. She has published research articles in international refereed journals.
92	NEUVILLE	Jessica NEUVILLE	Jessica NEUVILLE is an Associate Professor of Practice at the School of Optometry, The Hong Kong Polytechnic University. She received her Doctor of Optometry degree from the University of California, Berkeley. She has received the Faculty Prize for Outstanding Achievement in Teaching, and she is interested in developing eLearning strategies.
93	NG	Davy Tsz Kit NG	Davy Tsz Kit NG
94	NG	Paul NG	Paul NG is experienced lecturer specializing in pedagogy, technology, law, and ethics. Plays a leading role in shaping educational policy for digital learning and generative AI across the Polytechnic & ITE sector. At Ngee Ann Polytechnic, develops colleagues in digital flipped learning design, facilitation, and Gen-AI-integrated assessment.
95	NGAI	To NGAI	To NGAI
96	NURGISSAYEVA	Aru NURGISSAYEVA	Aru NURGISSAYEVA
97	OH	Jae-Eun OH	Jae-Eun OH
98	ONG	Caron ONG	Caron ONG is a Senior Academic Mentor at Temasek Polytechnic's School of Humanities & Social Sciences in Singapore. With a Master of Education in Mathematics, he specializes in integrating Generative AI in education, pioneering innovative assessment redesign and developing educational chatbots. He leads extensive professional development initiatives across educational institutions.
99	QI	Cong QI	Cong QI, Dr. Qi is a lecturer in the MM department, PolyU. She received her PhD degree in MIS at the University of Hong Kong in 2008. Her research areas cover new technology adoption, AI and GenAI, Blockchain, and IS pedagogical research, etc. She specializes in conducting quantitative and qualitative behavioral research. Ms Lei Yue graduated from the Hong Kong Polytechnic University in 2024. She then got a Master degree in Business Analytics from University of Illinois Urbana-Champaign, 2025. She was on Dean’s

			list at Polyu, and presented her capstone project in America Conference of Information Systems, 2024.
100	QIAN	Liang QIAN	Liang QIAN
101	QIN	Jiaqi QIN	Jiaqi QIN is an undergraduate student at National University of Defense Technology, Qian Xuesen Honors Class—an elite program of 40 selected students. She is interested in technology-enhanced L2 writing pedagogy and contributes to data processing and analysis in this research on AI-assisted collaborative writing.
102	SABETZADEH	Farzad SABETZADEH	Farzad SABETZADEH, Dr. Farzad Sabetzadeh is a full-time associate professor of Knowledge and Innovation Management at the faculty of business at the City University of Macau. Currently, he is serving as the program coordinator for the International Business Cohort (IBC) program at the City University of Macau.
103	SANTANDREU CALONGE	David SANTANDREU CALONGE	David SANTANDREU CALONGE, Dr. David Santandreu Calonge (SFHEA) is Head of Educational Program Development, Center for Teaching and Learning, Mohamed bin Zayed University of Artificial Intelligence (MBZUAI, United Arab Emirates). He is an award-winning academic and has previously worked in Hong Kong (HKU, CityU, Lingnan U, HKBU), South Korea (Sungkyunkwan U, Visiting Professor), Australia (RMIT University, The University of Adelaide), and Dubai (Canadian University Dubai, Dean and Associate Professor). He has written in the areas of education policy, Generative AI in Learning and Teaching, MOOCs, innovative curriculum design, and disruptive technologies in education.
104	SHEK	Daniel SHEK	Daniel SHEK, Prof. Daniel Shek is Associate Vice President (Undergraduate Programme), Dean of College of Undergraduate Studies, Chair Professor of Applied Social Sciences, and Li and Fung Endowed Professor in Service Leadership Education at The Hong Kong Polytechnic University. Prof. Shek has published more than 700 articles in international refereed journals.
105	SHEN	Daniel SHEN	Daniel SHEN is CEO of Soqqle, leading development of VR learning experiences powered by GenAI avatars. His work bridges immersive technology and education, focusing on how conversational agents support brainstorming, guidance, and engagement in virtual classrooms. His research interests span VR interaction design, human-AI collaboration, and learning analytics for deployment.
106	SHEN	Jianfu SHEN	Jianfu SHEN, Prof. Jianfu Shen

		Prof. Shen joined The Hong Kong Polytechnic University as an Assistant Professor in 2019 and was promoted to Associate Professor in 2025. Prof. Shen's research covers a broad range of topics in real estate economics and finance, green/climate finance, ESG, etc. He teaches business valuation and development finance. He is also a certificated FRM (Financial Risk Manager) and CFA (Chartered Financial Analyst) charterholder.
107 SHEN	Qianhui SHEN	Qianhui SHEN
108 SHUM	Johnson Chung Ming SHUM	Johnson Chung Ming SHUM
109 SINGH	Manpreet SINGH	Manpreet SINGH, Dr. Singh, an expert in immersive educational technologies, specializes in designing STEM lessons using interactive simulations, virtual reality, and GenAI. His research focuses on inquiry-based pedagogies to enhance STEM education and 21st-century skills. His work on immersive metaverse environments has been published in various peer-reviewed research articles in top journals.
110 SO	Chi Fuk SO	Chi Fuk SO, Dr So's research interests encompass a broad spectrum of fields, including AI Education, IT in education, STEM education, simulation-based learning, mobile learning, mobile app development, digital health, health informatics, adaptive filtering, digital signal processing, machine learning, and data analytics.
111 SO	Ming Hang SO	Ming Hang SO, Dr. So has over 20 years of experience in financial journalism as both practitioner and educator. A recipient of the HSUHK Teaching Excellence Award, his research and teaching interests focus on financial journalism education and AI in teaching and learning. He has presented his work at international conferences including the past three annual International Communication Association conferences.
112 SUN	Zhuo SUN	Zhuo SUN is a Master's candidate in English Translation and Interpreting at School of Foreign Studies, Harbin Engineering University. She specializes in AI-assisted translation and Area Studies.
113 TAM	Barbara TAM	Barbara TAM is a Senior Educational Development Manager at the Educational Development Centre, The Hong Kong Polytechnic University. She specialises in faculty development, particularly in training new academics and teaching assistants. Barbara supports

			academics in developing their pedagogy and drives initiatives that promote active learning and academic integrity.
114	TAN	Siew TAN	Siew TAN, Verily Tan is a Senior Education Specialist at the Centre for Teaching, Learning, and Technology (CTLT), where she supports faculty-facing initiatives in Generative AI and professional development. Her research interests focus on GenAI in teaching, learning, and blended education, and she also contributes to the PDPT Core programme and the Course Design Institute.
115	TANG	Chong TANG	Chong TANG, Mr. Chong Tang is an MBA graduate from the faculty of business at City University of Macau.
116	THADANI	Dimple THADANI	Dimple THADANI, SFHEA, PGCHE, is a Senior Lecturer specializing in Information Systems and Business Intelligence. Dimple is committed to blending technology, education, and business to equip students for the digital era. Her research spans AI in business education, digital transformation and platform strategies.
117	TORCUK	Yagiz TORCUK	Yagiz TORCUK
118	TSANG	Tsz Ning TSANG	Tsz Ning TSANG holds a BA (Hons) Digital Media from PolyU. Her research interests include media design, design education and generative AI. She is currently working as a project assistant in the integration of AI in media design education project at the School of Design, The Hong Kong Polytechnic University.
119	TSUI	Lik Hang TSUI	Lik Hang TSUI is an Associate Professor in the Department of Chinese and History at the City University of Hong Kong, where he is also the Associate Director of the Talent and Education Development Office. Prior to joining CityU, he worked as a Departmental Lecturer in Classical Chinese at the University of Oxford and as a Postdoctoral Fellow at Harvard University with the China Biographical Database (CBDB). He has published on middle period Chinese history, especially on epistolary culture and urban history, as well as on digital humanities methods for studying Chinese history. He is a Fellow of both the Royal Historical Society and the Royal Asiatic Society of Great Britain and Ireland. He has also held fellowships at Academia Sinica in Taiwan, Max Planck Institute for the History of Science in Berlin, and University of Western Australia in Perth. Recognized for his teaching, he received his

		university's Teaching Excellence Award and Hong Kong's UGC Teaching Award (Early Career Faculty Members) in 2023.
120 VU	Huong Nam VU	Huong Nam VU is an MA student in Applied Linguistics at National Sun Yat-sen University, Taiwan.
121 WANG	Kubert WANG	Kubert WANG
122 WANG	Simon WANG	Simon WANG, Dr Wang has been teaching as a lecturer in English at the Language Center since 2016. At the LC he has taught a number of credit-bearing and supplementary courses. He has also been active in developing AI-enhanced teaching and learning platforms with funding support from the university.
123 WANG	Xi WANG	Xi WANG received his PhD in Optics from the University of Science and Technology of China (USTC) in 2016. Since 2017, he has been a associate researcher at the College of Electronic Engineering, National University of Defense Technology (NUDT), China. His research focuses on laser irradiation effects.
124 WANG	Yu Hsuan WANG	Yu Hsuan WANG, Ms. Wang Yu Hsuan holds a Bachelor's degree in English Studies from HKU and a Master's degree in Applied Linguistics with distinction from University College London. An in-service English teacher, she is pursuing a Doctor of Philosophy in Education, focusing on language pedagogy, technology, and generative AI to enhance teaching.
125 WEI	Yingmei WEI	Yingmei WEI received the Ph.D. degree in computer science and technology from the National University of Defense Technology, China, in 2000, where she is currently a Professor with the College of Systems Engineering. Her research interests include virtual reality, information visualization, and visual analysis techniques.
126 WONG	Chi-Ming WONG	Chi-Ming WONG is currently an Associate Professor of the Department of Health Technology and Informatics at The Hong Kong Polytechnic University and Fellow of Higher Education Academy (FHEA), focusing on Course-Based Undergraduate Research Experience and GenAI-augmented flipped education.
127 WONG	Chun Fan WONG	Chun Fan WONG, Jacky Wong comments, broadcasts and writes on art, culture and education for Hong Kong newspapers and magazines and is art critic and contributor for newspaper columns. He is a member of Hong Kong Arts Administrators

		Association and Hong Kong Society for Education in Art. He is also a painter.
128 WONG	Debbie WONG	Debbie WONG
129 WONG	Macy Mei Chi WONG	Macy Mei Chi WONG, Dr Macy Wong is the Head of Employability Services and Senior Lecturer at the College of Professional and Continuing Education, the Hong Kong Polytechnic University. Her research interests include internationalization and pedagogy in higher education, employability, entrepreneurship. Her work has been published in Asian Business & Management, Career Development International, etc.
130 WONG	Paulina WONG	Paulina WONG
131 WONG	Rachel WONG	Rachel WONG
132 WONG	Tiffany WONG	Tiffany WONG
133 WU	Mephisto WU	Mephisto WU
134 WU	Tian, Talia WU	Tian, Talia WU, Dr. Wu holds a PhD in statistical genetics and is a lecturer in the Department of Mathematics. She teaches a faculty-wide course in the Faculty of Science, one general education capstone course and two master's courses. She serves as Associate Program Director of MSc in Operational Research and Business Statistics.
135 XIAO	Jialing Juley XIAO	Jialing Juley XIAO
136 XU	Kai XU	Kai XU
137 XU	Nathan XU	Nathan XU, Dr. Nathan Xu obtained a PhD degree from HKUST in 2023. He serves as a Teaching Fellow in the Department of Mechanical Engineering at PolyU. Before joining PolyU, Dr. Xu worked with the Division of Integrative Systems and Design at HKUST. His teaching expertise encompasses mechatronics, robotics, design thinking, and entrepreneurship.
138 YANG	Shuo YANG	Shuo YANG received the Ph.D. degree in software engineering from the National University of Defense Technology, Changsha, China, in 2022. He is currently an associate professor at the College of Systems Engineering, National University of Defense Technology. He has taught the courses of "computer programming" and "Simulation Software Technology".
139 YANG	Yang YANG	Yang YANG received the B.S. and Ph.D. degrees in Computer Science from the University of Science and

			Technology of China (USTC). He is currently a postdoctoral researcher at the College of Systems Engineering, National University of Defense Technology. His research focuses on deep learning.
140	YAO	Jiale YAO	Jiale YAO is an undergraduate in the Qian Xuesen Honors Class at National University of Defense Technology, an elite program selecting 40 top students. Intrigued by the application of L2 writing and Generative AI (GenAI) in education, she conducts research on AI-assisted collaborative writing via academic courses and faculty guidance.
141	YEUNG	Edmond YEUNG	Edmond YEUNG, Dr Edmond Yeung has extensive experience in teaching and learning, student development, programme management and marketing. His current teaching and research focus is on personal and leadership development, active learning, blended learning, experiential learning as well as innovation and entrepreneurship.
142	YIN	Quanjun YIN	Quanjun YIN
143	YOSHIZUKA	Hiroshi YOSHIZUKA	Hiroshi YOSHIZUKA has served in various leadership positions including at the EIKEN Foundation of Japan, as an advisor to The Japan Times. He currently serves as Director of the Institute for English Language Education at Seibido Publishing Co., Ltd. He has authored numerous English textbooks for university-level instruction.
144	YU	Wing Yan YU	Wing Yan YU, Dr Lydia Yu is an Assistant Professor of Practice at the School of Optometry, The Hong Kong Polytechnic University. With a robust background in optometry and vision science, she has dedicated her career to advancing eye care and optometric education. She is recognised for her contributions to both academic and clinical settings, mentoring students and collaborating with researchers. Her clinical and research interests focus on vision rehabilitation, myopia management and emerging new technologies in education.
145	ZENG	Jianbin ZENG	Jianbin ZENG, Prof. Jianbin Zeng holds a PhD degree in English Language and Literature and teaches English Academic Paper Writing at Fudan University, Shanghai, China. He currently serves as both a teaching supervisor of Fudan College, and an education supervisor at the Graduate School of Fudan University.
146	ZHANG	Miao ZHANG	Miao ZHANG
147	ZHANG	Qi ZHANG	Qi ZHANG received the Ph.D. in control science and engineering from the National University of Defense

		Technology, Changsha, China 2018. He is currently a lecturer at the College of Systems Engineering, National University of Defense Technology. His research interests include behavior tree modeling and Computer Generated Forces (CGF) modeling, learning and training.
148 ZHANG	Xuhua ZHANG	Xuhua ZHANG
149 ZHANG	Yi ZHANG	Yi ZHANG, Dr Yi Zhang received her PhD degree from the Hong Kong Polytechnic University in 2025. She serves as Research Associate for a government funded research project.
150 ZHANG	Emma ZHANG	Emma ZHANG, Dr. Emma H. Zhang is English Lecturer at the Language Centre of Hong Kong Baptist University. With extensive teaching experience in academic writing, creative expression, and literature, her research spans comparative literature, mythology, and e-learning. She has particular expertise in virtual exchange, having conducted multiple VE projects and serving as a facilitator for Soliya's Global Connect programme, a leading initiative in the field. Her publications appear in journals such as Life Writing and Cha.
151 ZHAO	Dapeng ZHAO	Dapeng ZHAO received his PhD in Signal and Information Processing from the College of Electronic Engineering, National University of Defense Technology (NUDT) in 2008. Since 2019, he has been an Associate Professor at the College of Electronic Engineering, NUDT, China. His research focuses on infrared physics.
152 ZHAO	Rong ZHAO	
153 ZHOU	Laura ZHOU	Laura ZHOU, Dr. Zhou is senior educational development manager (senior digital learning specialist) of Educational Development Center at HKPolyU. She is Fellow of Higher Education Academy (FHEA) and obtained CMALT qualification (Certified Membership of Association for Learning Technology). She is experienced in QA for eLearning, technology-enhanced learning and teaching and blended learning.
154 ZHOU	Xiaona ZHOU	Xiaona ZHOU, Xiaona Zhou is an EdD candidate at the Department of Linguistics and Modern Language Studies, the Education University of Hong Kong. Her research interests include speech perception and production as well as technology-enhanced pronunciation instruction in diverse linguistic contexts.

155 ZHU	Dianyong ZHU	Dianyong ZHU, Prof. Dianyong ZHU is the Dean and Professor at School of Foreign Studies, Harbin Engineering University. He serves as a council member of the China Association of Higher Education. His research interests focus on Second Language Acquisition, language testing, and pedagogy, with over 30 academic publications.
156 ZHU	Zhengqiu ZHU	Zhengqiu ZHU

Index

3D Holographic Projection 531
21st-century Skills 452, 454-6, 458, 581

Academic English Writing 150
Academic Integrity 15, 33, 70, 79, 80-3, 85, 151, 159, 185, 222-4, 226, 228-9, 270, 305-7, 343, 347, 385, 405, 543, 545
Activity Design 34, 407, 411
AI
-agents 110-2, 468, 549
-and Education 26
-assisted Feedback 517, 520
-assisted Grading 187-8, 195, 233, 240, 378-9
-assisted Language Learning 90
-assisted Learning 57-9, 65, 145, 273, 282, 442
-cognition 442-5, 447-9
-collaboration 157-8, 160, 320, 322, 506
-dependence 218, 491-2
-empowered writing 343
-ethics 159, 385, 434, 437-8, 487
-experience 427, 429, 431, 442-3
-feedback 39, 41-5, 207,212-3, 215, 217, 517-8, 521, 523-4
-grading 233-6, 238, 328
-in Education (AIEd) 119
-interaction 217, 442-9, 501-2
-interaction Knowledge 426-8
-literacy 272, 278, 326, 335, 340, 385, 387, 393-8, 407, 413, 434-9, 442, 452-5, 458, 463-8, 487, 491-2, 494, 496, 501, 503, 517, 523, 526, 528, 544
-mediated language learning 207
-Readiness 311, 351, 463-7, 492
-usefulness 491, 495
Artificial Intelligence (AI) 29, 49, 61, 81, 90, 104, 142, 240, 263, 434, 491, 517
Artificial Intelligence in Education (AIEd) 240
Artificial Intelligence Literacy 434, 452
Assessment Redesign 541-2, 544-6
Assessment Reform 13, 16, 546
Assessment 222, 233, 305, 360
Authentic Learning 57, 350, 423, 452-9
Automated Essay Scoring (AES) 328-9, 332
Automated Grading 190, 222, 233

Business Education 114, 272-3, 278
Business Management Education 163-4
Business pedagogy 114

Case-study Student-driven Practice 114
Chatbot 23, 50, 55, 75, 108, 115, 127, 130-6, 138, 255, 259, 290, 356-7, 505, 507, 510
ChatGPT 23, 50, 54, 61-2, 64, 66, 78, 81, 100, 133, 151-2, 163-6, 171,173-5, 178-180, 184-5, 222-3, 263, 306, 329, 335, 338, 340, 344, 349, 366, 369, 374-7, 387, 390, 442, 470-5, 481, 483, 541, 543, 545, 565-8
Chemical Education Research 187
Chemistry Education 187
Chinese EFL learners 48-9. 51-3
Classroom Discourse 119
Clinical Simulation 255
Cognitive Development 57, 70, 350-1, 463-5, 467, 577
Cognitive Offloading 19-20, 104, 278, 470-1, 476, 501
Collaborative Learning 163-6, 168-9, 178-9, 278, 356, 531, 536, 538
Collaborative Problem-solving 223, 289, 291-2, 299
Collaborative Writing 157
Communication 21, 48, 61-2, 67-8, 90-5, 100, 142-7, 150, 166-7, 171-2, 175-6, 199, 210, 227, 240-3, 282-3, 286, 289-290, 308, 345, 347, 353, 355-6, 360, 367, 378, 404, 456, 517, 541, 544, 546, 549, 562, 581
Connected Speech Training 48
Conversational AI 55-8, 136, 289, 481
Copywriting 61-2, 67
Course paper assessment 343-4, 346-7
Creative Displacement Anxiety (CDA) 526
Creative Problem-solving 55, 449, 455, 565, 567
Creative Resilience 526-8
Creative Thinking 442-9
Critical AI Literacy 150-1, 154, 207, 212, 487, 501
Critical Thinking 61-2, 66-7, 71, 77, 146, 151, 154, 161, 185, 223, 278, 309, 321, 335, 338, 348, 353, 355-6, 360, 391, 393, 405, 417-8, 420, 438, 470-1, 474-6, 497, 501-2, 507
Crystallographic Structures 531, 533, 536

Data Analysis Report 378
Deep Reading 470, 475
Design Cognition 55-6
Design Education 198, 240, 242, 478-9, 526
Design Thinking 565-6, 569, 571-3
Dialogue Analysis 119
Digital Humanities 505, 509
Digital Literacy 30, 178, 180, 185, 335, 438, 470

Digitally Assisted Intercultural Communication 90
Doubao 48, 50, 52, 344

Education 247
Educational Innovation 417, 423
Educational Materials 263-4
Educational Robotics 452
Effectual Entrepreneurship 55-6
EFL Learners 150-1, 215, 343-4, 346-7, 470, 475
EFL Pedagogy 61
EFL Writing 157
Engagement 374
English Connected Speech Processes 48
English Oral Presentations 39, 40, 42-3
Equity Valuation 272-7
Ethics and Integrity 100, 305-6, 350
Evaluation Frameworks 393
Expertise 13-5

Faculty Development 70-1, 76-8, 311, 400, 407, 431, 542
Feedback 374
Flipped Classroom 565
Flipped Learning 417-8, 420-3, 554
Future Work Skills 491-2, 495

GenAI 233, 296, 407
GenAI Assessment 335
GenAI Enabled Method 565
GenAI Tools 30-2, 53, 59, 99, 100, 110, 160, 178, 183, 185, 208-9, 222, 224-9, 305, 335-9, 356, 385-6, 400, 406, 422, 442, 471, 478-9, 480, 482, 541, 543-4, 567
Generative AI (GenAI) 13, 142, 150, 198, 240, 255, 272, 313, 320, 385, 400, 417, 470, 478, 501, 505, 541, 565, 571, 581
Generative AI Literacy 19
Generative Artificial Intelligence (GenAI) 127, 157, 163, 171, 178, 187, 207, 240, 247, 263, 272, 297, 314, 335, 400, 426, 442, 526, 531, 533
Generative Artificial Intelligence (GenAI) in Higher Education 163, 207, 426
Geomatics 247-9, 251-2
Grading Consistency and Fairness 313
Grading Platform 187

Healthcare Education 255
Health Science Education 335
Higher Education 13, 393, 400, 417, 491
Higher Education Assessment 328
Higher Education Pedagogy 565
History Education 505, 511
HiVE 575-8
Human–AI Collaboration 19, 67, 127, 158, 163-4, 169, 263, 267, 269, 270, 277, 428, 517-8, 526, 528
Human-AI Interaction 15, 160, 163, 169, 428, 528
Human Anatomy Education 178
Human-in-the-loop 189, 337, 340-1, 378, 380, 385
Hybrid Intelligence 39, 40, 44-6, 305-6, 517-9, 521-4, 528

IELTS 91, 127, 130-1, 133, 135-6, 207-13, 329
Immersive Learning 146, 247, 549-52, 557, 581
Immersive Visualisation 531, 538
Incorporating GenAI in Learning 100, 105, 111, 305-6, 351
Information Literacy 505-6, 510-1
Institutional Change 422, 541-2, 546
Interactive STEM Pedagogy 531
Interdisciplinary Education 296-7, 302
Interdisciplinary Learning 77, 408, 565
Intrinsic Motivation 374-6, 393-7

Journalism Education 282

Knowledge Co-creation 247, 263
Knowledge Graphs 247, 557, 563
Kotter's Model 541-3, 546

L2 speaking 289-90
Large-Enrollment Courses 187
Large Models 557
Laser Irradiation Effect 296
Leadership Education 80, 165, 169
Learner Autonomy 66, 215
Learning Analytics 119, 300, 302, 350
Learning Difficulties 198
Learning Outcomes 40, 76, 103, 106, 110, 121, 143, 146,164,176, 185, 200, 204, 223-4, 227, 248, 256, 259, 282-3, 287, 301, 305, 309, 324, 326, 337, 351, 355-6, 374-7, 385, 388, 401, 419-20, 501, 576-8

Media Design 242-3, 478-80, 482-4, 486-7
Metacognition 42, 465, 467, 501
Metaverse 422, 549-54, 575-8, 581
Meta-work 26-30, 32-4
Multimodal Learning 335
Multimodal Literacy 61, 66, 68

Multi-persona AI 289-90, 292

Negotiation 114-6
New Academics 400-6
Non-STEM 407-10, 413, 415
Nursing Education 142-3, 146

Occupational Therapy 171-2, 175
OpenImpactLab 463-4, 467-8
Open-source LLM 328

Pedagogical Design 20, 24, 78, 337, 464, 501-3, 506
Pedagogical Innovation 56, 62, 67, 80, 178, 272-3, 300, 344, 538
Pedagogy 39, 55
Pedagogy and technology integration 282
Peer Assessment 39-43, 45
Peer Collaboration 157-61, 348, 536
Personalised Learning Pathway 549
Polyadic Interaction 289-90
Positive youth Development 80, 85, 434, 435
Postgraduate Course 296
Problem-Based Learning 557
Process Assessment 215
Process-based Evaluation 343, 369
Professional Development 34, 105, 111, 240, 270, 393-4, 396-8, 400, 402, 406-8, 413-5, 418, 422, 426-7, 431, 544-5, 558
Professional Learning 26, 28, 397
Programming Exercise 320-1
Project-based Learning 297, 463-4
Prompt Engineering 23, 142-4, 146-7, 187, 189, 194, 266, 277-8, 299, 340, 408, 472, 476

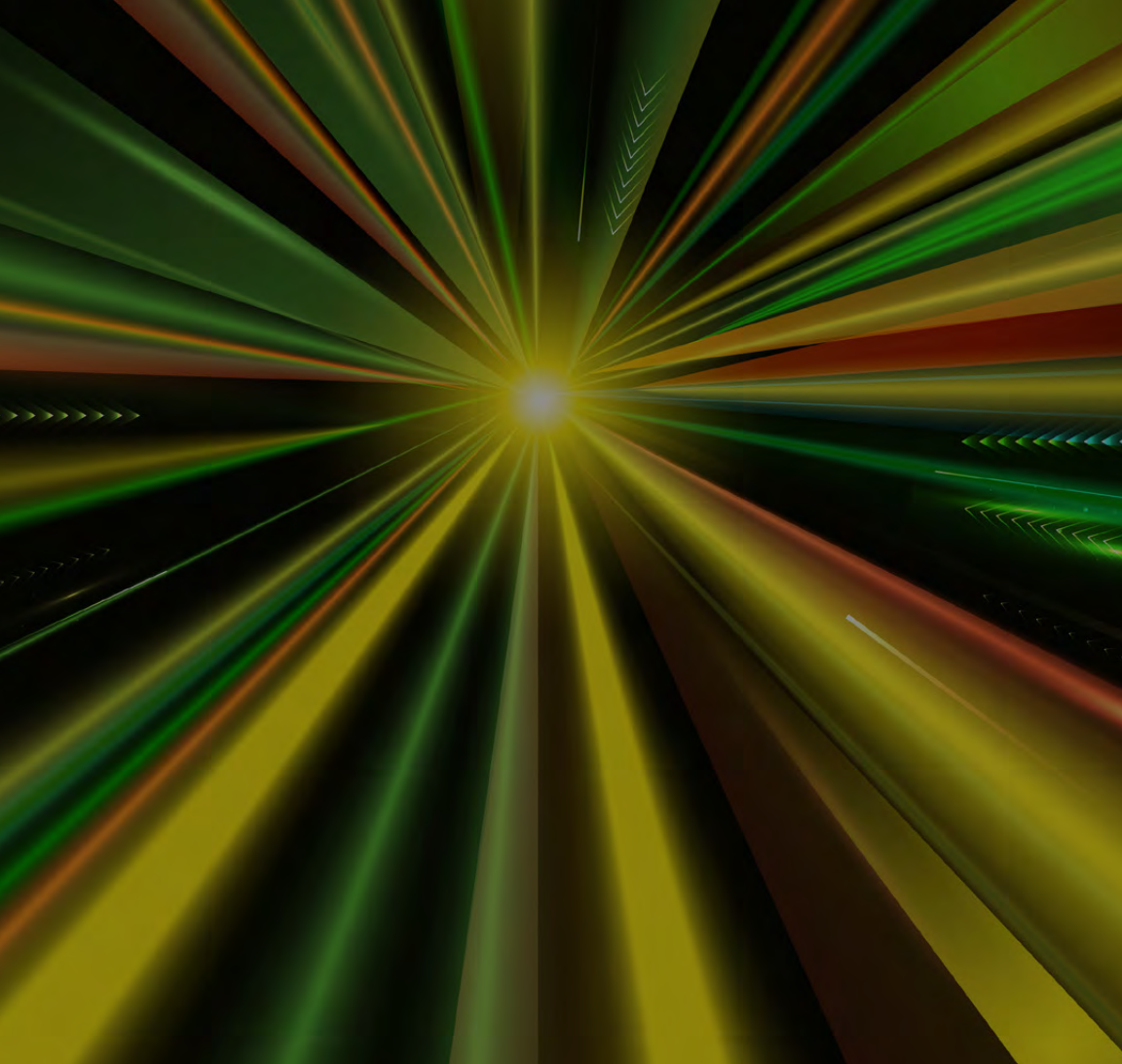
Read-write-comment 320-3
Reflective Essay Scoring 328
Reflective Writing 223, 233, 328, 332, 505, 510
Role-play Pedagogy 255
Rubric-based Assessment 378

Scale Development 426, 428
Science Education 222, 229, 248, 251
Second Language Speaking 127
Second Language Writing 517
Self-assessment 39, 41, 96-7, 234, 329, 346-7, 398
Self-efficacy 127-9, 132-3,135, 198-9, 202-3, 217, 393-7, 435-7, 491-3, 495-7
Self-regulated Learning 20, 40, 66, 188, 374, 501, 503
Semantic Similarity 119, 124
Service-Learning 73, 575-7
Simulated Patients 171
Skills Development 255
Smart Teaching 557-9, 563
Speaking Practice 127-32, 135-6, 207, 212-3, 215, 289-90, 294
Speaking Test Preparation 207
STEAM Competency 393-8
STEM 163, 169, 187, 195-6, 310, 407-11, 413, 415, 581
Student AI Literacy 282, 310, 350-1
Student Engagement 32, 110-1, 134, 163, 169, 185, 296, 335-6, 340, 385, 388-90, 417, 421-2, 458, 549
Students' Perception 85, 222, 339, 518, 521
Student Thriving 434

Task-based Learning 169, 557
Task Sequence 157
Teacher Feedback 39, 517, 522-4
Teacher Professional Development 105, 111, 356, 385, 426
Teaching and Assessment 70, 305, 391, 541-2
Team-based Learning 531-2
Technological Social Responsibilities 80-1, 83, 85
Technology Adoption 400-1, 403, 492
Textbook Development 263-5
Theory of Planned Behaviour (TPB) 240-1
Threshold Concepts 19-21, 24
TPACK 426-8, 430-1
Transfer of Learning 385, 391
Transparent Rubrics 313-4, 316

Venture Creation 55-6, 59
Video-based Learning 80-1, 85
Virtual Exchange 90
Virtual Reality 142, 532, 575

Workflow Formation 198-9, 202, 204
Writing Pedagogy 19, 215
Writing Performance 150-2, 157-8, 161, 518-22



Published in Hong Kong, China; by the project entitled “Exemplary learning, teaching and assessment practices in UGC-funded institutions in the age of GenAI [ELTA G-AI]”, funded by the Fund for Innovative Technology-in-Education (FITE) - Inter-institutional Collaborative Activities (IICA) from the Hong Kong University Grants Committee (UGC).

Title: Collection of Selected Papers from the International Conference on GenAI and Pedagogical Innovations 2026 / edited by Julia Chen, Alle Leung, Elza Tsang, Angela Ng, Jacky Chau, Roy Kam, Mitesh Patel, Dawn Lo, Barbara Tam, Leo Chon, Kevinia Cheung, Eunice Tang and Kelly Ho.

Identifiers: ISBN 978-988-19170-0-3 (PDF) | Handle http://hdl.handle.net/10397/121738

Produced by

Educational Development Centre

The Hong Kong Polytechnic University