



# 融合前沿科技： 成就新形态的 学习者

Embracing Frontier  
Technology:

Cultivating the New  
Paradigm of Learners

## 第二十九届 全球华人计算机 教育应用大会

GCCCE 2025

大会论文集(英文论文)  
Main Conference Proceedings  
[ English Paper ]

中国 无锡 | 江南大学

中国 香港 | 香港理工大学

Jiangnan University, Wuxi, China

The Hong Kong Polytechnic University, Hong Kong, China

2025年5月24日至28日

24th - 28th May 2025

学习科学与计算机支持协作学习

移动、泛在与情境化学习

优雅化学习、教育游戏与数字玩具

高等教育与成人学习的技术应用、教师专业发展

科技增强语言与人文学科学习

人工智能教育应用、智慧学习环境

学习分析与学习评估

STEM与创客教育

教育技术创新、政策与实践



Publisher: Global Chinese Society for Computers in Education

Title: Conference Proceedings (English Paper) of the 29th Global Chinese Conference on  
Computers in Education (GCCCE 2025).

Editors: SUN Daner, SHANG Junjie, WEN Yun, WU Sheng-yi, KONG Siu Cheung, MA Zhiqiang,  
HU Xiangen

Edition date: 2025.06

Edition: princeps

ISSN: 3005-3226

# **第 29 届全球华人计算机教育应用大会**

**The 29th Global Chinese Conference on Computers in Education**

## **GCCCE 2025 大会论文集（英文论文）**

### **GCCCE 2025 Main Conference Proceedings**

#### **(English Paper)**

#### **Editors**

SUN Daner, The Education University of Hong Kong

SHANG Junjie, Peking University

WEN Yun, Nanyang Technological University

WU Sheng-yi, Tsing Hua University

KONG Siu Cheung, The Education University of Hong Kong

MA Zhiqiang, Jiangnan University

HU Xiangen, The Hong Kong Polytechnic University

#### **Associate Editors**

Fang Jianwen, Wenzhou University

HSIA Lu-Hoia, Chin-Yi University of Technology, Taiwan

CHEN Zhi-hong, Taiwan Normal University, Taiwan

Yang Xian-Min, Jiangsu Normal University

AW Guat Poh, Nanyang Technological University, Singapore

Liu Xiangyong, Jiangnan University

HUNG Hui-Chun, Central University, Taiwan

CHIANG Feng-Kuang, Shanghai Jiao Tong University

Hsu Ting-Chia, Taiwan Normal University, Taiwan

Rustam SHADIEV, Zhejiang University

#### **Editing and Proofreading**

LIU Jingxi, Jasper, The Education University of Hong Kong

XU Zhiwen, Wynna, The Education University of Hong Kong

ZHENG Zhizi, The Education University of Hong Kong

WANG Chen, The Education University of Hong Kong

JIA Fenglin, The Hong Kong Polytechnic University

WANG Juan, The Education University of Hong Kong

TIAN Peiyao, The University of Hong Kong

**Cover Page Design**

Gu Yan, Jiangnan University

**Coordinating and Designing the Conference Certificate**

WANG Jing, Jiangnan University

## **Table of Contents**

1. Message from Organizer .....	1
2. Conference Organization .....	6
3. Keynote Speech .....	10

### **Main Conference English Paper Track**

Open Education and Global Competence: Advancing Glocalized Practices.....	17
Pre-service Teachers' Attitudes and Behavioral Intention Towards Generative Artificial Intelligence: A Structural Equation Modeling Investigation Based on TAM .....	25
Exploring the Effect of Scaffolding Strategies in GenAI Chatbot on Student Engagement and Programming Skill Development.....	32
Impact of Technology Enhanced Cooperative Learning based of Dialogic Feedback Strategy on Students' Learning Effectiveness .....	40
Exploring the Effects of VR-assisted Learning on Students' Learning Engagement in a University EFL Classroom .....	44
Developing an AI-empowered Chinese Composition Assessment System for Hong Kong Students	48
Investigating Learner Emotional Engagement in Synchronous Online Learning .....	54
Revolutionizing Language and Culture Education through Immersive 360-Degree Video and Artificial Intelligence .....	58
A Pilot Study on Bridging EFL Writing and Speaking Skills through AI-enhanced Authentic Short Video-Making .....	62
Evaluation of a Professional Development Course on Artificial Intelligence Literacy for Administrative Staff in Hong Kong .....	70
Exploring the Mechanisms for Implementing AR-based Learning Activities through the Lens of Motivational Design.....	78
A Self-regulated Learning Approach to Promoting Learners' Participation and Performance in Writing Education .....	87
The Study on the Implementation of Artificial Intelligence Technology in College English Instruction .....	91
Exploring the Potential of AI-Generated Lesson Designs Underpinned by the TPACK Framework for Educators in Higher Education: A Comparative Study.....	95

A study on the Impact of AI-Driven STEAM Curricula on Computational Thinking Development in the Chinese Elementary School .....	99
Research on the Design and Practice of Maker Educational Activities for the Cultivation of Computational Thinking .....	103
Evaluation of ChatGPT's Implementation in Undergraduate Java Programming Tutorials .....	107
Students' Perceptions and Usage Patterns of ChatGPT as an Automated Writing Evaluation (AWE) Tool.....	114
Interactive Effects of Graphic Organizer Strategy and Strategy Preference on Self-Paced Video Learning .....	118
Empowering Socratic Teaching Through Large Language Models: New Opportunities and Challenges in Intelligent Tutoring Systems .....	122
What Makes the Ideal AI Collaborator? Exploring Student and Teacher Perspectives on Roles, Support, and Challenges .....	125
Impact of Using a Smart Online Teaching Tool “Rain Classroom” on Sophomores’ Class Engagement in an Intercultural Communication Course.....	129
AI-Powered Socratic Learning for Psychological Statistics: Enhancing Flipped Classroom Practices with Generative Intelligent Tutoring Systems .....	134
Neither Schoolgirls nor Schoolboys but STEM Learners: Develop and Evaluate Training Course for Hong Kong Secondary School STEM Teachers to Cultivate Gender Equality in the Learning Environment: Case Study .....	137
Exploring EFL Learners’ Academic Emotions and Emotion Regulation Strategies in AI-Assisted Collaborative Academic Writing Tasks.....	145
ChatGPT Usage Patterns in Essay Writing: A Case Study of Advanced, Intermediate, and Low-Proficiency English Learners .....	147

## 1. Message from Organizer

The Global Chinese Conference on Computers in Education (GCCCE) was founded in May 1997. Over the years, it has developed into a significant annual gathering for Chinese and international scholars, educators, and policymakers to share and discuss the latest research developments in the application of computers in education. The conference program typically includes keynote addresses, research paper presentations, workshops, doctoral student forums, and forums for primary and secondary school teachers. Since its inception in 1997, the GCCCE has been hosted in various cities, including Beijing (five times), Guangzhou (four times), Hong Kong (three times), Singapore (twice), Taipei (twice), Macau, Zhongli, Nanjing, Honolulu (Hawaii), Lansing (Michigan), Hangzhou, Kenting, Shanghai, Wuhan, Lanzhou, and Chongqing.

The 29th Global Chinese Conference on Computers in Education (GCCCE 2025) was held from May 24 to 28, 2025, at Jiangnan University in Wuxi and The Hong Kong Polytechnic University in Hong Kong SAR, China. The theme of GCCCE 2025 is "**Embracing Frontier Technology: Cultivating the New Paradigm of Learners**".

The GCCCE conference brings together education policy makers, scholars, educators, principals, and frontline teachers from around the world to exchange and share the latest research work and achievements in computerized education applications. The conference program includes keynote speeches, paper presentations, workshops, doctoral forums, K-12 teacher forums, and corporate exhibitions. Meanwhile, the EPT (English Paper Track) was established for the sixth time in this conference, and a total of 42 article submissions were received. This demonstrates the growing attention from non-Chinese researchers to the GCCCE Conference over the past six years, attracting numerous authors from English-speaking countries to participate and exchange ideas. Furthermore, this year's GCCCE will continue the tradition of the previous conference by arranging two English keynote speeches and inviting two top international scholars to deliver the talks.

The GCCCE 2025 conference includes ten conference themes:

- C1: Learning Sciences & Computer-Supported Collaborative Learning
- C2: Mobile, Ubiquitous & Contextual Learning
- C3: Joyful Learning, Educational Games & Digital Toys
- C4: Technology Applications in Higher Education and Adult Learning, Teacher Professional Development
- C5: Technology-Enhanced Language and Humanities Learning
- C6: Artificial Intelligence in Education Applications and Practices, Intelligent Learning

---

## Environments

C7: The Learning Analytics and Learning Assessment

C8: STEM and Maker Education

C9: Educational Technology: Innovations, Policies & Practice

EPT: English Paper Track

Within EPT and each sub-conference, an Executive Chair, Co-Chairs, and Program Committee (PC) Members were appointed to manage the review and programming processes. Some sub-conferences also have review committee members and advisors. GCCCE 2024 invites Chinese paper submissions from Chinese scholars worldwide and English paper submissions globally. This conference has received a total of 440 submissions from 874 authors. These paper submissions come from nine countries and regions, including Mainland China, Taiwan, Hong Kong, Macau, United States, Uzbekistan, Kazakhstan, Vietnam and Singapore. The statistical data of the regions of the submitting authors are shown in Table 1.

Table 1. Statistics of the regions of authors submitting to the nine sub-conferences and EPT of GCCCE 2025

Region	MCN	TW	HK	SG	MO	UZB	KAZ	US	VNM
No. of author	699	130	29	8	3	2	1	1	1
No. of papers	338	68	10	4	1	1	1	1	1
No. of Accepted papers	261	61	10	4	1	1	1	1	1

Note: MCN - Mainland China, TW - Taiwan, HK - Hong Kong, MO - Macau, US - United States, UZB - Uzbekistan, KAZ - Kazakhstan, VNM - Vietnam and SG - Singapore

Each submission was assigned to at least three PC members for the first round of review. The results were then meta-reviewed by the chair and co-chairs of the corresponding sub-conference or EPT before a final decision was made. Through this rigorous review process, 336 papers were accepted (see Table 2), with an acceptance rate of 76.4%. Among them, 7 papers were nominated for the Best Chinese Research Paper Award (limited to long papers accepted by the sub-conferences), 4 were nominated for the Best English Research Paper Award (limited to long papers accepted by the EPT), 6 were nominated for the Best Student Paper Award (limited to long papers accepted by the sub-conferences and the EPT), 7 were nominated for the Best Technical Design Paper Award (limited to long or short papers accepted by the sub-conferences and the EPT), and 8 were nominated for the Best K-12 Teachers' Paper Award (limited to long or short papers accepted by the K-12 Teachers' Forum).

Table 2. GCCCE 2025 The status of submissions and acceptance for each sub-session

Sub-conference	Submission	Accept	Full paper	Short paper	Poster	Elimination	Acceptance Rate
<b>C1: Learning Sciences &amp; Computer-Supported Collaborative Learning</b>	63	45	11	33	1	18	<b>71.4%</b>
<b>C2: Mobile, Ubiquitous &amp; Contextual Learning</b>	20	16	3	7	6	4	<b>80.0%</b>
<b>C3: Joyful Learning, Educational Games &amp; Digital Toys</b>	34	25	7	9	9	9	<b>73.5%</b>
<b>C4: Technology Applications in Higher Education and Adult Learning, Teacher Professional Development</b>	53	41	12	19	10	12	<b>77.4%</b>
<b>C5: Technology-Enhanced Language and Humanities Learning</b>	14	11	3	7	1	3	<b>78.5%</b>
<b>C6: Artificial Intelligence in Education Applications and Practices, Intelligent Learning Environments</b>	107	87	20	65	2	20	<b>81.3%</b>
<b>C7: The Learning Analytics and Learning Assessment</b>	28	20	5	11	4	8	<b>71.4%</b>
<b>C8: STEM and Maker Education</b>	30	22	6	11	5	8	<b>73.3%</b>
<b>C9: Educational Technology: Innovations, Policies &amp; Practice</b>	49	37	8	29	0	12	<b>76.0%</b>
<b>EPT</b>	42	32	9	17	6	10	<b>76.2%</b>
<b>Total</b>	<b>440</b>	<b>336</b>	<b>84</b>	<b>208</b>	<b>44</b>	<b>104</b>	<b>76.4%</b>

This conference has invited four academic experts to deliver keynote speeches. They are Prof.Chen Mingxuan from Jiangnan University, Prof. Tak-Wai Chan from National Central University, Prof. Art Graesser from University of Memphis, Prof. Hui-Chun Chu from Soochow University. The keynote speeches are as follows:

### **Keynote 1: Why Education? What's Education For? ‘Global Harwell’ and ‘General Artificial Companions’ Co-Shape the Future**

Speaker: Professor Tak-Wai Chan, Chair Professor at National Central University in Taiwan

### **Keynote 2: Empowering Intelligence: The Reconstruction of Education from Knowledge Transmission to Future Wisdom**

Speaker: Professor Chen Mingxuan, the second level professor at Jiangnan University, the renowned teacher and doctoral supervisor in Jiangsu Province

### **Keynote 3: Conversations with Computer Agents in the Era of Generative AI**

Speaker: Professor Art Graesser, professor in the Department of Psychology and the Institute of Intelligent Systems at the University of Memphis, the Honorary Research Fellow at University of Oxford.

---

**Keynote 4: Research Design and Innovative Teaching Models Integrating Generative AI with Educational Technology**

Speaker: Professor Hui-Chun Chu, Distinguished Professor of the Department of Computer Science and Information Management; The chairman of the Institute of Human-Centered Artificial Intelligence

These four keynote speeches, nine sub-conferences, EPT, five forums (including the Main Conference Invited High-level Forum in the Jiangnan University session, two Main Conference sub-forums, and the Higher Education Forum and Sessions Forum in the Hong Kong Polytechnic University session), and the two new Journal Forums added this year, together make up the main conference of the GCCCE 2025. The main conference invited a topical discussion panel theme, "Why Education? What's Education For? 'Global Harwell' and 'General Artificial Companions' Co-Shape the Future (教育为何？教育何为？“全球和幸”与“通智同伴”共塑未来)." Subforum 1 in Jiangnan University is titled “Generative AI - Driven Teacher Learning and Professional Development: Practical Exploration and Theoretical Reconstruction (生成式人工智能驱动的教师学习与专业发展：实践探索与理论重构)", subforum 2 in Jiangnan University is titled “Learning with 'Wisdom': The Application of Smart Technologies in Contextualized, Authentic, and Communicative Language Learning (学语有“智”：智慧科技在情境化、真实性和交际性语言学习中的应用)", the Higher Education Forum in the Hong Kong Polytechnic University is titled “Higher Education in the Era of Advanced Reasoning Models (先进推理模型时代中的高等教育)", Subforum 1 in the Hong Kong Polytechnic University is titled “Generative AI-Powered Immersive Learning: Unlocking the Future of Education (生成式人工智能驱动的沉浸式学习：开启教育的未来). The above forums are highly significant topics for discussion in the era of artificial intelligence. In order to respond to the participants' need for publication and to share the most recent development, this conference featured a special session of dialogues between the editors-in-chief of SSCI English journals and Chinese journals, which provided an efficient communication platform for academic advancement and sharing.

The conference would like to thank all those who collaborated to make this session possible. We express our sincere gratitude to the Executive Chairs, Co-chairs, committee members, and volunteers of the nine sub-conferences, the EPT, the workshops, the K-12 Teachers' Forum, and the Doctoral Student Forum, and the Local Organizing Committee members for their help during the preparation of the conference. We would like to extend a special thank you to the conference chair for his guidance and support in the conference coordination.

We sincerely hope that you will enjoy the various conference activities of GCCCE 2025 and gain rich inspiration and wonderful experiences from them. Let us work together to build a more resilient and internationalized GCCCE academic community and pass on the spirit of GCCCE.

KONG Siu Cheung, The Education University of Hong Kong, Hong Kong  
Conference Chair

SUN Daner, The Education University of Hong Kong, Hong Kong  
International Program Coordination Committee Chair

SHANG Junjie, Peking University  
International Program Coordination Committee Co-Chairs

WEN Yun, Nanyang Technological University, Singapore  
International Program Coordination Committee Co-Chairs

WU Sheng-yi, Tsing Hua University, Taiwan  
International Program Coordination Committee Co-Chairs

## 2. Conference Organization

### Organizer

Global Chinese Society for Computers in Education (GCSCE)

Jiangnan University

The Hong Kong Polytechnic University

### Host

School of Humanities Jiangnan University

Jiangsu Research Center of “Internet Plus Education”, Jiangnan University

Institute for Higher Education Research and Development, The Hong Kong Polytechnic University

### Conference Chair

KONG Siu Cheung, The Education University of Hong Kong, Hong Kong

### Conference Consultants

CHAN Tak-Wai, Central University, Taiwan

CHU Hui-Chun, Soochow University, Taiwan

WONG Lung Hsiang, Nanyang Technological University, Singapore

LIU Geping, Southwest University

### International Program Coordination Committee

#### Chair

SUN Daner, The Education University of Hong Kong, Hong Kong

#### Co-Chairs

SHANG Junjie, Peking University

WEN Yun, Nanyang Technological University, Singapore

WU Sheng-yi, Tsing Hua University, Taiwan

### Local Organizational Committee

#### Chairs

MA Zhiqiang, Jiangnan University

HU Xiangen, The Hong Kong Polytechnic University, Hong Kong

### Members of the Local Organizational Committee

#### Conference Coordination Leadership Team

Wang Hui

Wu Honglin

Chen Mingxuan

(Ma Zhiqiang

Liu Xiangyong

**Guarantee Team**

Liu Xiangyong	Qian Yizhou	Yan Dahu
Quan Guolong	Tian Na	Shu Hang

**Venue Team**

Wang Zhijun	Mou Zhijia	Wang Meng
Gao Hong	Liu Wenhui	

**Material Promotion Team**

Wang Jing	Zhang Hongying	Gu Yan
Yao Jiajia	Zhu Fangfang	

**Contact Team**

Cai Huiying)	Jin Xinquan)
--------------	--------------

**Guidance and Support Team**

Zhang Chunmei	Xu Lu	Wang Xiaogen
Sun Li	Zhang Qun	Chen Genyan
Wei Jianxin	Liu Xueqin	Xu Yue

**International and Local Coordination Team)****The Education University of Hong Kong**

LIU Jingxi, Jasper	XU Zhiwen, Wynna	ZHENG Zhizi
WANG Chen	WANG Juan	WANG Xiaolan

**The Hong Kong Polytechnic University**

Lee Wai Lun, Aka

**Sub-Conference Program Committee****C1: Learning Sciences & Computer-Support Collaborative Learning****Executive Chair**

Fang Jianwen, Wenzhou University

**Co-Chairs**

Ma Zhiqiang, Jiangnan University

Pei Leisi, The Education University of Hong Kong, HongKong

Tu Yunfang, Soochow University, Taiwan

Yuan Guangji, Nanyang Technological University, Singapore

**C2: Mobile, Ubiquitous & Contextual Learning****Executive Chair**

HSIA Lu-Hoia, Chin-Yi University of Technology, Taiwan

**Co-Chairs**

Fu Qingke, Huzhou University

HWANG Gwo-Haur, Yunlin University of Science and Technology, Taiwan

---

Liang Hsin-Yi, I-Shou University, Taiwan

Ma Ning, Beijing Normal University

**C3: Joyful Learning, Educational Games and Digital Toys**

**Executive Chair**

CHEN Zhi-hong, Taiwan Normal University, Taiwan

**Co-Chairs**

Bai Shurui, The Education University of Hong Kong, HongKong

Li Xiuhan, Central China Normal University

LIAO Chang-Yen, Central University, Taiwan

**C4: Technology Applications in Higher Education and Adult Learning, Teacher Professional Development**

**Executive Chair**

Yang Xian-Min, Jiangsu Normal University

**Co-Chairs**

Lijia Lin, University of Macau, Macau

Wang Junjie, Pingtung University, Taiwan

Zhang Danyang, Shenzhen University

**C5: Technology-Enhanced Language and Humanities Learning**

**Executive Chair**

AW Guat Poh, Nanyang Technological University, Singapore

**Co-Chairs**

Huang Fang, Shanghai International Studies University

LIEN Yujen, Chung Yuan Christian University, Taiwan

SIM Seok Hwa, Ministry of Education, Singapore, Singapore

Wu Junjie, Macau University of Science and Technology, Macau

**C6: Artificial Intelligence in Education Applications and Practices, Intelligent Learning Environments**

**Executive Chair**

Liu Xiangyong, Jiangnan University

**Co-Chairs**

Hu Xiangen, The Hong Kong Polytechnic University, HongKong

CHIN Kai Yi, Soochow University, Taiwan

---

Wu Longkai, Central China Normal University

**C7: Learning Analytics and Learning Assessments**

**Executive Chair**

HUNG Hui-Chun, Central University, Taiwan

**Co-Chairs**

Chen Fu, University of Macau, Macau

WANG Shu-Ming, Cultural University, Taiwan

Khor Ean Teng, Nanyang Technological University, Singapore

Yang Tzu-Chi, Yang Ming Chiao Tung University, Taiwan

**C8: STEM and Maker Education**

**Executive Chair**

CHIANG Feng-Kuang, Shanghai Jiao Tong University

**Co-Chairs**

CHEN Ying-Chih, Arizona State University, USA

HOU Huei-Tse, Taiwan University of Science and Technology, Taiwan

Wang Minhong, The University of Hong Kong, HongKong

Zhan Zehui, South China Normal University

**C9: Educational Technology: Innovations, Policies & Practice**

**Executive Chair**

Hsu Ting-Chia, Taiwan Normal University, Taiwan

**Co-Chairs**

CHIU Min-Chi, Taichung University of Science and Technology, Taiwan

Wang Jingyun, Durham University, UK

Yu Xiaohua, East China Normal University

### 3. Keynote Speech

#### **Keynote Speech 1**

- May 26, 2025 (Monday morning) 09:40-10:40 AM
- Location: Wen Ho Hall Auditorium

**Title of keynote speech: Why Education? What's Education For? ‘Global Harwell’ and ‘General Artificial Companions’ Co-Shape the Future**



Professor Tak-Wai Chan  
Chair Professor  
National Central University

#### **Summary**

The chaotic state of the world today resembles the ancient Spring and Autumn and Warring States periods. While these historical conflicts were once confined to the Chinese nation, they now encompass the entire globe. As educators, we must ask: Why education? What is the fundamental reason behind its existence? What is its essence? What values does it hold? After more than two thousand years of accumulated wisdom from China's history, have we found any answers—or even just temporary ones?

If we go further: What's education for? What are its ultimate purposes and goals? What should we strive for? In which direction should we move? Is it for ourselves, for others, for our own country—or also for the entire world? How should we implement and actualize these goals? Can we provide some possible answers?

As for ‘General Artificial Companions,’ like all other technologies, they are merely tools to assist in achieving educational goals—in this case, the ‘Global Harmony and Wellbeing Goal’—making the process more effective and excellent. However, due to the immense power of AI, in the foreseeable future, both the concept of ‘Global Harwell’—a term blending harmony and wellbeing—and that of ‘General Artificial Companions’ will co-shape the future of education and humanity.

The questions Why education? and What's education for? compel us to reflect deeply on the nature and purpose of education. Ultimately, they are no different from the fundamental human inquiries: Why life? and What's life for?

#### **Short Bio**

Tak-Wai Chan is Chair Professor at National Central University in Taiwan. A visionary in the field of digital learning, he originated the concept of AI learning companions and developed a prototype in 1988. By 1992, he built the world's first dedicated networked learning system for collaborative learning and competitive learning games. In 2000, he established EduCity, the largest

---

online learning community in the world. In the 2000s, he was at the forefront of research on intelligent classrooms, one-to-one technology-enhanced learning, and mobile learning. Collaborating with international researchers, he proposed the notion of Seamless Learning in 2006, describing the crucial impact of the Internet on learning. To support the long-term transformation of Asian education from examination-driven to interest and creation-driven, he developed the Interest-Driven Creator (IDC) Theory with Asian scholars in 2018. Recently, in response to the rising frequency of global conflicts and unprecedented challenges confronting humankind, he has been promoting the concept of Global Harwell (a term combining 'harmony' and 'wellbeing') alongside international researchers, as a possible universally shared educational goal. Over his career, Chan has been building various platforms for researchers and practitioners. He was a key co-founder of the ICCEs conference series in 1993 and the GCCCEs in 1997. He also founded the Association for Reading for Tomorrow in 2016 and a mini-experimental school based on IDC Theory in 2017, as the model for the future Asian education.

## **Keynote Speech 2**

- May 26, 2025 (Monday) 14:00-15:00 PM
- Location: Wen Ho Hall Auditorium

**Title of keynote speech: Empowering Intelligence: The Reconstruction of Education from Knowledge Transmission to Future Wisdom**



**Professor Chen Mingxuan**

**The second level professor at Jiangnan University, the renowned teacher and doctoral supervisor in Jiangsu Province  
Jiangnan University**

### **Summary**

The report discusses the core proposition of where education should head in the era of digital intelligence and how AI can empower educational innovation. It argues that the widespread application of artificial intelligence technology has diminished the value of "knowledge memorization," while higher-order cognitive abilities and adaptive wisdom have become the key competencies for humans to thrive in the future. As AI continues to surpass human skills, education must undergo a disruptive transformation, focusing on cultivating students' future wisdom. The report offers a unique perspective on the new implications of smart education, proposing that collective intelligence can be enhanced through human-computer collaborative models, such as large-scale AI systems. It advocates for a teaching paradigm that emphasizes "problem-led, multi-dimensional interaction and technology-empowered instruction" to shift from knowledge transmission to the cultivation of future wisdom. Finally, the report presents typical teaching cases supported by large AI models, showcasing the practical process and outcomes.

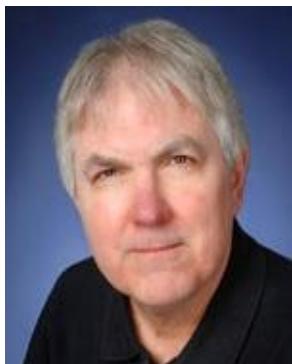
### **Short Bio**

Chen Mingxuan, a second level professor at Jiangnan University, is a renowned teacher and doctoral supervisor in Jiangsu Province; Former deans of Normal College, Education College, and Humanities College at Jiangnan University; The current Vice Chairman of the Teaching Guidance Committee for Educational Technology of the Ministry of Education, a member of the Expert Committee for Basic Education Curriculum Textbooks of the Ministry of Education, the Vice Chairman of the Information Technology Education Professional Committee of the China Education Technology Association, and the Vice Chairman of the Film and Television Media Professional Committee of the China Education Technology Association. Main research areas: Theory and Practice of Digital Education, Understanding Learning and Cognition Technical support for learning and teaching

## **Keynote Speech 3**

- May 27, 2025 (Tuesday) 9:00-10:00 AM
- Location: 209 Tian Jia Ping Building

### **Title of keynote speech: Conversations with Computer Agents in the Era of Generative AI**



Professor Art Graesser

professor at the University of Memphis, Honorary Research Fellow at University of Oxford  
University of Memphis

### **Summary**

The first half of this presentation provides an update on my recent work on computer agents in learning and assessment environments. This includes a version of AutoTutor that helps adults learn comprehension skills, electronics, and various subject matters. These systems have three-party conversations, called trialogues, where two agents (such as a tutor and a peer) interact with the adult. This work is compatible with the intelligent tutoring system architecture of the Generalized Intelligent Framework for Tutoring (GIFT, gifttutoring.org). The second half of the presentation provides observations, reflections, and questions about the use of Generative AI (large language models, ChatGPT) in the development of learning and assessment applications with conversational agents. I will comment on some current projects of other researchers that are exploring how Gen-AI can enhance the development, quality, and scope of conversation-based learning and assessment.

### **Short Bio**

Art Graesser is a professor in the Department of Psychology and the Institute of Intelligent Systems at the University of Memphis, as well as an Honorary Research Fellow at University of Oxford. He received his Ph.D. in psychology from the University of California at San Diego. His research interests question asking and answering, tutoring, text comprehension, inference generation, conversation, reading, problem solving, memory, emotions, artificial intelligence, computational linguistics, and human-computer interaction. He served as editor of the journal Discourse Processes (1996–2005) and Journal of Educational Psychology (2009-2014), as well as presidents of 4 societies, including Society for Text and Discourse (2007-2010), the International Society for Artificial Intelligence in Education (2007-2009), and the Federation of Associations in the Behavioral and Brain Sciences (2012-13). He and his colleagues have developed and tested software in learning, language, and discourse technologies, including those that hold a conversation in natural language and interact with multimedia (such as AutoTutor) and those that analyze text on multiple levels of language and discourse (Coh-Metrix and Question Understanding Aid -- QUAID).

He has served on five panels with the National Academy of Sciences and four OECD expert panels on problem solving, namely PIAAC 2011 Problem Solving in Technology Rich Environments, PISA 2012 Complex Problem Solving, PISA 2015 Collaborative Problem Solving (chair), and PIAAC Adaptive Problem Solving 2021

## **Keynote Speech 4**

- May 28, 2025 (Wednesday) 10:20-11:20 AM
- Location: 209 Tian Jia Ping Building

### **Title of keynote speech: Research Design and Innovative Teaching Models Integrating Generative AI with Educational Technology**



Professor Hui-Chun Chu

Distinguished Professor, The chairman of the Institute of Human-Centered Artificial Intelligence  
Soochow University

### **Summary**

The development of generative AI brings new possibilities to education, and its integration with emerging educational technologies serves as a crucial driver for teaching innovation. However, effectively designing research to explore the practical benefits of applying generative AI in educational settings and creating impactful teaching models remains a challenge. This speech will focus on the application of generative AI in educational technology, analyzing its impact on learners, educators, and educational policymakers. The content will include examples of innovative self-regulated learning models, specific methods for educational research design, and strategies for evaluating research or teaching performances. It aims to provide practical guidance for scholars in the fields of education and technology, inspiring further teaching innovation and applied research opportunities.

### **Short Bio**

Dr. Hui-Chun Chu's academic specialties include mobile and ubiquitous learning, game-based learning, flipped learning, technology-assisted healthcare education, AI in medical diagnosis and education, and knowledge engineering in education. Dr. Chu has published more than 167 academic papers, including 72 academic journal papers, in which 40 papers are published in well-recognized SSCI journals. Owing to the distinguished academic performance and service in e-learning, she received the Annual Young Scholars Outstanding Researcher Award--Ta-You Wu Memorial Award from the Ministry of Science and Technology in 2014. In addition, she received the Outstanding ICT Elite Award in 2020. She also served as the Associate Editor of IEEE Transactions on Learning Technologies (SSCI, Q1) since 2015. She has also served as the guest editor of Interactive Learning

Environments (SSCI, Q1) in 2013 and 2016. Moreover, Dr. Chu received the reward of “The top 50 Flipped Learning leaders in higher education worldwide” in 2018.

# Open Education and Global Competence: Advancing Glocalized Practices

Wen-Li Chang<sup>1</sup>, Jerry Chih-Yuan Sun<sup>2\*</sup>,

<sup>1,2</sup>Institute of Education, National Yang Ming Chiao Tung University

\* jerrysun@nycu.edu.tw

**Abstract:** In response to the educational demands of the 21st century, integrating open education into global competence development has evolved into a necessary consideration for educators and administrators. A comprehensive learning framework proposed by the Organization for Economic Cooperation and Development (OECD) in 2019 refines competence with focus on a cyclical process of anticipation, action, and reflection. By adopting glocalized strategies that address the Sustainable Development Goals (SDGs) through local perspectives, an undergraduate general education program utilizes open resources and massive open online courses as access to understanding global issues, meanwhile engaging students in self-initiated actions towards relevant local challenges. Under structured guidance in stages, students explore target issues on social networking sites, probe further with MOOCs, and share action plans using new media platforms. A mixed-methods approach combines quantitative data from students' summative self-assessment and qualitative insights from forum posts collected throughout one semester. Preliminary findings reflect the positive impact of open educational practices in cycles on growing globally competent learners who confidently anticipate and act with openness, though expanded reflective opportunities are necessary to ensure meaningful learning support and to drive continuous improvements in impactful local actions on global issues.

**Keywords:** 21st century competencies, sustainable development goals, cyclical learning process, open education, glocalized approach

## 1. Introduction

To enhance human well-being by 2030, the Organization for Economic Cooperation and Development (OECD) (2019a) proposed an educational framework prioritizing multi-faceted competencies—knowledge, skills, attitudes, and values—across personal, local, societal, and global aspects. The dynamic relationship between the varied aspects and the desired transformative competencies is characterized by a continuous cycle of improvement. In this cyclical progression of anticipation (A) and actions (A) constantly improved via reflection (R), the OECD learning framework aims to prepare students for adaptability in today's changing world. The transformative competencies in the OECD Learning Compass 2030 (2019b) emphasize sustainability for students themselves, for others, and for the planet to achieve. Students are encouraged to contribute to the real world by creating new value, reconciling tensions and dilemmas, and taking responsibility in action. While the OECD Learning Compass promotes student-centered and multi-faceted competency building for 2030, ethnic variations across and within nations remain integral to the changing environment and therefore deserve a culturally responsive approach. To counter the long-standing assimilationist notion of global citizenship, responsible citizens must refine their competencies via self-initiated actions for bringing justice on both national and global scales (Banks, 2004). This ideal combination of global and multicultural aspects merges inter- and intra-national ethnic diversity and enhances students' interrelated identifications across cultural, national, and global dimensions (Banks, 2015), most ideally realized through the anticipation-action-reflection (A-A-R) cycle of continuous improvement.

## 2. Theoretical Assumption and Educational Practice

### 2.1. Focused on the Glocalized Perspective of Core Competencies

The cyclical OECD framework for building global competencies adopts a *glocalized* approach, aimed at “blending and connecting local and global contexts while maintaining the significant contributions of the different cultural communities and contexts” (Patel & Lynch, 2013: 223). Within the highlighted higher education curriculum core was the researchers’ clear reference to Boyd’s (2006) and Khondker’s (2004) description of glocalization where both point toward the necessary blending approach, and the latter further emphasizes the negative mentality removed (i.e., “the fear of difference [erased] but not the differences”). In a broader sense aligning with a necessary shift towards the glocalization of learning (Niemczyk, 2019), higher education programs should prepare competent and culturally responsible 21<sup>st</sup> century citizens with not only knowledge and skills for consciousness and creation, but also attitudes and values, evolving from perceptive to appreciative. As shown in Figure 1, this glocalized attempt is clearly captured in the local framework for building core competencies through general education on the local tertiary level. By widening disciplinary coverage and merging intra- and inter-personal communication, a balanced approach is introduced to local and international perspectives, intended to connect local concerns with global thinking, meanwhile responding to the spirit of “connecting with local and global communities and creating a better future,” as stated by the Ministry of Education (Education in Taiwan, 2023/2024) in depicting a lifelong, ongoing learning path for all.

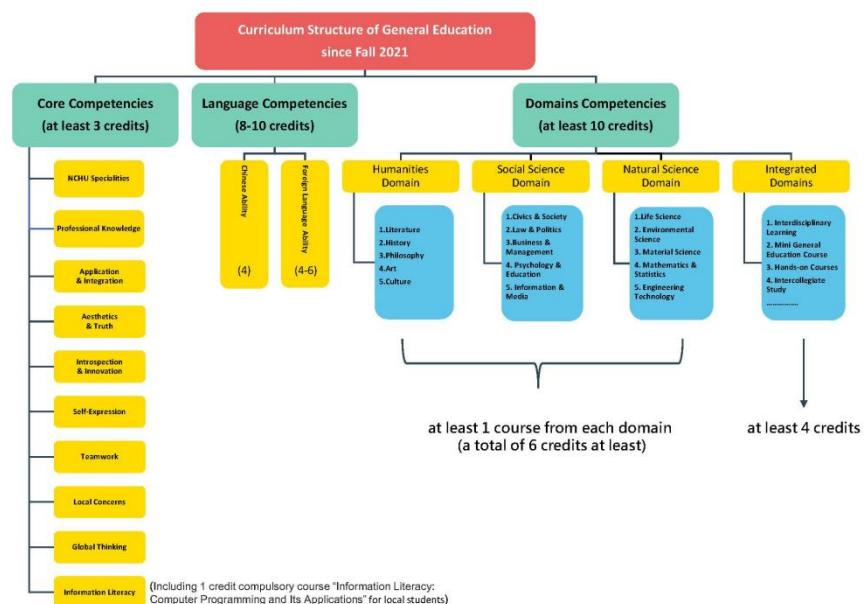


Fig.1 2021/2022 Taiwan framework for college-level general education (Curriculum Structure of General Education, 2021)

## 2.2. Integrating Glocalized Learning and Open Education for SDG Competence Development

The OECD Learning Compass strengthens a glocalized approach in general education while encouraging a personalized, cyclical A-A-R progression for achieving the Sustainable Development Goals (SDGs) proposed by the United Nations (UN) in 2015. The UN’s call for transformative action by 2030 supports a step-by-step educational process—understanding SDGs, defining priorities, establishing goals, integrating actions, and communicating outcomes—positioning global competence as a lifelong commitment to socially responsible values, up-to-date SDGs knowledge, and effective problem-solving skills. Central to this effort, as noted in prior research (e.g., Ossianilsson, 2024), is the rise of open education whose milestone in 2012 redefines educational practices with technology-enhanced open courses and inclusive learning opportunities. From word-spreading to online collaboration, and to digital content sharing, inclusivity in education has been significantly expanded with the widely varied practice options, including social networking platforms, massive open online courses or MOOCs, and new media channels (Euler & Poupart, 2018; Kramer & Bosman, 2018; Steiner, 2022), in contrast to traditional classroom practices. Open education’s close alignment with SDG 4 is recognized by the recent 2<sup>nd</sup> World Open Educational Resources Congress and is expected to help “ensure inclusive and equitable quality education and promote lifelong learning opportunities for all” (United

Nations, 2015), meanwhile laying the groundwork for testing and advancing planned actions toward development goals that include the culturally diverse and emerging SDG 18 (e.g., Visseren-Hamakers, 2020).

To evaluate the role of open education within the A-A-R cycle and its impact on global competence development, the study addresses the following research questions:

1. How does students' self-evaluation reflect the effects of cyclical open educational practices on their global competence?
2. How does class participation demonstrate students' engagement with the learning cycle in open education settings?

### **3. Course Description**

#### **3.1. Learning Objectives and Instructional Design**

The course focuses on SDGs and global sustainability, guiding students with the OECD Learning Compass (2019) through a cyclical A-A-R process for continuous improvement. On an open learning path toward whole-scale well-being, students develop global competencies by anticipating action plans toward world issues, acting on sustainable development goals, reflecting to evaluate action plans, contributing solutions to SDG-related challenges, and utilizing open learning tools in various positions (careful planner, topic explorer, open learner, idea contributor, critical thinker). Structured guidance supports students as they navigate diverse open learner positions and engage in individualized development of global competencies through stage-by-stage web-based practices, including investigation, focused exploration, and participatory actions. Students follow a progression of learning stages: (1) social networking for issues as conflict causes in target regions, (2) self-directed open learning for deep understanding, and (3) collaborative contributions as citizen journalists. By the end of the semester, students are expected to grow effective competencies in balancing different perspectives on global issues, in communicating ideas with diverse audiences, as well as in proposing appropriate ongoing actions to reinforce the SDGs framework.

#### **3.2. Class Schedule and Learner-Centered Activities**

To encourage the use of OECD Learning Compass, the course is structured into stages of work in alignment with each student's self-developed action plan that in general, engages personal anticipation and ongoing reflection. Table 1 presents an overview of the 18-week semester, highlighting three primary stages of work. The first four weeks are dedicated to building familiarity with the course, while the final two weeks focus on self-assessment, encouraging students to evaluate their overall learning progress and the effectiveness of their SDG action plans.

Table 1. Tentative timetable for class arrangements

Week		Content	Activity	Platform
1-2		Course overview		School's class/learning management system
3-4		Open learning and global sustainability	Pick the SDG you care	United Nations on SDGs (at un.org)
5-8		Social networking for relevant topics and issues	Explore with one tool and check for readiness: Reflection (1)	Social media: Facebook, Instagram, Twitter or else
9-12		Self-directed open learning for deep understanding	Proceed with one tool and check progress: Reflection (2)	MOOCs: Coursera, edX, FutureLearn or else
13-16		Collaborative contribution as a citizen journalist	Contribute with one tool and check success: Reflection (3)	New media: Blog, vlog, podcast or else
17-18		Self-assessment		School's class/learning management system

The flowchart in Figure 2 summarizes class activities in time order, underlined by a learner-centered approach, as well as the clear attempts to position students in multiple roles in their open learning process: careful planner, topic explorer, open learner, idea contributor, and critical thinker.

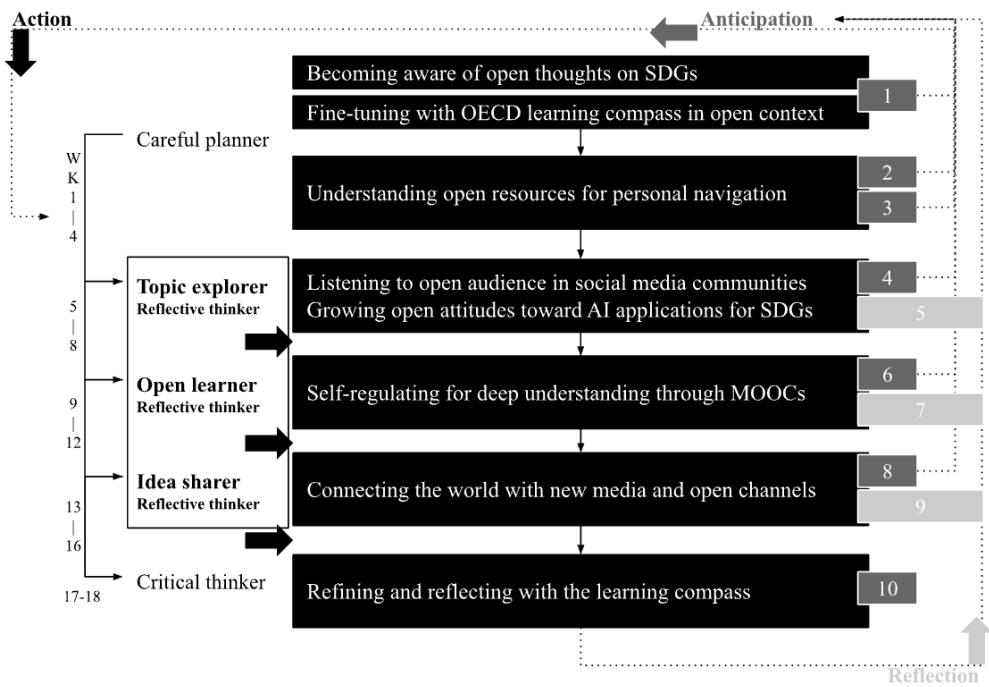


Fig.2 Learner position and activity design in anticipation-action-reflection cycle

Aggregated by the class portal (as shown in Figure 3), the learner-centered activities are designed to engage learners and continuously evaluate outcomes through various forms, purposes, and interaction patterns. These include Activity 1, a whole-class polling session for class warm-up; Activities 2 and 3, individual online surveys for personal reflection and action plan preparation, respectively; Activities 4, 6, and 8, whole-class open forum discussions centered on Open Educational Practices (OEPs 1-3); Activities 5, 7, and 9, individual online surveys for personal reflections (1-3); and Activity 10, an individual online self-assessment to conclude the course. Figure 4 displays a sample radar chart depicting students' self-report on their alignment with the core dimensions (anticipation, action, reflection, openness, and contribution) as they progress through the structured activities, develop action plans, assess their impact on SDG-related initiatives, and apply open learning tools in various roles.

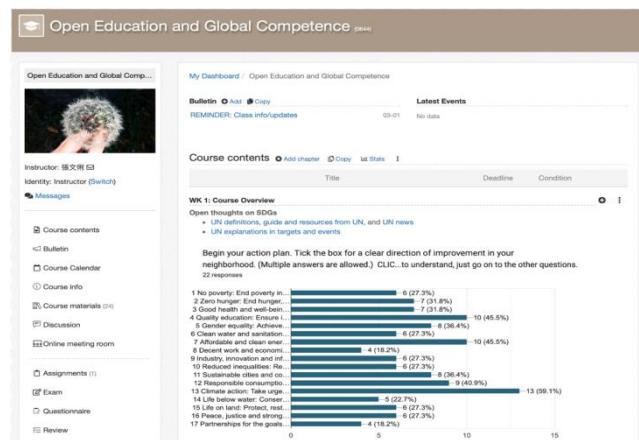


Fig.3 Screenshot of class/learning management system welcome page

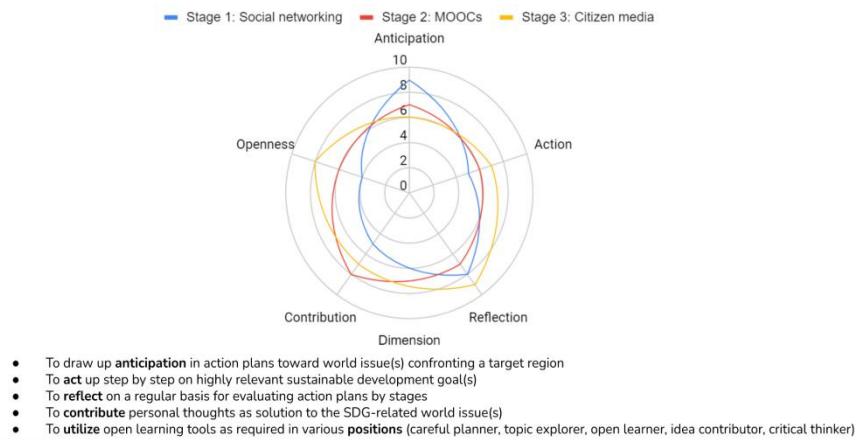


Fig.4 Self-assessment of global competencies and personal sustainability: Sample radar chart

### 3.3. Data Collection and Analysis

Within the school-wide general education framework, 47 valid responses were collected through the enrolled university students' self-assessment on a 10-point Likert Scale, based on pre-announced evaluation criteria derived from the OECD Learning Compass dimensions and course objectives (anticipation, action, reflection, contribution, and openness) (Figure 4). The scale enabled finer distinctions in student perceptions. Statistical analysis of the grading points included identifying overall trends and calculating Pearson correlations between the core dimensions. Additionally, the students' forum posts, which focused on perceived open elements in their action plans, were coded and analyzed within the reflective cycle framework for common themes and further compared with quantitative findings. This mixed-methods approach integrates both quantitative and qualitative data to provide a comprehensive understanding of the topic.

## 4. Student Perception and Feedback

With the core dimensions representing the cycle of localized open educational practices toward global issues, results of the students' self-assessment reveal a relatively high satisfaction in their general performance across all three stages. As shown in Figure 5, the students perceived Openness as the most satisfying dimension ( $M = 7.96$ ,  $SD = 1.00$ ), while Contribution, the least satisfying ( $M = 6.51$ ,  $SD = 1.16$ ). In line with prior research on open education as an accessible and inclusive ecosystem aligning with SDG 4 (e.g., Ossianilsson, 2024), the course's stage-by-stage preparation builds the students' familiarity with open learning tools and enhances their confidence in associating with various learning positions. Their comparatively weaker perception and recognition of Anticipation ( $M = 7.60$ ,  $SD = 0.88$ ), Reflection ( $M = 7.11$ ,  $SD = 1.07$ ), Action ( $M = 6.66$ ,  $SD = 1.26$ ), and Contribution (in a descending order) exhibit potential to grow with targeted practice in cycles, given that the students have demonstrated strong awareness of their ability to anticipate in action plans and to regularly reflect for improvement by stage. The higher standard deviations in Action and Contribution, compared to the low variability in Openness, suggest diverse experiences and perceptions in these areas that require tailored interventions such as structured reflection prompts, guided discussions, and iterative feedback mechanisms.

By further examining the correlations between the core dimensions, Action and Contribution ( $r = 0.48$ ) demonstrate a moderate positive relationship in Figure 6, while Anticipation and Openness ( $r = -0.17$ ) reveal a weak inverse relationship. The other dimension pairs generally show low correlations (e.g., Reflection and Action, Reflection and Contribution). The correlation differences reflect that the students' perceived openness and effectiveness of their action plans builds on their active engagement in taking actions to address issues. To ensure that student contributors continuously advance toward impactful actions, structured interventions for well-guided anticipation and meaningful reflection practice should be introduced in an ongoing learning cycle. This approach will contribute to the underlying

glocalized framework, as highlighted in prior research (e.g., Patel & Lynch, 2013), with strengthened integration into core dimensions and improved global perspectives on localized action.

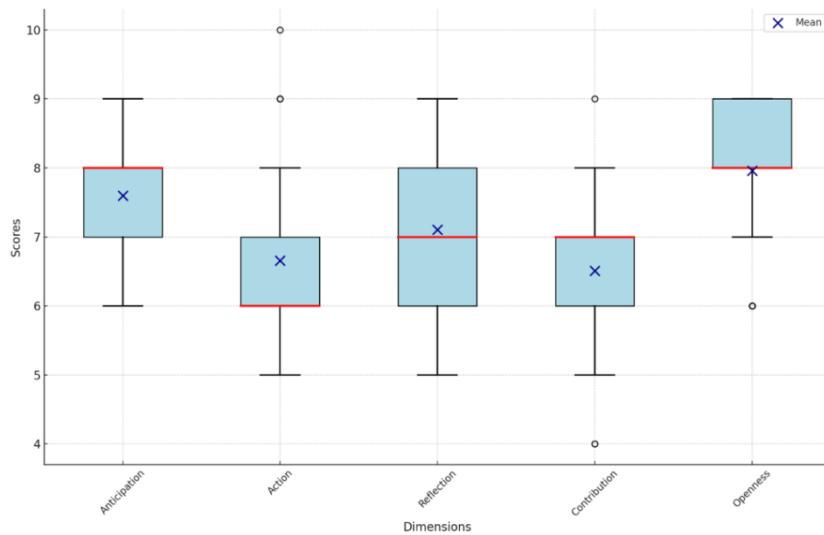


Fig.5 Average score by core dimension

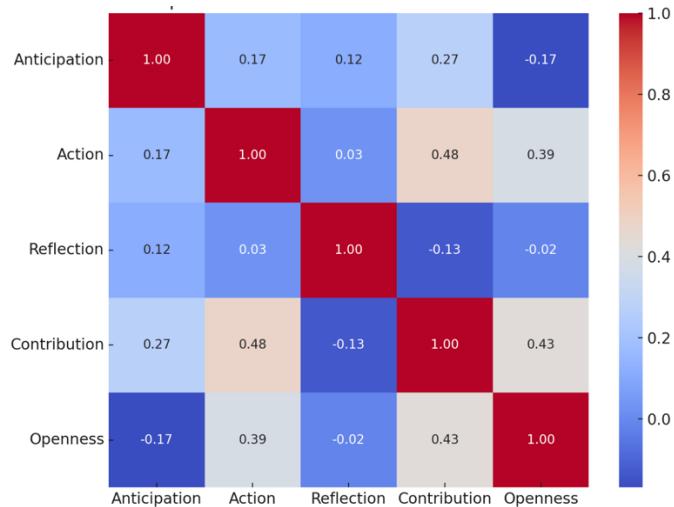


Fig.6 Heatmap of correlations between core dimension

The students' open forum posts were analyzed within the framework of a reflective cycle that comprises the core dimensions. In a close alignment with the learning cycle, Anticipation, Action, Contribution, and Openness – all these dimensions were well-represented in their posts; the students explicitly planned, executed, and openly shared their initiatives while promoting inclusive sustainability. By contrast, Reflection was not consistently performed and demonstrated a sheer focus on past experiences, as exemplified by one student's post, "About SDG 7, I am going to consume energy by taking bus and walking more. Furthermore, I can use products which have less carbon footprint." To ensure a more robust implementation of the A-A-R cycle and the OECD Learning Compass (2019), the emphasis on the Reflection stage is crucial and should be enhanced with not only structured interventions (e.g., prompts, questions) but also explicit reflection tasks (e.g., forums, journals, presentations), meanwhile integrating peer or community feedback for authentic glocalization.

## 5. Conclusion

Students' self-evaluations generally reflect positive impact of cyclical open educational practices on global competence, though implementing the core dimensions of an A-A-R process require additional support and structured

guidance. Students' active class engagement highlights their participation in anticipation and action during the process, while inconsistent reflection suggests the need to expand reflective practices in meaningful cyclical progression and continuous improvement. Social networking sites, MOOCs, and citizen media – these web-based channels have certainly brought the world closer to students, especially with the embedded open learning opportunities that connect global perspectives with local action. In higher education settings, the design and implementation of this course may challenge teachers whose perspective on global competencies can be biased or limited, therefore affecting their expertise in operating and incorporating open learning tools. Given that being globally competent encompasses the necessary knowledge, skills, attitudes, and values for effective engagement with the world, teachers who lack awareness or fail to recognize the benefits of global competencies and open learning may be reluctant to invest in or continue their development of the required expertise. Considering actions towards SDGs, possible resistance to change, insufficient professional development, and limited exposure to diverse perspectives - all these can further hinder their ability to utilize these tools effectively, and certainly their ability to provide meaningful guidance. To mitigate the challenges, teachers should seek ongoing professional development opportunities for digital pedagogy and reflective assessment, while also broadening their contextual understanding of global competencies for the purpose of enabling their incorporation of open learning practices that prepare students and themselves for a globalized society. Future studies should address self-reporting bias, limited generalizability, and cultural adaptability through advanced statistical analyses, along with triangulated assessments and cross-context research to enhance the credibility and applicability of findings.

### Acknowledgements

This research was supported by the National Science and Technology Council (formerly the Ministry of Science and Technology) in Taiwan through grant numbers NSTC 112-2410-H-A49-019-MY3 and MOST 111-2410-H-A49-018-MY4. We would like to thank National Yang Ming Chiao Tung University (NYCU)'s ILTM (Interactive Learning Technology and Motivation, see: <http://ILTM.lab.nycu.edu.tw>) lab members and the teachers for participating in the study, as well as the reviewers for providing valuable comments.

### References

- Banks, J. A. (2004). Teaching for social justice, diversity, and citizenship in a global world. In *The educational forum* (Vol. 68, No. 4, pp. 296-305). Taylor & Francis Group.
- Banks, J. A. (2015). *Cultural diversity and education: Foundations, curriculum, and teaching*. Routledge.
- Boyd, D. (2006, March). G/localization: When global information and local interaction collide. The O'Reilly Emerging Technology Conference. San Diego, CA.
- Curriculum Structure of General Education. (2021). National Chung Hsing University Office of Academic Affairs. <https://oaa.nchu.edu.tw/en-us/unit-page-p.237/page-detail.1808>
- Education in Taiwan. (2023/2024). Ministry of Education (MOE). Republic of China. [https://stats.moe.gov.tw/files/ebook/Education\\_in\\_Taiwan/2023-2024\\_Education\\_in\\_Taiwan.pdf](https://stats.moe.gov.tw/files/ebook/Education_in_Taiwan/2023-2024_Education_in_Taiwan.pdf)
- Euler, E., & Poupart, R. (2018). Spectrum of open education practices. Zenodo. <https://doi.org/10.5281/zenodo.1164799>
- Khondker, H. H. (2004). Glocalization as globalization: Evolution of a sociological concept. *Bangladesh E-journal of Sociology*, 1(2), 1-9.
- Kramer, B., & Bosman, J. (2018). Rainbow of open science practices. Zenodo. <https://doi.org/10.5281/zenodo.1147025>
- Niemczyk, E. K. (2019). Glocal education in practice: Teaching, researching, and citizenship. *Glocal education in practice: Teaching, researching, and citizenship*, 17, 11-18.
- OECD. (2019a). *Future of education and skills: OECD learning Compass 2030*. <http://www.oecd.org/education/2030-project/teaching-and-learning/learning/learning-compass-2030/>

- 
- OECD. (2019b). OECD future of education and skills 2030: Concept note.  
[https://www.oecd.org/education/2030-project/teaching-and-learning/learning/learning-compass-2030/OECD\\_Learning\\_Compass\\_2030\\_concept\\_note.pdf](https://www.oecd.org/education/2030-project/teaching-and-learning/learning/learning-compass-2030/OECD_Learning_Compass_2030_concept_note.pdf)
- Ossiannilsson, E. (2024). Open online learning and well-being: Ecosystem of open online, blended digital learning. *Socialinis ugdymas*, 61(1).
- Patel, F., & Lynch, H. (2013). Glocalization as an alternative to internationalization in higher education: Embedding positive glocal learning perspectives. *International Journal of Teaching and Learning in Higher Education*, 25(2), 223-230.
- Steiner, T. (2022). Open educational practices (OEP) – collection of scenarios. Wikimedia Commons.  
<https://doi.org/10.5281/zenodo.1183806>
- United Nations. (2015). 2030 Agenda for sustainable development. New York, NY: United Nations.
- Visseren-Hamakers, I. J. (2020). The 18th sustainable development goal. *Earth System Governance*, 3, 100047.

# Pre-service Teachers' Attitudes and Behavioral Intention Towards Generative Artificial Intelligence: A Structural Equation Modeling Investigation Based on TAM

Baoxin Guo<sup>1\*</sup>, Yue Feng<sup>1</sup>, Yifan Wang<sup>1</sup>, Sijie Zhang<sup>1</sup>, Xiaohong Liu<sup>2</sup>, Yingbin Zhang<sup>2</sup>

<sup>1</sup>School of Information Technology in Education, South China Normal University, Guangzhou

<sup>2</sup>Institute of Artificial Intelligence in Education, South China Normal University, GuangZhou

\*2958714843@qq.com

**Abstract:** This study uses the Technology Acceptance Model (TAM) to explore factors influencing pre-service teachers' intention to adopt generative artificial intelligence (GenAI). A total of 715 pre-service teachers participated in a questionnaire survey. Structural equation modeling was used to analyze the relationships among perceived ease of use (PEU), perceived usefulness (PU), attitude (ATT), and behavioral intention (BI). The results show that PEU positively affects PU, PU positively affects ATT, and ATT positively affects BI. Additionally, three significant mediating effects are identified. The findings provide valuable insights into the complex relationship between pre-service teachers' attitudes toward GenAI and their intention to adopt it.

**Keywords:** Pre-service teachers, generative artificial intelligence, technology acceptance model

## 1. Introduction

Generative Artificial Intelligence (GenAI), with its transformative capabilities, has emerged as a powerful tool in education, offering personalized learning, enhanced teaching resources, and innovative assessment methods (Jauhiainen & Guerra, 2023). For pre-service teachers, understanding and adopting such technologies is crucial to preparing for digitally driven classrooms. The Technology Acceptance Model (TAM) serves as a valuable framework for exploring the factors influencing technology adoption, such as perceived ease of use, perceived utility, and attitude (Davis, 1989). Prior study examined teachers' knowledge, attitudes, usage, and acceptance of GenAI, concluding that most teachers can actively adapt to and learn to use this technology (Zhai, 2024), but did not delve deeply into the underlying mechanisms and influence pathways. To address the gap, this study investigates the relationships among pre-service teachers' perceived ease of use, perceived utility, attitudes, and behavioral intention towards GenAI through a TAM-based structural equation modeling approach. The findings provide actionable insights to into enhancing pre-service teachers' AI literacy.

## 2. Literature review

### 2.1. Generative AI in Education

GenAI refers to a category of AI systems that leverage advancements in generative modeling and deep learning to create diverse forms of content using pre-existing media such as text, images, audio, and video (Alier et al., 2024; Fernández-Llorca et al., 2024). A prominent example is ChatGPT, an intelligent chatbot developed by OpenAI and built on large language models (Chiu, 2023; OpenAI, 2023). GenAI can respond effectively to questions and generate relatively accurate results quickly, leading to its growing popularity in education (Alier et al., 2024). For pre-service teachers, the adoption of GenAI necessitates not only technical proficiency but also a positive attitude toward its utility and ease of use (Wang et al., 2024). Understanding these factors is critical to ensuring that GenAI is effectively integrated into teacher education programs.

### 2.2. The Technology Acceptance Model

TAM provides a robust framework for understanding individuals' acceptance of new technologies (Davis, 1989). TAM posits that perceived ease of use (PEU) and perceived utility (PU) are the primary determinants of users' attitudes

toward technology, which in turn influence behavioral intention and actual usage. Over the years, TAM has been extensively applied in educational contexts to examine the adoption of learning management systems, mobile applications, and digital tools (Al-Adwan et al., 2023; Strzelecki, 2024). In the context of GenAI, TAM offers valuable insights into how pre-service teachers perceive and adopt this technology, particularly in understanding its potential to enhance teaching effectiveness and reduce workload. By leveraging TAM, this study aims to identify key factors influencing attitudes and behavioral intentions toward GenAI in teacher education.

### 2.3. Hypothesis Development

Based on the study by Scherer et al. (2019) on the TAM, this study proposes three hypothesis, displayed by Figure 1. Additionally, we also examined the indirect effects of PEU on ATT through PU and on BI through PU and ATT, as well as the indirect effect of PU on BI through ATT.

Hypothesis 1: Pre-service teachers' PEU will positively influence their PU.

Hypothesis 2: Pre-service teachers' PEU and PU will positively influence ATT toward GenAI.

Hypothesis 3: Pre-service teachers' PU and ATT will positively influence their BI toward GenAI.

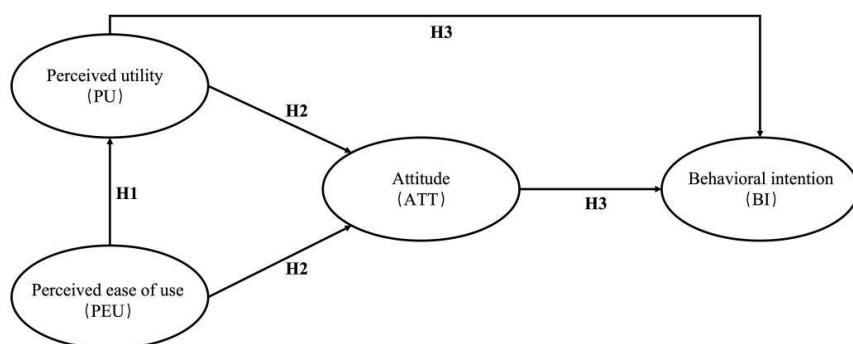


Fig.1 The hypothesized model of PU, PEU, ATT, and BI.

## 3. Methods

### 3.1. Participants and procedures

Between late September and early October 2023, this study surveyed pre-service teachers from six colleges and universities offering undergraduate teacher education programs in Guangdong Province, China. In this study, an electronic questionnaire was distributed to undergraduate teacher education students in these universities through social networks, and a total of 715 students voluntarily and anonymously participated in the survey. With "the sample must be teacher trainees" as the selection criterion, 616 valid samples were obtained. The samples cover a wide range of teacher education majors, including education, literature, history, political science, science, engineering, art, etc. The distribution of the samples is shown in Table 1.

Table 1. The gender, major, type of school, and year of study of the participants of the survey.

Variable	Values	n	%
Type of school	Double First-Class	432	70.10
	Other Universities	184	29.90
Gender	Male	103	16.70
	Female	513	83.30
Major	STEM	206	33.40
	Humanities and Social Sciences	410	66.60
Year of Study	Freshman	155	25.20
	Sophomore	175	28.40
	Junior	271	44.00
	Senior	15	2.40

### **3.2. Instruments**

The questionnaire has five parts: the first part is demographic information, the experience of use, and university implementation, the second part is a perceived utility scale, the third part is a perceived ease of use scale, the fourth part is an attitude scale, and the fifth part is a behavioral intention scale. The second, third, fourth, and fifth parts are adapted from the scales related to teachers' acceptance and behavioral intention toward technology (Joo et al., 2018; Scherer & Siddiq, 2015). Perceived utility scale (3 items), perceived ease of use scale (4 items), and attitude scale (6 items) items were scored on a five-point Likert scale ranging from 1 (very non-conformant) to 5 (very conformant). The Cronbach's alpha coefficients were examined to account for the internal consistency of the instruments. The Behavioral Intention Scale (2 items) was scored using a yes/no question with yes and no options and corresponding scores of 1 and 0. Pearson correlation coefficients between the two questions were examined.

In this study, the alpha coefficients of the Perceived Ease of Use Scale ( $\alpha = 0.84$ ), Perceived Utility Scale ( $\alpha = 0.87$ ), and Attitude Scale ( $\alpha = 0.95$ ) were all higher than 0.70, and the Pearson's correlation coefficient for two questions of the Behavioral Intention Scale was 0.71. And the standardized factor loadings of all items were higher than 0.6, which supported the structural validity of the questionnaire. To measure convergent validity, the extracted average variance (AVE) and composite reliability (CR) values were tested against a minimum criterion of 0.5 and 0.7, respectively (Hair et al., 2010), and the AVE and CR values of all four factors exceeded these thresholds, and thus convergent validity was accepted. In addition, discriminant validity was assessed by comparing the square root of the AVE of the four factors with the inter-factor correlation coefficients.

### **3.3. Data analysis**

The data analysis of this study consisted of three main stages. First, exploratory factor analysis (EFA) and validation factor analysis (CFA) were conducted to confirm the validity and reliability of the measurement scales. Second, descriptive statistical analyses, as well as analysis of variance (ANOVA), were conducted using SPSS to further understand the current status of GenAI knowledge, concepts, and behavioral intentions of the pre-service teacher population. For example, an independent samples t-test was used to examine the influence of factors such as gender, specialty, and experience of GenAI use, and one-way ANOVA was used to examine the influence of factors such as grade level, the status of GenAI-related learning activities carried out in the school, and the AI literacy-related requirements of the training program of their specialty. Finally, structural equation modeling was performed using the R language to examine the structural relationships among the variables in this study.

## **4. Results**

### **4.1. Descriptive results**

Perceived utility ( $M = 3.97$ ,  $SD = 0.69$ ) and attitude ( $M = 3.98$ ,  $SD = 0.69$ ) of pre-service teachers toward the educational application of GenAI were high, while perceived ease of use ( $M = 2.96$ ,  $SD = 0.80$ ) was low. In terms of behavioral intention, 77.30% of the pre-service teachers indicated that they would use GenAI for teacher education knowledge acquisition and 81.80% of the pre-service teachers indicated that they would use GenAI for teacher education skill enhancement.

The differences in PU, PEU, ATT, and BI of the pre-service teachers were analyzed in terms of gender, discipline, GenAI use, GenAI lectures offered by the school, and professional requirements. The results showed that there was a significant difference in pre-service teachers' PU in terms of GenAI use ( $p = 0.00$ ), school offering GenAI lectures ( $p = 0.00$ ), and professional requirements ( $p = 0.00$ ), a significant difference in pre-service teachers' PEU in terms of GenAI use ( $p = 0.00$ ), school offering GenAI lectures ( $p = 0.00$ ), and professional requirements ( $p = 0.00$ ), and a significant difference in pre-service teachers' ATT in terms of GenAI. Significant differences were found in pre-service teachers' ATT on GenAI use ( $p = 0.00$ ), professional requirements ( $p = 0.02$ ), and pre-service teachers' BI on GenAI use ( $p = 0.00$ ), professional requirements ( $p = 0.00$ ), and other than that, no other significant differences were found.

### **4.2. Structural model**

Structural equation modeling (SEM) was utilized to examine the structural relationships among the variables. First, the model fit was examined. As shown in Table 2, this model has a still good fit with  $\chi^2 = 1263.21$ ,  $df = 360$ ,  $\chi^2/df = 3.51 < 5$  (Kang & Ahn, 2021), root mean square error of approximation (RMSEA) = 0.06 < 0.08 (MacCallum et al., 1996), tucker-lewis index (TLI) = 0.92 > 0.9 (Hu & Bentler, 1999), comparative fit index (CFI) = 0.93 > 0.9 (Hu & Bentler, 1999).

Next, a path analysis was conducted. Figure 2 shows that pre-service teachers' PEU positively predicted PU; PU positively predicted ATT; PEU did not predict ATT; PU did not predict BI; and ATT positively predicted BI. Table 3 displays the results of the tests of the three hypothesis. Since PU had a positive effect on PEU ( $\beta = 0.36$ ,  $p < 0.001$ ), H1 was accepted. Since PU had a positive effect on ATT ( $\beta = 0.69$ ,  $p < 0.001$ ) and PEU did not predict ATT ( $\beta = -0.03$ ,  $p > 0.05$ ), H2 was partially accepted. Since ATT had a positive effect on BI ( $\beta = 0.35$ ,  $p < 0.001$ ), PU did not predict BI ( $\beta = 0.07$ ,  $p > 0.05$ ) and H3 was partially accepted. In total, the model explained 44.1% ATT variance and 19.5% BI variance.

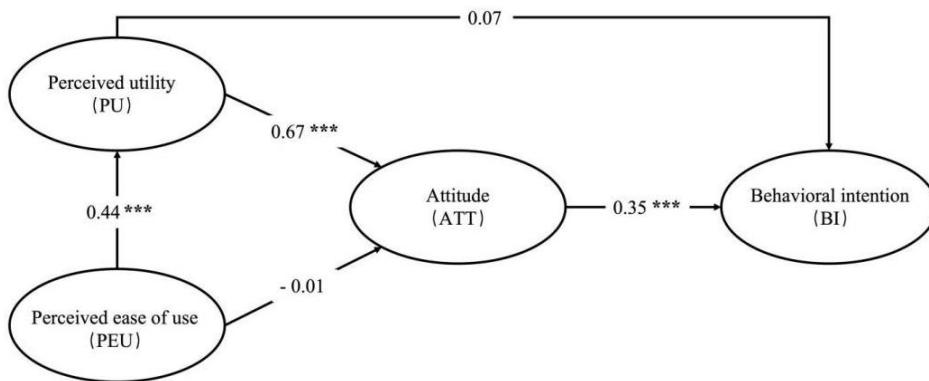


Fig.2 TAM of pre-service teachers' attitude towards GenAI.

Table 2. Model fitting analysis results.

Fitting index	$\chi^2$	$df$	$\chi^2/df$	RMSEA	TLI	CFI
Acceptable value	-	-	< 5	< 0.08	> 0.9	> 0.9
Results	491.55	144	3.41	0.06	0.94	0.95

Table 3. The results of hypothesis testing.

Hypothesis	Path	$\beta$	Supported?
H1	PEU → PU	0.44***	Yes
H2	PU → ATT	0.67***	Partially Supported
	PEU → ATT	-0.01	
H3	PU → BI	0.07	Partially Supported
	ATT → BI	0.35***	

Note. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

Table 4. The results of mediation effect test.

Path	Indirect effect	Lower bound	Upper bound
PEU → PU → ATT	0.29***	0.22	0.36
PU → ATT → BI	0.06***	0.04	0.09
PEU → ATT → BI	0.12***	0.08	0.16

Note. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

Finally, mediation effects were tested. Because the direct effects of PEU on ATT and PU on BI were not significant, the mediation effects involving these paths were not tested. Table 4 displays the mediation effect and 95% bootstrapped confidence intervals. PEU → PU → ATT indirect effect path is calculated as 0.29, and the lower limit and upper limit

are 0.22 and 0.36 respectively. The statistical significance of this indirect effect is very robust ( $P < 0.001$ ). The indirect effect path of PU → ATT → BI is calculated as 0.06, and the lower limit and upper limit are 0.04 and 0.09 respectively. The statistical significance of this indirect effect is very robust ( $P < 0.001$ ). PEU → ATT → BI indirect effect path is calculated as 0.12, the lower limit and upper limit are 0.08 and 0.16 respectively. The statistical significance of this indirect effect is very robust ( $P < 0.001$ ).

## 5. Discussion and Conclusion

This study applied the Technology Acceptance Model (TAM) to explore pre-service teachers' views on generative AI (GenAI) and its potential impact on their behavioral intentions to adopt it. The results reveal important insights regarding the factors that influence pre-service teachers' attitudes and behavioral intentions toward GenAI.

First, the study found that PEU of GenAI positively influenced its PU. This is consistent with the findings of Joo et al. (2018), suggesting that when pre-service teachers find GenAI easy to use and intuitive, they are more likely to perceive it as useful in educational contexts. The improved ease of use helps reduce technology-related anxiety, allowing pre-service teachers to focus more on the educational benefits of the technology (Wang et al., 2024).

Second, PU of GenAI positively influenced pre-service teachers' ATT toward it. This implies that when pre-service teachers recognize the practical advantages GenAI offers in enhancing their teaching practices, they develop a more favorable attitude toward using it.

Third, the study also highlighted that the PU did not have a direct effect on BI. Instead, it influenced BI indirectly through ATT. This suggests that while pre-service teachers may acknowledge the usefulness of GenAI, their actual intention to adopt it is primarily shaped by their attitudes. The complexity and newness of GenAI might hinder pre-service teachers from fully realizing its potential, which in turn affects their motivation to adopt the technology (Lan et al., 2024). This emphasizes the importance of shaping positive attitudes toward GenAI, as attitudes serve as a key mediator between perceived usefulness and behavioral intention.

Additionally, the mediating effects of PEU, PU, and ATT on BI were explored. The analysis revealed significant indirect effects for the pathways PEU → PU → ATT, PEU → ATT → BI, and PU → ATT → BI. Among these, the indirect effect of PEU → PU → ATT was the strongest, suggesting that improving the ease of use of GenAI can significantly enhance pre-service teachers' perceptions of its usefulness and, consequently, their attitudes toward using it.

Moreover, the indirect effect of PU → ATT → BI, though smaller, was still statistically significant. This indicates that positive attitudes play a crucial role in translating the perceived usefulness of GenAI into actual behavioral intentions. These results contradict the findings of Ramnarain et al. (2024), who concluded that pre-service teachers' attention to GenAI has no significant effect on their intention to use it. On the contrary, the conclusion of this study proves that pre-service teacher training should pay attention to the introduction and promotion of new technology, so as to cultivate the identification and positive attitude towards technology, after all, attitude is the key intermediary for the adoption of new educational technology.

In conclusion, the findings suggest that integrating GenAI into teacher education programs requires not only focusing on its perceived ease of use and usefulness but also actively fostering positive attitudes among pre-service teachers. Educational institutions and teacher training programs should aim to reduce technology anxiety, offer hands-on experiences, and provide real-world examples of successful GenAI applications in educational contexts (Blonder et al., 2024). By addressing these aspects, pre-service teachers' behavioral intentions to adopt GenAI can be enhanced, ultimately enriching the teaching and learning experience.

## 6. Limitation

Despite the study's contributions, limitations must be acknowledged. First, the sample was limited to pre-service teachers, potentially constraining the generalizability of the findings to in-service teachers. Future studies could include

diverse participant groups across different educational contexts to provide a more comprehensive perspective. Second, this study did not explore other potential factors, such as cognitive load or ethical concerns, that might influence attitudes and intentions toward GenAI. Future research could investigate these factors and their interplay with TAM constructs.

### Acknowledgements

This research was supported by Guangdong Philosophy and Social Science Foundation (GD24YJY06).

### References

- Al-Adwan, A. S., Li, N., Al-Adwan, A., Abbasi, G. A., Albelbis, N. A., & Habibi, A. (2023). "Extending the technology acceptance model (TAM) to predict university students' intentions to use metaverse-based learning platforms." *Education and Information Technologies*, 28(11), 15381–15413.  
<https://doi.org/10.1007/s10639-023-11816-3>
- 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).  
<https://doi.org/10.9781/ijimai.2024.02.011>
- Blonder, R., Feldman-Maggot, Y., & Rap, S. (2024). Are They Ready to Teach? Generative AI as a Means to Uncover Pre-Service Science Teachers' PCK and Enhance Their Preparation Program. *Journal of Science Education and Technology*. <https://doi.org/10.1007/s10956-024-10180-2>
- Chiu, T. K. F. (2023). The impact of generative AI (GenAI) on practices, policies and research direction in education: A case of ChatGPT and Midjourney. *Interactive Learning Environments*, 32(10), 6187–6203.  
<https://doi.org/10.1080/10494820.2023.2253861>
- 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>
- Fernández-Llorca, D., Gómez, E., Sánchez, I., & Mazzini, G. (2024). An interdisciplinary account of the terminological choices by EU policymakers ahead of the final agreement on the AI Act: AI system, general purpose AI system, foundation model, and generative AI. *Artificial Intelligence and Law*.  
<https://doi.org/10.1007/s10506-024-09412-y>
- Hair, J., Black, W., Babin, B., & Anderson, R. (2010). *Multivariate data analysis* (7th ed.). New Jersey: Prentice Hall.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55.
- Jauhiainen, J. S., & Guerra, A. G. (2023). Generative AI and ChatGPT in school children's education: Evidence from a school lesson. *Sustainability*, 15(18). <https://doi.org/10.3390/su151814025>
- Joo, Y. J., Park, S., & Lim, E. (2018). Factors influencing preservice teachers' intention to use technology: TPACK, teacher self-efficacy, and technology acceptance model. *Journal of Educational Technology & Society*, 21(3), 48–59.
- Kang, H., & Ahn, J. W. (2021). Model setting and interpretation of results in research using structural equation modeling: A checklist with guiding questions for reporting. *Asian Nursing Research*, 15(3), 157–162.
- Kline, R.B. (2023). *Principles and practice of structural equation modeling*. Guilford Publications.
- Lan, Y.-J., Chen, Y.-H., & Tu, Y.-F. (2024). Pre-service CFL teachers' conceptions of and attitudes toward ICT and image-GAI in Chinese teaching: A drawing perspective. *Educational Technology & Society*, 27(4), 431–453.  
[https://doi.org/10.30191/ETS.202410\\_27\(4\).TP01](https://doi.org/10.30191/ETS.202410_27(4).TP01)
- MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1(2), 130–149.

- 
- OpenAI. (2023). *ChatGPT: Optimizing Language Models for Dialogue*. <https://openai.com/index/chatgpt/>. Accessed January 10, 2025.
- Ramnarain, U., Ogegbo, A. A., Penn, M., Ojetunde, S., & Mdlalose, N. (2024). Pre-Service Science Teachers' Intention to use Generative Artificial Intelligence in Inquiry-Based Teaching. *Journal of Science Education and Technology*. <https://doi.org/10.1007/s10956-024-10159-z>
- Scherer, R., & Siddiq, F. (2015). Revisiting teachers' computer self-efficacy: A differentiated view on gender differences. *Computers in Human Behavior*, 53, 48–57.
- Scherer, R., Siddiq, F., & Tondeur, J. (2019). The technology acceptance model (TAM): A meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education. *Computers & Education*, 128, 13–35. Sciencedirect. <https://doi.org/10.1016/j.compedu.2018.09.009>
- Strzelecki, A. (2024). To use or not to use ChatGPT in higher education? A study of students' acceptance and use of technology. *Interactive Learning Environments*, 32(9), 5142–5155. <https://doi.org/10.1080/10494820.2023.2209881>
- Wang, K., Ruan, Q., Zhang, X., Fu, C., & Duan, B. (2024). Pre-service teachers' GenAI anxiety, technology self-efficacy, and TPACK: Their structural relations with behavioral intention to design GenAI-assisted teaching. *Behavioral Sciences*, 14(5), 373–373. <https://doi.org/10.3390/bs14050373>
- Zhai, X. (2024). Transforming teachers' roles and agencies in the era of generative AI: Perceptions, acceptance, knowledge, and practices. *Journal of Science Education and Technology*. <https://doi.org/10.1007/s10956-024-10174-0>

# Exploring the Effect of Scaffolding Strategies in GenAI Chatbot on Student Engagement and Programming Skill Development

Ean Teng Khor\* & Leta Chan

National Institute of Education, Nanyang Technological University, Singapore

\* eanteng.khor@nie.edu.sg

**Abstract:** Generative artificial intelligence (GenAI) technology has been integrated into various educational contexts since its introduction. However, research in this field often focuses more on GenAI's impact on students' learning outcomes rather than the instructional strategies employed to support learning. This study explores the effects of scaffolding strategies used by a GenAI educational chatbot on two aspects of learning: behavioral engagement with the chatbot and improvements in programming skills. 52 students interacted with different versions of a GenAI programming educational chatbot, varying in the extent of scaffolding provided. Their programming skills were assessed and compared before and after the intervention. A thematic analysis of the topics discussed in, and cognitive complexity of students' questions was also conducted. The study found that strategies which focused on enhancing conceptual understanding, as well as those that guided reflective practices, effectively fostered engagement, critical thinking, and programming skill development. These results underscore the need to align GenAI tools' functionality with students' needs to support meaningful learning. This study offers insights into the design of GenAI educational tools, opening paths for future research on training GenAI models to implement teaching pedagogies effectively.

**Keywords:** Generative Artificial Intelligence, Scaffolding Strategies, Behavioral Engagement, Programming Skill Development, Learning Technologies

## 1. Introduction

In the 21st century, GenAI has emerged as a hot topic in learning technologies, sparking widespread efforts to integrate it into modern educational practices. In programming education, GenAI shows considerable potential by generating code and explanations of programming concepts (Bahroun et al., 2023). A review of the literature on GenAI in education by Yusuf et al. (2024) identified three research focuses: the potential benefits and risks of GenAI integration, user perceptions and experiences, and the adoption of this technology by students and educational institutions. However, there has been less emphasis on the specific pedagogical strategies GenAI employs when tutoring students and the impact of these teaching techniques on learning outcomes. To address this gap, the present study examines the effects of scaffolding strategies employed by the GenAI programming education chatbot, *MyBotBuddy* (Khor et al., 2024a), on students' learning. Specifically, it examines their behavioral engagement during learning and their performance in programming tasks.

## 2. Literature Review

Modern programming teaching strategies often incorporate approaches such as direct instruction, collaborative learning, situated learning, and self-directed learning (Djenic & Mitic, 2017). GenAI promises to enhance these approaches by alleviating some of the challenges faced by students and teachers, such as comprehending complex content or preparing lessons. For example, GenAI can support direct instruction by generating lesson outlines or rubrics, creating demonstration code to illustrate programming concepts, or producing visual representations that simplify complex ideas (Cooper, 2023; Liu et al., 2024). Moreover, GenAI has proven effective in generating accessible explanations of programming concepts (Lee & Song, 2024), which can empower students to engage in self-directed

learning. These explanations scaffold students' understanding, equipping them to explore more advanced applications of their knowledge (Cooper, 2023). GenAI also facilitates collaborative learning by enhancing group activities such as pair programming: AI-generated reflection prompts can encourage students to share perspectives and reflect on their experiences after completing tasks (Naik et al., 2024). This process fosters cooperation and enables groups to tackle more complex programming challenges. Besides, GenAI has been employed to design programming practice exercises tailored to students' interests. Logacheva et al. (2024) found that enabling students to use GenAI to apply classroom content to real-world contexts of personal interest significantly boosted their motivation and engagement in programming learning.

However, the literature often lacks transparency regarding the teaching strategies GenAI tools use when guiding students. McGrath et al. (2024) found that most studies on GenAI chatbots in higher education made no reference to theories of educational practices. Just as teaching strategies are critical in human instruction, the strategies employed by educational GenAI tools merit equal attention. One such strategy that can be effectively implemented by GenAI tools is scaffolding. Scaffolding refers to the practice of providing guidance to help students complete tasks or understand concepts beyond their current level of expertise. Over time, the level of support is gradually reduced, empowering students to independently tackle more complex tasks or engage in higher-order thinking (Wood et al., 1976). Scaffolding has proven effective in motivating programming learners by helping them overcome initial challenges in learning a programming language, understanding problems, and devising solutions (Lin et al., 2021). Chen et al. (2024) illustrated the application of scaffolding in a GenAI coding assistant designed for elementary school students learning to code with Scratch. Recognizing the cognitive challenges younger learners face, the GenAI tool provided visual prompts to spark ideas, vivid images to represent project concepts, and a voice-guided assistant to explain coding steps, answer queries, and generate foundational code to help students progress when stuck. Students using the tool produced better code, retained more programming knowledge, and demonstrated higher engagement and motivation while learning. Similarly, Liao et al. (2024) developed a programming scaffolding system utilizing ChatGPT to enhance students' computational thinking. This system provided feedback on students' code, guided their problem-solving approaches, and addressed their questions, showcasing GenAI's potential to effectively implement scaffolding techniques to improve programming education.

Building on this foundation, the present study seeks to explore the impact of a GenAI educational chatbot on secondary school students' programming learning. Specifically, it addresses the following questions: (1) What is the impact of different scaffolding strategies used by a GenAI chatbot on students' programming ability? (2) What impact do different scaffolding strategies used by a GenAI chatbot have on students' questioning behavior when interacting with the chatbot? (3) Do students' interactions with a GenAI chatbot predict improvements in their programming abilities?

### **3. Research Design and Methods**

A total of 60 students from four secondary schools in Singapore were recruited for this study. However, data from 8 students were excluded due to technical issues, resulting in a final sample of 52 participants ( $n = 52$ ), of which six were female. The participants, aged 15 to 16 years, were students enrolled in the GCE 'O' Level Computing course. Ethics approval was obtained from the authors' institution and the Ministry of Education, along with consent from students, their parents, and their respective schools.

The study comprised three segments: a 30-minute pre-test, a one-hour intervention, and a 30-minute post-test. In the pre- and post-tests, students independently completed a Python programming task validating the check digits of ISBN-13 (pre-test) and ISBN-10 (post-test) numbers to assess their programming proficiency. During the intervention, students engaged with *MyBotBuddy* (Khor et al., 2024b), a chatbot developed based on GenAI model. Prompt engineering was leveraged in the design and development of *MyBotBuddy* which involves programming large language models through tailored prompts. The tailored prompts include breaking down a programming problem into smaller

problems and giving step-by-step instruction on completing the task. The training data was preprocessed and filtered to exclude damaging or biased language. The training data sources were diversified, and bias detection and mitigation approaches were included. *MyBotBuddy*'s capabilities were enriched using a knowledge base that included the students' Computing syllabus and an API to process and handle students' requests effectively.

The students were introduced to and briefed on using *MyBotBuddy*, then encouraged to freely explore programming topics or collaborate with the chatbot to refine their pre-test task code. To examine the effects of scaffolding strategies on students' learning, *MyBotBuddy*'s feedback was continuously refined by modifying the instructional prompts provided to the chatbot to enhance its responses. Throughout the study, it underwent iterative improvements to better scaffold students' learning, resulting in four distinct versions. Each version incorporated more detailed instructions on guiding and supporting students. Table 1 highlights the changes made to improve *MyBotBuddy*'s educational impact. Each school interacted with a different version of *MyBotBuddy*, in order of recruitment.

Table 1. Iterations of MyBotBuddy

Version	<i>n</i>	Improvements
1.0 (As of Nov 2023)	12	NIL
2.0 (As of May 2024)	21	The model was instructed to ask for address students by name when responding to any queries, creating a more personalized and friendly interaction. Besides, the scope of queries it could address was expanded to include not only programming-related topics but also other relevant areas of computing, further enhancing the students' learning experience.
3.0 (As of Jul 2024)	8	The model was instructed to engage students in more dialogue by asking guiding or follow-up questions, one at a time. These questions were designed to help students better understand the nature of the programming problem they were working on.
4.0 (As of Oct 2024)	11	The chatbot's tone was adjusted to be more supportive, with a clearer focus on helping students learn programming tasks and concepts. The model was given specific goals of fostering computational thinking, critical thinking, and reflection while guiding students through programming challenges. The model was also given a framework, complete with examples, on how to encourage students to think critically, reflect on their work, debug their code, promote self-regulation in their learning, and draw connections to real-world applications. Besides, it was instructed to gradually reduce its guidance and encourage students' independence.

The first version of *MyBotBuddy*, which served as the control, was provided only basic interaction instructions. It was designed to be a helpful and friendly AI assistant, limited to computing-related enquiries. The chatbot was instructed to guide students by breaking down problems and providing step-by-step assistance without directly offering answers. Subsequent iterations incorporated increasingly detailed instructions, with the final version employing a comprehensive framework focused on promoting critical thinking, reflection, and boosting students' programming knowledge instead of simply helping students complete increasingly complicated programming tasks (Hobert, 2019). The final version was deliberately designed to allow students to utilize GenAI's generative capabilities to practice analysis and evaluation. We leveraged the cognitive domain of Bloom's taxonomy (Anderson et al., 2001) to design guiding questions that scaffolded students' critical thinking abilities, as recommended by Lim and Makany (2023). In alignment with Bloom's taxonomy, the chatbot initially focused on lower cognitive complexity levels, such as remembering and understanding, to ensure students' foundational understanding. Questions were designed to confirm students' grasp of key concepts and familiarize them with the problem at hand (De Jesus et al., 2003). For example, *MyBotBuddy* prompted students to break down programming problems into essential elements and explain their logic.

Once foundational knowledge was established, *MyBotBuddy* progressed to more advanced levels, such as analysis and application. At these levels, the chatbot prompted students to explore relationships between concepts, consider factors influencing program outcomes, and propose solutions to potential errors. *MyBotBuddy* prompted students to reflect on their problem-solving strategies and consider broader applications of their solutions. Finally, the chatbot led students towards the synthesis and evaluation levels. It encouraged them to code independently and reflect on their solutions' quality. By gradually moving from lower to higher levels of cognitive complexity, *MyBotBuddy* fostered students' critical thinking and independent learning skills alongside strengthening their foundational programming knowledge. This structured approach allowed *MyBotBuddy* to evolve from a basic assistant to an effective scaffolding educational tool aligned with established pedagogical principles.

Students' pre- and post-test tasks were graded by subject matter experts using a standardized marking rubric. The rubric awarded up to 10 points for each task for accurate application of programming functions or concepts. To evaluate the overall effectiveness of *MyBotBuddy* in enhancing students' programming performance, a paired samples *t*-test was conducted to assess significant changes in participants' pre- and post-test scores. In addition, a Kruskal-Wallis H test was used to identify any significant score differences between different iterations of the chatbot. Students' interactions with *MyBotBuddy* were recorded and transcribed for thematic analysis. The number of queries was recorded, then the topics discussed were coded as educational or non-educational. Distractions, general inquiries, and non-academic questions were considered non-educational, while educational topics included the ISBN task and questions about programming in general or other academic subjects. The cognitive complexity of the interactions was also assessed to gauge students' depth of engagement, with questions classified as either confirmation or transformation (De Jesus et al., 2003). Confirmation questions, which were further coded as remembering or understanding questions based on Bloom's taxonomy (Anderson et al., 2001), intended to promote understanding of the topic. These included requests for explanations for a concept or code, solution generation, error identification and explanation, and clarification of students' understanding of code. Transformation questions involving higher cognitive processes like experimenting or reflecting, were grouped as application, analysis, evaluation, or creating questions according to Bloom's taxonomy (Anderson et al., 2001). Questions coded under these groups included attempts to apply learned content to other contexts, experimenting with the structure of a program, evaluating the efficiency or accuracy of a program or functions, and attempts to use their existing programming knowledge and *MyBotBuddy*'s generative capabilities to create new products. Asking more transformation questions implied more cognitively complex discussions with the chatbot, which indicated deeper engagement. The frequency of each question type, the total questions asked, and the total educational engagements with the chatbot were analyzed using the Kruskal-Wallis H test and quantile regression.

#### **4. Findings**

The paired samples *t*-test revealed that students improved significantly from the pre-test ( $M = 6.83$ ,  $SD = 3.15$ ) to post-test ( $M = 7.94$ ,  $SD = 2.49$ ) following the intervention with *MyBotBuddy*,  $t(51) = 2.61$ ,  $p < .001$ . GenAI educational chatbots may thus effectively enhance secondary school students' programming learning. Tests of normality and homogeneity of variance revealed that non-parametric tests were more appropriate for analysing relationships between chatbot versions, students' questioning behavior, and their score improvements. A Kruskal-Wallis H test comparing the effects of *MyBotBuddy*'s scaffolding techniques across the different iterations on students' score improvements found significant differences,  $\chi^2 (3, N = 52) = 8.96$ ,  $p = .030$ . While post-hoc comparisons using Dunn's method with a Bonferroni correction for multiple tests (adjusted  $p < .05$ ) showed no significant differences, unadjusted p-values revealed that students using the third ( $M = 2.75$ ,  $SD = 3.33$ ) and fourth ( $M = 0.91$ ,  $SD = 2.21$ ) versions of *MyBotBuddy* significantly outperformed those using the second version ( $p = .020$ ;  $p = .014$ ). The small sample size and conservative nature of the Bonferroni correction may have concealed potential differences. The scaffolding techniques employed in the third and fourth versions may thus have had a more pronounced effect on students' programming performance.

The Kruskal-Wallis H test was also employed to investigate differences in students' behavioral engagement with the chatbot across each version of *MyBotBuddy*. Across the four versions, no significant differences were detected in the mean number of questions asked,  $\chi^2 (3, N = 52) = 6.76, p = .080$ , the mean incidences of educational topics discussed with the chatbot,  $\chi^2 (3, N = 52) = 5.82, p = .121$ , the mean incidences of discussing the ISBN task with the chatbot,  $\chi^2 (3, N = 52) = 4.67, p = .197$ , and the mean incidences of asking remembering questions,  $\chi^2 (3, N = 52) = 6.57, p = .087$ . However, the analyses revealed a significant difference in frequency of discussion of non-educational topics across the four versions of *MyBotBuddy*,  $\chi^2 (3, N = 53) = 26.08, p < .001$ . The final group ( $M = 0.73, SD = 2.41$ ) discussed non-educational topics significantly less frequently than the first ( $M = 9.92, SD = 19.91, p = .016$ ) and second group ( $M = 2.72, SD = 5.74, p = .000$ ), while the third group ( $M = 5.5, SD = 14.38$ ) discussed these topics significantly less often than the second group ( $p = .015$ ). There was also a significant difference in frequency of discussion of general programming topics,  $\chi^2 (3, N = 52) = 11.97, p = .007$ , and other non-programming academic topics,  $\chi^2 (3, N = 52) = 24.89, p < .001$ , across the four versions of *MyBotBuddy*. Students interacting with the third ( $M = 4.5, SD = 5.32$ ) and final versions of *MyBotBuddy* ( $M = 8.18, SD = 6.68$ ) discussed general programming topics significantly more often than the second ( $M = 2.29, SD = 3.05, p = .019$ ) group. The final group ( $M = 0.27, SD = 0.47$ ) of students also discussed non-programming educational topics significantly less often than the first group of students ( $M = 1.58, SD = 3.42, p = .000$ ), while the second group ( $M = 0.05, SD = 0.21$ ) of students engaged in the same topic significantly less often than the first ( $p = .001$ ) and third ( $M = 2, SD = 3.55, p = .026$ ) group of students. A significant difference in the frequency of asking confirmation,  $\chi^2 (3, N = 52) = 10.70, p = .013$ , and transformation questions,  $\chi^2 (3, N = 52) = 12.35, p = .006$ , was also uncovered between the four versions of *MyBotBuddy*. Students using the final version ( $M = 3.63, SD = 2.06$ ) asked significantly more confirmation questions than those using the second version ( $M = 2.38, SD = 2.31, p = .044$ ), while students using the third version ( $M = 3.5, SD = 2.27$ ) asked significantly more transformation questions than those using the first ( $M = 9.67, SD = 8.22, p = .030$ ) and second version ( $M = 2.76, SD = 2.49, p = .004$ ). Lastly, a significant difference was detected between the four versions of *MyBotBuddy* in the mean number of understanding,  $\chi^2 (3, N = 52) = 12.53, p = .006$ , application,  $\chi^2 (3, N = 52) = 38.24, p < .001$ , analysis,  $\chi^2 (3, N = 52) = 15.19, p = .002$ , creating,  $\chi^2 (3, N = 52) = 10.88, p = .012$ , and evaluation questions asked,  $\chi^2 (3, N = 52) = 11.35, p = .010$ . Students using the final version ( $M = 0.73, SD = 0.79$ ) asked significantly more understanding questions ( $p = .003$ ) than those using the first version ( $M = 3.33, SD = 2.81$ ). Students using the second version of *MyBotBuddy* ( $M = 0, SD = 0$ ) also asked significantly fewer application questions than the first ( $M = 1.08, SD = 2.61, p = .000$ ) and fourth group of students ( $M = 2, SD = 1.34, p = .000$ ). The third ( $M = 0.625, SD = 0.74$ ) and final group of students ( $M = 0.27, SD = 0.65$ ) asked significantly more analysis questions than those using the first version ( $M = 1.5, SD = 1.73, p = .003; p = .049$ ). Students using the third version of *MyBotBuddy* ( $M = 1, SD = 1.60$ ) also asked significantly more creating questions than the second ( $M = 0.62, SD = 1.28, p = .048$ ) and first groups ( $M = 2.08, SD = 2.97, p = .038$ ). Lastly, students using the third version of *MyBotBuddy* ( $M = 0.625, SD = 1.26$ ) asked evaluation questions significantly more than the first ( $M = 5, SD = 3.74, p = .019$ ) and second groups of students ( $M = 1.71, SD = 1.42, p = .014$ ). Thus, the results indicate that the scaffolding strategies used in the third and fourth versions of *MyBotBuddy* are most successful in inducing more focused questioning behavior and a diverse range of cognitively complex questions.

The study also investigated how students' behavioral engagement with *MyBotBuddy*, measured by the total number of questions asked, the topics discussed, and the types of questions posed, influenced their post-test score improvement. To account for high heteroscedasticity and data outliers, quantile regression was employed to split students into three groups along the 25th, 50th, and 75th quantiles. This facilitated a more comprehensive analysis of the data's wide distribution, enabling us to examine differences in how students with varying levels of improvement interacted with the chatbot. At the 25th quantile, a significant negative relationship was observed between students' score improvement and their discussion of general programming topics ( $\beta = -.18, p = .026$ ), suggesting that students who discussed general programming topics more frequently experienced comparatively smaller score improvement. In contrast, at the 50th quantile, score improvement was significantly positively associated with asking remembering questions ( $\beta = .28, p$

= .019). Students who asked more remembering-related questions exhibited moderate improvements in their scores. Interestingly, a significant negative relationship was found between asking evaluation questions and students' score improvement at the 25<sup>th</sup> ( $\beta = -.60, p < .001$ ), 50<sup>th</sup>, ( $\beta = -.55, p < .001$ ), and 75<sup>th</sup> quantiles ( $\beta = -.83, p = .004$ ). This suggests that students who posed more evaluation questions generally showed smaller improvements compared to their peers at the same quantile. Thus, it is important to consider the specific type of engagement with *MyBotBuddy* to enhance learning outcomes.

## 5. Discussion

The study aimed to explore the effects of scaffolding techniques used by a GenAI educational chatbot, *MyBotBuddy*, on students' behavioral engagement, particularly questioning behavior, and their programming skills. The results indicated that using *MyBotBuddy* significantly improved students' scores, reinforcing the potential of GenAI chatbots to enhance students' critical thinking and programming abilities. Similarly, Yilmaz and Yilmaz (2023) observed gains in students' computational thinking, programming self-efficacy, and motivation after using ChatGPT, while Hobert (2023) demonstrated the effectiveness of a GenAI tutor that provided personalized feedback. This study offers insights into using GenAI in high school education, a less explored context, and examines student-chatbot engagement and interactions.

Comparisons across the four versions of *MyBotBuddy*, each incorporating more detailed scaffolding techniques, revealed that students who used the third and final versions benefitted most. These students were more focused and asked higher quality questions, demonstrating GenAI's effectiveness in enhancing learning through scaffolding. Students may thus benefit from responses that chunk information and include guided reflection questions. Structuring information into smaller sections reduces cognitive load to help students encode information more efficiently and retain it better (Thalman et al., 2019). This frees cognitive resources for critical evaluation and application of learned content. Later versions of *MyBotBuddy*, which emphasized supporting students' task understanding, also effectively promoted critical thinking, focus, and improved programming skills. This finding aligns with Lee and Song's (2024) finding that both students and teachers value explanations that enhance conceptual understanding, supporting this approach. Moreover, later versions of *MyBotBuddy* which included guiding and reflective questions, successfully encouraged sophisticated questioning behavior, as students asked many questions with balanced cognitive complexity. This finding contradicts concerns of complacency or overreliance due to GenAI use (Bailey, 2023), suggesting that GenAI chatbots can foster critical thinking and questioning. For teachers planning to integrate GenAI into classrooms, it may be beneficial to instruct the chatbot to break down responses into smaller parts and employ guiding and reflection questions. Students also need guidance to interact with GenAI tools appropriately, framing them as brainstorming companions instead of answer keys. Encouraging students-GenAI collaboration in classroom exercises can help instill this attitude and demonstrate its effective use.

Interestingly, the study found that engaging in more general programming discussions or asking evaluation questions was associated with smaller improvements. Selby (2015) explored an inverse relationship between the complexity of computational thinking skills and levels of Bloom's taxonomy where higher-level cognitive skills, such as evaluation, were mapped to lower-level computational thinking skills. Consequently, while students may demonstrate advanced cognitive skills, they may not possess the computational thinking skills required to break down the post-test task effectively. Scaffolding critical thinking alone may thus be insufficient to improve programming performance. GenAI programming tools could benefit from an approach that scaffolds computational thinking skills progressively, enabling students to become more well-rounded and capable of tackling complex programming problems.

## 6. Conclusion

Overall, the findings suggest that there is no universal solution for integrating GenAI chatbots into programming education. Simply adopting existing chatbot models may not be enough to facilitate students' learning. Instead, this study advocates a collaborative approach where educators work with GenAI to align its use with effective pedagogies that are most suited to the subject matter and students' needs. By adapting the chatbot to the learning context, educators can maximize its potential to enhance student outcomes. This study acknowledges certain limitations. The small sample size, as few schools offer GCE 'O' Level Computing, may have limited the statistical power of the analyses to detect significant group differences. Since students only interacted with *MyBotBuddy* once, many may have also approached the tool to test its capabilities, rather than allowing themselves to be guided by its prompts. Some students sought to evaluate the chatbot's response quality or how the chatbot functioned. Long-term interactions might more accurately portray how students engage with *MyBotBuddy* and its long-term effects on students' programming and critical thinking skills.

### Acknowledgements

This research is supported by the Ministry of Education, Singapore, under its Academic Research Fund (RG89/22).

### References

- Anderson, L. W., Krathwohl, D. R. D. R., & Bloom, B. S. (2001). A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. New York: Longman.
- Bahroun, Z., Anane, C., Ahmed, V., & Zacca, A. (2023). Transforming education: A comprehensive review of generative artificial intelligence in educational settings through bibliometric and content analysis. *Sustainability*, 15(17), 12983. <https://doi.org/10.3390/su151712983>
- Bailey, J. (2023). AI in Education: The leap into a new era of machine intelligence carries risks and challenges, but also plenty of promise. *Education Next*, 23(4), 28-35.
- Chen, L., Xiao, S., Chen, Y., Song, Y., Wu, R., & Sun, L. (2024). ChatScratch: An AI-augmented system toward autonomous visual programming learning for children aged 6-12. In F. F. Mueller, P. Kyburz, J. R. Williamson, C. Sas, M. L. Wilson, P. T. Dugas, & I. Shklovski (Eds.), *CHI '24: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (pp. 1-19). Association for Computing Machinery. <https://doi.org/10.1145/3613904.3642229>
- Cooper, G. (2023). Examining science education in ChatGPT: An exploratory study of generative artificial intelligence. *Journal of Science Education and Technology*, 32(3), 444–452. <https://doi.org/10.1007/s10956-023-10039-y>
- De Jesus, H. P., Teixeira-Dias, J. J. C., & Watts, M. (2003). Questions of chemistry. *International Journal of Science Education*, 25(8), 1015-1034. <https://doi.org/10.1080/09500690305022>
- Djenic, S., & Mitic, J. (2017). *Teaching strategies and methods in modern environments for learning of programming* [Paper presentation]. 14th International Conference on Cognition and Exploratory Learning in Digital Age (CELDA 2017), Algarve, Portugal. <https://eric.ed.gov/?id=ED579455>
- Hobert, S. (2019). Say hello to 'coding tutor'! design and evaluation of a chatbot-based learning system supporting students to learn to program. *Proceedings of the International Conference on Information Systems (ICIS)*. [https://aisel.aisnet.org/icis2019/learning\\_environ/learning\\_environ/9](https://aisel.aisnet.org/icis2019/learning_environ/learning_environ/9)
- Hobert, S. (2023). Fostering skills with chatbot-based digital tutors – training programming skills in a field study. *i-com*, 22(2), 143-159. <http://dx.doi.org/10.1515/icom-2022-0044>
- Khor, E. T., Chan, L., Koh, E., & Seow, P. (2024a). Student Perceptions of Using Generative AI Chatbot in Learning Programming. In *Proceedings of the 32nd International Conference on Computers in Education*. Asia-Pacific Society for Computers in Education.

- Khor, E. T., Koh, E., & Seow, P. (2024b). Leveraging generative artificial Intelligence to support students in learning programming: A chatbot pilot study. In *Proceedings of the 18th International Conference on Learning Sciences*. International Society of the Learning Sciences.
- Lee, S., & Song, K. S. (2024). Teachers' and students' perceptions of AI-generated concept explanations: Implications for integrating generative AI in computer science education. *Computers and Education: Artificial Intelligence*, 7, Article 100283. <https://doi.org/10.1016/j.caear.2024.100283>
- Liao, J., Zhong, L., Zhe, L., Xu, H., Liu, M., & Xie, T. (2024). Scaffolding computational thinking with ChatGPT. *IEEE Transactions on Learning Technologies*, 17, 1628-1642. <https://doi.org/10.1109/TLT.2024.3392896>
- Lim, S. S., & Makany, T. (2023). Deploying chatbots to build students' critical thinking skills: Leveraging generative AI effectively and purposefully in higher education. In M. A. Peters, & R. Heraud (Eds.), *Encyclopedia of Educational Innovation* (pp. 1-6). Springer. [http://dx.doi.org/10.1007/978-981-13-2262-4\\_299-1](http://dx.doi.org/10.1007/978-981-13-2262-4_299-1)
- Lin, S., Meng, N., Kafura, D., & Li, W. (2021). PDL: Scaffolding problem solving in programming courses. In C. Schulte, B. A. Becker, M. Divitini, & E. Barendsen (Eds.), *ITiCSE '21: In Proceedings of the 26th ACM Conference on Innovation and Technology in Computer Science Education*. Association for Computing Machinery. <https://doi.org/10.1145/3430665.3456360>
- Liu, Z., & Guo, R., & Jiao, X., & Gao, X., & Oh, H., & Xing, W. (2024), *How AI assisted k-12 computer science education: A systematic review*. [Paper presentation] 2024 ASEE Annual Conference & Exposition, Portland, Oregon. <https://peer.asee.org/47532>
- Logacheva, E., Hallas, A., Prather, J., Sarsa, S., & Leinonen, J. (2024). Evaluating contextually personalized programming exercises created with generative AI. In *Proceedings of the 2024 ACM Conference on International Computing Education Research*. Association for Computing Machinery. <https://doi.org/10.1145/3632620.3671103>
- McGrath, C., Farazouli, A., & Cerratto-Pargman, T. (2024). Generative AI chatbots in higher education: A review of an emerging research area. *Higher Education*, 1-17. <https://doi.org/10.1007/s10734-024-01288-w>
- Naik, A., Yin, J. R., Kamath, A., Ma, Q., Wu, S. T., Murray, C., Bogart, C., Sakr, M., & Rose, C. P. (2024). *Generating situated reflection triggers about alternative solution paths: A case study of generative AI for computer-supported collaborative learning*. In *Proceedings of the International Conference on Artificial Intelligence in Education*. Springer Nature Switzerland. [https://link.springer.com/chapter/10.1007/978-3-031-64302-6\\_4](https://link.springer.com/chapter/10.1007/978-3-031-64302-6_4)
- Selby, C. C. (2015). Relationships: Computational thinking, pedagogy of programming, and Bloom's Taxonomy. In *Proceedings of the Workshop in Primary and Secondary Computing Education* (pp. 80-87). Association for Computing Machinery. <https://doi.org/10.1145/2818314.2818315>
- Thalman, M., Souza, A. S., & Oberauer, K. (2019). How does chunking help working memory?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(1), 37-55.
- Wood, D. J., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89-100. <http://dx.doi.org/10.1111/j.1469-7610.1976.tb00381.x>
- Yilmaz, R., & Yilmaz, F. G. K. (2023). The effect of generative artificial intelligence (AI)-based tool use on students' computational thinking skills, programming self-efficacy and motivation. *Computers in Education: Artificial Intelligence*, 4, Article 100147. <https://doi.org/10.1016/j.caear.2023.100147>
- Yusuf, A., Pervin, N., Román-González, M., & Noor, N. M. (2024). Generative AI in education and research: A systematic mapping review. *Review of Education*, 12(2), Article e3489. <https://doi.org/10.1002/rev3.3489>

# Impact of Technology Enhanced Cooperative Learning based of Dialogic Feedback Strategy on Students' Learning Effectiveness

Ngoc-Diep Dao<sup>1</sup>, Shao-Chen Chang<sup>1\*</sup>

<sup>1</sup>Department of Information Communication, Yuan Ze University, Taiwan  
changsc@saturn.yzu.edu.tw

**Abstract:** As technology advances, essay writing is critical in higher education, but its incorporation into writing processes is undervalued. Students struggle with properly utilizing technology, which limits its potential for collaboration and better writing performance. This study investigates how dialogic peer feedback, paired with tech-enhanced collaborative writing, improves critical thinking, metacognition, and motivation. Based on Dialogic Feedback Theory, it uses a mixed-methods approach that includes pre/post-study questionnaires and collaboratively edited writings. Paired t-tests show that combining technology and peer feedback considerably improves these skills, which are closely correlated with improved writing skills.

**Keywords:** Tech-enhanced collaborative writing (TECW), Dialogic Feedback Theory, Critical Thinking, Learning Motivation, Metacognition.

## 1. Introduction

Academic writing is a complicated skills that requires critical thinking, metacognition, and motivation to achieve academic achievement (Stephen, 2022). Critical thinking is vital in writing since it involves analyzing and improving arguments rather than simply following writing strategies (Rahmat, 2020). Peer feedback improves this skill by allowing students to analyze and refine their ideas and assessments (Zou et al., 2023). Metacognition is crucial for enhancing argumentative writing because it improves self-assessment and problem-solving skills, allowing students to successfully control their learning (Rios, 2020). Peer feedback increases metacognitive awareness by prompting students to think about their writing and make informed revisions (Carless, 2013). It also improves motivation for academic writing (Sügülü et al., 2019). Technology-enabled peer feedback broadens these benefits by encouraging collaboration, improving critical thinking, metacognition, and motivation (Halpern, 2013). As technology advances, its role in academic writing becomes increasingly important. However, concerns about creativity and originality often hinder its adoption. While researchs into TECW have grown, the significance of dialogic feedback in tech-enhanced collaborative writing, particularly its impact on critical thinking, motivation, and metacognition, remains unexplored. This study address these shortcomings through three key questions: (1) Does integrating dialogic peer feedback with tech-enhanced collaborative writing improve students' critical thinking? (2) Does this method enhance metacognition in academic writing? (3) Does this method boost learning motivation?

## 2. Literature Review

Dialogic feedback includes dialogues that negotiate meaning, clarify expectations, and foster mutual understanding (Winstone et al., 2022). It is most effective when teachers and students build trust and share learning procedures and quality standards. Dialogic feedback, which is frequently utilized in the classroom, has been noticed to improve writing skills. Bouwer et al. (2024) showed that students who received dialogic feedback had greatly improved their semantic writing and revision skills. Dialogic peer feedback combines dialogue and peer feedback to promote two-way communication in which students actively construct meaning from input (Steen, 2017). It is especially useful in second language (L2) writing, enhancing writing skills and critical thinking (Hu, 2019). Gielen (2020) expands on this

perspective by offering a three-phase collaborative framework: planning, feedback and task involvement, which improves students' ability to process and use feedback through self-, shared-, and cooperative regulation.

Collaborative writing involves two or more students working together to create a single text (Storch, 2019). This process consists of planning, drafting, writing, and reviewing, with complete participation from all participants and no task division (Zhang & Zhang, 2022). Tech-enhanced collaborative writing has been extensively researched and identified as an effective way to second language (L2) learning (Su & Zou, 2022). Technology enhances the process's flexibility, engagement, and efficiency, resulting in better interaction and higher-quality writing than traditional techniques (Zou et al., 2022). Over the last decade, studies have verified TECW's effectiveness in L2 learning (Zou et al., 2022). It allows students to apply their knowledge, master their skills, share resources, exchange ideas, and engage in self-reflection and peer evaluation, all of which improve the quality of collaborative writing.

### 3. Research Methodology

The experimental process starts with an Introduction and Pre-questionnaire that assesses students' baseline skills and introduces them to dialogic peer feedback and TECW. During the Brainstorming ideas stage, students develop their ideas through peer discussions, embracing various perspectives. Then, in Creating a Detailed Outline stage, students structure their essays collectively while addressing logical flaws through systematic feedback. During the Writing Draft and Feedback stage, students engage in iterative peer review to improve their individual and group work, ensuring coherence and alignment with the essay's overall argument. Finally, in the Final Submission and Post-Questionnaire stage, students incorporate feedback, finalize their essays, and reflect on how dialogic collaboration influenced their writing and critical thinking. This organized procedure encourages active participation, peer-driven modifications, and in-depth analytical abilities, ultimately improving students' overall writing skills.



Fig.1 The experimental process

This study focuses on students who have completed their final reports for the Introduction to Database Systems course, which requires them to work in groups to choose one of five database-related themes. The research technique is divided into three stages: brainstorming ideas, developing a detailed outline, and composing the writing with feedback, all using an online collaborative writing platform with standardized templates. The study included 43 students (23 males and 20 females) ages 20 to 22, who were randomly assigned to seven groups. Based on the pre-questionnaires, all individuals volunteered and shared similar characteristics in critical thinking, metacognition, and learning motivation. Students used an online collaborative writing platform to track their progress and receive peer feedback from their groups. The groups used three templates: brainstorming idea, detail outline, and draft writing with feedback. The five criteria for feedback are accuracy, fluency, complexity, mechanics, and content, with participants responding based on these criteria.

### 4. Result

The study used a paired t-test to evaluate students' critical thinking skills in essay writing before and after using an integrated learning method that included dialogic peer feedback and tech-enhanced collaborative writing. The study found a modest pretest score ( $M = 4.03$ ,  $SD = 0.56$ ) and a considerable posttest improvement ( $M = 4.36$ ,  $SD = 0.49$ ). The t-test result ( $t = -3.22$ ,  $p < 0.001$ ) indicated a statistically significant increase in critical thinking skills.

Table 1. Critical thinking skills paired t-test

Variable	N	Mean	SD	t
Pretest	43	4.03	0.56	-3.22***
Posttest	43	4.36	0.49	

$P^{***}<0.001$

As a result, after using the integrated essay writing learning technique, students' self-assessed metacognition writing skills tend to improve. In the pretest phase, students self-assessed their skills with  $M = 3.89$ ,  $SD = 0.59$ , but in the posttest phase, these two values reached a little, with  $M = 4.25$  and  $SD = 0.50$ . The t-value of -3.29, with a mean difference of -0.36, indicates a significant difference between the pretest and posttest phases.

Table 2. Metacognition skills paired t-test

Variable	N	Mean	SD	t
Pretest	43	3.89	0.59	-3.29***
Posttest	43	4.25	0.50	

$P^{***}<0.001$

Based on the results, we found that students' learning motivation skills improved after the learning activity. According to Table 3, the mean of the pre-test was  $M = 3.98$ ,  $SD = 0.55$ ; the post-test scores were  $M = 4.34$  and  $SD = 0.43$  ( $t=-4.19$ ). It can be found that students can effectively improve their learning motivation skills after using dialogic peer feedback with the tech-enhanced collaborative writing method.

Table 3. Learning motivation skill paired t-test

Variable	N	Mean	SD	t
Pretest	43	3.98	0.55	-4.19***
Posttest	43	4.34	0.43	

$P^{***}<0.001$

To verify the effectiveness of this essay writing process, we compared draft templates through 4 weeks using an online collaborative platform. The results revealed a considerable improvement in both content quantity and quality. In weeks 1-2, essays averaged 300 words, mostly containing preliminary thoughts and outlines. By weeks 3-4, essays grew to 700-1200 words, featuring deeper analysis and greater consistency. Students demonstrated increasing confidence in expressing ideas as well as considerable progress in critical thinking skills, as they identified merits and shortcomings in their own and their classmates' essays.

## 5. Discussion and conclusion

The current study created an integrated learning model that blends tech-enhanced collaborative writing with dialogic peer feedback to help students improve their essay writing skills while simultaneously investigating the impacts on metacognition, learning motivation, and critical thinking. Students evaluated peer feedback using platform features such as "Change History" to establish its validity and utility, and the findings revealed a considerable improvement in critical thinking. Furthermore, the technique was helpful in improving students' metacognitive skills and learning motivation. The length and depth of essays increased through 4 weeks, indicating a significant improvement in their ability to organize thoughts, build arguments, and present evidence-based analysis. Furthermore, the feedback got more extensive and helpful, with students citing logical aspects and academic theories to explain and defend their points of view. Feedback no longer just pointed out flaws but also provided precise advice on how to improve the essay's structure and links between ideas, assisting peers in improving both substance and expression. Despite its benefits, there are still problems with applying this method, such as variances in student engagement, which are sometimes caused by

a lack of motivation or familiarity with peer evaluation. Furthermore, variances in study habits and knowledge backgrounds can result in variable feedback quality. The study is also limited by its small sample size, as it is the first experimental group. However, additional study with more diverse sample size will assist to improve the results' generalizability and dependability. Overall, the TECW with peer feedback has demonstrated to improve essay quality, critical thinking, metacognition, and motivation. Further research could look into the long-term effect and the incorporation of more tailored feedback methods for different learning demands.

## References

- Bouwer, C., & van Braak, J. (2024). Educational technologies and their impact on student outcomes in online learning environments. *Computers & Education*, 178, 104379.
- Carless, D. (2013). *Trust and its role in facilitating dialogic feedback*. In D. Boud & E. Molloy (Eds.), *Feedback in higher and professional education: Understanding it and doing it well* (pp. 100–114). Routledge.
- Gielen, M., Peeters, E., & Dochy, F. (2020). A collaborative learning approach to dialogic peer feedback: A theoretical framework. *Assessment & Evaluation in Higher Education*, 45(5), 657–668.
- Halpern, D. F. (2013). *Thought and knowledge: An introduction to critical thinking*. Psychology Press.
- Li, D., & Qian, J. (2024). Engaging in dialogic peer feedback in L2 writing development: A microgenetic approach. *System*, 126, 103513.
- Rahmat, N. H. (2020). Conflict resolution strategies in class discussions. *International Journal of Education*, 12(3), 49–66.
- Rios, J. A., Ling, G., Pugh, R., Becker, D., & Bacall, A. (2020). Identifying critical 21st-century skills for workplace success: A content analysis of job advertisements. *Educational Researcher*, 49(2), 80–89.
- Stephen, J. S. (2022). Persistence of nontraditional undergraduate online students: Towards a contemporary conceptual framework. *Journal of Adult and Continuing Education*, 29, 438–463.
- Steen-Utheim, A. T., & Wittek, L. (2017). How to improve the effectiveness of feedback in higher education: Understanding the students' perspective. *Studies in Higher Education*, 42(2), 254–267.
- Storch, N. (2019). *Collaborative writing*. *Language Teaching*, 52(2), 162–179.
- Sügümlü, Ü. (2019). The impact of peer feedback on writing skills development in higher education. *International Electronic Journal of Elementary Education*, 11(1), 37–45.
- Su, F., & Zou, D. (2022). Technology-enhanced collaborative language learning: Theoretical foundations, technologies, and implications. *Computer Assisted Language Learning*, 33(10), 1237–1263.
- Winstone, N. E., & Carless, D. (2020). Dialogic feedback and the importance of trust. *Studies in Higher Education*, 45(5), 830–841.
- Zhang, Y., & Zhang, J. (2022). Exploring the role of feedback in fostering EFL writing competence. *Journal of English for Academic Purposes*, 44, 100847.
- Zou, D., Xie, H., & Wang, F. L. (2022). Effects of technology enhanced peer, teacher and self-feedback on students' collaborative writing, critical thinking tendency and engagement in learning. *Journal of Computing in Higher Education*, 35(1), 166–185.

# Exploring the Effects of VR-assisted Learning on Students' Learning Engagement in a University EFL Classroom

Jiarong Chen<sup>1</sup>, Xiaoshuang Zhang<sup>1</sup>, Bowen Jing<sup>1</sup>, Lin Luan<sup>1\*</sup>

<sup>1</sup>School of Humanities, Beijing University of Posts and Telecommunications, Beijing, China

\*luanlin@bupt.edu.cn

**Abstract:** Learning engagement is vital to the learning outcomes of English as a foreign language (EFL) learners. Virtual reality (VR) can create an immersive and interactive environment for users. Nevertheless, evidence is scarce regarding whether VR can boost EFL learners' learning engagement. This study, thus, aims to explore the effects of VR technology on EFL learners' learning engagement. A quasi-experiment was designed for this study with the self-developed Situational English in Virtual Reality platform. 64 EFL learners were randomly divided into experimental and comparison groups. The experimental group with 31 students received English learning with the assistance of VR on the designed platform, while the comparison group with 33 students watched instructional videos on PC. Data was collected using an adapted VR-assisted learning engagement questionnaire. The quantitative results demonstrated that VR is beneficial in enhancing EFL learners' learning engagement, especially cognitive, emotional, and social engagement. This study offers empirical evidence of the positive impact of VR technology on EFL learners' learning engagement. Pedagogically, VR-assisted language learning could be incorporated in future language learning class as an effective instructional approach to increase students' learning engagement. This study provided some suggestions for both language instructors and technical experts in terms of task design, platform optimization, and material development.

**Keywords:** virtual reality (VR), EFL learners, learning engagement

## 1. Introduction

Active learning is a key concern to second language learning (Dörnyei & Kormos, 2000). Many researchers try to find out how to improve the learning engagement of English as a foreign language (EFL) learner and thus promote educational progress in language learning. With the development of technology, virtual reality (VR) has been introduced into the educational sector. Researchers have explored "the distinguishing characteristics of VR that provide unique opportunities for educational use" (Petersen et al., 2022). However, the former research does not suffice to check VR's effects on EFL learners' learning engagement. To shed more light on this issue, the paper is set to investigate the influence VR-based learning has on EFL learners from four aspects, that is, cognitive engagement, behavioral engagement, emotional engagement and social engagement.

## 2. Literature Review

### 2.1. Learning Engagement

An engaged EFL learner is actively involved in and committed to their own learning, and without engagement meaningful learning is unlikely (Hiver et al., 2024). Researchers have delved into many dimensions about EFL learning engagement, which can be summarized as cognitive, behavioral, emotional and social dimensions. Kuhlmann et al (2024) explored the importance of active cognitive engagement in STEM learning. They discovered that how active students' cognitive engagement was would influence students' learning outcomes. Doo and Kim (2024) found a small-to-medium effect of learning engagement to learning outcomes and that there was no statistical difference among behavioral, cognitive, emotional, and general engagement in affecting learning outcomes. The study of Ratan et al.

(2022) suggested that students' subjective course gains were significantly related to social engagement. In this study, behavioral engagement refers to the time and energy EFL learners spend on study; cognitive engagement indicates their mental efforts; emotional engagement equals the positive and negative emotions; and social engagement comprises the EFL learner's social interaction with other people.

## 2.2. VR and Learning Engagement

VR has been proven to be an efficient tool for education which can provide immersive and natural face-to-face interaction (Naylor, 2023). Dubovi (2022) explored students' emotional engagement and cognitive engagement with a VR-based simulation. It indicates that the virtual environment is able to promote students' positive emotions like happiness and enhance their cognitive engagement. Another study of Dubovi (2024), though believing that there could be "excessive" affections, provided further evidence for the positive effects of VR on learning engagement, which found a complex emotional mechanism behind VR learning. However, researches on the effects of VR on learning engagement are still not adequate. Therefore, the paper aims to investigate whether VR has a positive impact on EFL learner's learning engagement.

## 3. Method

### 3.1. Research Context and Participants

The study was conducted in Beijing University of Posts and Telecommunications from two intact English classes with 64 sophomores (53% males, aged 18-20). All of the participants are native Chinese speakers who have at least learned English for 8 years and have passed National College English Test Band 4 (CET-4). Two classes were randomly assigned to either the VR-assisted experimental group ( $N=31$ ) or the computer-assisted comparison group ( $N=33$ ). Both experimental group and comparison group learned the same content, while the former learning with the VR platform and the latter watching the instructional videos playing on the personal computer screens.

### 3.2. Instruments

#### 3.2.1. The VR learning platform

The self-developed *Situational English in Virtual Reality* platform was employed in the study (Figure 1). The VR platform allows users to have an immersive check-in experience in an international airport, helping them to enhance their language learning performance.

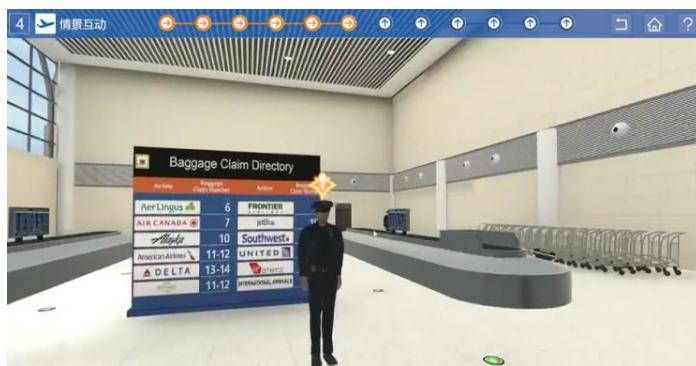


Fig.1 The self-developed Situational English in Virtual Reality platform

#### 3.2.2. VR-assisted learning engagement questionnaire

The VR-assisted learning engagement questionnaire was revised on the questionnaire of Luan et al. (2020). It adopted a 5-point Likert scale format, comprising a total of 16 questions, with 4 questions for each of the following aspects: behavioral (e.g. I stay focused in the virtual environment), emotional (e.g. I enjoy learning new things about English), cognitive (e.g. I try to understand my mistakes when I get something wrong), and social engagement (e.g. I like working with virtual characters).

### 3.3. Research Procedure

The experiment lasted for six weeks with two 45-minute classes each week. In the first week, all participants were required to complete the VR-assisted learning engagement pre-questionnaire. In the second week, the VR group was trained to use the VR platform, while the comparison group received basic instructions about PC operation. From the third week to the fifth week, the VR players took part in English learning on the VR platform, whereas the video watchers learned English by watching the instructional videos on a PC. In the final week, the VR-assisted learning engagement post-questionnaire was administered to all participants.

### **3.4. Data Collection and Analysis**

This study employed the VR-assisted learning engagement questionnaires to collect quantitative data. A total of 64 students' responses to questionnaires were collected and measured to analyze participants' learning engagement when the two groups were respectively learning in the virtual environment and the traditional language classroom. ANCOVA was performed to compare the effects of two different learning approaches on learners' engagement.

## **4. Results**

ANCOVA was conducted to identify between-group differences. The pretest was the covariate, the post-test the dependent variable, and the groups the fixed factor. Regarding learning engagement, the non-significant interaction of the independent variable and the covariate was verified ( $F = 2.64, p = 0.11 > 0.05$ ), which suggests that the assumption of homogeneity of the regression slope was fulfilled. Levene's test of the homogeneity of variance was also met ( $F = 0.95, p = 0.34 > 0.05$ ), indicating that the null hypothesis was tenable and the variance was equal across groups.

Afterwards, ANCOVA was employed for overall learning engagement, which revealed significant differences ( $F = 5.83, p = 0.02 < 0.05$ ). The adjusted means and standard error of the experiment group were 67.08 and 0.94 and were higher than those of the comparison group of 63.93 and 0.91. Therefore, the experiment group had a better performance than the comparison group for overall learning engagement. ANCOVA was also employed to explore the difference between the scores of the pre-test and the post-test of the two groups. Results showed that the cognitive engagement ( $F = 4.60, p = 0.04 < 0.05$ ), emotional engagement ( $F = 4.87, p = 0.03 < 0.05$ ) and social engagement ( $F = 5.85, p = 0.02 < 0.05$ ) of the experimental group was better than those of the comparison group, while existing no significant difference in behavioral engagement ( $F = 0.80, p = 0.38 > 0.05$ ).

## **5. Discussion and Conclusion**

The study investigated the effects of VR-assisted learning on EFL learners' learning engagement. Analysis of the VR-assisted learning engagement questionnaires indicated that, compared to traditional classrooms, VR could promote EFL learners' overall engagement, especially cognitive, emotional, and social engagement, while the impact on EFL learners' behavioral engagement is relatively small. The result echoed that of Dubovi (2022) who found that VR was beneficial for students' cognitive engagement and emotional engagement and Dubovi's (2024) finding of the positive effects of VR on learning engagement.

The study has significant pedagogical implications for instructors and educators. First, since VR can enhance the learning engagement of EFL learners, instructors and educators are encouraged to further incorporate VR into classrooms. Second, as the study witnessed no significant improvement of behavioral engagement, it is recommended to design various learning scenarios, interaction elements as well as rewarding mechanisms in order to further enhance students' behavioral engagement. It should be noted that several limitations need to be acknowledged for future studies. Researchers could utilize qualitative data (e.g., interviews) to better understand student behaviors in VR environments, and quantitative data (e.g., speaking tests) to evaluate learning outcomes. Moreover, the relatively brief duration of the experiment compromises its representativeness when factors like the novelty of VR could influence the result of the study. Future research needs to extend the experimental period for more in-depth results. Also some technical issues such as feeling dizzy when operating the platform still need optimization.

## **Funding**

This paper was part of the project of “Research on Innovation and Application of Foreign Language Teaching Mode Empowered by Virtual Simulation Experiments in Colleges and Universities in the Context of New Liberal Arts” supported by 2024 Higher Education Research Project of China Association of Higher Education (Grant No. 24WY0303).

## References

- Doo, M. Y., & Kim, J. (2024). The relationship between learning engagement and learning outcomes in online learning in higher education: A meta-analysis study. *Distance Education*, 45(1), 60–82.
- Dörnyei, Z., & Kormos, J. (2000). The role of individual and social variables in oral task performance. *Language Teaching Research*, 4, 275–300.
- Dubovi, I. (2022). Cognitive and emotional engagement while learning with VR: The perspective of multimodal methodology. *Computers & Education*, 183, 104495.
- Dubovi, I. (2023). Facial expressions capturing emotional engagement while learning with desktop VR: the impact of emotional regulation and personality traits. *Interactive Learning Environments*, 32(9), 5041–5057.
- Hiver, P., Al-Hoorie, A. H., vitta, j. p., & Wu, J. (2024). Engagement in language learning: A systematic review of 20 years of research methods and definitions. *Language Teaching Research*, 201-230.
- Kuhlmann, S. L., Plumley, R., Evans, Z., Bernacki, M. L., Greene, J. A., Hogan, K. A., Berro, M., Gates, K., & Panter, A. (2024). Students' active cognitive engagement with instructional videos predicts STEM learning. *Computers & Education*, 216, 105050.
- Luan, L., Hong, J.-C., Cao, M., Dong, Y., & Hou, X. (2020). Exploring the role of online EFL learners' perceived social support in their learning engagement: a structural equation model. *Interactive Learning Environments*, 1-12.
- Naylor, W., Debattista, K., & Chalmers, A. (2023). Perception-based high quality distributed virtual reality. *Virtual Reality*, 27 (3), 2529-2539.
- Petersen, G. B., Petkakis, G., & Makransky, G. (2022). A study of how immersion and interactivity drive VR learning. *Computers & Education*, 179, 104429.
- Ratan, R., Ucha, C., Lei, Y., Lim, C., Triwibowo, W., Yelon, S., Sheahan, A., Lamb, B., Deni, B., & Chen, V. H. H. (2022). How do social presence and active learning in synchronous and asynchronous online classes relate to students' perceived course gains? *Computers & Education*, 191.

# Developing an AI-empowered Chinese Composition Assessment System for Hong Kong

## Students

Hiu Laam Naomi Lee<sup>1</sup>, Yik Elinor Wong<sup>2</sup>, Wai Yin Koey Chung<sup>1</sup>, Chi Fuk Henry So<sup>\*3</sup>

<sup>1,3</sup>Centre for Learning, Teaching and Technology, <sup>2</sup>Department of Chinese Language Studies, <sup>3</sup>Department of Mathematics and Information Technology

The Education University of Hong Kong

<sup>\*</sup> hcfso@edu.hk

**Abstract:** With the advancements in natural language processing and the growing adoption of Large Language Models (LLMs) in education, AI-based automated essay evaluation has emerged as a significant topic in research on writing literacy assessment. Despite this, there is limited research on Chinese automated essay scoring systems, leaving the subject largely underexplored. This paper introduces an AI-empowered Chinese composition assessment system specifically designed for primary and secondary schools in Hong Kong. The system features customizable prompts based on marking criteria, employs multiple AI models to enhance grading accuracy, and enables rapid report generation. We will outline the system's design, workflow, and future directions for field testing and eventual implementation.

**Keywords:** Automated Essay Evaluation, AI in education, Chinese composition assessment, educational technology

## 1. Introduction

Writing literacy is an essential skill that enhances one's ability to express ideas and communicate effectively. It is vital for students' lifelong development and capacity to adapt to societal changes in the future (Genlott & Grönlund, 2013). With advancements in natural language processing (NLP) technology, AI-based automated essay evaluation has become a prominent area of research in assessing writing literacy. The repetitive task of evaluating students' writing requires a significant investment of time and effort from teachers. Consequently, automated essay evaluation significantly alleviates the grading workload for educators. This project aims to develop an AI-empowered Chinese composition assessment system for Hong Kong primary and secondary school teachers and students. The system evaluates writing proficiency through automated analysis of compositions written by native Chinese students and provides targeted feedback for improvement. The system features several key elements:

1. Customized Experience: Each participating school will have a unique experience with no waiting time. Teachers can adjust the system's requirements and customize mark allocations to fit their teaching schedules and marking schemes. The system accommodates any topics and prerequisites, automatically generating prompts based on the specified criteria for immediate use by the AI models.

2. Enhanced Grading Accuracy: The system employs multiple AI models to assess assignments, which improves grading accuracy. Two AI models will collaborate to evaluate each assignment, while a third model will act as a separate verifier if the first two models produce significantly different scores.

3. User-Friendly Interface and Rapid Feedback: A dedicated interface for teachers will be established as the system develops. Teachers can easily upload compositions and receive comprehensive AI-generated reports within minutes, all within a unified application framework.

This paper reviews the need for an AI-empowered composition assessment system in Hong Kong. It provides an overview of the design and development of the system, highlighting the integration of pre-built AI models, particularly

focusing on the prompt framework and workflow associated with the system. Lastly, the paper concludes by discussing the upcoming field testing and the anticipated limitations of the system.

## **2. The use of AI and NLP in educational assessments**

The study of automated essay assessment began in 1966 with Page's pioneering research on the Project Essay Grader system (Page, 1966). This innovative system utilized a multiple regression program to replicate the evaluations of human markers. It predicted essay scores by analyzing factors such as the frequency of uncommon words, the use of prepositions and commas, and the overall length of the essay. The program then compared these scores with those of randomly selected essays, and surprisingly, it was found to be indistinguishable from the assessments made by English teachers. Since then, automated essay scoring (AES) has remained an important area of research (Ke & Ng, 2019).

The advent of generative artificial intelligence models that can comprehend and produce human language marks a significant advancement in the application of AI for educational assessments. Prior to the emergence of pre-built AI models such as ChatGPT and Gemini, AES systems were primarily developed using traditional machine learning techniques. Since the 1990s, these systems have focused on identifying patterns within various features derived from extensive datasets (Ramesh & Sanampudi, 2022).

Past achievements in AES include notable systems such as the Intelligent Essay Assessor (IEA) by Foltz et al. (1999), e-rater V2 by Attali and Burstein (2006), and IntelliMetric by Rudner et al. (2006). These systems are characterized as handcrafted feature AES, whose effectiveness primarily depends on the quality of the features designed by experts. In contrast, more recent systems developed over the last decade employ neural network models that automatically learn features for essay scoring, as highlighted by Hussein et al. (2019). A significant advancement in this field was introduced by Alikaniotis et al. (2016), who presented a deep neural network (DNN) model capable of autonomously extracting essay scoring features without relying on predefined templates. Recently, DNN-based models have garnered considerable attention. Uto (2021) has compiled a list of new models from 2016 to 2021 that underscores this emerging trend.

The introduction of ChatGPT in 2022, followed by the releases of Gemini and Claude in 2023, has dramatically enhanced the accessibility of AI language models. According to Mizumoto and Eguchi (2023), these AI language models hold significant potential as tools for AES. Furthermore, Latif and Zhai (2024) highlight the importance of conducting additional research on the capabilities of various generative AI models in the context of automatic scoring in education.

## **3. The current needs of the system in Hong Kong**

Over the past few decades, numerous AES platforms have been developed, primarily aimed at assessing the writing skills of second language (L2) learners. Most of these platforms are tailored for evaluating English writing, with significantly fewer systems available for other languages, especially Chinese. Consequently, the potential of LLMs in the domain of Chinese AES remains largely untapped. In Hong Kong, only a limited number of studies have examined automated essay evaluation using datasets derived from local students. Notably, these datasets are exclusively composed of English writings (Chan et al., 2023; Lee et al., 2009). We resonate with other researchers in the field and acknowledge the potential of leveraging prebuilt LLMs for automated essay evaluation. Our concept for the AI-empowered Chinese Composition Assessment System is groundbreaking in its application of prebuilt LLM platforms for automated grading. Moving forward, we will explore the feasibility of customizing assessment criteria to generate prompts on the user side, allowing the system to provide essay scores and reports within a minute.

## **4. Development of the AI-empowered Chinese Composition Assessment System**

The AI-empowered Chinese Composition Assessment System is designed to assess writing proficiency through automated analysis of student compositions, delivering personalized developmental advice using pre-trained AI models.

This system is supported by a prebuilt LLM studio from the Centre for Learning, Teaching, and Technology (LTTC) at The Education University of Hong Kong (EduHK). It offers a Model-as-a-Service (MaaS) platform that enables developers to tailor intelligent workflows for educational applications.

The system is designed to incorporate an intuitive application interface that facilitates user-friendly navigation. Users can configure assessment criteria as needed and upload their compositions on a page-by-page basis. The development of the system and application interface is currently underway. Figure 1 shows a detailed illustration of the design and structure of the AI-empowered Chinese Composition Assessment System, highlighting its innovative features and capabilities.

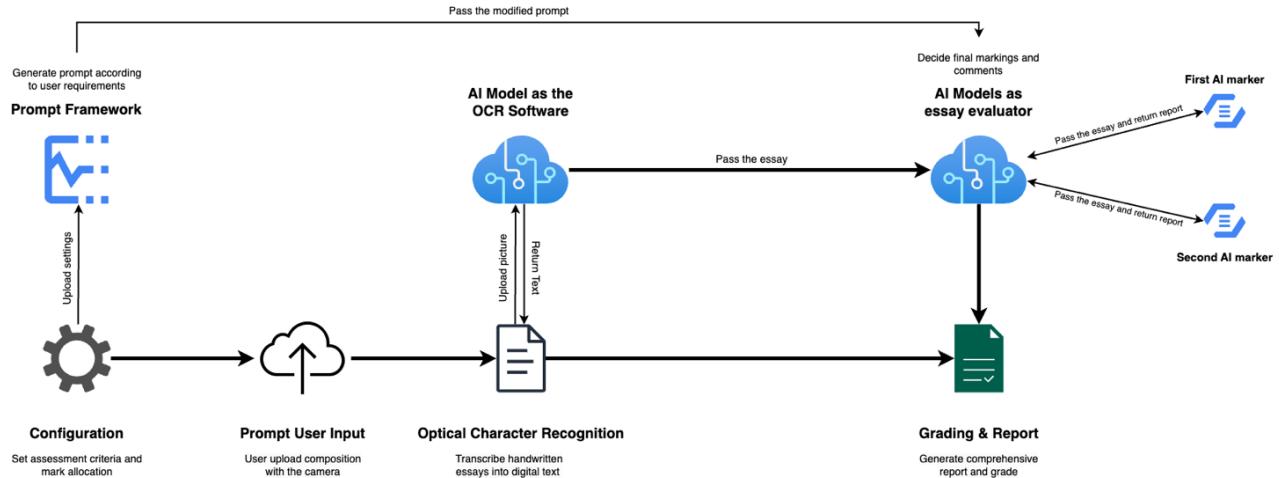


Fig.1 The structure of the AI-empowered Chinese Composition Assessment System

類別	框架	只可以自訂黃色格內容	
#任務	#角色	你是一名香港中學中文科老師。 我會提供有關課堂的資料，作業的細節以及評分準則。	
	#指令	你需要根據所提供的指令認真思考並作出評分。繁體字和簡體字相通，並不當作有錯別字。如果一個詞中部分是簡體字，部分是繁體字，則簡體部分算錯別字。	
#規則	#技能	###技能1: 評價作文 ###技能2: 提高作文能力 ###技能3: 反饋及評價 ###技能4: 批注及修改 ###技能5: 符號, 字體, 字數	根據上傳的學生作文，總結並理解作文內容並做出整體評價。 根據寫作能力學習重點提出建議。舉例說明如何寫作。 提供正面反饋並讚賞學生優點。 根據評分的結果，對學生作文中細節做批註。 評估標點符號和錯別字並給出建議。
	#回覆格式	1: 使用段落格式評價學生論文。 2: 使用評分標準以表格形式做出評估，在每項評分後根據評分重點說明評分的依據。 3: 引用學生作文部分論文原文，並對此部分原文做出批註和改善建議。 4: 使用繁體中文進行回覆。	
#內容	科目 作文題目 體裁 類型 年級 字數限制 總分	中文 投訴無牌小販擺賣 應用文 投訴信 中二 沒有 34	
#總分	內容的總分為 格式的總分為 標點字體的總分為 錯別字每個扣分 ...的總分為 ...的總分為	16 15 3 0.5 ... ...	最多扣 3 分
#評分指引	###內容: 格式部分總分15分：啟首語寫上「啟啟者」，得3分；結束語在正文後的新一行靠左空兩格寫上「此致」，得3分；在結束語「此致」的下一行靠左頂格寫上題目要求的受文者，得3分；在受文者下一行靠右寫上「投訴人」及投訴人的姓名，加上啟告語「啟」或「謹啟」，得3分；最後一行靠左頂格寫上日期，包括年、月、日，得3分。 ###格式: ... ###標點字體: 按能否適切使用標點符號、適當地分段、字體秀麗整齊程度綜合評分，最高3分。如有寫字，最少給1分。 ...	內容部分總分16分：表明寫信原因，得2分；交代最少2項投訴原因，每項投訴原因，根據描述的清晰程度、合理程度評分，每項滿分為4分，共計8分；提出要求收件人交代或解釋或處理投訴狀況，得4分；提供自己聯絡方法供處理部門聯絡自己，得2分。	

Fig.2 Prompt Framework

The prompt generation framework for the system will be grounded in expert consultations and assessment rubrics established by the Hong Kong Education Bureau. Teachers can input their specific requirements, including but not limited to the subject, composition topic, text type, genre, grade level, word limit, and total marks. Furthermore,

educators can customize the marking criteria based on various aspects, such as content, format, punctuation, spelling errors, or any elements they wish to prioritize. Figure 2 depicts the concept of prompt generation within the backend of the system, with the customizable elements highlighted in yellow. The requirements entered by users will be processed to generate prompts that will be supplied to the AI model. Figure 3 showcases examples of these prompts.

<b>指令 (第一部分)</b>	<pre> #角色 你是一名香港中學中文科老師。我會提供有關課堂的資料，作業的細節以及評分準則。你需要根據所提供的指令認真思考並作出評分。 繁體字和簡體字相通，並不當作有錯別字。如果一個詞中部分是簡體字，部分是繁體字，則簡體部分算錯別字。  #技能 ###技能1: 評價作文 根據上傳的學生作文，總結並理解作文內容並做出整體評價。 ###技能2: 提高作文能力 根據寫作能力學習重點提出建議。舉例說明如何寫作。 ###技能3: 反饋及評價 提供正面反饋並讚賞學生優點。 ###技能4: 批注及修改 根據評分的結果，對學生作文中細節做批注。 ###技能5: 符號，字體，字數 評估標點符號和錯別字並給出建議。  #回覆格式 1: 使用段落格式評價學生論文。 2: 使用評分標準以表格形式做出評估，在每項評分後根據評分重點說明評分的依據。 3: 引用學生作文部分論文原文，並對此部分原文做出批注和改善建議。 4: 使用繁體中文進行回覆。 </pre>
<b>指令 (第二部分)</b>	<pre> #內容 科目為中文。 作文題目為投訴無牌小販擺賣。 體裁為應用文。 類型為投訴信。 年級為中二。 字數限制為沒有。 總分為34。  #總分 內容的總分為16。 格式的總分為15。 標點字體的總分為3。 錯別字每個扣分0.5。 最多扣3分。  #評分指引 ###內容: 內容部分總分16分：表明寫信原因，得2分；交代最少2項投訴原因，每項投訴原因，根據描述的清晰程度、合理程度評分，每項滿分為4分，共計8分；提出要求收件人交代或解釋或處理投訴狀況，得4分；提供自己聯絡方法供處理部門聯絡自己，得2分。  ###格式: 格式部分總分15分：啟首語寫上「敬啟者」，得3分；結束語在正文後的新一行靠左空兩格寫上「此致」，得3分；在結束語「此致」的下一行靠左頂格寫上題目要求的受文者，得3分；在受文者下一行靠右寫上「投訴人」及投訴人的姓名，加上啟告語「啟」或「謹啟」，得3分；最後一行靠左頂格寫上日期，包括年、月、日，得3分。  ###標點字體: 接能否適切使用標點符號、適當地分段、字體秀麗整齊程度綜合評分，最高3分。如有寫字，最少給1分。 </pre>

Fig.3 Prompts generated for the AI model

The AI-empowered system is designed to transcribe handwritten essays into digital text and deliver a comprehensive report that includes grades and comments within one minute. This process involves various tasks, such as Optical Character Recognition (OCR) of handwritten work (refer to Figure 4), marking, and grading (as shown in Figure 5), all facilitated by the collaboration of multiple AI models like Gemini and ChatGPT. We use pre-trained models for their efficiency, given that the number of essays we can gather in Hong Kong is considerably smaller than the extensive labelled and unlabelled data utilized by large-scale pre-trained AI models. Han et al. (2021) underscore the advantage of using pre-trained AI models as a foundation rather than developing a new model from scratch.

\* (一) 作文原文 :

--- OCR Start ---

啟啟者：

本人為太子道花園大廈住客，近日大廈門口出現許多無牌小販擺賣，情況日趨嚴重，本人深感困擾，特此致函投訴。

首先，無牌小販霸佔大廈門口行人通道，阻礙居民出入。在上下班繁忙時間，市民出入受阻，需繞道而行，十分不便。

其次，他們整天大聲叫賣，十分嘈吵。有些學生需要溫習，他們的叫喊聲令學生無法專心，亦妨礙大廈居民休息。

此外，自從無牌小販出現後，街道變得十分骯髒。人們棄置食物殘渣和垃圾，吸引和滋生蚊蟲、老鼠等，令衛生環境惡化，嚴重影響社區整潔。

大廈附近的無牌小販日益猖獗，嚴重影響居民生活，本人促請大廈物業管理處迅速處理此事，給居民一個交代，還我們一個乾淨整潔的居住環境。如有任何問題，請致電 5364798 。

此致

太子道花園大廈物業管理處負責人 住客

2024 年 5 月 20 日

投訴人趙思懿敬上

--- OCR End ---

Fig.4 A random sample of OCR by the system

### 三、作文三評分表

評分項目	最高分	實際得分	評分依據
內容	16	15	清晰表明寫信原因 (2/2); 列舉三項投訴原因, 描述清晰且具體, 每項給予 5 分, 共計 15 分 (15/12)。
格式	15	12	啟首語 (3/3); 結束語及署名「投訴人」正確, 但「台鹽」位置錯誤, 扣 3 分 (6/9); 日期格式正確 (3/3)。
標點字體	3	3	標點符號使用正確, 段落劃分合理, 字體工整。
錯別字	-3	-1	電話號碼缺失數字, 扣 1 分。
總分	34	29	

### 內容分析及修改建議 (三篇作文共通)

三篇作文都清楚地表達了投訴的理由。然而, 可以透過加入更生動的例子、更精確的數據, 以及更強而有力的詞語來提升文章的說服力。例如, 可以將「十分吵雜」改為「叫賣聲此起彼伏, 音量巨大, 嚴重影響居民睡眠」, 或加入具體數據, 例如「小販佔據了約五米長的通道」。此外, 可以更具體地說明噪音和衛生問題對居民的影響, 例如, 影響睡眠質量, 引致精神不佳, 影響工作效率, 甚或影響身心健康。

### 格式分析及修改建議 (三篇作文共通)

大部分同學都能掌握投訴信的基本格式, 但部分細節需要改進, 例如「啟啟」或「台鹽」的位置應在收件人姓名之下, 而非並列。此外, 投訴人的姓名應寫完整, 避免造成聯絡上的不便。

### 標點及錯別字

同學們應仔細檢查標點符號的使用, 並避免出現錯別字, 確保文字表達準確無誤。

### 總結

三篇作文各有優劣, 同學們已初步掌握投訴信的寫作技巧, 但仍需進一步空間, 希望同學們在日後的寫作中, 多加練習, 注意細節, 力求做到內容充實, 邏輯清晰, 格式正確、文字精煉, 寫出更具說服力的投訴信。特別是作文三, 其內容描述較為詳盡, 同學應繼續保持此良好的寫作習慣。

Fig.5 Random samples of marking reports by the system

The initial testing of the system involved evaluating a sample set of essays to assess grading accuracy and feedback quality. The results indicate that the current AI model and its configurations require further refinement. The system frequently struggles to recognize certain words in various compositions accurately or delivers incorrect corrections. Other researchers have identified similar issues. Several studies indicate that ChatGPT alone achieves a disappointing accuracy rate, ranging from 51% to 57% (Mizumoto & Eguchi, 2023; Kim et al., 2024). Furthermore, some researchers contend that AI may not yet possess the sophistication required to evaluate complex writing elements, apply scoring criteria effectively, or consider emotional nuances as adeptly as humans (Bui & Barrot, 2024). As a result, we are currently conducting tests to assess the accuracy of integrating various pre-trained AI models to enhance reliability.

In the second phase, we will concentrate on fine-tuning the AI model with more intricate prompts and incorporating the capability to upload scoring samples marked by humans for reference. This will enable the AI to utilize these samples when evaluating students' essays, and is the most effective for ensuring that the AI aligns closely with human evaluations. Lastly, we intend to collect sample essays from local primary and secondary schools and invite teachers to provide feedback to facilitate further improvements. The system's validity and reliability will be thoroughly assessed and customized for the context of Hong Kong before its official release.

## 5. Conclusion

The AI-empowered Chinese Composition Assessment System significantly advances automated essay evaluation, particularly in Chinese writing assessment. Our pilot testing indicates that the system can substantially reduce grading workloads for educators while providing timely and constructive feedback to students, which is crucial for fostering an engaging learning environment. By employing a multi-model grading approach, we have enhanced the reliability and validity of evaluations, ensuring fair and accurate assessments. While the system shows promise, further refinements are essential to fully capture the intricacies of human writing and ensure that assessments truly reflect student capabilities.

We have integrated existing LLMs like ChatGPT, Gemini, and Llama into a MaaS framework. Our user-friendly prompt framework allows educators to customize assessment criteria, aligning automated grading with human evaluations for more relevant feedback. As the landscape of AES in Chinese writing evolves, our work addresses a notable gap in the literature. We aim to enhance the sophistication and reliability of essay evaluations by optimizing

prompting techniques that better reflect human assessment, which could ultimately inform best practices in writing instruction.

In conclusion, while the AI-empowered Chinese Composition Assessment System shows great potential, it also underscores the complexities of automating writing assessment. Future research will focus on refining the system with diverse human-scored samples, ultimately striving to create an assessment tool that not only streamlines grading but also enriches the educational experience for educators and students.

### Acknowledgements

The project was supported by the President's Development Fund, EdUHK. Ethical approval was granted by the Human Research Ethics Committee of The Education University of Hong Kong with the reference number [2023-2024-0451].

### References

- Alikaniotis, D., Yannakoudakis, H., & Rei, M. (2016). Automatic text scoring using neural networks. arXiv preprint arXiv:1606.04289.
- Attali, Y. & Burstein, J. (2006). Automated Essay Scoring With e-rater® V.2. *Journal of Technology, Learning, and Assessment*, 4(3).
- Bui, N. M., & Barrot, J. S. (2024). ChatGPT as an automated essay scoring tool in the writing classrooms: how it compares with human scoring. *Education and Information Technologies*, 1-18.
- Chan, K. K. Y., Bond, T., & Yan, Z. (2023). Application of an automated essay scoring engine to English writing assessment using many-facet rasch measurement. *Language Testing*, 40(1), 61-85.
- Foltz, P. W., Laham, D., & Landauer, T. K. (1999). The intelligent essay assessor: Applications to educational technology. *Interactive Multimedia Electronic Journal of Computer-Enhanced Learning*, 1(2), 939-944.
- Genlott, A. A., & Grönlund, Å. (2013). Improving literacy skills through learning reading by writing: The iWTR method presented and tested. *Computers & education*, 67, 98-104.
- Han, X., Zhang, Z., Ding, N., Gu, Y., Liu, X., Huo, Y., ... & Zhu, J. (2021). Pre-trained models: Past, present and future. *AI Open*, 2, 225-250.
- Ke, Z., & Ng, V. (2019). Automated Essay Scoring: A Survey of the State of the Art. In *IJCAI* (Vol. 19, pp. 6300-6308).
- Kim, H., Baghestani, S., Yin, S., Karatay, Y., Kurt, S., Beck, J., & Karatay, L. (2024). ChatGPT for writing evaluation: Examining the accuracy and reliability of AI-generated scores compared to human raters. *Exploring artificial intelligence in applied linguistics*, 73-95.
- Latif, E., & Zhai, X. (2024). Fine-tuning ChatGPT for automatic scoring. *Computers and Education: Artificial Intelligence*, 6, 100210.
- Lee, C., Wong, K. C., Cheung, W. K., & Lee, F. S. (2009). Web-based essay critiquing system and EFL students' writing: A quantitative and qualitative investigation. *Computer Assisted Language Learning*, 22(1), 57-72.
- Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics*, 2(2), 100050.
- Page, E. B. (1966). The imminence of... grading essays by computer. *The Phi Delta Kappan*, 47(5), 238-243.
- Ramesh, D., & Sanampudi, S. K. (2022). An automated essay scoring systems: a systematic literature review. *Artificial Intelligence Review*, 55(3), 2495-2527.
- Rudner, L. M., Garcia, V., & Welch, C. (2006). An evaluation of IntelliMetric™ essay scoring system. *The Journal of Technology, Learning and Assessment*, 4(4)

# Investigating Learner Emotional Engagement in Synchronous Online Learning

Wang Qiyun <sup>1\*</sup>, Quek Choon Lang <sup>2</sup>, Wen Yun <sup>3</sup> and Chin Yi Jie <sup>4</sup>

<sup>1, 2, 3, 4</sup>National Institute of Education, Nanyang Technological University, Singapore

<sup>1</sup>Co-Director, Centre for Research and Development in Learning, Nanyang Technological University, Singapore

\* qiyun.wang@nie.edu.sg

**Abstract:** Emotional engagement is highly important in synchronous online learning (SOL). However, few studies have investigated what learning activities can emotionally engage online learners. This study was conducted in a course with 16 graduate learners. The instructor deliberately designed some learning activities/events. The study aimed to investigate if these activities/events could emotionally engage the learners. Morphcast was used to analyze the learners' emotions in the Zoom session. Results showed that a well-prepared introduction, artifact sharing, taking a break, and giving peer feedback could emotionally engage online learners. However, the learners showed negative emotions in the ending period. Implications for teachers to design engaging lessons are discussed.

**Keywords:** Synchronous online learning, Emotional engagement, Facial emotion recognition, Online learning

## 1. Introduction

Synchronous online learning (SOL) was highly effective during the pandemic (Wang et al., 2023). Its flexibility allows instructors to teach virtually in real-time, overcoming the challenges of transactional distance and enabling students to participate in geographically diverse locations. However, online students often have lower engagement than those in the physical classroom settings (Wang et al., 2023). Among the dimensions of cognitive, behavioral, and emotional engagement (Martin & Borup, 2022), emotional engagement appears most crucial for online learning, as students may quickly become bored or frustrated when they lack social support (Dewaele et al., 2022). Additionally, research suggests that emotional engagement has greater effects on learner outcomes than other types of engagement (Deng, 2021). Therefore, the purpose of this study was to investigate what learning activities can emotionally engage online students in SOL. The research question of this study aimed to answer is: What instructional activities/events can emotionally engage online learners in SOL?

## 2. Emotional Engagement and Learning Activities

Emotional engagement refers to students' emotional reactions to their academic pursuits (Martin & Borup, 2022). Scholars typically classify emotions into positive, negative, and neutral categories (Martin & Borup, 2022). Positive emotions, such as happiness, are often correlated with better learning outcomes. Negative emotions often hinder learning. However, certain neutral emotions, like confusion, which can either be positive or negative, may also be useful for learning (Halverson & Graham, 2019).

Pekrun et al. (2023) highlighted the significance of arousal, valence (positivity), and objects of interest for emotional engagement. An object of interest refers to any activity associated with a relevant emotion. Some typical learning activities and events in SOL include instructor-led presentations, autonomous learning, collaborative learning, and structured breaks (Prayogo et al., 2024).

Instructor-led presentations are common in lectures. Exemplary instructors leverage on positive emotions and minimize negative emotions to capture students' attention (Martin & Borup, 2022). Likewise, investing in autonomous learning activities, such as artifact development, enhances emotional engagement. Positive emotions emerge when learners take ownership of their education. Their confidence increases as they receive guidance and refine their work through feedback (Prayogo et al., 2024). Furthermore, collaborative learning can deepen emotional engagement as

learners foster camaraderie and friendly competition with peers (Volet et al., 2019). Moreover, taking breaks during learning helps alleviate negative emotions, which supports knowledge retention.

### 3. Methodology

This study was conducted in a course at the National Institute of Education, Singapore, where 13 Masters of Education and 3 Doctorate candidates were enrolled. One of the course assignments was a group assignment that required students to develop a 1 hour e-learning package in groups of four. Learners presented their theoretical framework and showcased their prototypes in a SOL session, where student emotional engagement was monitored using a facial emotion recognition analysis (FERA) tool: Morphacst.

Morphcast is an FERA tool utilized for monitoring student emotional engagement. Its web-based version could analyze a student's expressions during each minute of a Zoom meeting. Attention, arousal, and positivity were measured. Seven basic emotions (i.e., happiness, surprise, neutrality, disgust, fear, anger, and sadness) were analyzed. Additionally, four quadrants were used to characterize learners' valence and arousal. The operational description of each metric is presented in Table 1.

Table 1. Operation definitions of Morphcast metrics (adopted from Zignego et al., 2023)

Metric	Description
Student engagement	Attention
	Arousal
	Positivity
Basic emotions	Happiness, surprise, neutrality, disgust, fear, anger, and sadness
Polar quadrant	Measure the level of focus students have on the content
	Measure the intensity of emotional responsiveness
	Measure the degree of positive response
	Measure the degree of seven basic emotions as a percentage
High control	Measure student level of control during interaction flow.
Low control	Measure a lesser degree of student control.
Conductive	Represents behaviours that facilitate smooth interaction.
Obstructive	Assesses behaviours that hinder the interaction process.

With regard to the study's learning activities/ events that were conducted during the last session of the course using Zoom, the instructor began the session with an *introduction*, where he implemented strategies such as a warm welcome, breaking the ice, and positive encouragement. Each group was then required to present two components of their work: *theoretical framework* and *artifact sharing*. In the middle of the session, a 15-minute *break* was provided. Also, *technical issues* were experienced by presenters during certain time periods. Nevertheless, after each group's presentation, students were invited to provide *peer feedback* using an online sheet. At the end of the session, the instructor also held a *debriefing* and offered feedback to the learners.

### 4. Results

Morphcast results are presented in the Appendix. The descriptive data were collected from the students, and these were then averaged to produce a mean score for each metric. The students were most attentive (=68.34%) and happy (=13.58%) during the instructor's introduction. A comparison between artifact sharing and theory presentation also showed that students appeared more attentive (=66.40% vs. 64.95%), less positive (=53.7% vs. 55.51%), and had lower control (=67.56% vs. 64.95%) during artifact sharing. Furthermore, Figure 1 illustrates increased arousal (in blue) and increased frustration (in green) when three PhD students presented their artifacts. During break time, learners demonstrated the highest arousal (=39.42%) and the lowest neutral scores (=19.55%). Learners were also the least angry (=11.98%) and most surprised (=18.15%) when facing technical issues. Additionally, during peer feedback, students appeared most positive (=57.73%) and displayed the highest control (=6.10%). Moreover, learner engagement was the lowest (arousal = 33.85% & attention = 60.15%), while negative emotions, including anger (=14.38%) and sadness (=20.64%), were the highest during the debriefing phase.

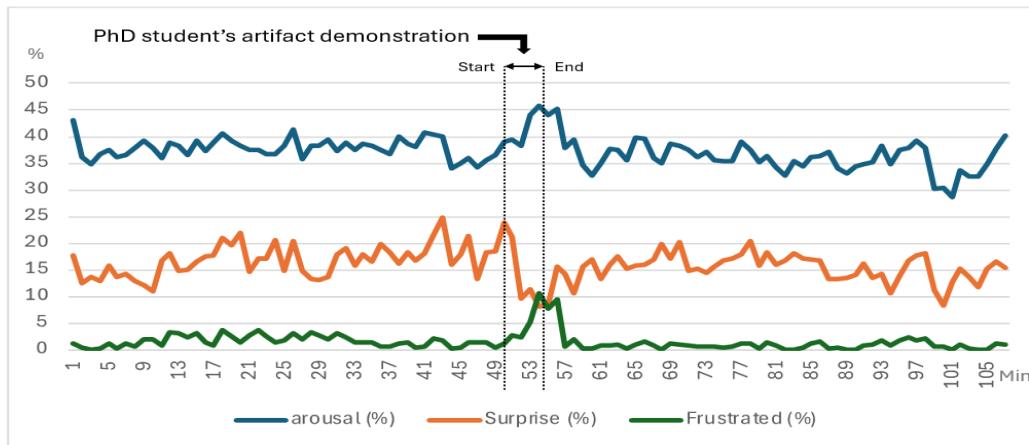


Fig.1 Time period when PhD students presented their

## 5. Discussion and Conclusion

Learners were most attentive and happy during the introduction. This suggest that students are most receptive during the period. Therefore, this is a good time for teachers to establish ground rules, make important announcements, and establish a positive relationship with learners. The study also revealed that, unlike theory presentation, learners were more attentive but less confident during artifact sharing. This may stem from their desire to see how their peers have performed (Jansen et al., 2022). Learners felt less confident when others did better, as evidenced by the increase in arousal and frustration when the PhD students presented their artifacts. Nevertheless, the surge in negative emotion has the potential to enhance learning as students strive for improved outcomes.

Students were more expressive during the break, as their arousal was the highest and neutrality lowest. Implementing such breaks can reduce negative feelings in a class (Prayogo et al., 2024). Also, students were not angry but rather surprised by technical faults. This finding was inconsistent with the result of Federman (2019). Additionally, students appeared most confident and in control during peer feedback. The increased autonomy may be attributed to students assuming the role of assessors to evaluate their peers (Prayogo et al., 2024). Student engagement was lowest during the debriefing phase, where negative emotions such as anger and sadness were detected. This finding implies that teachers should not use this period to convey important information.

In conclusion, a well-prepared introduction, artifact sharing, having a time break, giving peer feedback, and solving technical issues could increase learners' emotional engagement. However, learners showed negative emotions in the ending period.

## References

- Deng, R. (2021). Emotionally engaged learners are more satisfied with online courses. *Sustainability*, 13(20), 11169.
- Dewaele, J. M., Albakistani, A., & Ahmed, I. K. (2022). Levels of foreign language enjoyment, anxiety and boredom in emergency remote teaching and in in-person classes. *The Language Learning Journal*, 52(1), 117–130.  
<https://doi.org/10.1080/09571736.2022.2110607>
- Federman, J. E. (2019). Interruptions in online training and their effects on learning. *European Journal of Training and Development*, 43(5/6), 490-504.
- Halverson, L. R., & Graham, C. R. (2019). Learner engagement in blended learning environments: A conceptual framework. *Online learning*, 23(2), 145-178.
- Jansen, M., Boda, Z., & Lorenz, G. (2022). Social comparison effects on academic self-concepts—which peers matter most? *Developmental Psychology*, 58(8), 1541.

- Martin, F., & Borup, J. (2022). Online learner engagement: Conceptual definitions, research themes, and supportive practices. *Educational Psychologist*, 57(3), 162–177. <https://doi.org/10.1080/00461520.2022.2089147>
- Pekrun, R., Marsh, H. W., Elliot, A. J., Stockinger, K., Perry, R. P., Vogl, E., Goetz, T., Tilburg, W.A.P.V., Lüdtke, O., & Vispoel, W. P. (2023). A three-dimensional taxonomy of achievement emotions. *Journal of Personality and Social Psychology*, 124(1), 145.
- Prayogo, A., Khotimah, K., Istiqomah, L., & Maharsi, I. (2024). Students' emotional engagement in online classes: a conceptual framework. *The International Journal of Information and Learning Technology*, 41(1), 61-72.
- Volet, S., Seghezzi, C., & Ritchie, S. (2019). Positive emotions in student-led collaborative science activities: Relating types and sources of emotions to engagement in learning. *Studies In Higher Education*, 44(10), 1734-1746.
- Wang, Q., Wen, Y., & Quek, C. L. (2023). Engaging learners in synchronous online learning. *Education and Information Technologies*, 28(4), 4429-4452.
- Zignego, M. I., Bertirotti, A., Gemelli, P. A., & Pagani, L. (2023). Neurodesign: A Game-Changer in Educational Contexts.

#### **Appendix:** Student engagement data during different learning activities

Metrics	Learning Tasks	Introduction	Theory	Artifact	Break	Technical	Peer	Debriefing
		explanation	sharing	issues	feedback			
Student engagement	Arousal (%)	37.38	37.14	37.29	39.42	37.43	37.76	33.85
	Attention (%)	68.34	64.95	66.40	60.52	64.15	62.34	60.15
	Positivity (%)	54.98	55.51	53.70	49.72	55.80	57.73	56.27
Basic emotions	Angry (%)	13.12	12.50	12.16	12.24	11.98	13.46	14.38
	Disgust (%)	7.68	9.72	10.20	18.04	10.28	10.00	8.00
	Fear (%)	5.93	6.93	7.37	8.24	7.04	7.61	5.30
	Happy (%)	13.58	10.43	10.76	12.10	11.07	13.47	13.33
	Neutral (%)	25.76	26.35	26.51	19.55	24.01	24.42	24.64
	Sad (%)	19.78	17.59	16.14	16.65	17.62	15.71	20.64
	Surprise (%)	14.37	16.47	17.01	13.44	18.15	15.54	13.86
Polar quadrant	Conductive (%)	17.11	18.64	17.88	13.35	19.37	19.38	22.52
	High_Control (%)	4.27	3.82	3.24	2.16	3.96	6.10	2.31
	Low_Control (%)	68.67	64.95	67.56	62.55	64.08	62.01	69.25
	Obstructive (%)	9.95	12.40	11.32	21.94	12.59	12.51	5.92

# Revolutionizing Language and Culture Education through Immersive 360-Degree Video and Artificial Intelligence

Rustam Shadiev<sup>1\*</sup>, Nurassyl Kerimbayev<sup>2</sup>, Narzikul Shadiev<sup>3</sup>, Mirzaali Fayziev<sup>3</sup>, Fahriye Altinay<sup>4</sup>, Zhanargul Beisembayeva<sup>5</sup>, Roza Zhusupova<sup>5</sup>, Elena Gaevskaya<sup>6</sup>, and Asror Muhamedov<sup>7</sup>

<sup>1</sup> Zhejiang University, China

<sup>2</sup> Al-Farabi Kazakh National University, Kazakhstan

<sup>3</sup> Uzbekistan-Finland Pedagogical Institute, Uzbekistan

<sup>4</sup> Near East University, North Cyprus, Turkey

<sup>5</sup> L.N. Gumilyov Eurasian National University, Kazakhstan

<sup>6</sup> Saint Petersburg State University, Russia

<sup>7</sup> Alfraganus University, Uzbekistan

\* rustamsh@gmail.com

**Abstract:** This study explores the influence of incorporating virtual reality (VR) supported by 360-degree video and artificial intelligence (AI) technologies into cross-cultural learning initiatives to improve language proficiency, intercultural understanding, and communication skills. The research addresses the difficulties associated with authentic cultural engagement, often hindered by language barriers and limitations on travel. A series of sequential investigations was conducted, each refining the VR learning experience powered by 360-degree video and AI tools. Participants took part in cultural exchange activities such as self-introductions, cultural sharing, and reflective discussions on foreign cultures, facilitated by innovative educational technologies. Results indicated notable advancements in participants' linguistic skills, cultural awareness, and communicative abilities, emphasizing the value of immersive, technology-enhanced learning environments. The findings highlight the capacity of cutting-edge educational tools to bridge geographical and linguistic divides, foster intercultural competence, and enrich technology-mediated language learning strategies.

**Keywords:** Virtual Reality, 360-degree video, artificial intelligence, cross-cultural education, language learning.

## 1. Introduction

Recognizing and understanding cultural differences are vital components of contemporary education and society (Çiftçi, 2016; Tarihoran, 2020; Wang & Zhang, 2022). Cross-cultural learning, underpinned by cultural convergence theory, where individuals from varied cultural backgrounds exchange information to gain knowledge and develop intercultural attitudes (Çiftçi, 2016; Shadiev & Huang, 2016; Wang & Zhang, 2022), plays a critical role in this realm. This approach is further supported by contextual learning theory, which emphasizes the importance of situating learning within real-world activities, contexts, and cultures (Hwang et al., 2021; Hwang et al., 2022; Gu et al., 2017).

However, authentic cultural learning is often impeded by challenges such as travel restrictions and language barriers (Çiftçi, 2016; Wang & Zhang, 2022). Language obstacles, in particular, can significantly hinder communication and cultural exchange in cross-cultural programs, especially when participants have limited proficiency in foreign languages (Çiftçi, 2016; Wang & Zhang, 2022).

Innovative educational technologies offer promising solutions to these challenges. VR enables immersive and engaging cross-cultural learning experiences (Shadiev et al., 2021), while artificial intelligence facilitates the creation and enhancement of content and supports communication for individuals with limited language abilities (Lee, 2021). By integrating 360-degree video technology with AI tools in VR environments, seamless interaction and cultural exchange

become possible, overcoming both geographical and linguistic barriers (Akdere et al., 2021; Lee, 2021; Rupp et al., 2019; Steigerwald et al., 2022).

Despite these technological advancements, there is a noticeable gap in research exploring the role of 360-degree video and AI tools in enhancing cross-cultural learning. This study seeks to address that gap by designing activities that support language and cross-cultural learning through these technologies. Earlier studies often lacked interactive features within 360-degree video content and did not facilitate collaborative engagement among learners. In contrast, our research incorporated interactive components, such as text, images, hints, and embedded questions, within the video content and enabled real-time communication among participants viewing 360-degree videos.

This study investigates the impact of educational designs integrating 360-degree video and AI tools on language acquisition and cultural learning. It examines how technology-enhanced cross-cultural learning activities can foster language development, intercultural understanding, and communicative competence. This paper presents the findings from four sequential studies, each showcasing innovative approaches to cross-cultural education.

## 2. Methodology

In the initial study, we developed a VR learning environment leveraging 360-degree video technology. The second study introduced AI tools, such as translation systems, to bridge language barriers and facilitate participant interaction. The third study further enhanced the VR environment by adding features for real-time communication, content creation, and interactive viewing. In the fourth study, participants contributed by creating interactive VR content using multimedia elements, significantly improving the interactivity of the platform. Finally, in the last study, generative AI (GAI) was utilized to inspire content creation, identify and address issues or language challenges, and enhance understanding of the material.

Participants in the studies were university students from China, Indonesia, Russia, and Uzbekistan. They were provided with detailed instructions on the learning activities and technologies used. Before commencing the activities, we evaluated their baseline language proficiency, intercultural understanding, and communicative abilities. The cultural exchange activities were divided into the following distinct phases: *Personal Introductions*. Participants created 360-degree videos to introduce themselves, their hobbies, and daily routines. These videos were uploaded to the activity platform for sharing. *Cultural Presentations*. Participants selected cultural topics, such as traditional crafts, architecture, markets, or customs, and provided historical and informational content in 360-degree videos. They then viewed their partners' videos to gain insights into other cultures. *Reflections on Foreign Cultures*. After viewing the cultural presentations, participants created video-recorded reflections comparing the foreign cultures they had learned about with their own. These reflective reports were shared among the group. Interactive online meetings facilitated further discussion, where students viewed and discussed each other's videos, posed questions, and exchanged ideas about the cultural topics. Following these activities, we conducted post-tests to evaluate language skills (focused on speaking), intercultural understanding, and communicative competence. Interviews were also conducted to gather additional qualitative evidence supporting the findings.

The intervention was designed to develop language proficiency, enhance cross-cultural understanding, and improve communicative abilities through activities like personal introductions, cultural presentations, and reflective reports, all supported by VR supported by 360-degree video and AI technologies. To measure its effectiveness, various tools were used, including language proficiency tests crafted by experienced EFL instructors, reflective reports (Shadiev et al., in press), a questionnaire (Fantini, 2000), interviews, and video content analysis. Data were analyzed by two researchers. Language tests were evaluated using standardized rubrics assessing fluency, pronunciation, vocabulary, and communicative effectiveness. Reflective reports, interview transcripts, and video content were analyzed using open coding (Strauss & Corbin, 1990). Questionnaire responses were assessed on a 5-point Likert scale. Key technologies employed in the studies included Insta360 cameras (<https://www.insta360.com>) for recording, Oculus VR head-mounted displays for viewing (<https://www.oculus.com/gear-vr>), iFlytek translation tools

(<https://global.iflytek.com>), ChatGPT as a generative AI tool (<https://chatgpt.com/>), Tencent Meeting for communication (<https://meeting.tencent.com>), and Wonda for creating interactive video content (<https://www.wondavr.com>).

### **3. Results and discussion**

The study demonstrated favorable outcomes, showing significant improvements in language proficiency ( $p<0.05$ ), intercultural understanding ( $p<0.05$ ), and communicative competence ( $p<0.05$ ).

In interviews, numerous participants expressed that the technology-supported activities effectively enhanced their language abilities and cross-cultural learning. They reported improvements in speaking skills, cultural knowledge, comprehension, competence, and communication. Some students noted that the value of the immersive environments extended beyond the shared content, particularly in acquiring cultural insights. Translation tools were also praised for addressing language-related challenges, enabling participants to fully engage with and understand the video content. Generative AI proved valuable in offering creative inspiration, delivering feedback on created content, and enhancing the understanding of cultural materials shared by foreign peers.

The immersive VR environments, combined with interactions with peers from diverse backgrounds, significantly contributed to participants' cognitive development, aligning with the principles of cultural convergence theory. Many participants remarked on the benefits of these activities in enhancing speaking skills, thereby meeting the objective of fostering communicative competence through authentic and immersive learning experiences.

Students also emphasized that their exchanges with foreign peers deepened their intercultural understanding, consistent with the contextual learning theory. These interactions provided meaningful experiences, enabling them to acquire cultural knowledge and develop attitudes through engagement in real-world contexts, activities, and environments.

Regarding communicative competence, the integration of multimodal content and interactive elements in 360-degree videos enriched learners' perspectives and comprehension. These features streamlined the learning process by helping learners focus on key areas while offering greater autonomy in navigating their learning paths. Translation tools further supported participants by reducing language barriers, thereby enhancing their engagement with video content and improving overall communication skills.

In summary, the findings indicated significant advancements in language proficiency, cross-cultural understanding, and communicative competence. This underscores the efficacy of incorporating cultural convergence and contextual learning theories into technology-enhanced cross-cultural learning activities, achieving the study's goals.

### **4. Conclusions**

Our research introduces an innovative approach to language and culture learning, demonstrating how modern technologies can create more immersive, authentic, and interactive learning experiences. This evolution in technology-assisted learning signifies a shift towards utilizing VR and AI tools, surpassing traditional methods. The results highlight the transformative potential of VR and AI technologies in advancing language skills, cultural understanding, and communication abilities, marking a significant step forward in the evolution of technology-assisted learning methodologies.

Based on the study's positive outcomes, we recommend embedding intercultural learning activities within VR-based interactive environments supported by 360-degree video and AI technologies. This approach proved effective in fostering language skills, intercultural understanding, and communicative competence. Promoting active engagement in language and cognitive interactions is essential for developing these skills. The study underscores the importance of ample interaction within intercultural activities to enhance their effectiveness. Researchers are encouraged to explore diverse strategies or methods to provide learners with increased opportunities for cultural engagement.

However, the study faced certain limitations. The sample size was relatively small, and the duration was short, which may affect the generalizability of the findings. Additionally, a gender imbalance, with most participants being female, could influence the outcomes. These factors warrant cautious interpretation of the results. Furthermore, the lack of a control group limits the robustness of the conclusions.

### Acknowledgements

This research was supported by the Fundamental Research Funds for the Central Universities of the People's Republic of China.

### References

- Akdere, M., Acheson, K., & Jiang, Y. (2021). An examination of the effectiveness of virtual reality technology for intercultural competence development. *International Journal of Intercultural Relations*, 82, 109-120.
- Çiftçi, E.Y. (2016). A review of research on intercultural learning through computer-based digital technologies. *Educational Technology & Society*, 19, 313–327
- Tarihoran, N. A. (2020). Using Facebook Group as a Blended Learning Medium in Teaching Cross-Cultural Understanding in Islamic Higher Education. *International Journal of Research in STEM Education*, 2(1), 13-25.
- Fantini, A. E. (2000). *A Central Concern: Developing Intercultural Competence. About our institution, SIT Occasional Papers Series*. Palo Alto, CA: Spring
- Gu, X., Wang, H., & Mason, J. (2017). Are they thinking differently: A cross-cultural study on the relationship of thinking styles and emerging roles in computer-supported collaborative learning. *Educational Technology & Society*, 20(1), 13-24.
- Hwang, W.Y., Hariyanti, U. Chen, N.S., & Purba, S.W.D (2021). Developing and validating an authentic contextual learning framework: promoting healthy learning through learning by applying. *Interactive Learning Environments*.
- Hwang, W.Y., Guo, D.C., Hoang, A., Chang, C.C., & Wu, N.T. (2022). Facilitating authentic contextual EFL speaking and conversation with smart mechanisms and investigating its influence on learning achievements. *Computer Assisted Language Learning*. DOI: 10.1080/09588221.2022.2095406
- Lee, S.-M. (2021). The effectiveness of machine translation in foreign language education: a systematic review and meta-analysis. *Computer Assisted Language Learning*, 1-23.
- Rupp, M. A., Odette, K. L., Kozachuk, J., Michaelis, J. R., Smither, J. A., & McConnell, D. S. (2019). Investigating learning outcomes and subjective experiences in 360-degree videos. *Computers & Education*, 128, 256-268.
- Shadiev, R. & Huang, Y.M. (2016). Facilitating cross-cultural understanding with learning activities supported by speech-to-text recognition and computer-aided translation. *Computers & Education*, 98, 130-141.
- Shadiev, R., Wang, X., Chen, X., Gayevskaya, E. & Borisov, N. (in press). Research on the impact of the learning activity supported by 360-degree video and translation technologies on cross-cultural knowledge and attitudes development. *Education and Information Technologies*.
- Shadiev, R., Wang, X., Wu, T. T., & Huang, Y. M. (2021b). Review of Research on Technology-Supported Cross-Cultural Learning. *Sustainability*, 13(3), 1402.
- Steigerwald, E. et al. (2022). Overcoming Language Barriers in Academia: Machine Translation Tools and a Vision for a Multilingual Future. *BioScience*, 72(10), 988–998.
- Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Sage.
- Wang, J. & Zhang, S. (2022). Cross-Cultural Learning: A Visualized Bibliometric Analysis Based on Bibliometrix from 2002 to 2021. *Mobile Information Systems*, 2022, 7478223.

# A Pilot Study on Bridging EFL Writing and Speaking Skills through AI-enhanced Authentic Short Video-Making

Yi-Fan Liu<sup>1</sup>, Muhammad Irfan Luthfi<sup>2,4\*</sup>, Wu-Yuin Hwang<sup>2,3</sup>

<sup>1</sup> Research Center for Testing and Assessment, Academy for Educational Research, Taiwan.

<sup>2</sup> Graduate Institute of Network Learning Technology, Central University Taiwan.

<sup>3</sup> Department of Computer Science and Information Engineering, Dong Hwa University, Taiwan.

<sup>4</sup> Department of Electronics and Informatics Engineering Education, Universitas Negeri Yogyakarta, Indonesia

\* iir.irfan02@gmail.com

**Abstract:** The integration of technology in English as a Foreign Language (EFL) instruction offers significant potential to enhance writing and speaking proficiency. Among technological applications, authentic tasks like short video production have proven effective in promoting real-world language use and learner engagement. While prior research highlights the benefits of video-based learning for specific skills such as speaking fluency and writing coherence, limited attention has been paid to the interplay between these competencies or the role of video quality in comprehensive language outcomes. This study investigates the impact of Artificial Intelligence-enhanced authentic short video-making (AI-eSV) on the integrated development of EFL writing and speaking skills. Conducted over two weeks, the study engaged 20 first-year university students from a public university in central Taiwan in a structured process of collaborative scriptwriting and video recording in authentic contexts. Using a pretest-posttest design, video quality assessments, and participant interviews, the study provided a holistic evaluation of the intervention. Results showed significant improvements in writing and speaking skills, including enhanced creativity, content organization, and communicative confidence. The collaborative and authentic nature of the tasks further fostered learner engagement and skill integration. These findings highlight the pedagogical value of AI-enhanced short video-making as a bridge between theoretical instruction and practical language application, emphasizing the importance of incorporating technology-driven, authentic activities in EFL curricula. Future research should explore the long-term impact and scalability of this innovative approach.

**Keywords:** EFL writing and speaking, AI-enhanced authentic short video-making, writing and speaking skills, technology-enhanced learning, collaborative learning

## 1. Introduction

The integration of technology into EFL instruction has demonstrated significant potential in enhancing writing and speaking skills through innovative pedagogical approaches. Authentic tasks, such as video-making, establish real-world contexts for language use, enabling learners to bridge the gap between classroom instruction and practical communication (Ouyang, 2024). The use of multimedia tools has further supported iterative language practice, promoting vocabulary acquisition, grammatical accuracy, and communicative confidence (Abdolmanafi-Rokni & Qarajeh, 2014). Learner-generated content, such as short videos, has been linked to increased engagement and motivation, fostering creativity and collaboration during language practice (Huang, 2021). Video-based tasks, particularly those incorporating scriptwriting and performance, provide structured opportunities for meaningful language application, improving both writing and speaking skills (Hafner & Miller, 2021).

Existing research underscores the efficacy of video-based learning in specific language domains. Menggo et al. (2022) observed that video-making tasks enhance speaking performance through practical language use, while Göktürk (2016) reported improvements in fluency and grammatical accuracy via digital video recordings. Similarly, Jiang et al. (2021) highlighted gains in speaking and listening skills through automatic speech recognition, though these outcomes were not consistently reflected in overall proficiency metrics. Cahyono and Astuti (2019) emphasized the creative and organizational benefits of video tasks in writing, while Yeh et al. (2020) demonstrated that project-based video-making enhanced critical thinking, collaboration, and holistic language development. Despite these advancements, gaps persist in understanding how video quality correlates with integrated skill development and overall proficiency, highlighting the need for further investigation.

To address these gaps, this study developed the AI-enhanced Short Video-making (AI-eSV) system, a learning tool based on communicative language teaching (CLT) principles (Rahman, Singh, & Pandian, 2018). AI-eSV uses artificial intelligence to give real-time, personalized feedback during scriptwriting and video production. It helps improve fluency, coherence, and grammar by providing feedback that allows students to make changes and improve their work (Huang, 2021; Yeh et al., 2020). By engaging students in video-making with real-world contexts, AI-eSV supports the development of both writing and speaking skills. This study explores two main research questions: (1) Is there any improvement in students' EFL writing and speaking skills after using AI-eSV? and (2) Is there any correlation between students' short video quality and their EFL writing and speaking skills? These questions guide the study's goals of evaluating AI-eSV's effectiveness and understanding the link between video quality and language learning.

## **2. Literature Review**

### ***2.1. The Role of AI-Driven Technology in Enhancing EFL Skills***

Artificial Intelligence (AI) has increasingly influenced EFL education, particularly in enhancing writing and speaking skills. AI-powered tools, such as chatbots and intelligent tutoring systems, provide learners with interactive platforms to practice language skills, offering immediate feedback and personalized learning experiences. For instance, AI chatbots have been shown to improve oral communication skills by simulating authentic conversational contexts, thereby increasing learners' engagement and proficiency (Wu & Li, 2024). Additionally, AI-driven writing assistants help students develop their writing abilities by providing real-time corrections and suggestions, which can reduce anxiety and improve overall writing quality (Song & Song, 2023). Despite these advancements, the integration of AI in EFL teaching presents challenges, including educators' skepticism and the need for effective implementation strategies (Zhang & Umeanowai, 2024). Moreover, while AI tools like ChatGPT offer personalized learning experiences, concerns about academic dishonesty and over-reliance on technology have been raised (Lo et al., 2024). Therefore, while AI offers promising tools for EFL instruction, further research is necessary to fully understand its impact and optimize its application in language learning environments.

### ***2.2. EFL Writing and Speaking Enhancement with Video-Making***

Video-making has emerged as an effective pedagogical tool for enhancing English as a Foreign Language (EFL) learners' skills, particularly in writing and speaking. Engaging students in video production fosters active learning and authentic language use, leading to improved proficiency. For instance, Yeh et al. (2021) found that EFL students participating in video-making projects demonstrated significant advancements in writing skills, attributed to increased motivation and the necessity to organize content coherently. Similarly, Göktürk (2016) reported that digital video recordings positively impacted learners' oral fluency, providing opportunities for self-assessment and iterative improvement. Moreover, the use of English-subtitled videos has been shown to enhance listening and vocabulary skills, contributing to overall language competence (Bahtiar, 2023). Implementing animated videos in instruction also resulted in higher vocabulary retention among learners, indicating the versatility of video-based methods in language education (Almurashi, 2016). Furthermore, utilizing platforms like YouTube for video-making projects has facilitated improvements in speaking skills, offering learners authentic contexts to practice and refine their language abilities

(Melisa et al., 2023). These studies collectively underscore the multifaceted benefits of incorporating video-making into EFL curricula to bolster language proficiency.

### **2.3. AI-Driven Video-Making in Authentic Contexts for EFL Skill Enhancement**

Integrating AI into video-making within authentic contexts has shown significant potential for enhancing EFL skills. AI-driven tools, such as intelligent tutoring systems and virtual reality environments, provide immersive and interactive platforms that facilitate authentic language use and contextual learning. For instance, Jiang (2022) highlights the application of AI in creating realistic conversational agents, enabling learners to practice speaking in lifelike scenarios. Similarly, Lee et al. (2024) demonstrate that immersive virtual reality can situate learners in authentic settings, promoting natural language acquisition through contextualized experiences. Moreover, Zhang and Umeanowai (2024) discuss the transformative influence of AI in EFL contexts, emphasizing its role in providing personalized feedback and adaptive learning pathways. However, challenges such as educators' skepticism toward AI integration and the need for effective implementation strategies persist, as noted by Alshumaimeri and Alshememry (2023). Despite these challenges, the convergence of AI and video-making in authentic contexts offers promising avenues for EFL skill enhancement, warranting further exploration and empirical validation.

### **3. AI-Enhanced Authentic Short Video Making**

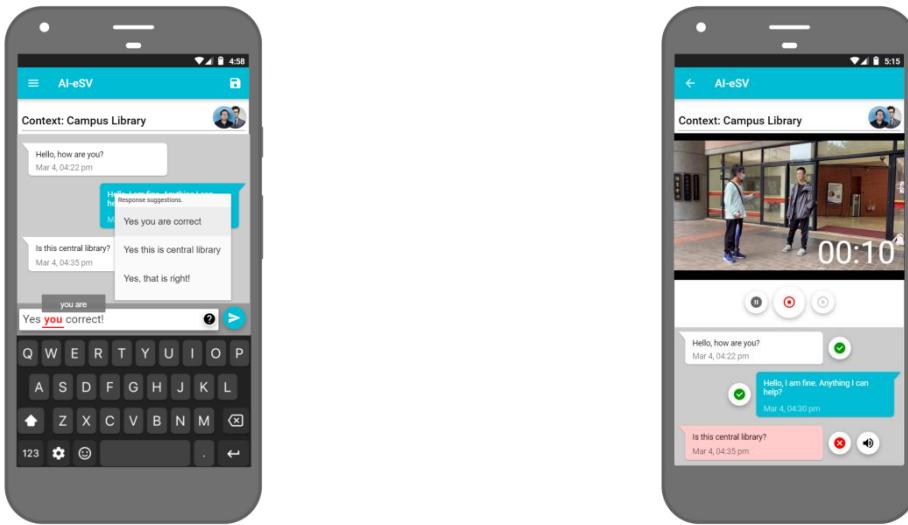


Fig.1 User Interface of AI-eSV to help Students Enhance Their EFL Writing and Speaking Through Creation of Short Video in Authentic Context; Left Side: Short Video Script Writing; Right Side: Authentic Short Video Recording

The AI-enhanced Authentic Short Video-Making (AI-eSV) system is explicitly designed based on the principles of Communicative Language Teaching (CLT), emphasizing interaction, real-world context, and learner engagement as key components of language acquisition. By integrating these principles, the AI-eSV system creates an effective and authentic language learning environment that aligns with CLT methodologies. Initially, students worked in pairs to plan the content of their videos, with a focus on scriptwriting as a foundational step. During this planning phase, students utilized AI-driven writing assistance to refine their scripts collaboratively (Figure 1). The AI-eSV system provided real-time feedback on grammatical accuracy, sentence coherence, and contextual relevance, allowing students to iteratively enhance their writing. The AI system flagged syntactic errors, suggested alternative phrasing, and analyzed vocabulary choices to help students improve the clarity and expressiveness of their scripts. This integration of AI technology ensured that the scripts met high linguistic standards while maintaining authenticity and contextual relevance to the video projects. The collaborative planning phase further promoted critical thinking and meaningful language practice, strengthening the foundation for the subsequent stages of video production.

Following the planning stage, students produced 30-second videos in authentic contexts, leveraging their campus environment to align their content with academic themes. The use of real-world settings, such as libraries, lecture halls, and study areas, reinforced the practical application of their language skills while integrating their academic environment into the task (Figure 2). During the recording process, students were required to speak loudly and clearly, ensuring their oral communication could be effectively evaluated by the teacher. The AI-eSV further supported this process by analyzing speech clarity and pronunciation in real-time. If the system detected unclear speech, mispronunciations, or unnatural pauses, it automatically paused the recording and prompted students to listen to a Text-to-Speech (TTS) playback of the correct pronunciation. Students were then required to practice and re-record their lines until the system confirmed their speech clarity and accuracy. The AI-eSV system utilized speech recognition algorithms to assess intonation, rhythm, and articulation, helping students refine their pronunciation and fluency. This iterative learning process ensured that students produced high-quality spoken content while reinforcing proper pronunciation, fluency, and confidence in their spoken language. The integration of AI-driven feedback allowed students to self-correct, make multiple attempts, and progressively refine their speaking skills, making the learning process more interactive and adaptive.



Fig.2 Students Making Short Videos in Authentic Context; Left Side: Two Students Discuss about Teaching Center Building;  
Right Side: Two Students Discuss about Campus Book Store.

#### 4. Method

##### 4.1. Participant

The study involved 20 first-year university students from a public university in central Taiwan, encompassing diverse educational backgrounds, linguistic proficiency, and cultural perspectives. This diversity provided a robust foundation for evaluating the impact of the short video-making intervention on learners with varying levels of preparedness. Integrated into their first-year academic curriculum, the intervention aligned closely with their educational context, while voluntary participation demonstrated students' genuine interest in improving their EFL writing and speaking skills. This participant composition offered critical insights into how collaborative and authentic tasks can address challenges faced by diverse learners, ensuring the relevance and applicability of the findings to similar educational settings.

##### 4.2. Data Collection and Analysis

The study implemented a structured research procedure consisting of a pretest, study introduction, video planning (script writing), short video creation in authentic contexts, video sharing, posttest, and participant interviews. The pretest and posttest assessed students' EFL writing and speaking skills using a short conversation test and a video description writing task, with evaluations conducted by two English teachers achieving a Cohen's kappa reliability coefficient of 0.782. Creativity, organization, and confidence were assessed using detailed rubrics developed in collaboration with experienced EFL instructors. The rubrics included criteria such as the originality of ideas, logical flow and structure, and self-assurance in delivery, each rated on a 5-point Likert scale. These assessment tools were piloted prior to the study to establish validity and reliability, ensuring alignment with the study's objectives. During the

video-making phase, students collaboratively produced 30-second videos in authentic academic contexts, integrating practical language use, while video sharing facilitated collaborative learning. Video quality was assessed through a questionnaire focusing on EFL writing development, speaking development, and video-making engagement, with a Cohen's kappa value of 0.814, indicating high inter-rater reliability. Combined with qualitative data from participant interviews, these evaluations provided a comprehensive analysis of the intervention's effectiveness in enhancing EFL writing and speaking skills through AI-supported authentic video-making tasks.

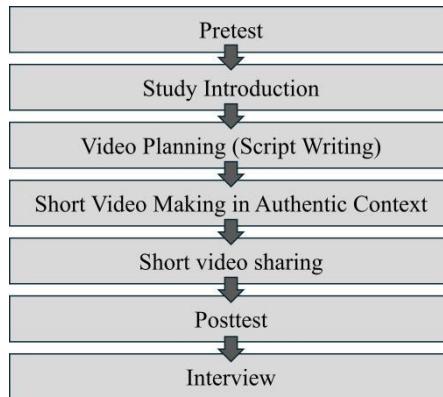


Fig.3 Research Design

The data analysis utilized multiple methods to ensure reliability and accuracy. Descriptive statistics summarized data from pretests, posttests, video scripts, and video quality evaluations. Independent sample t-tests revealed significant improvements in students' EFL writing and speaking skills between pretest and posttest scores. Pearson correlation analysis examined relationships between video quality metrics and skill development, identifying significant correlations between video-making engagement and specific skill areas. The reliability of the creativity, organization, and confidence measures was verified through inter-rater agreement, while the validity was established by aligning the rubrics with established pedagogical frameworks. Cohen's kappa confirmed consistency in assessment and questionnaire ratings. This comprehensive approach, integrating quantitative analysis with qualitative insights, provided a robust evaluation of the impact of authentic short video-making tasks on EFL skill enhancement.

Table 1. Research Variable

#	Variable	Description
1	Pretest/Posttest Score	Measures students' EFL writing and speaking skills before and after the intervention.
2	Development of Writing Skills	Assesses the quality and content of the written script created during video planning.
3	Demonstration of Speaking Skills	Evaluates students' spoken performance during the video-making process.
4	Engagement in Collaborative and Creative Language Use	Analyzes the video quality based on participant engagement.

#### 4. Analysis of Students EFL Writing and Speaking Skills

The analysis of students' EFL writing and speaking skills reveals significant improvement following the implementation of the short video-making intervention (Table 2). Pretest scores ( $M = 51.00$ ,  $SD = 2.34$ ) reflected basic sentence structures, frequent grammatical errors, and limited vocabulary in writing, as well as speaking marked by hesitation, monotony, and restricted lexical choices. Posttest scores ( $M = 95.95$ ,  $SD = 2.28$ ) demonstrated substantial progress, with a statistically significant mean difference of 44.95 ( $p < 0.001$ ). Writing skills showed notable enhancements in complexity, grammatical accuracy, and lexical diversity, while speaking performance improved in fluency, confidence, and expressive range. These results underscore the effectiveness of the intervention in fostering integrated language skills through authentic, iterative video-making tasks.

Table 2. Comparison of Pretest and Posttest Score

	N	Mean	Std. Dev	Sig. (2-tailed)
Pretest	20	51.00	2.340	<0.001
Posttest	20	95.95	2.282	

These findings are consistent with prior research in EFL education. Ouyang (2024) reported significant enhancements in writing and speaking skills through video-based tasks, emphasizing how authentic activities facilitate vocabulary acquisition and communication fluency. Kessler (2018) observed that collaborative video production fosters speaking confidence and grammatical accuracy in writing, with iterative processes such as planning and drafting reinforcing language acquisition. Similarly, Abdolmanafi-Rokni and Qarajeh (2014) demonstrated that video-making improves writing coherence and speaking fluency, supported by multimodal elements that enhance retention and the application of language structures. Huang (2021) highlighted the importance of collaboration and peer feedback in refining writing organization and speaking fluency, while Hafner and Miller (2021) emphasized the role of contextual video projects in fostering purposeful and precise language use. Yeh et al. (2020) further illustrated that project-based video-making promotes critical thinking, writing depth, and speaking clarity, aligning with this study's findings of improved grammatical precision, coherence, and expressiveness. Qualitative data from participant interviews corroborate these findings, with students reporting that scriptwriting expanded their vocabulary and improved sentence structure, while repeated rehearsals boosted confidence and fluency. Peer feedback during rehearsals further refined pronunciation and delivery, demonstrating the effectiveness of video-making in promoting linguistic accuracy and communicative competence, consistent with observed test score improvements.

## 5. Analysis of Students' Video Quality Toward EFL Writing and Speaking Skills

The analysis of students' video quality in relation to EFL writing and speaking skills revealed significant relationships between key variables, while also highlighting notable limitations (Table 3). A strong correlation was observed between speaking skills demonstrated in video tasks and writing skill development in authentic contexts ( $r = .739$ ,  $p < 0.01$ ), indicating that students who performed well in oral presentations also exhibited higher proficiency in writing. Additionally, engagement in collaborative and creative efforts during video production significantly correlated with speaking skills ( $r = .646$ ,  $p < 0.01$ ), emphasizing the positive impact of collaboration on oral communication abilities. However, no significant correlations were found between posttest scores and video quality metrics, including writing and speaking skill development, collaborative engagement, and creativity. These findings suggest that while video-making activities effectively enhanced isolated skill areas, they did not directly translate into measurable improvements in overall posttest performance, potentially due to external factors such as test design or other mediating influences.

Table 3. Analysis of Students' Video Quality Toward EFL Writing and Speaking Skills

#		(1)	(2)	(3).	(4)
1	Posttest Score		1		
2	Development of Writing Skills	0.190		1	
3	Demonstration of Speaking Skills	0.257	0.739**		1
4	Engagement in Collaborative and Creative Language Use	0.352	0.250	0.646**	1

\*\*. Correlation is significant at the 0.01 level (2-tailed).

These findings align with prior research on the benefits and limitations of video-based tasks in EFL learning. Menggo et al. (2022) highlighted that video-based tasks significantly enhance speaking skills through meaningful language use, consistent with this study's findings on improved speaking and collaboration. Göktürk (2016) observed that digital video recordings improved overall speaking performance but had limited impact on fluency, supporting the conclusion that video tasks target specific skills rather than comprehensive outcomes. Similarly, Jiang et al. (2021)

noted that ASR-supported video activities enhanced perceived speaking abilities, though these gains were not consistently reflected in standardized assessments, reinforcing the lack of correlation between video quality and posttest scores. Cahyono and Astuti (2019) emphasized that video tasks fostered creativity and content development in writing, despite limited improvements in grammatical accuracy, highlighting the complementary role of video-based learning alongside traditional methods. Interviews corroborated these findings, with participants reporting improvements in writing organization, vocabulary use, fluency, and confidence. One student stated, "Writing for the videos made me think about how to connect my ideas better," while another noted, "Practicing my lines repeatedly helped me become more confident and fluent." Peer feedback further refined pronunciation and delivery. Some participants, however, cited challenges in balancing creativity and accuracy, which may explain the limited correlation with test scores. These results demonstrate how video-making tasks promote targeted skill development through iterative practice and collaboration.

## 6. Conclusion

This study demonstrates the efficacy of AI-eSV in enhancing EFL writing and speaking skills, with quantitative results showing significant improvements in grammatical accuracy, lexical diversity, and fluency in both written and oral communication. Significant correlations between video quality metrics and specific language skills further highlight the effectiveness of video-making tasks in fostering targeted linguistic development. Qualitative insights from participant interviews corroborate these findings, with students reporting increased confidence, creativity, and collaborative engagement. However, the small sample size and reliance on standardized assessments limit the generalizability of the findings and may overlook nuanced improvements, such as creativity and authentic language use. Participants also faced technical challenges, such as time constraints and limited equipment access, underscoring the need for enhanced support in future implementations. To address these limitations, future research should include larger, more diverse samples, multidimensional assessment tools, and longitudinal designs to evaluate the sustained impact of such interventions. Integrating advanced technologies, such as automated feedback systems and adaptive learning platforms, could further optimize instructional design, ensuring scalability and accessibility across varied educational contexts. By fostering integrated language development through real-world engagement, AI-driven authentic video-making aligns with communicative language teaching principles and equips learners with essential skills for effective communication, paving the way for innovative, collaborative, and learner-centered methodologies in EFL education.

## Acknowledgments

I extend my gratitude to the reviewer, advisors, the High Interaction in Multimedia Lab team, the Research Assistants in Video-to-Text Recognition, and the EFL Study group.

## References

- Abdolmanafi-Rokni, S. J., & Qarajeh, M. (2014). Digital storytelling in EFL classrooms: The effect on oral performance. *International Journal of Language and Linguistics*, 2(4), 252–257.  
<https://doi.org/10.11648/j.ijll.20140204.12>
- Almurashi, W. A. (2016). The effective use of YouTube videos for teaching English language in classrooms as supplementary material at Taibah University in Alula. *International Journal of English Language and Linguistics Research*, 4(3), 32–47.
- Alshumaimeri, Y. A., & Alshememry, A. K. (2023). The extent of AI applications in EFL learning and teaching. *IEEE Transactions on Learning Technologies*. <https://doi.org/10.1109/TLT.2023.3322128>
- Bahtiar, I. (2023). Exploring English subtitles on YouTube video for the enhancement of English as a foreign language in students' listening comprehension. *Journal of Languages and Language Teaching*, 11(1), 1–10.

- Cahyono, B. Y., & Astuti, U. P. (2019). Effect of process writing approach combined with video-based mobile learning on Indonesian EFL learners' writing skill across creativity levels. *International Journal of Instruction*, 12(3), 325–340. <https://doi.org/10.29333/iji.2019.12320a>
- Göktürk, N. (2016). Examining the effectiveness of digital video recordings on oral performance of EFL learners. *Teaching English with Technology*, 16(2), 71–96.
- Hafner, C. A., & Miller, L. (2021). Project-based learning and the development of oral proficiency. *Language Learning & Technology*, 25(3), 123–145.
- Huang, H. W. (2021). Effects of smartphone-based collaborative vlog projects on EFL learners' speaking performance and learning engagement. *Australasian Journal of Educational Technology*, 37(6), 18–40. <https://doi.org/10.14742/ajet.6623>
- Jiang, M. Y. C., Jong, M. S. Y., Lau, W. W. F., Chai, C. S., & Wu, N. (2021). Using automatic speech recognition technology to enhance EFL learners' oral language complexity in a flipped classroom. *Australasian Journal of Educational Technology*, 37(2), 110–131. <https://doi.org/10.14742/ajet.6798>
- Jiang, R. (2022). How does artificial intelligence empower EFL teaching and learning nowadays? A review on artificial intelligence in the EFL context. *Frontiers in Psychology*, 13, Article 1049401. <https://doi.org/10.37905/jetl.v4i2.24446>
- Kessler, G. (2018). Technology and the future of language teaching. *Foreign Language Annals*, 51(1), 205–218. <https://doi.org/10.1111/flan.12318>
- Lee, S.-M., Wang, X., Park, I., & Lestiono, R. (2024). It feels so real! Situated authentic language learning in immersive virtual reality. *Education and Information Technologies*, 29, 24023–24045. <https://doi.org/10.1007/s10639-024-12807-8>
- Melisa, M., Abid, A., & Muhamad, F. (2023). Utilising YouTube to facilitate English speaking skills improvement among EFL learners. *Jambura Journal of English Teaching and Literature*, 4(2), 85–91. <https://doi.org/10.37905/jetl.v4i2.24446>
- Menggo, S., Basir, A., & Halum, Y. S. (2022). Video-based tasks in strengthening speaking skills of EFL college students. *Indonesian Journal of EFL and Linguistics*, 7(2), 279. <https://doi.org/10.21462/ijefl.v7i2.510>
- Ouyang, C. (2024). Video-based instruction: Effects on EFL learners' spoken learning. *Creative Education*, 15(12), 2609–2618. <https://doi.org/10.4236/ce.2024.1512159>
- Rahman, M. M., Singh, M. K. M., & Pandian, A. (2018). Exploring ESL teacher beliefs and classroom practices of CLT: A case study. *International Journal of Instruction*, 11(1), 295–310. <https://doi.org/10.12973/iji.2018.11121a>
- Song, H., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2023.1260843>
- Wu, Y., & Li, X. (2024). Unraveling effects of AI chatbots on EFL learners' language skill development: A meta-analysis. *Springer Nature*. <https://doi.org/10.1007/s40299-024-00853-2>
- Yeh, H. C., Heng, L., & Tseng, S. S. (2020). Exploring the impact of video making on students' writing skills. *Journal of Research on Technology in Education*, 53(4), 446–456. <https://doi.org/10.1080/15391523.2020.1795955>
- Zhang, L., & Umeanowai, A. (2024). Exploring the transformative influence of artificial intelligence in EFL context: A comprehensive bibliometric analysis. *Springer Nature*. <https://doi.org/10.1007/s10639-024-12937-z>
- Zhang, X., & Umeanowai, K. O. (2024). Exploring the transformative influence of artificial intelligence in EFL context: A comprehensive bibliometric analysis. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-024-12937-z>

# Evaluation of a Professional Development Course on Artificial Intelligence Literacy for Administrative Staff in Hong Kong

Siu Cheung Kong<sup>1,2\*</sup>, Yin Yang<sup>1</sup>, Yue Wu<sup>2</sup>

<sup>1</sup> Department of Mathematics and Information Technology, The Education University of Hong Kong, China

<sup>2</sup>Artificial Intelligence and Digital Competency Education Centre, The Education University of Hong Kong, China

\*sckong@eduhk.hk

**Abstract:** This study examines the implementation and outcomes of an artificial intelligence (AI) literacy course for administrative staff. One hundred and twelve administrative staff from schools, small to medium-sized enterprises, and universities in Hong Kong participated in a 30-hour course, which included the introduction to machine learning concepts, AI tools, and AI ethics. The learning outcomes culminated in a group presentation showcasing the participants' learning achievements. Using the attention, relevance, confidence, and satisfaction model, the motivation survey revealed high participant engagement, with strong correlations between motivation factors. The participants demonstrated positive acceptance of AI tools. Thematic analysis of pre- and post-course reflections highlighted significant improvements in their understanding and application of AI tools, particularly in terms of efficiency and work quality. This study contributes to the underexplored area of AI literacy development among administrative professionals, highlighting the need for tailored workplace training initiatives in future study.

**Keywords:** administrative staff, artificial intelligence literacy, motivation survey, professional development, thematic analysis

## 1. Introduction

The AI Index Report 2024 revealed that 52% of Americans express more concern than excitement about artificial intelligence (AI), an increase from 38% in 2022 (Stanford University, 2024), worrying primarily about AI's impact on their jobs. According to McKinsey & Co. (2023), AI is currently used in 55% of organisations in at least one business unit or function, up from 50% in 2022 and 20% in 2017. Manyika et al. (2017) indicated that administrative roles are among the most vulnerable to AI-driven job elimination. Furthermore, McKinsey & Co.'s May 2024 report projected continued declines in demand for workers in food services, production, customer services, sales, and office support, sectors that already experienced a downturn from 2012 to 2022, through 2030. Thus, although AI is poised to reshape the labour market, it is also expected to enhance productivity and bridge the gap between low- and high-skilled workers (Stanford University, 2024).

Most AI literacy initiatives primarily target K–12 and higher education, focusing on teaching skills (Ahmad et al., 2022; Casal-Otero et al., 2023; Kong et al., 2023). However, there remains a gap in AI literacy education tailored to administrative staff. This study fills this gap by equipping administrators with essential AI knowledge, fostering deeper and more reliable thinking about their career development, and preparing them for an uncertain future. The course is designed to help them (1) use AI to work more efficiently; (2) stay updated on emerging AI tools; (3) evaluate AI tools for their work; and (4) develop motivation to adopt AI tools in the future.

The study answered the following research questions: (1) What factors motivate administrative staff to participate in an AI literacy course? (2) How do administrative staff's perceptions of AI in the workplace evolve throughout the course?

## 2. Literature Review

## **2.1. AI Literacy at Work**

AI literacy refers to the competencies needed by workers to use AI and establish a synergistic relationship with it (Kong et al., 2021). Despite various definitions, Laupichler et al. (2023) provided a concise definition, emphasising understanding, using, monitoring, and critically thinking about AI applications, independent of the ability to develop AI models.

As the AI literacy framework remains undefined, we adopted the model of Kong et al. (2021) and its three key components: understanding AI concepts, evaluating AI applications, and applying AI concepts to real-world problem-solving. These components align with the four dimensions of AI literacy: (1) cognitive (understanding of AI concepts); (2) metacognitive (use of AI concepts for problem-solving); (3) affective (psychological readiness to use AI); and (4) social (ethics of problem-solving with AI) (Kong et al., 2024).

## **2.2. The Attention, Relevance, Confidence, Satisfaction Model**

The attention, relevance, confidence, satisfaction (ARCS) model is a motivational design framework used for decades in various countries and educational settings. It is a fundamental model of instructional design (Keller, 1987; Li & Keller, 2018). The ARCS model has applications beyond traditional educational environments, including workforce training and professional development. It has been employed to assess motivation and engagement in diverse learning contexts, as demonstrated in studies such as Chang et al. (2019). By focusing on fostering attention, demonstrating relevance, building confidence, and ensuring satisfaction, the ARCS framework offers a robust approach to designing and evaluating instructional interventions.

## **2.3. Project-Based Learning**

Project-based learning (PBL) is an instructional approach that engages learners by immersing them in tasks that involve problem-solving, inquiry, and collaboration. Through PBL, learners are encouraged to seek solutions, ask critical questions, debate ideas, design actionable plans, and effectively communicate with others (Choi et al., 2019). A growing body of research highlights the benefits of PBL in enhancing learners' motivation, problem-solving abilities, teamwork, and communication skills (Zhang & Ma, 2023). For professional training contexts, especially those with limited course durations and diverse participant backgrounds, PBL offers a practical and impactful learning strategy. By simulating real-world challenges, PBL effectively mirrors the complexities of professional environments, equipping learners with the practical skills and decision-making capabilities that traditional teaching methods may not adequately address.

## **3. Methodology**

### **3.1. Research Design**

The course consisted of five lessons, each lasting six hours. The course included three face-to-face teaching sessions, one self-study day (with materials provided), and one project-work day where participants prepared a presentation on their AI product (Table 1).

### **3.2. Participants**

Administrative staff were recruited from K–12 schools, small to medium-sized enterprises (SMEs), and universities in Hong Kong using convenience sampling. We emailed administrative staff of all Hong Kong schools and used available channels to contact SMEs and our university for participant recruitment. Interested individuals submitted an online application alongside self-declaration and consent forms to enrol in the course. No programming experience was required.

A total of 112 administrative staff participated, including 41 men (37%) and 71 women (63%). The age distribution included 24 participants aged 20–30 (21%), 36 aged 31–40 (32%), 34 aged 41–50 (30%), 16 aged 51–60 (14%), and two aged over 60 (2%). Regarding educational attainment, 18 held a diploma or certificate (16%), 67 a bachelor's degree (60%), 26 a master's degree (23%), and one a doctoral degree (1%). In terms of background, 49 participants came from K–12 schools (44%), 37 from SMEs (33%), and 26 from universities (23%).

Table 1. Course content and data sources

Time	Teaching Mode	Topics	Surveys
Week 1	Face-to-face teaching	(1) Concepts of Generative AI (GenAI), data security and ethics (2) Techniques: prompt engineering, material generation (speech-to-text transcription & PowerPoint generation), chatbot creation	Written reflection (pre-course)
Week 2	Self-study	(1) Prompt engineering (2) AI regulatory policies	N/A
Week 3	Face-to-face teaching	(1) Concepts of machine learning and deep learning (2) Techniques: machine learning model building, material generation (MS Word, Excel, Copilot, Images), chatbot creation	N/A
Week 4	Project preparation	Creation of AI artefacts	N/A
Week 5	Face-to-face teaching Presentation	Learning about AI using robots Project presentation and peer assessment	(1) Written reflection (post-course) (2) Motivation survey (3) Acceptance survey (4) Course evaluation

### 3.3. Data Collection and Analysis

A mixed-methods approach combining qualitative and quantitative data collection and analysis was used. Pre-course, the participants wrote reflections (50–100 words) on their attitudes towards the AI course and their expectations. Post-course, they reflected on whether their expectations were met (50–100 words).

We adapted a motivational survey based on the ARCS model, consisting of 12 items rated on a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree) to assess the participants' views after the course. Additionally, an AI tools acceptance survey with 18 items rated on a 5-point Likert scale was designed, incorporating six constructs: perceived usefulness, perceived ease of use, attitude towards usage, behavioural intention, self-efficacy, and subjective norm (Chow et al., 2012; Compeau & Higgins, 1995; Davis, 1989; Rafique et al., 2020; Watson & Rockinson-Szapkiw, 2021). Finally, a nine-item course evaluation survey was administered, including six items rated on a 5-point Likert scale and three open-ended questions, to obtain detailed participant feedback on the most beneficial content and areas for improvement.

To address the first research question, descriptive data from the motivation survey were analysed. Qualitative and quantitative data were used to answer the second research question. Descriptive data from the acceptance survey were examined to determine course satisfaction, while a thematic analysis of pre- and post-course reflective writings was conducted to understand participant expectations and learning outcomes. The researchers categorised each participant's reflections based on whether their expectations were met as follows: (1) failed (0 marks): not satisfied and perceived the course as unhelpful; (2) partially satisfied (0.5 marks); (3) satisfied (0.75 marks): happy with unexpected gains despite not meeting initial expectations; (4) fully satisfied (1 mark): fully met expectations and gained additional knowledge.

## 4. Results

### 4.1. Motivation

The PBL for Motivation Survey had high reliability (Cronbach's alpha = 0.967) (Table 2). The Attention component had the highest mean of 4.33 (SD = 0.76), indicating that most responses were positive, with some variability. Satisfaction followed closely, with a mean of 4.32 (SD = 0.78), suggesting that the participants were generally satisfied with the course. Relevance had a mean of 4.05 (SD = 0.83), indicating that the participants found the course relevant to their professional development. Confidence received the lowest mean of 3.99 (SD = 0.82), indicating room for improvement in building confidence in using AI at work. Overall, the data reflect a positive perception of all course aspects among the participants.

Table 2. Descriptive statistics for motivation survey variables.

Item	Min	Max	M	SD
Attention	1	5	4.33	0.76
Relevance	1	5	4.05	0.83
Confidence	1	5	3.99	0.82
Satisfaction	1	5	4.32	0.78

#### 4.2. Acceptance

The GenAI Tools Acceptance Survey also had high reliability (Cronbach's alpha = 0.960), with reliability above 0.850 for all six variables. Overall, the survey indicated strong acceptance and positive attitudes towards GenAI tools. Behavioural intention had the highest mean ( $M = 4.15$ ,  $SD = 0.83$ ), suggesting that the participants were likely to adopt and continue using GenAI tools in their activities. Perceived usefulness followed closely ( $M = 4.13$ ,  $SD = 0.76$ ), indicating that the participants felt that these tools effectively supported their tasks. They showed positive attitudes towards using GenAI tools at work ( $M = 4.04$ ,  $SD = 0.76$ ). The mean of self-efficacy was 3.88 ( $SD = 0.85$ ), indicating that the participants generally felt confident in their ability to use GenAI tools. The mean of subjective norm was 3.76 ( $SD = 0.90$ ), reflecting the participants' perceptions of social pressure or expectations from peers and authorities regarding the use of GenAI tools. Perceived ease of use received the lowest score ( $M = 3.40$ ,  $SD = 0.84$ ), highlighting that although the participants recognised the usefulness of GenAI tools, they found them difficult to use, indicating a need for further professional development.

#### 4.3. Reflection

A qualitative thematic analysis was performed on the reflective writings of the participants, examining their pre- and post-course insights. Initially, 10 codes grouped under four themes were identified after reviewing all reflective writings. Post-course, two new themes and one new code (in *italics* in Table 3) emerged, expanding the classification to 6 themes and 13 codes.

The course had a substantial impact on the participants' perceptions and intentions to use AI tools. Initially, 53.92% of the participants wanted to enhance their work and personal lives by focusing on efficiency (31.34%), quality (12.90%), and creativity (5.53%), expecting AI tools to automate tasks, optimise workflows, and reduce skills barriers (e.g., 'I hope to improve my work efficiency by using at least the simplest AI tools to minimise repetitive and labour-intensive tasks to save time for more important and strategic tasks'). Post-course reflections showed a decline, with only 27.82% of the participants viewing AI as beneficial for work or life, while mentions of efficiency dropped to 16.54%, highlighting application challenges (e.g., 'AI tools are useful for productivity, but practical implementation is complicated'). Mentions of quality and creativity also declined, with the participants recognising the need for a nuanced understanding of AI's capabilities and limitations.

Table 3. Comparison of pre- and post-course reflective writings.

Theme	Code	Pre-Course		Post-Course	
		Frequency	Percentage	Frequency	Percentage
1. Better Work or Life	Total	117	53.92%	74	27.82%
	Efficiency	68	31.34%	44	16.54%
	Quality	28	12.90%	14	5.26%

	Creativity	12	5.53%	12	4.51%
	Management	5	2.30%	2	0.75%
	Information Source	4	1.84%	2	0.75%
2. Self-development	Total	77	35.48%	99	37.22%
	Skills	27	12.44%	23	8.65%
	Understanding	26	11.98%	51	19.17%
	Competency	24	11.06%	11	4.14%
	<i>Confidence &amp; Planning</i>			14	5.26%
3. Sharing with Others		7	3.23%	25	9.40%
4. Thoughts about AI		16	7.37%	21	7.89%
5. Course Suggestions				9	3.38%
6. Challenges & Considerations				38	14.29%
<b>Total</b>		<b>217</b>	<b>100%</b>	<b>266</b>	<b>100%</b>

*Note.* The frequencies reflect the number of codes, as the participants could refer to multiple aspects in their responses, leading to overlapping frequencies and percentages across different codes and themes.

Mentions of self-development rose from 35.48% to 37.22%, and understanding of AI concepts increased from 11.98% to 19.17%, indicating a deeper conceptual grasp (e.g., ‘Learning about GenAI tools was enlightening, enhancing my understanding of AI’s potential’). However, mentions of acquiring new skills decreased slightly to 8.65%, suggesting that the participants valued theoretical knowledge over immediate skill acquisition.

Confidence and planning to use and learn about AI tools were mentioned in 5.26% of the post-course reflections. Several participants wanted to stay up to date with AI trends (e.g., ‘Taking an AI course was enlightening... It opened my eyes to the ethical implications and societal impact of AI technologies’). This evolution indicates that although the course provided fundamental knowledge, ongoing support is essential to translate learning into practical applications.

Interest in sharing AI knowledge with others increased, from 3.23% pre-course to 9.40% post-course. The participants wanted to educate their colleagues and family members (e.g., ‘I’m thrilled to teach others about AI’s transformative potential’). Thoughts about AI also increased slightly, from 7.37% to 7.89%, highlighting ongoing discussions about AI’s role and ethical implications in broader contexts.

The participants faced challenges in applying AI tools, with 14.29% expressing concerns post-course, such as unsatisfactory results, complex implementation processes, and ethical considerations. One participant highlighted a major issue: ‘Senior management scepticism limits access to AI platforms, hindering practical use’. This underscores the importance of promoting AI literacy at all organisational levels.

Participant feedback on course improvement suggested adjustments to the schedule and a reduction in theoretical content. Some requested advanced modules to deepen their understanding, while others hoped for executive participation to facilitate organisational adoption of AI tools. As one participant suggested, ‘Promoting AI literacy should start at the executive level’.

Overall, the course effectively enhanced the participants’ understanding of AI, as evidenced by their increased focus on understanding rather than immediate practical application. However, the decline in perceived work improvement highlights the need for continued support to translate AI knowledge into practice. Addressing these challenges in future iterations of the course will enhance its impact, enabling administrative staff to better navigate the complexities of AI at work.

To evaluate changes in the participants’ reflections, a scoring system was used (2 = significant improvement, 1 = some improvement, 0 = no improvement). The statistical average of these scores was 0.86. The score distribution showed that 4% of the participants (4 responses) demonstrated no improvement, 18% (20 responses) showed some improvement, and 74% (83 responses) exhibited significant improvement.

These results underscore the effectiveness of the AI literacy course in enhancing participants' understanding of AI and its potential applications. However, they also highlight the need for additional support and resources to facilitate the practical implementation of AI tools at work. Addressing these challenges in future courses could enhance the impact of AI literacy training for administrative staff.

#### **4.4. Evaluation**

The evaluation survey had high reliability (Cronbach's alpha = 0.916). Items 1–6 aimed to provide a macroscopic view of the participants' experiences. Their descriptive statistics are presented in Table 4. Items 7–9 used open-ended questions to further explore the participants' perspectives on the most and least useful content, as well as suggestions for improvement.

The most useful content was categorised into three main areas: (1) Underlying Concepts, such as machine learning and neural networks, with 23% of the participants finding that this content provided a clearer understanding of fundamental AI concepts; (2) Impact and Evolution, which included ethics, trends, and future directions of AI and their professional development, with 16% of the participants mentioning that this content broadened their horizons in the field; and (3) Tool Use, which encompassed introductions of new tools, available resources, and hands-on practice sessions.

The participants also identified several areas for improvement, such as some content perceived as too difficult (33%), a tight schedule during the teaching process (29%), and challenges in applying AI knowledge to real-world situations (4%). Many participants wanted more support, such as a platform for continued learning and communication or advanced courses offering deeper knowledge and up-to-date information. These suggestions aligned with those mentioned in previous reflections before the course.

Table 4. Descriptive analysis of items.

Item	M	SD
1. I understand GenAI better after attending the course.	4.48	0.60
2. The hands-on activities in the workshop helped me better understand GenAI.	4.48	0.61
3. I like the blended learning mode of this course (self-study and workshop participation).	4.38	0.69
4. Overall, the course is worth taking.	4.50	0.64
5. Overall, the course is well organised.	4.36	0.67
6. I will recommend this course to my colleagues.	4.36	0.73

#### **5. Conclusions and Discussion**

This study highlights the positive impact of an AI literacy course on administrative staff, showcasing both motivations and evolving perceptions. The participants were primarily motivated by attention and satisfaction, finding the course content engaging and relevant. Their acceptance of AI tools was high, as they recognised their usefulness in enhancing productivity. Yet, challenges with ease of use indicated the need for user-friendly tools and comprehensive training. Addressing these aspects is essential to promote wider AI adoption at work.

The participants' self-reflections revealed a shift from immediate improvements in work-life balance to a deeper understanding of AI concepts. Although the course provided fundamental AI knowledge, translating it into practical skills remains a challenge. This underscores the need for ongoing support to help participants apply their learning effectively. Additionally, the participants showed growing interest in sharing AI knowledge, suggesting potential for broader community impact. Encouraging this could extend the course's benefits to other professional networks.

Course feedback indicated that while some content was enlightening, it was sometimes too advanced. Future courses should offer tailored content, including field-specific modules and case studies, to better align with professional contexts.

In conclusion, although the course enhanced AI understanding, addressing content customisation and practical application challenges could increase its effectiveness. Conducting a needs analysis before course development would

ensure alignment with participants' specific needs, maximising its impact. Future research should explore tailored AI literacy programmes and evaluate long-term outcomes to refine and enhance training approaches.

### Acknowledgements

The authors acknowledge financial support for this project from the Li Ka Shing Foundation and the Research Grants Council, University Grants Committee of the Hong Kong Special Administrative Region, China (Project No. EdUHK CB326, CB357).

### References

- Ahmad, S. F., Alam, M. M., Rahmat, M. K., Mubarik, M. S., & Hyder, S. I. (2022). Academic and administrative role of artificial intelligence in education. *Sustainability*, 14(3), 1101.
- Casal-Otero, L., Catala, A., Fernández-Morante, C., Taboada, M., Cebreiro, B., & Barro, S. (2023). AI literacy in K-12: A systematic literature review. *International Journal of STEM Education*, 10(1), 29.
- Chang, Y. S., Hu, K. J., Chiang, C. W., & Lugmayr, A. (2019). Applying mobile augmented reality (AR) to teach interior design students in layout plans: Evaluation of learning effectiveness based on the ARCS model of learning motivation theory. *Sensors*, 20(1), 105.
- Choi, J., Lee, J. H., & Kim, B. (2019). How does learner-centered education affect teacher self-efficacy? The case of project-based learning in Korea. *Teaching and Teacher Education*, 85, 45–57.
- Chow, M., Herold, D. K., Choo, T. M., & Chan, K. (2012). Extending the technology acceptance model to explore the intention to use Second Life for enhancing healthcare education. *Computers & Education*, 59(4), 1136–1144.
- Compeau, D. R., & Higgins, C. A. (1995). Computer self-efficacy: Development of a measure and initial test. *MIS Quarterly*, 19(2), 189–211.
- 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>
- Keller, J. M. (1987). Development and use of the ARCS model of instructional design. *Journal of Instructional Development*, 10(3), 2–10
- Kong, S. C., Cheung, M. Y. W., & Tsang, O. (2024). Developing an artificial intelligence literacy framework: Evaluation of a literacy course for senior secondary students using a project-based learning approach. *Computers and Education: Artificial Intelligence*, 6, 100214. <https://doi.org/10.1016/j.caai.2024.100214>
- Kong, S. C., Cheung, M. Y. W., & Zhang, G. (2021). Evaluation of an artificial intelligence literacy course for university students with diverse study backgrounds. *Computers & Education: Artificial Intelligence*, 2, 100026. <https://doi.org/10.1016/j.caai.2021.100026>
- Kong, S. C., Cheung, W. M. Y., & Zhang, G. (2023). Evaluating an artificial intelligence literacy programme for developing university students' conceptual understanding, literacy, empowerment and ethical awareness. *Educational Technology & Society*, 26(1), 16–30. [https://doi.org/10.30191/ETS.202301\\_26\(1\).0002](https://doi.org/10.30191/ETS.202301_26(1).0002)
- Laupichler, M. C., Aster, A., & Raupach, T. (2023). Delphi study for the development and preliminary validation of an item set for the assessment of non-experts' AI literacy. *Computers & Education: Artificial Intelligence*, 4, 100126. <https://doi.org/10.1016/j.caai.2023.100126>
- Li, K., & Keller, J. M. (2018). Use of the ARCS model in education: A literature review. *Computers & Education*, 122, 54–62.
- Manyika, J., Lund, S., Chui, M., Bughin, J., Woetzel, J., Batra, P., ... & Sanghvi, S. (2017). Jobs lost, jobs gained: Workforce transitions in a time of automation. *McKinsey Global Institute*, 150(1), 1–148.
- McKinsey & Co. (2023). *The state of AI in 2023: Generative AI's breakout year*. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year>

- 
- McKinsey & Co. (2024). *A new future of work: The race to deploy AI and raise skills in Europe and beyond.* <https://www.mckinsey.com/mgi/our-research/a-new-future-of-work-the-race-to-deploy-ai-and-raise-skills-in-europe-and-beyond>
- Rafique, H., Almagrabi, A. O., Shamim, A., Anwar, F., & Bashir, A. K. (2020). Investigating the acceptance of mobile library applications with an extended technology acceptance model (TAM). *Computers and Education*, 145, 103732. <https://doi.org/10.1016/j.compedu.2019.103732>
- Stanford University. (2024). *Artificial Intelligence Index Report 2024*. <https://aiindex.stanford.edu/report/>
- Watson, J., & Rockinson-Szapkiw, A. (2021). Predicting preservice teachers' intention to use technology-enabled learning. *Computers & Education*, 168, 104207. <https://doi.org/10.1016/j.compedu.2021.104207>
- Zhang, L., & Ma, Y. (2023). A study of the impact of project-based learning on student learning effects: A meta-analysis study. *Frontiers in Psychology*, 14, 1202728.

# Exploring the Mechanisms for Implementing AR-based Learning Activities through the Lens of Motivational Design

Yun Wen\*, Sujin He

1 National Institute of Education, Nanyang Technological University, Singapore

\* yun.wen@nie.edu.sg

**Abstract:** The application of Augmented Reality (AR) in science and its potential have been widely recognized. Research generally suggests that AR can enhance students' motivation in learning, though findings on whether AR improves academic performance remain inconsistent. This mixed methods quasi-experimental study, which examined the use of a developed AR-assisted learning app in science classrooms, confirmed the critical role that teachers play in implementing AR-enhanced inquiry activities. Additionally, the study analyzed teaching mechanisms for the effective use of AR-enabled learning, offering specific suggestions for AR instructional designers and teachers to effectively design and implement AR activities into teaching practices.

**Keywords:** Augmented Reality, ARCS, Motivation, Primary School, Science

## 1. Introduction

The past decade has seen a development in the number of Augmented Reality (AR) applications being used in education. Although studies have shown the benefits in using AR for students (Chang et al., 2023), there are certain implications that educators should be mindful of. The effective use of AR also requires the adoption of suitable pedagogical approaches and teaching strategies (Garzón et al., 2020; Wen et al., 2023). Pedagogical models that provide practical strategies for designing teaching processes and resources, such as Keller's ARCS model (1987), have been used to promote learners' engagement in AR-supported learning environments (e.g., Hao, 2023; Li et al., 2023). While teachers envision the advantages in adopting AR-enhanced learning, they may lack the necessary competencies for AR implementation (Nikou et al., 2023). Furthermore, managing students are challenges that teachers face, and if not given adequate support, may result in the teacher feeling overwhelmed and not able to fully capitalise on AR's affordances, ultimately missing out on potential teaching opportunities. However, few studies focus on understanding the prerequisites for teachers or specific mechanisms to implement AR activities effectively in classrooms. Therefore, the purpose of this study is to investigate which specific real-time mechanisms employed by teachers in implementing AR-supported learning activities may influence students' learning motivation and their learning performance.

## 2. Literature Review

This study is grounded in Keller's ARCS model (Attention, Relevance, Confidence, Satisfaction) (Keller, 1987). It has been effectively applied to technology-enhanced learning environments, serving the needs of educational research and practice (Ma & Lee, 2021). Motivation has been identified as a crucial factor in learning that affects students' academic achievement and emotion (Shapiro et al., 2017), so motivation is an important component from which instructional design can be understood. There are also other motivational design models; however, earlier models often concentrate on specific motivational characteristics, such as achievement motivation. In contrast, the ARCS model takes a more holistic approach to designing motivating learning environments.

Keller's model (1987), comprising four components of attention, relevance, confidence, and satisfaction, demonstrates a motivational design strategy by first capturing and sustaining learners' attention, while triggering their curiosity. Next, relevance requires learning activities to be aligned with the individual goals and needs of learners for it

to be perceived as meaningful. The model also emphasized that learners' confidence can be fostered by their sense of control and their expectations of success, as it impacts the effort, they are willing to invest in the activities. Lastly, learners' completion of learning goals will impact their level of satisfaction and cultivate a sense of reward and pride (Li & Keller, 2018). Keller outlined these four major components that need to be present for learners to feel motivated and sustain that motivation.

Studies have shown that the use of AR technology in educational settings can help enhance motivation (e.g. Silva et al., 2023). For instance, as reported in the systematic review by Garzón et al. (2019), motivation was the second most frequently reported benefit of AR use in classrooms. The overlay of 3D virtual elements in a real-world setting provides an immersive quality that captures learners' attention, creating a learning environment that can enhance motivation (Xu et al., 2022). Moreover, the contextualized scenarios provided by AR technology can bridge the gap between the theoretical content and real-world application (Lin et al., 2023). The AR elements can provide authentic learning scenarios which help students see the relevance of their learning. This can engage students further, motivating them through a sense of purpose and meaning. Although a growing number of studies have paid attention to the design of AR learning environments guided by ARCS model (e.g., Hao & Lee, 2021; Laurens-Arredondo, 2022), there is scarcity of research on teachers' implementation strategies that reflect the ARCS model in AR environments to ensure the effectiveness of design.

1. What are the effects of AR-based learning approach guided by ARCS design on students' academic performance and learning motivation?
2. What mechanisms used by teachers during the implementation of AR-based learning may affect students' learning motivation?

### **3. Methods**

#### **3.1. Participants**

This study was carried out in three Singapore primary schools on the topic of digestion. It involved grade 4 students (10 – 12 years) from 10 classes, along with their science teachers. The duration for the study was for a period of over three weeks. The learning objectives for all the classes were similar, based on the curriculum. However, 6 classes from the 3 schools used the AR app “The Doctor’s Digest” to learn about digestion, while 4 control classes from the same schools did not.

For the experimental classes, the teachers used the AR Doctor’s digest app together with AR designed worksheets. Additionally, they also covered worksheets from the school’s official Textbook. Each lesson began with a brief introduction to the Digestive system and a demonstration of how to use the AR app. Students then used the ipads to complete the AR activities and worked in groups of three or four. For the control classes, students were taught using traditional teaching methods without the use of AR. The control class lessons included teacher-led lectures and teacher-led class discussions. Teachers presented the concepts using visual aids such as PowerPoint slides and showed videos to complement their lectures. Both the experimental and control classes completed the same worksheets from the school’s official Textbook.

#### **3.2. AR-based learning guided by ARCS design on the topic of digestive system**

The app’s storyline is designed for students to assume the identity of a doctor. The task is to diagnose and treat patients who are facing various digestive problems. This interactive approach aims to introduce students about the digestive system and related health issues in authentic real-life scenarios. The app comprises four activities – “Training simulation”, “What happens?”, “Case Files” and “Create an alien”. Figure 1a shows the interface of the “what happens” activity, and Figure 1b shows the teacher helping the group take a photo with the AR alien they created in the last activity. All the AR learning activities are aligned with students’ science syllabus. A teacher’s dashboard, designed alongside with the AR app, enables teachers to monitor students’ group work progress and performance. The dashboard can also support students in monitoring and reflecting on their group’s progress.

“Training simulation” gives an overview of the digestive system, where students have to position the different digestive organs (mouth, gullet, stomach, small intestine, large intestine, anus) correctly, in order to activate a simulation of how digestion occurs. For “What happens?”, students are given situations such as “What do you think will happen to the digestive system if the small intestine is too short.” Students need to apply what they have learnt and give reasons for their explanations. In “Case Files”, students need to diagnose digestive issues that the patients are experiencing, based on the symptoms presented in the case files. They will observe a simulation of the digestive problem, apply their knowledge, and select the appropriate remedy from the provided options for the specific digestive issue. The final activity, “Create an Alien”, is meant to consolidate and provide a synthesis of the digestion concepts taught. In this activity, students are given hypothetical scenarios in imaginary planets (e.g., climate with little rainfall) where the aliens have access to only certain types of food (e.g., only vegetables). Based on these conditions, they are tasked with designing a digestive system that is specifically adapted to the alien’s environment and available food sources (e.g., sharp teeth, long gullet).



(a) students reviewing the simulation



(b) students taking a photo with the alien they created

Fig.1 Students working on the “what happens” and “create an alien” activity in groups

### 3.3. Data Collection

To investigate students’ learning performance, pre-and post-tests were administered before and after the intervention. There were three multiple-choice questions (6 marks) and two open-ended questions (6 marks) for the pre-test, with a total score of 12 marks. For the post-test, the questions were different but of a similar difficulty level. To explore students’ learning motivation, a motivation survey questionnaire was adapted and rewritten to fit the learning content of this study, based on the measurement scale used by Hao and Lee (2021). The questionnaire consisted of 18 items in total. All the students from the 10 classes were invited to complete both the pre- and post-surveys, as well as the pre- and post-tests. After cleaning the raw data, 287 valid responses were keyed into the SPSS.

The reliability of the questionnaire was tested and showed that the questionnaire and its four subscales of attention (Cronbach’s  $\alpha = .91$ ), relevance (Cronbach’s  $\alpha = .83$ ), confidence (Cronbach’s  $\alpha = .84$ ), and satisfaction (Cronbach’s  $\alpha = .91$ ) were reliable in terms of internal consistency. Then, the data were subjected to a confirmatory factor analysis (CFA) to test the construct validity of responses to the questionnaire. The values of composite reliability (CR) for the four subscales of the questionnaire were acceptable with values greater than 0.7 (Fornell & Larcker, 1981), and the convergent validity (AVE) was acceptable with values greater than 0.5 (Peterson, 2000). A satisfactory model fit was obtained ( $\chi^2(126) = 256.6$ ,  $p < .001$ , CFI = .968, RMSEA = .059, SRMR = .032) (Hair et al., 2010).

We also video recorded all the AR classes to analyse how teachers conducted the AR lessons. During the classroom interventions, the research team set up 2 video cameras to collect classroom observation data. Focus-group students’ interviews were conducted after all the AR lessons.

### 3.4. Data Analysis

The data was analysed in two phases. In the first phase statistical analysis was done to investigate the effect of the AR-based learning approach. In the next phase, the case classes were chosen based on their learning performance, and content analysis was conducted to investigate teachers' instructional events. A one-way factorial ANCOVA analysis was conducted to examine the academic differences between the AR and non-AR groups by taking the students' pre-test results as a covariate variable. To understand students' perceptions towards motivational design upon using different teaching approaches, MANCOVA was conducted to examine motivations from four aspects. The pre-survey results served as the covariant to eliminate the differences in students' motivation.

Content analysis was conducted to analyse the instructional events by systematically categorizing code and identification of themes (Hsieh & Shannon, 2005). Content analysis was conducted in the order of (1) identifying all class-level instructional events, (2) categorising the mechanisms reflected instructional motivation. The coding was done with the Behavioural Observation Research Interactive Software (Boris).

Based on students' learning performance, two case classes were selected to investigate how the teachers implemented the AR lessons. The teacher's class level instructions were first segmented by themes for further analysis. Teachers' discourse within small groups were excluded. The class-level instructions were segmented in terms of different themes, for instance, the instructional events for introducing learning objective, providing feedback etc. The unit of analysis of this study refers to every meaningful teacher-to-class instructional event. In the second phase of video-based content analysis, the case classes' instructional events were explored in detail to detect enactment mechanisms for promoting learning motivation (Tabel 1). The second author helped code the video data in terms of the coding scheme. To check coding reliability, the first author coded the data separately, and the interrater reliability was over 0.92 (Cohen's kappa). Additionally, students' post-interview data were cited to triangulate the findings of content analysis.

Table 1. Coding scheme for mechanisms to stimulate motivation

ARCS Dimensions	Observed Strategies
Attention	A1. Initiating the topic by asking questions A2. Introducing AR functions A3. Introducing AR activities
Relevance	R1. Introducing learning objectives R2. Explaining the linkage between AR tasks and worksheets
Confidence	C1. Providing details on expectations of activities and role assignment
Satisfaction	S1. Providing feedback about group work progress S2. Providing feedback about students' understanding S3. Providing feedback about the group of students

#### 4.Findings

##### 4.1. *The effects of the designed AR-based learning environment on students' academic performance*

As shown in Table 2, the ANCOVA result showed a nonsignificant result  $F(2, 280) = 2.53, p = .111, \eta^2 = .009$ , indicating that no differences on academic performance were found between the AR and non-AR approaches. Nevertheless, when we paid attention to the pre-test and post-test performance of each AR class, we found the class's performance were diverse (see Table 3). In this study, we selected AR2 and AR4 classes as case classes as their post-test scores were the highest and lowest respectively after the intervention.

Table 2. the one-way ANCOVA result of the academic performance of the two conditions

Group	N	Adjusted Mean	Adjusted SD	F	P	$\eta^2$
AR Classes	154	7.35	.24	2.53	.112	.009
Non-AR Classes	133	6.79	.26			

Table 3. Students' academic performance in each AR class

Class	N	Pre-test		Post-test		<i>t</i>	<i>p</i>
		Mean	SD	Mean	SD		
School A_AR1	31	4.60	2.01	8.73	1.78	8.33	<.001
School A_AR2	34	6.79	3.30	9.79	1.67	5.40	<.001
School B_AR3	37	3.27	2.06	6.10	2.91	5.85	<.001
School B_AR4	20	2.30	1.45	3.40	2.67	2.05	.055
School C_AR5	21	4.85	2.48	8.29	2.10	5.93	<.001
School C_AR6	26	4.62	1.92	6.27	2.82	2.66	.014

#### 4.2. The effects of the designed AR-based learning environment on students' learning motivation

The group means on attention, relevance, confidence, and satisfaction are shown in Table 4. The Levene's tests of equality of error variances were not significant, indicating that the assumption of equal error variance of the dependent variables across groups was met. The multivariate result for the effect of the AR approach showed statistical significance,  $F(4, 278) = 2.94$ ,  $p = .021$ ,  $\eta^2 = .041$ , indicating that students from the AR classes showed higher learning motivation compared to those from the non-AR classes. Univariate tests showed an effect of the AR approach on students' attention perception,  $F(1, 281) = 5.38$ ,  $p = .021$ ,  $\eta^2 = .019$  (see Table 5). Students in the AR classes reported to have higher attention perception than those in the non-AR classes ( $p < 0.05$ ). However, the AR approach did not impact on students' perceptions of relevance, confidence, and satisfaction significantly.

Table 4. Descriptive findings of students' learning motivation

Motivation	AR (n = 154)		Non-AR (n = 133)	
Subscales	Mean	SD	Mean	SD
Attention	3.92	.94	3.67	.92
Relevance	3.96	.85	3.93	.77
Confidence	3.85	.94	3.73	.94
Satisfaction	4.13	.87	4.02	.82

Table 5. MANCOVA summary of students' motivation

Motivation Subscales	<i>F</i> (1, 281)	<i>P</i> Value	Partial $\eta^2$
Attention	5.38	.021	.019
Relevance	.07	.79	.000
Confidence	.56	.46	.002
Satisfaction	.73	.38	.003

In the post-interview, students from the AR groups all shared that they were particularly drawn to the realistic simulations and interactive elements in the AR app. These features likely captured their attention more effectively than traditional teaching methods due to their immersive and hands-on qualities, allowing students to explore and manipulate the content independently rather than merely observing a teacher's demonstration.

Table 6. Descriptive findings of students' learning motivation about the two cases

Motivation	School A_AR2		School B_AR4	
Subscales	Pre	Post	Pre	Post
Attention	4.04 (.68)	4.35 (.62)	3.75 (.67)	3.48 (1.10)
Relevance	4.46 (.45)	4.44 (.53)	3.78 (.62)	3.53 (.87)
Confidence	3.93 (.80)	4.23 (.73)	3.73 (.67)	3.45 (1.03)
Satisfaction	4.51 (.43)	4.65 (.47)	4.03 (.65)	3.68 (.83)

As shown in Table 6, we further examined the learning motivation data from the two case classes and observed improvements across all four motivation subscales in the AR2 class. However, no significant improvements were detected in the AR4 class, in which students' pre- and post-test scores had no significant improvement.

#### 4.3. Observed implementation mechanisms in the case classes

Based on the results of the content analysis, we compared the strategies employed by teachers while enacting AR lessons (see Table 7). A notable difference was observed in the approach taken by the teacher of the AR2 class, who engaged students in the digestive system topic at the beginning of all activities by asking questions (14min 42sec). Aligning with inquiry-based learning, this strategy effectively sparked students' interest and prompted them to generate their own set of questions related to the topic. In terms of attracting students' attention, another key difference was that the AR2 teacher spent time thoroughly explaining the features of the AR app including the user interface and the function of buttons (8min 26sec), rather than only explaining the design of AR activities (6min). Regarding learning objectives, the AR2 teacher did not repeatedly highlight them (1mins28sec), as the AR4 teacher did (6min 26sec), but both teachers reminded students of the connection between AR app activities and the worksheet.

During the activities, the AR2 teacher occasionally encouraged students to collaborate with one another (18sec). More importantly, the teacher used the dashboard to collect students' responses to the open-ended questions and provide content-related feedback accordingly (33min 44sec). For example, the AR2 teacher organized and projected all groups' responses and commented on the completeness of their answers, thereby promoting a better understanding and application of targeted concepts. However, because such summarization took time, she did not provide this feedback at the end of each lesson, but at the beginning of the following AR lesson, to address issues identified based on students' inputs.

In contrast, this strategy was less observed in the AR4 class (6min 50sec). The AR4 teacher tended to provide content-related summaries and feedback at the end of each activity. The AR4 teacher frequently used the dashboard to monitor group progress during the activity (13min 15sec). He occasionally praised faster groups and reminded slower groups to catch up, offering assistance when necessary. The AR2 teacher, however, did not intentionally praise specific groups for their progress. Furthermore, during AR activities, the AR2 teacher chose not to project her teacher dashboard onto the screen for students, contrary to our suggestion. She explained that her students were typically highly engaged in learning activities, and projecting group progress might unnecessarily heighten their sense of competition.

Table 7. a comparison of used strategies for promoting students' learning motivation in the two case classes

ARCS dimensions	Observed Strategies of Enactment	School A_AR2		School B_AR4	
		f	Length (mm:ss)	f	Length (mm:ss)
Attention	A1. Initiating the topic by asking questions	1	14:42	0	0
	A2. Introducing AR functions	4	8:26	5	5:16
	A3. Introducing AR activities	4	6:00	8	13:51
Relevance	R1. Introducing learning objectives	2	1:28	3	6:26
	R2. Explaining the link between AR tasks and worksheets	6	4:19	2	3:32
Confidence	C1. Providing details on expectations of activities and role assignment	2	0:18	0	0
Satisfaction	S1. Providing feedback about group work progress	3	2:39	19	13:15
	S2. Providing feedback about students' understanding	7	33:34	2	6:50
	S3. Providing feedback about the group of students	0	0	2	1:36

The interview data from the two classes also highlighted obvious differences in their experiences during the AR lessons. These differences were mainly reflected in two aspects: (1) familiarity with the AR app and (2) students' reflection on their group collaboration. However, students from both classes were highly consistent in their responses regarding whether the AR activities were engaging compared to teacher-led lessons. They all agreed that the visualization and hands-on approach were more interesting.

AR4 students mentioned that they sometimes struggled with operating the app and expressed dissatisfaction with their group's performance during the activities. Students from AR4 Group 7 shared that they faced challenges during the

first AR lesson. Although the teacher showed them how to use the AR app, they still did not fully understand it until the second AR lesson, when the teacher used the PowerPoint slides to guide them and explain the functions. AR4 Group 2 students noted that while the teacher's explanation helped them understand how to carry out the AR activities, their bigger challenge was not knowing how to collaborate. This hindered the smooth progress of their activities, making the experience less enjoyable. AR4\_G2S3 said "The first activity was quite fun until everyone started to get angry and didn't get along with each other." AR4\_G2S1 added "in our training simulation (the first activity), everyone started going out of control. We have no teamwork." AR4\_G2S3 further explained "Everyone got different answers. We don't know who has the correct answer, the problem is still a problem and then we don't know what to do, how are we supposed to know." Regarding the usefulness of the dashboard, AR4 Group 2 mentioned that, initially, we paid attention to it and noticed their team was working somewhat slowly. After a few AR lessons, they noticed other groups had improved, but their own team had not. This left them feeling frustrated.

On the contrary, when asked their feedback on group work. AR2\_G3S1 said "quite good, we build team bonding. It was quite difficult to handle the iPad and the pieces of paper together so in a group it would be easier". When asked how they resolved differences in opinions, they added that "... just try out both ways and see which one is more correct." AR2 Group 4 students also agreed that working in a group helped them learn better.

## 5. Discussion

There was no significant difference in academic performance between AR and non-AR classes across the 10 classes analyzed. However, from the perspective of learning motivation, AR classes showed significant improvement in the attention dimension. In some AR classes, both students' learning motivation and academic performance showed significant improvement, attributed to the effective use of AR. Our data suggest that these outcomes are closely related to how teachers implemented AR activities. Based on the results from the comparative cases, we identified three implementation mechanisms: (1) Thoroughly introducing AR functions at the start of lessons; (2) Tailoring feedback on regulation based on students' readiness to pedagogical approaches; (3) Using AR activities to promote a new cycle of inquiry.

First, when introducing AR apps into the class, especially for the first time, it is essential to allocate sufficient time to explain its features comprehensively. While AR with 3D simulations can naturally capture students' attention and boost their motivation, the novelty effect may diminish as students encounter challenges in continuing the activities. As AR environments are relatively unfamiliar to most students, teachers need to focus initially on explaining AR's functions rather than diving directly into the activities.

Second, feedback during AR activities should consider students' readiness for the pedagogical approach. In our study, classes with less observable improvement included students lacking collaborative learning skills. Socially shared regulation of learning is critical for students' collaborative learning engagement (Järvelä et al., 2020), and affects their cognitive, motivational, and emotional states (Edwards et al., 2024). Projecting dashboards to students may help them to monitor and reflect on their group work progress. Nevertheless, teachers should clearly articulate their expectations and take into account the specific characteristics of the pedagogical approach if it is unfamiliar to their students. Additionally, they need to spend additional time coordinating group progress. Teachers' feedback should be provided beyond reminding students' keep monitoring their group progress, but how to negotiate with one another and make consensus.

Last, in AR classes, teachers need to focus on identifying students' knowledge gaps through their performance in the activities and promoting deeper understanding via further inquiry, so as to improve student' satisfaction with learning effectively. This finding aligns with the motivational AR-based learning approach proposed in Li et al.'s study (2023), which highlights the importance of discussion between teachers and students in fostering students a sense of accomplishment and satisfaction. Overemphasizing feedback on students' work progress while neglecting content-related feedback may leave students feel struggled or unsupported. Studies (Carless & Winstone, 2023) have

emphasized the importance of avoiding delays in providing post-task feedback. However, this does not mean that such feedback must always be provided within the same lesson. Feedback can also be provided in subsequent lessons, as ICT tools often record students' inputs, enabling teachers to prepare more tailored feedback based on students' performance. Therefore, when integrating AR apps into classroom teaching, it is crucial to allocate sufficient time for completing activities and deepening discussion to ensure their effectiveness.

## 6. Conclusion

This study confirmed the positive impact of the AR-based learning approach, guided by the ARCS design, on students' academic performance and learning motivation in primary science classes. It highlighted the pivotal role of teachers in the implementation process and identified three key mechanisms to enhance student motivation. However, the study also has limitations. The most significant limitation lies in the need to account for differences in class contexts. The mechanisms we identified are based on two comparative cases. However, the students of these classes differed notably in their academic performance levels and familiarity with collaborative learning before the intervention. The intent of this comparison is not to evaluate teachers' enactment but uncover potentially effective approaches and shed light on challenges that can be addressed. Future research could take class contexts into account to further validate and refine these mechanisms.

## References

- Carless, D., & Winstone, N. (2023). Teacher feedback literacy and its interplay with student feedback literacy. *Teaching in Higher Education*, 28(1), 150-163.
- Chang, C. Y., Kuo, H. C., & Du, Z. (2023). The role of digital literacy in augmented, virtual, and mixed reality in popular science education: a review study and an educational framework development. *Virtual Reality*, 27(3), 2461-2479.
- Edwards, J., Nguyen, A., Lämsä, J., Sobociński, M., Whitehead, R., Dang, B., ... & Järvelä, S. (2024). Human-AI collaboration: Designing artificial agents to facilitate socially shared regulation among learners. *British Journal of Educational Technology*.
- Garzón, J., Baldiris, S., Gutiérrez, J., & Pavón, J. (2020). How do pedagogical approaches affect the impact of augmented reality on education? A meta-analysis and research synthesis. *Educational Research Review*, 31, 100334.
- Garzón, J., Pavón, J., & Baldiris, S. (2019). Systematic review and meta-analysis of augmented reality in educational settings. *Virtual Reality*, 23(4), 447-459.
- Hair, J. F. J., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis: A Global perspective* (7th ed.). Upper Saddle River, NJ: Pearson Education.
- Hao, K. C. (2023). Creating a DGBL integrating ARCS motivation theory with animation, narrative story, fun, and usability to enhance learning motivation and effectiveness. *Interactive Learning Environments*, 31(9), 5698–5714.
- Hao, K. C., & Lee, L. C. (2021). The development and evaluation of an educational game integrating augmented reality, ARCS model, and types of games for English experiment learning: an analysis of learning. *Interactive Learning Environments*, 29(7), 1101–1114.
- Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative health research*, 15(9), 1277-1288.
- Järvenoja, H., Järvelä, S., & Malmberg, J. (2020). Supporting groups' emotion and motivation regulation during collaborative learning. *Learning and Instruction*, 70, 101090.
- Keller, J. M. (1987). Development and use of the ARCS model of instructional design. *Journal of Instructional Development*, 10(3), 2–10.

- Laurens-Arredondo, L. (2022). Mobile augmented reality adapted to the ARCS model of motivation: a case study during the COVID-19 pandemic. *Education and Information Technologies*, 27(6), 7927-7946.
- Li, K., & Keller, J. M. (2018). Use of the ARCS model in education: A literature review. *Computers & Education*, 122, 54-62.
- Li, M., Chen, Y. T., Huang, C. Q., Hwang, G. J., & Cukurova, M. (2023). From motivational experience to creative writing: A motivational AR-based learning approach to promoting Chinese writing performance and positive writing behaviours. *Computers & Education*, 202, 104844.
- Lin, X. F., Hwang, G. J., Wang, J., Zhou, Y., Li, W., Liu, J., & Liang, Z. M. (2023). Effects of a contextualised reflective mechanism-based augmented reality learning model on students' scientific inquiry learning performances, behavioural patterns, and higher order thinking. *Interactive Learning Environments*, 31(10), 6931-6951.
- Ma, L., & Lee, C. S. (2021). Evaluating the effectiveness of blended learning using the ARCS model. *Journal of Computer Assisted Learning*, 37(5), 1397–1408. <https://doi.org/10.1111/jcal.12579>
- Nikou, S. A., Perifanou, M., & Economides, A. A. (2023). Development and validation of the teachers' augmented reality competences (TARC) scale. *Journal of Computers in Education*.
- Shapiro, H. B., Lee, C. H., Roth, N. E. W., Li, K., Çetinkaya-Rundel, M., & Canelas, D. A. (2017). Understanding the massive open online course (MOOC) student experience: An examination of attitudes, motivations, and barriers. *Computers & Education*, 110, 35-50.
- Silva, M., Bermúdez, K., & Caro, K. (2023). Effect of an augmented reality app on academic achievement, motivation, and technology acceptance of university students of a chemistry course. *Computers & Education: X Reality*, 2, 100022.
- Wen, Y. Wu. L., He. S. J., Ng., H-E., Teo., B. C., Looi, C-K. & Cai, Y. (2023) Integrating Augmented Reality into Inquiry-based Learning Approach in Primary Science Classrooms, *Educational Technology Research & Development*. 71, 1631–1651.
- Xu, X., Kang, J., & Yan, L. (2022). Understanding embodied immersion in technology-enabled embodied learning environments. *Journal of Computer Assisted Learning*, 38(1), 103-119.

### Acknowledgments

This study is funded by MOE SSHRF 8/22 WY. The views expressed in this paper are the author's and do not necessarily represent the views of the host institution.

# A Self-regulated Learning Approach to Promoting Learners' Participation and Performance in Writing Education

Qianqin Liu<sup>1</sup>, Xinxiang Wang<sup>1</sup>, Yuting Chen<sup>2\*</sup>, Youlin Yang<sup>1</sup>, Caixia Li<sup>1</sup>, Ming Li<sup>3,1</sup>

<sup>1</sup>Zhejiang Key Laboratory of Intelligent Education Technology and Application, Zhejiang Normal University, Jinhua 321004, China; <sup>2</sup>The Chinese University of Hong Kong, Hong Kong SAR, China; <sup>3</sup>Zhejiang Institute of Optoelectronics, Jinhua 321004, China

(\*Corresponding author: yuting.chen@cuhk.edu.hk)

**Abstract:** Augmented reality (AR) is seen as a useful tool for improving student learning in writing education, but the limitations of its application and its effectiveness in combination with self-regulated learning (SRL) remain questionable. This study explores the impact of the AR-SRL teaching approach on primary school students' Chinese writing performance and learning engagement. 115 students were divided into three groups: the traditional SRL approach (C-SRL), the AR-based SRL approach (AR-SRL), and the AR technology under the traditional approach (AR). Results showed that the AR-SRL group outperformed others in writing, cognition, emotion, behavior, and social engagement, with stronger autonomy and engagement, particularly among those with high self-regulation. In contrast, the AR group without SRL support showed more random behavior and lower task efficiency, while the C-SRL group had less interaction and engagement. These findings emphasize the role of SRL in technology-enhanced learning and the limitations of relying solely on technology, indicating a need for further investigation into the integration of AR and SRL in education.

**Keywords:** Augmented reality, Self-regulated learning, Learning engagement, Behavior sequence analysis, Writing education

## 1. Introduction

This study addresses challenges in primary school Chinese writing, such as limited expression, insufficient feedback, and low motivation, which result from traditional teaching methods that emphasize knowledge transmission over writing interest and self-regulation. While self-regulated learning (SRL) strategies can improve writing skills (Shen & Bai, 2024), primary students often lack guidance in applying them effectively. Augmented reality (AR) shows promise for enhancing engagement and writing performance but may also introduce cognitive load and distractions. This study proposes a model integrating AR and SRL to examine the effects on writing outcomes, engagement, and behavior through three teaching approaches: C-SRL, AR and AR-SRL. The following research questions:

**RQ1:** How do different learning approaches affect students' writing outcomes?

**RQ2:** How do different learning approaches influence students' learning engagement?

**RQ3:** How does self-regulation level impact students' learning engagement across different approaches?

**RQ4:** How do learning behavior patterns vary across different learning approaches and self-regulation levels?

## 2. The self-regulated learning based AR approach for Chinese article writing

This study, conducted in a fourth-grade writing course on "My Adventure", aims to enhance students' perception of hard-to-observe scenes and improve realism. Using the "Wild Animal AR" system, students interact with 3D scenes featuring voice explanations, questions, and virtual touch. The AR-SRL approach, based on Zimmerman's model, structures writing into three stages (Figure 1 on left). The "I Am a Writing Doctor" learning sheet acts as an external memory aid to support self-regulation in writing.

## 3. Methods

### 3.1. Experimental design and procedure

The study used a mixed-method quasi-experimental design, combining quantitative (ANOVA) and qualitative (behavioral sequence analysis) methods. The four-week experiment (Figure 1 on right) involved three groups: (a) C-SRL group: Traditional video-based learning with simple task-oriented sheets and general teacher feedback. (b) AR group: Immersive AR-based learning with traditional task sheets and immediate teacher feedback. (c) AR-SRL group: Writing tasks within AR, supported by self-regulated learning sheets guiding goal setting, self-diagnosis, peer evaluation, and self-assessment, with teacher support. Students' writing levels were comparable before the experiment.

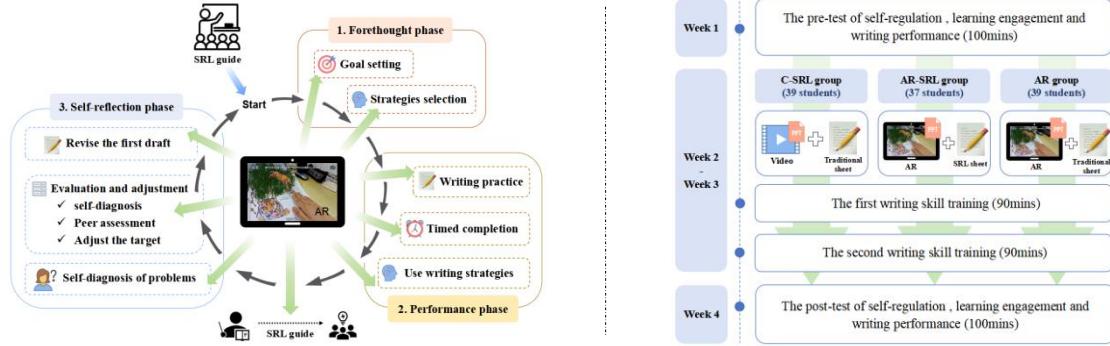


Fig.1 Structure of the AR-SRL approach ( left) and experiment procedure (right)

### 3.2. Measurements

This study assessed students' writing ability, self-regulated learning (SRL), and engagement using multiple methods: (a) Writing Evaluation Criteria (Huang et al., 2020), focusing on Accuracy, Organization, Expression, and Creativity, totaling 100 points. (b) SRL Ability Questionnaire (Barnard et al., 2009), with 21 items across six dimensions, classifying students into high (high-SRL) and low (low-SRL) groups. (c) Learning Engagement Questionnaire (Wang et al., 2016), with four dimensions (cognition, behavior, emotion, social). (d) Learning Behavior Coding, analyzed via 90-minute video recordings, coded every 15 seconds using Li et al. (Li et al., 2023) scheme.

## 4. Results

In question 1, the homogeneity test of variance revealed that all dimensions except organization failed to meet homogeneity, so the Welch test and Games method were used. Results showed significant differences in writing performance, with the AR-SRL group outperforming the C-SRL group in total score, accuracy, organization, expression, and creativity, and surpassing the AR group in accuracy, organization, and expression. In question 2, one-way ANOVA assessed the impact of learning approaches on engagement. Pre-test results showed no significant differences, ensuring comparability. The post-test results indicated that both AR-SRL and AR groups outperformed C-SRL in all engagement dimensions, with the AR-SRL group showing greater improvement in cognitive and social engagement, and the AR group excelling in emotional engagement. In question 3, the study analyzed engagement differences between high- and low-SRL students across three learning approaches. In the C-SRL group, high-SRL students excelled in cognitive and emotional engagement but not behavioral engagement. In the AR-SRL group, post-test results showed higher cognitive, emotional, and social engagement for high-SRL students. In the AR group, pre-test differences were seen in emotional engagement, and post-test differences in cognitive engagement, with high-SRL students performing better.

In question 4, C-SRL students mainly listened (SL, 40.37%), completed tasks (SC, 34.79%), and observed materials (SO, 10.75%). AR-SRL students prioritized task completion (SC, 43.19%), listening (SL, 28.80%), and peer discussion (SDP, 10.22%). AR students focused on listening (SL, 36.34%), task completion (SC, 35.67%), and interacting with learning tools (SO, 14.10%). These results suggest AR-SRL students engage more in tasks and discussions, C-SRL students focus on listening and tasks, and AR students interact more with tools. Figure 2 (left) visualizes the behavioral paths of the three groups. In the AR-SRL group, students often transitioned from listening to answering or sharing (SL → SR; SR → SA) and formed circular behaviors through discussion or observation (SO → SDP; SDP → SO). C-SRL students focused on task lists or teacher explanations (SC → SC; SL → SL), while AR

students alternated between observation and discussion (SO → SDP; SDP → SO) with fewer behavioral transitions, often repeating peer list viewing or question-raising (SLP → SLP; SR → SR).

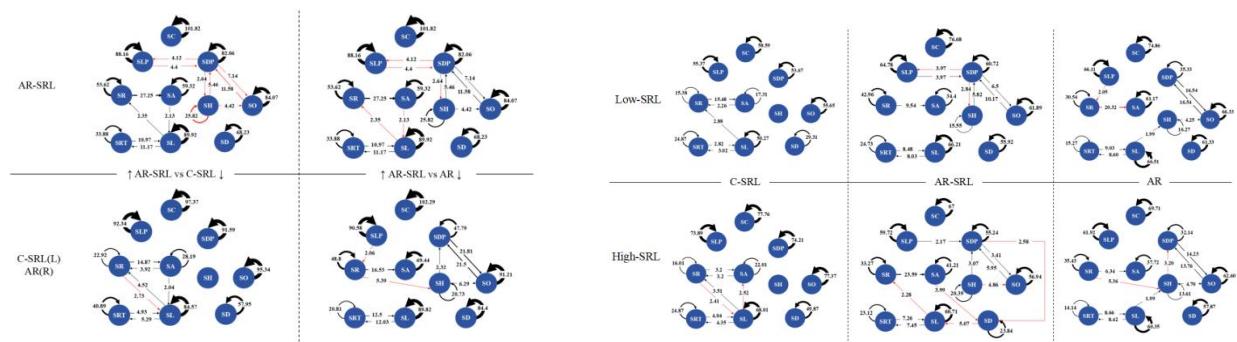


Fig.2 Learning behavior patterns of the three groups (left) and behavior patterns at two SRL levels in each group (right).

Figure 2 (right) shows that in the AR-SRL group, high-SRL students were less disorganized and listened more compared to low-SRL students. In contrast, the high-SRL students in C-SRL were more disorganized, listening less while observing picture books. The AR group resembled C-SRL, with high-SRL students showing more disorder but engaging more in peer discussions. High-SRL students in the AR-SRL group exhibited less disorder and followed a distinct behavior pattern: moving from answering questions and peer discussion, to classroom disorder, then back to listening and answering questions again. This pattern was absent in C-SRL and AR groups.

## 5. Discussion and conclusion

RQ1 examined the impact of AR-SRL on writing performance. The results showed the AR-SRL group outperformed the C-SRL group in overall writing, accuracy, organization, expression, and creativity, and exceeded the AR group in expression and creativity. This supports Li et al. (Li et al., 2023), who found that AR enhances learning through multimedia and interaction. Combining AR's immersive experience with SRL's structured approach boosts engagement and writing skills. However, AR alone may be ineffective without clear learning goals and strategies. RQ2 explored the impact of different groups on students' learning engagement. The results showed that the AR-SRL group had significantly higher cognitive, behavioral, emotional, and social engagement than the C-SRL group, particularly in cognitive and social engagement. Studies suggest that greater autonomy in writing enhances motivation and engagement, especially for lower-performing students (Li et al., 2023). The AR-SRL approach boosts engagement by combining AR's interactive experience with SRL strategies, providing multiple feedback channels. While the AR group showed higher emotional engagement than the C-SRL group, the AR-SRL group showed minimal differences in emotional engagement, possibly due to the cognitive load of SRL. RQ3 examined learning engagement differences among students with varying levels of SRL. Results showed that high SRL students in the AR-SRL group outperformed low SRL students in cognitive, emotional, and social engagement, as high SRL students were better able to utilize the resources and strategies of the AR-SRL approach. Low SRL students may face cognitive load and reduced emotional engagement in AR environments. The addition of AR enhanced cognitive engagement for high SRL students and emotional engagement for low SRL students. Furthermore, low SRL students in the AR group showed greater improvement in emotional engagement, suggesting AR is more effective in mobilizing emotions for lower-regulated learners. RQ4 analyzed classroom behavior differences under different teaching approaches. Results showed that AR-SRL students exhibited more active behaviors, including frequent self-monitoring, reflection, communication, and task-focused actions, reflecting their high engagement. This suggests that integrating SRL strategies in AR environments promotes self-directed learning. In contrast, AR students without SRL showed random, passive behaviors, highlighting the importance of instructional design. Traditional SRL group behaviors were fragmented with low interaction. The AR-SRL group showed significantly higher self-monitoring (42.9%) and social communication (23.9%) compared to the C-SRL (23.2%, 24%) and AR groups (34.2%, 22%), demonstrating its ability to enhance autonomous learning. High SRL students in the AR-SRL group exhibited better performance with less disorder and more

participation, while C-SRL students displayed more disorder despite being attentive. The AR-SRL approach motivated high SRL students and enhanced task focus. In the AR group, high SRL students engaged more in peer discussions but showed disordered behavior due to the disorganized AR teaching mode. Both the AR-SRL and AR groups displayed behavior patterns suggesting that AR technology fosters interaction and social engagement, supporting the AR-SRL approach's effectiveness in enhancing engagement and optimizing behavior.

This study examines the impact of combining augmented reality (AR) with self-regulated learning (SRL) strategies on primary school students' engagement in Chinese writing. The results show that the AR-SRL approach significantly enhances cognitive, emotional, behavioral, and social engagement. AR motivates students through an immersive environment, while SRL strategies help manage learning, especially for those with strong self-regulation skills. However, students with weak self-regulation may struggle without structured guidance, such as clear objectives and real-time feedback. Over time, the AR-SRL approach can foster sustained motivation, improve independent learning, and strengthen writing skills. By encouraging active learning and self-regulation, this approach helps students develop lifelong learning habits, boosting their adaptability and long-term academic success. The study provides a foundation for using AR and SRL in writing education and suggests further research to adapt this model for other subjects.

### Acknowledgements

This study was Supported by the "Empirical Research on the Construction of Self-regulatory Pathways in AR Supported Writing" of a Project Supported by Scientific Research Fund of Zhejiang Provincial Education Department. (No.Y202353535). Ming Li acknowledged the supports from Jinhua Science and Technology Plan (No. 2023-3-003a).

### References

- Barnard, L., Lan, W. Y., To, Y. M., Paton, V. O., & Lai, S. L. (2009). Measuring self-regulation in online and blended learning environments. *The Internet and Higher Education*, 12(1), 1-6.
- Huang, H. L., Hwang, G. J., & Chang, C. Y. (2020). Learning to be a writer: A spherical video-based virtual reality approach to supporting descriptive article writing in high school Chinese courses. *British Journal of Educational Technology*, 51(4), 1386-1405.
- Li, M., Chen, Y. T., Huang, C. Q., Hwang, G. J., & Cukurova, M. (2023). From motivational experience to creative writing: A motivational AR-based learning approach to promoting Chinese writing performance and positive writing behaviours. *Computers & Education*, 202, 104844.
- Shen, B., & Bai, B. (2024). Chinese university students' self-regulated writing strategy use and EFL writing performance: Influences of self-efficacy, gender, and major. *Applied Linguistics Review*, 15(1), 161-188.
- Wang, M. T., Fredricks, J. A., Ye, F., Hofkens, T. L., & Linn, J. S. (2016). The math and science engagement scales: Scale development, validation, and psychometric properties. *Learning and Instruction*, 43, 16-26.
- Yang, G., Chen, Y. T., Zheng, X. L., & Hwang, G. J. (2021). From experiencing to expressing: A virtual reality approach to facilitating pupils' descriptive paper writing performance and learning behavior engagement. *British Journal of Educational Technology*, 52(2), 807-823.
- Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal*, 45(1), 166-183.

# The Study on the Implementation of Artificial Intelligence Technology in College English Instruction

Tianchi Xia<sup>1</sup>, Juan Wang<sup>1</sup>

<sup>1</sup> School of Smart Education, Jiangsu Normal University

\*2020240567@jsnu.edu.cn

**Abstract:** As artificial intelligence (AI) technology continues its rapid evolution, its integration into college English instruction is reshaping traditional teaching methods. This paper reviews the development and core characteristics of AI, examines its current application in listening, speaking, reading, and writing through intelligent voice recognition and automated essay scoring systems, and compares international and domestic practices. In doing so, it highlights both the benefits—such as personalized instruction and enhanced efficiency—and challenges, including high implementation costs, mismatched resource allocation, and insufficient teacher training. Strategic recommendations are proposed to better integrate AI into course content and enhance teacher competencies, thereby improving overall teaching quality and student learning outcomes.

**Keywords:** Artificial Intelligence, College English Teaching, Application Cases, Strategic Recommendations

## 1. Introduction

### 1.1. Research Background and Significance

#### 1.1.1. Research Background

The maturation of AI technology has led higher education institutions worldwide to explore its potential in English instruction. UNESCO data reveals that more than 70% of universities have incorporated AI-assisted teaching methods, reflecting an ongoing shift toward innovative, technology-enhanced educational practices that support personalized learning in a globalized academic environment.

#### 1.1.2. Significance of the Research

This study is significant in several respects:

**Educational Innovation:** It illustrates AI's capacity to disrupt and transform traditional teaching models by introducing adaptive, data-driven methods.

**Enhanced Learning Outcomes:** By analyzing AI's impact on listening, speaking, reading, and writing, the paper demonstrates how personalized feedback and automated processes can boost student performance.

**Practical Applications:** Detailed case studies—such as intelligent voice recognition and automated essay scoring—provide concrete evidence of AI's efficacy in real classroom settings.

**Future Directions:** The research identifies key challenges in integrating AI into English instruction and offers strategic recommendations to address these issues, paving the way for future educational reforms.

### 1.2. Content and Methodology of the Research

This study employs a literature review combined with case analysis to evaluate the practical use and effectiveness of AI in college English teaching. The research focuses on the application of intelligent voice recognition for listening and automated essay scoring systems for writing, supported by empirical data and comparative studies. This approach provides both theoretical insights and practical guidelines for enhancing English instruction through AI.

## 2. Overview of AI Technology

### 2.1. Definition and Development of AI Technology

### **2.1.1. Definition of AI Technology**

Artificial Intelligence refers to the ensemble of computational techniques—including machine learning, deep learning, natural language processing (NLP), and computer vision—that simulate human cognition and enable machines to perform tasks such as data analysis, pattern recognition, and decision-making.

### **2.1.2. Developmental History of AI Technology**

From its inception at the Dartmouth Conference in 1956, AI has evolved through several distinct phases. Early research focused on symbolic computation and problem solving. The subsequent “AI winter” period saw slowed progress due to technological and economic hurdles. The emergence of expert systems in the 1980s, followed by breakthroughs in machine learning and big data analytics in the 21st century, has culminated in today’s widespread and sophisticated AI applications that influence diverse fields—including education.

## **2.2. Main Characteristics of AI Technology**

### **2.2.1. Intelligentization**

AI systems are characterized by their ability to self-learn and adapt based on large datasets. In educational contexts, these systems analyze student behavior to offer personalized learning paths, adjust instructional content in real time, and provide immediate corrective feedback, thereby enhancing teaching precision and efficacy.

### **2.2.2. Automation**

Automation allows AI to perform repetitive tasks such as grading and data analysis with minimal human intervention. Despite requiring high initial investments in technology and teacher training, automation significantly reduces workload and accelerates the feedback cycle, ultimately leading to more efficient educational processes.

## **3. Current Status of AI Technology Application in College English Teaching**

### **3.1. Comparison of Application Status Internationally and Domestically**

#### **3.1.1. International Application Status**

In countries such as the United States, the United Kingdom, and Canada, AI has become a critical component of college English instruction. Advanced intelligent systems adjust course materials dynamically based on student progress, and AI-driven assessment tools provide rapid, accurate feedback on both oral and written language skills. The use of immersive technologies like Virtual Reality (VR) further enhances learning by simulating real-life communication scenarios.

#### **3.1.2. Domestic Application Status**

In China, universities are rapidly embracing AI in English education. Applications such as intelligent voice recognition for listening and automated essay scoring for writing are increasingly common. However, challenges persist—ranging from limited teacher expertise in AI tools to uneven distribution of technological resources—hindering the full potential of these innovations in enhancing teaching outcomes.

### **3.2. Problems in Application**

#### **3.2.1. Issues with Technological Costs**

High initial investments, continuous maintenance expenses, and the cost of specialized teacher training represent significant barriers to widespread AI adoption. For many institutions, especially those with limited budgets, these financial challenges restrict the effective implementation of AI technologies.

#### **3.2.2. Problems with Matching Educational Resources**

There is often a mismatch between available educational resources and the technological requirements needed for successful AI integration. Even institutions with state-of-the-art facilities may face difficulties in aligning these resources with instructional needs, largely due to inadequate training and support systems for educators.

## **4. Application Strategies of AI Technology in College English Teaching**

### **4.1. Strengthening the Integration of AI Technology with Teaching Content**

#### **4.1.1. Optimization of Teaching Content**

To fully harness the potential of AI, curricula must be restructured to integrate advanced technologies. For example, incorporating tools such as Google Speech-to-Text in listening and speaking exercises can create realistic learning environments that mimic real-world interactions. Furthermore, updating teaching materials to include contemporary topics and AI-related content can significantly enhance student engagement and relevance.

#### **4.1.2. Strategies for Technology Integration**

Effective integration requires that educators develop a deep understanding of AI functionalities. Teachers should be trained to design AI-driven instructional modules that align with course objectives while maintaining data security and ethical standards. Establishing robust protocols for data collection and privacy is essential for creating a trustworthy and effective learning environment.

#### **4.2. Enhancing Teachers' AI Skills Literacy**

##### **4.2.1. Importance of Skill Training**

Enhancing the AI literacy of educators is crucial for the successful deployment of these technologies in the classroom. Adequate training can enable teachers to utilize AI tools effectively, leading to more interactive and personalized teaching methods that improve student performance.

##### **4.2.2. Training Strategies and Methods**

A comprehensive training approach is recommended, encompassing foundational courses, hands-on workshops, and advanced seminars focused on innovative applications. Forming teacher learning communities and collaborating with AI technology providers can foster continuous professional development. Additionally, institutions should allocate sufficient resources—both in terms of time and finances—to support ongoing teacher training and technical assistance.

### **5. Conclusions and Recommendations**

#### **5.1. Overview of Research Findings**

This study confirms that AI technology has a significant positive impact on college English instruction by facilitating personalized feedback and efficient teaching practices in listening, speaking, reading, and writing. However, challenges such as high implementation costs and the need for extensive teacher training remain major hurdles to its broader adoption.

#### **5.2. Suggestions and Prospects**

##### **5.2.1. Recommendations for College English Teaching**

Colleges should leverage existing online platforms and incorporate advanced AI tools—such as iFlytek's speech recognition and Pigai.org's automated essay scoring systems—to create a more interactive and efficient learning environment. Updating course content and promoting active teacher-student interactions are critical to maximizing AI's benefits.

##### **5.2.2. Prospects for Future Research Directions**

Future research should focus on enhancing the accuracy of AI in handling diverse accents and creative writing, while exploring big data analytics to further personalize learning. Addressing ethical issues such as data privacy and fairness will also be essential to ensure that AI technologies are implemented equitably across educational institutions.

#### **5.3. Limitations and Shortcomings of the Research**

Although this study provides a comprehensive overview of AI applications in college English teaching, its scope is limited by sample size and reliance on secondary data. Future studies should incorporate a broader range of methodologies and larger sample sizes to capture the full diversity and complexity of AI integration in educational environments.

### **References**

Crompton, H., Burke, D. Artificial Intelligence In Higher Education: the State of the Field[J] International

---

Journal of Educational Technology in Higher Education, 2023 , 20:22 .

Yu, J., Wang, L. Research on the Integration Path and Practice of AI Intelligent Technology and English Teaching Reform in Higher Vocational Colleges and Universities[J]. Applied Mathematics and Nonlinear Sciences, 2024, 9(1) .

Lucas, K. , Di, Z. Techno Stress and English Language Teaching in the Age of Generative AI[J] .Educational Technology Society, 2024, 27(2):306-320 .

Wang, B. An Intelligent Integration Method of AI English Teaching Resources Information Under Multi-agent Cooperation[J] .International Journal of Continuing Engineering Education and Life-Long Learning, 2024, 34(1):88-99 .

Wu, X. Dynamic evaluation of college English writing ability based on AI technology[J] . Journal of Intelligent Systems, 2022 , 31(1):298-309 .

# Exploring the Potential of AI-Generated Lesson Designs Underpinned by the TPACK

## Framework for Educators in Higher Education: A Comparative Study

Yanjie Song<sup>1\*</sup>, Jiachen Fu<sup>1</sup>, Thomas K.F. Chiu<sup>2</sup>, Irwin King<sup>2</sup>, Huamin Qu<sup>3</sup>

<sup>1</sup>The Education University of Hong Kong, Hong Kong, PRC

<sup>2</sup>The Chinese University of Hong Kong, Hong Kong, PRC

<sup>3</sup>The Hong Kong University of Science and Technology, Hong Kong, PRC

\* ysong@edu.hk

**Abstract:** This paper investigates the potential of integrating generative AI into lesson design underpinned by the Technological Pedagogical Content Knowledge (TPACK) framework for educators in higher education. The study introduces a novel workflow that uses multiple AI agents to generate tailored lesson designs in two learning environments: one in a traditional classroom and another in the metaverse learning environment. A comparative study was conducted in a university-level product design course using three lesson designs under three conditions: Lesson design (LD) 1 – Manually designed lesson design in a traditional classroom, LD 2 – AI-generated lesson design in a traditional classroom, and LD 3 – AI-generated lesson design in the metaverse environment. The TPACK framework with the same lesson objectives underpinned all the lesson designs. Through qualitative analysis, the paper compares the three lesson designs, examining the potential of AI-generated lesson designs in supporting teacher professional development. The findings indicate that AI-generated lesson designs have great potential to enhance lesson design underpinned by the TPACK framework for educators.

**Keywords:** Generative AI, Lesson design, TPACK framework, Educational technology

### 1. Introduction

Technology integration into educational settings has profoundly transformed teaching practices, providing novel avenues for personalizing learning and fostering curriculum innovation. Despite these advancements, traditional approaches to lesson design often fall short in adaptability, struggling to keep pace with the rapidly evolving technological landscape. This gap underscores the need for robust frameworks that facilitate a more effective integration of these critical elements. The Technological Pedagogical Content Knowledge (TPACK) framework, introduced by Mishra and Koehler (2006), offers a comprehensive model that synergizes technology, pedagogy, and content knowledge, thereby enhancing the design of educational practices. This framework has been extended into various models, such as BOPPPS-TPACK and TPCK-W, which assist educators in acquiring and applying integrated knowledge in their teaching endeavors (Lee & Tsai, 2010; Zhang & Zhou, 2023).

However, the continuous evolution of educational technology necessitates ongoing adaptations of the TPACK framework to embrace new technological innovations and meet diverse educational needs effectively. In this context, recent developments in generative AI represent a significant opportunity, offering the capability to automate the creation of lesson designs that are contextually relevant and pedagogically coherent. Research by Durmus (2024) demonstrates the potential of tools like ChatGPT to reduce the time required for lesson design significantly. However, some studies also highlight limitations, such as the rudimentary nature of AI-generated lesson designs, which often require substantial refinement (Duha, 2023). These challenges stem from the general limitations of large language models, which may not adequately address the specificities of diverse educational contexts. Against this background, this paper aims to explore a novel workflow based on the TPACK framework, integrating multiple generative AI agents to bridge this gap.

## 2. Development of a lesson design workflow with multiple AI agents underpinned by the TPACK framework

Generative AI, while powerful in automating many aspects of educational design, still faces inherent limitations. These include the inability to handle particular, nuanced requests or customize lesson designs that require deep domain expertise. For instance, customizing lesson designs necessitates deep knowledge of subject matter, which generative AI lacks, potentially resulting in inaccuracies or oversimplifications (Giannakos et al., 2024). To address these challenges and enhance the flexibility of AI-generated lesson designs, the integration of multiple agents and Retrieval-Augmented Generation (RAG) technology can be employed.

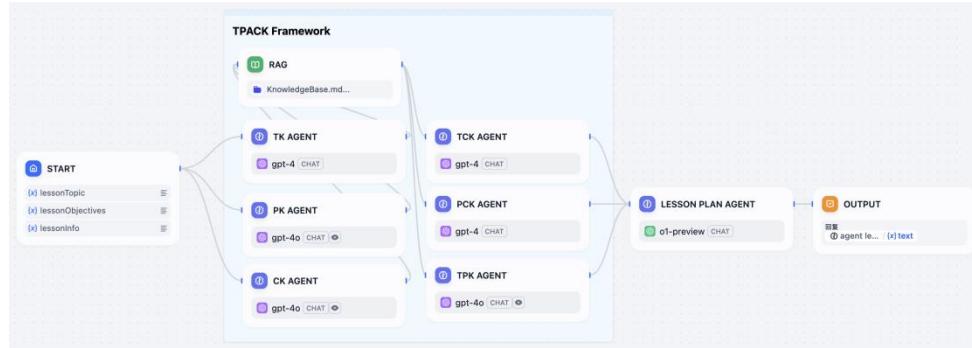


Fig.1 A lesson design workflow with multiple agents underpinned by the TPACK framework

Based on these technologies, the study developed a workflow integrating six distinct agents (see Figure 1). This workflow commences with the input of basic course information, followed by a two-stage lesson design process designed based on the TPACK framework. In the first stage, the TK (Technological Knowledge), PK (Pedagogical Knowledge), and CK (Content Knowledge) agents utilize RAG technology to retrieve the most relevant information based on the input. This information is then fed into the second stage, which includes the TCK (Technological Content Knowledge), PCK (Pedagogical Content Knowledge), and TPK (Technological Pedagogical Knowledge) agents. These agents simulate a teacher's cognitive process, deliberating across the three dimensions of TPACK to refine the lesson design.

## 3. Method

### 3.1. Research design

To evaluate the potential of AI-generated lesson designs, a comparative study was conducted in a product design course in higher education. The research question is: What are the differences among the three lesson designs: Lesson design (LD 1, designed manually, LD 2 and LD 3 created by generative AI? A teacher provided a manually designed lesson design (LD 1), in which essential information from this course (e.g., topics, learning objectives, necessary knowledge, and duration) was extracted. The information was then input into the AI workflow to generate two distinct lesson designs: AI-generated for the traditional classroom (LD 2) and AI-generated lesson design in metaverse-based classroom (LD 3).

### 3.2. Data collection

The data collection involved qualitative data from three lesson designs aligned with the same learning objectives. In addition to the lesson designs, semi-structured interviews were conducted with the teacher who facilitated the lessons. These interviews aimed to collect the educator's insights and reflections on the utility, practicality, and pedagogical potential of using AI-generated lesson designs.

### 3.3. Data analysis

The qualitative analysis involved a detailed comparison of the three plans. Content analysis was utilized to examine the extent to which each design is aligned with the six categories of content delivery, technology integration,

pedagogical approach, student engagement, flexibility, and ease of implementation adapted from Savage (2014). Teacher interviews were coded based on the six categories as a framework.

#### 4. Findings and Discussions

Table 1 demonstrates a progressive integration of technology and pedagogical innovation across three lesson designs (LDs). LD1 employs basic digital tools (slides, videos) and traditional lectures, offering simplicity but limited interactivity and flexibility. LD2 integrates structured digital content, AI assistance, and project-based learning, achieving moderate technological integration while enhancing engagement without compromising feasibility. LD3, built on the metaverse platform Learningverse (Song et al., 2023) and inquiry-based methods, creates an immersive learning environment, demanding advanced technical expertise but significantly deepening student participation and cognitive development.

Comparative analysis reveals a hierarchical evolution across three dimensions: technological advancement, pedagogical methodology, and interactivity. Implementation complexity aligns directly with technological sophistication, positioning LD1 as ideal for conventional educational contexts, LD2 as a transitional model for incremental digital adoption, and LD3 as a frontier framework for advanced technology-enabled learning environments.

Table 1. Comparison of three lesson designs adapted from Savage (2014)

Category	LD 1	LD 2	LD 3
Content Delivery	Basic digital tools (slides, videos)	Enhanced digital tools and structured content	Highly interactive and immersive tools
Technology Integration	Minimal, with basic presentation tools	Moderate, with digital tools and AI assistance	High, with AI assistance and the metaverse-Learningverse
Pedagogical Approach	Traditional lectures and discussions	More dynamic, Project-based with AI-scaffolding strategies	Inquiry-based, highly interactive methods
Student Engagement	Limited interactivity	Improved interaction and participation	Extensive engagement with immersive technology
Flexibility	Low	Moderate; allows some customization	High; adaptable to real-time feedback
Ease of implementation	High; familiar environment	Medium; requires familiarity with AI tools	Low; needs technological proficiency

Feedback from the teacher during interviews was positive towards the AI-generated lesson designs. The teacher noted the time-saving benefits and enhanced adaptability of educational content provided by AI, which also offered a broader range of perspectives and richer content. This facilitated greater critical thinking and creativity in problem-solving among educators.

#### 5. Implications and limitations

The analysis shows that lesson designs generated based on the workflow align more closely with student-centered, innovative teaching philosophies and incorporate emerging technologies such as AI and the Metaverse. This integration opens up new avenues for teachers to expand their instructional technology, offering novel ways to enhance pedagogical practices. While the study highlights significant benefits of integrating AI into lesson design, several limitations were identified that suggest areas for future improvement. One notable limitation is the dependency on a well-prepared knowledge base for the RAG technology to produce high-quality lesson designs. This preparation requires educators to provide comprehensive and detailed content in advance, which can be time-consuming and demands a high level of subject expertise. Therefore, there is a pressing need to enhance teacher competency in digital tools and technologies. Providing adequate training and support is essential for teachers to successfully adopt and integrate these innovative educational technologies into their teaching practices.

## 6. Conclusion

In conclusion, this study highlights the significant potential of AI-generated lesson designs to enhance educational practices through tailored content, dynamic pedagogical strategies, and advanced technological integration. However, the successful implementation of such plans requires not only sophisticated technological setups but also substantial teacher preparation. Addressing the practical challenges of these implementations, particularly in terms of the technological proficiency of educators, is crucial. Thus, enhancing teacher digital competencies is essential to fully leverage the capabilities of AI in education (Scarci et al., 2024). This approach will not only facilitate the seamless integration of cutting-edge technologies into classrooms but also ensure that educational outcomes are maximized in an era of rapid technological advancement.

## Acknowledgements

The study was funded by the Fund for Innovative Technology-in-Education (FITE) (Ref. 05404 and 05408) (2023-26), University Grant Committee, Hong Kong.

## References

- Chai, C. S., Koh, J. H. L., & Tsai, C. C. (2013). A Review of Technological Pedagogical Content Knowledge. *Educational Technology & Society*, 16(2), 31–51.
- Durmus, C. (2024). From AI-Generated Lesson designs to the Real-Life Classes: Explored by Pre-Service Teachers.
- Duha, M. S. U. (2023). ChatGPT in Education: An Opportunity or a Challenge for the Future? *TechTrends*, 67(3), 402-403.
- Giannakos, M., Azevedo, R., Brusilovsky, P., Cukurova, M., Dimitriadis, Y., Hernandez-Leo, D., ... & Rienties, B. (2024). The promise and challenges of generative AI in education. *Behaviour & Information Technology*, 1-27.
- Johnson, L., Adams Becker, S., Cummins, M., & Estrada, V. (2016). *NMC Horizon Report: 2016 Higher Education Edition*. The New Media Consortium.
- Karaduman, T., & Akman, B. (2024). A Comprehensive Review of Technological Pedagogical Content Knowledge (TPACK). *e-Kafkas Journal of Educational Research*, 11(1), 141-159.
- Lee, M. H., & Tsai, C. C. (2010). Exploring teachers' perceived self efficacy and technological pedagogical content knowledge with respect to educational use of the World Wide Web. *Instructional Science*, 38, 1-21.
- Mishra, P., & Koehler, M. J. (2006). Technological Pedagogical Content Knowledge: A Framework for Teacher Knowledge. *Teachers College Record*, 108(6), 1017–1054.
- Savage, J. (2014). Lesson design: Key concepts and skills for teachers. Routledge.
- Scarci, A. S., Teixeira, T. M., & Dal Forno, L. F. (2024). Artificial Intelligence and its relations with digital competencies and Education.
- Song, Y., Cao, J., Wu, K., Yu, P. L. H., & Lee, J. C. K. (2023). Developing “Learningverse”—A 3-D Metaverse Platform to Support Teaching, Social, and Cognitive Presences. *IEEE Transactions on Learning Technologies*, 16(6), 1165-1178.
- Zhang, S., & Zhou, A. (2023). The Construction and Practice of a TPACK Development Training Model for Novice University Teachers. *Sustainability*, 15(15), 11816.

# A study on the Impact of AI-Driven STEAM Curricula on Computational Thinking Development in the Chinese Elementary School

Bowen Li<sup>1</sup>, Kaiwen Pan<sup>2</sup>, Qing Zhang<sup>3</sup>

<sup>1,2,3</sup> Xi'an Jiaotong-Liverpool University

**Abstract:** This study investigates integrating artificial intelligence (AI) into computational thinking (CT) curricula for primary school students in China. A 9-week after-school course was designed for 15 Grade 4-6 students, combining STEAM education, project-based learning (PBL), and co-design methodologies. Students used the Machine Learning for Kids platform to build and train models, integrating them into Scratch to create AI-powered applications. Pre- and post-surveys were designed based on the Computational Thinking Scale (CTS) to measure students' CT dimensions, while the Bebras Challenge test was adopted to evaluate students' performance in CT test, both showing significant improvement. Additionally, focus groups were conducted to collect students' feedback on this learning experience. Students reported significant improvement in engagement, problem-solving skills, and awareness of AI applications, despite some challenges about technology and language. These findings demonstrate the potential of integrating AI education into primary school curricula to enhance students' CT skills, contributing to the development of AI and CT education frameworks while providing practical implications for implementing AI education in K-12 schools.

**Keywords:** AI Education, Computational Thinking, Co-design, Project-based Learning, STEAM Education

## 1. Introduction

Computational Thinking (CT), as defined by Wing (2006), emphasizes representing and solving problems using computer science concepts. With AI's growing role in daily life, understanding and leveraging AI to address complex problems is essential (Brynjolfsson & McAfee, 2017). Recognizing this need, China has established CT and AI literacy as core competencies in its K-9 Information Technology curriculum, providing a strong foundation for students to succeed in the digital age (Ministry of Education of China, 2022). Early exposure to CT and AI is crucial, because children's cognitive abilities are highly malleable, making it easier to cultivate strong problem-solving skills (Bers et al., 2019). This study examines how AI education impacts CT skill development among Chinese primary school students, highlighting the intersection of CT and AI in early education.

## 2. Literature Review

Integrating AI education with STEAM and project-based learning (PBL) offers a multidisciplinary framework that promotes computational thinking (CT) more effectively than traditional methods by connecting learning to real-world problems and encouraging iterative solution refinement (Huang & Qiao, 2024; Shin et al., 2021). Moreover, co-design, as a collaborative learning approach, further enhances CT by positioning students as creators and innovators, emphasizing active participation and hands-on engagement to foster creativity, critical thinking, and problem-solving skills (Sunday et al., 2024). However, assessing CT remains a challenge, with qualitative methods like interviews being underutilized compared to traditional tests, portfolios, and surveys, which often focus narrowly on algorithmic thinking and problem decomposition while overlooking creativity, collaboration, and reflection (Cutumisu et al., 2019; Tang et al., 2020; Brennan & Resnick, 2012).

## 3. Research Questions

Existing research underscores the potential of pedagogical approaches in fostering computational thinking (CT) yet critical gaps remain. These include understanding how AI education can be integrated into STEAM through PBL and co-design, as well as the prevailing reliance on quantitative methods that fail to capture qualitative insights and the interplay between CT assessment tools. To address these gaps, this study explores the following research questions:

RQ1: How does the AI-machine learning curriculum impact pupils' CT skills?

RQ2: What are the relationships between CT dimensions measured through scale surveys, and students' performance on CT tests?

RQ3: What are students' experiences and perceptions of this AI-integrated STEAM curriculum?

## 4. Methodology

### 4.1. Course Design

In this 9-week course, students explored the fundamentals of ML through PBL. Projects were developed on the ML for Kids platform teaching students to build ML models which can be exported to Scratch, allowing students to build AI-powered applications. The course began in Week 1 with a pre-test to assess students' prior knowledge, followed by an introduction to the basic syntax of Scratch. In Week 2, students learned the basics of supervised learning, while Weeks 3-7 guided them through ML projects like text recognition ("Make me Happy"), image recognition ("Pokemon Images"), and chatbot creation ("Owls Chatbots"). Week 8 encouraged creativity and collaboration through co-designing individualized chatbots, and Week 9 concluded with a review and post-test to measure learning outcomes.

### 4.2. Data Collection

#### 4.2.1. Survey (scales)

The Computational Thinking Scale (CTS) developed by Korkmaz and Bai (2019) was adopted in this study as it has been validated in the Chinese context. This reliable tool assessed five key dimensions: creativity, algorithmic thinking, cooperativity, critical thinking, and problem-solving, aligning well with the objectives of this AI-ML course. Participants completed the CTS both before (pre-test) and after (post-test) the course to evaluate the impact of the course on their CT development.

#### 4.2.2. Bebras tests

A set of pre- and post-tests consisting of six CT questions was used to evaluate students' CT skills, such as algorithmic thinking and pattern recognition abilities. These questions were selected from Bebras Challenge tests. Question 1 and Question 6 are shown as examples in Figure 1. Question 1 assesses students' pattern recognition by requiring students to identify the most often borrowed book through analyzing patterns and frequencies in records, while Question 6 examines encryption algorithm analysis based on a provided example.



Fig.1 Question 1 and question 6 of CT test from Bebras Challenge

#### 4.2.3. Interviews

The interviews, conducted in focus groups each involving 4-5 students, explored participants' experiences and perspectives regarding their ML projects, while also examined their CT skills. It began by investigating participants' favorite aspects of working on projects, including what they found most engaging. Next, they were asked to explain ML in simple terms to assess their communication skills. The discussion then shifted to real-world ML applications, focusing on its practical benefits. Finally, participants shared how they tackled project challenges, revealing their problem-solving strategies and resilience—key elements of computational thinking.

#### **4.3. Data Analysis**

This study employed multiple analyses to evaluate the learning program's impact. Paired t-test was used to compare pre- and post-survey data of Computational Thinking Scales (CTS) and test scores to identify significant changes. Additionally, thematic analysis was implemented to explore students' perceptions of a STEAM curriculum integrating ML, project-based learning, and co-design, revealing insights into their learning experiences. Furthermore, Pearson correlation analysis was used to examine the relationship between CTS results and Bebras Challenge scores, assessing the correlation between these two assessment tools. Overall, this mixed-methods approach offered a comprehensive understanding of the program's impact on students' learning experience and capability development.

### **5. Findings**

#### **5.1. Improvements in CT skills by The Course**

The paired-samples t-test results showed significant improvements in all five CT dimensions: Creativity, Collaboration, Critical Thinking, Problem-solving Thinking, and Algorithmic Thinking. Additionally, the Bebras Challenge Score increased from an average of 3.84 to 4.08. While this improvement was not statistically significant ( $p = 0.217$ ), it still shows improvement in students' performance. These findings indicate that the course was effective in enhancing students' CT skills across all measured dimensions (as shown in Table 1).

Table 1. Results of paired-samples t-test across computational dimensions

Dimension	Mean (pre)	Mean (post)	Mean Difference	Std. Deviation	t	df	p-value
Creativity	2.91	4.53	-1.62	.41	-15.12	14	<.001
Algorithmic Thinking	3.63	4.18	-.55	.72	-2.9	14	0.011
Collaboration	3.81	4.61	-.80	.65	-4.7	14	<.001
Critical Thinking	3.45	4.45	-1.00	.59	-6.48	14	<.001
Problem-solving Thinking	2.58	4.81	-2.22	.34	-24.97	14	<.001
Bebras Challenge Score	3.84	4.08	-.24	.72	-1.2	14	.217

#### **5.2. Lack of Correlation Between CT Dimensions Survey Results and Bebras Challenge Scores**

The Pearson correlation analysis of the pre-test data shows no significant relationship between the dimensions of Computational Thinking (CT) as measured by Korkmaz's scale and the Bebras Challenge scores. Specifically, the correlations between the Bebras Challenge score and the CT dimensions—Creativity (0.327), Collaboration (-0.051), Critical Thinking (-0.397), Problem-solving Thinking (-0.091), and Algorithmic Thinking (-0.464)—are either weak or negligible, with none reaching statistical significance. This lack of significant correlation suggests that Korkmaz's scale may not effectively capture or prioritize key CT abilities such as pattern recognition, decomposition, and algorithm design, which are central to the Bebras Challenge.

#### **5.3. Participant Experiences with AI-integrated STEAM Curriculum**

Thematic analysis of interview data revealed recurring themes centered on Engagement and Interest, Challenges and Problem-Solving, Perceived Benefits, and Suggestions for Improvement. Participants expressed enthusiasm for ML projects, particularly those fostering creativity and autonomy, with their favorites like the owl chatbot and co-designed chatbot. One participant noted, "I felt like the chatbot was quite versatile after we finished programming it."

Programming was a highlight, described as a way to make machines "smarter and more capable," while also enhancing logical thinking and English skills. Challenges included technical issues like "nested structures", language barriers, and unreliable internet connections. Students often solved these independently, using logical step-by-step analysis. Suggestions for improvement included introducing more advanced programming languages, providing better English support, and interdisciplinary integration. Overall, the project was seen as transformative, improving problem-solving and programming skills, and inspiring excitement for AI's real-world applications. As one participant summarized, "AI is not far from us—it's already embedded in our daily lives, and learning this can change how we see the world."

## 6. Conclusion and Discussion

This study demonstrated the effectiveness of the AI-ML course in enhancing students' CT skills, as evidenced by the CTS survey and the Bebras Challenge test. Yet, the lack of correlation between these two assessment tools highlights the complementary nature of these assessment tools, suggesting that their combined use provides a more holistic and comprehensive evaluation of students' computational thinking skills. Furthermore, students reported increased engagement, skill development, and awareness of AI's real-world applications through ML projects despite encountering challenges like technical difficulties, language barriers, and infrastructure issues. Addressing these challenges through refined project designs, language support, and technical solutions could optimize the learning experience. Overall, the course successfully fostered students' CT skills and AI literacy, preparing them to adapt to and excel in an AI-driven world.

## References

- Bers, M. U., González-González, C., & Armas-Torres, M. B. (2019). Coding as a playground: Promoting positive learning experiences in early childhood classrooms. *Computers & Education*, 138, 130–145.
- Brennan, K., & Resnick, M. (2012, April). New frameworks for studying and assessing the development of computational thinking. In *Proceedings of the 2012 annual meeting of the American Educational Research Association*, Vancouver, Canada (Vol. 1, p. 25).
- Brynjolfsson, E., & McAfee, A. (2017). *Machine, platform, crowd: Harnessing our digital future*. WW New York: Norton & Company, 564.
- Cutumisu, M., Adams, C., & Lu, C. (2019). A scoping review of empirical research on recent computational thinking assessments. *Journal of Science Education and Technology*, 28(6), 651-676.
- Huang, X., & Qiao, C. (2024). Enhancing computational thinking skills through artificial intelligence education at a STEAM high school. *Science & Education*, 33(2), 383–403.
- Korkmaz, Ö., & Bai, X. (2019). Adapting computational thinking scale (CTS) for Chinese high school students and their thinking scale skills level. *Participatory Educational Research*, 6(1), 10-26.
- Ministry of Education of the People's Republic of China. (2022). *Information Technology Curriculum Standards for Compulsory Education (2022 Edition)*. People's Education Press.
- Shin, N., Bowers, J., Krajeik, J., & Damelin, D. (2021). Promoting computational thinking through project-based learning. *Disciplinary and Interdisciplinary Science Education Research*, 3, 1-15.
- Sunday, A. O., Agbo, F. J., & Suhonen, J. (2024). Co-design pedagogy for computational thinking education in K-12: A systematic literature review. *Technology, Knowledge and Learning*, 1–56.
- Tang, X., Yin, Y., Lin, Q., Hadad, R., & Zhai, X. (2020). Assessing computational thinking: A systematic review of empirical studies. *Computers & Education*, 148, 103798.
- Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, 49(3), 33–35.

# Research on the Design and Practice of Maker Educational Activities for the Cultivation of Computational Thinking

Zhangze Chen<sup>1\*</sup>, Jia Zhu<sup>2</sup>,

<sup>1</sup> School of Education, Zhejiang Normal University, Jinhua, China

<sup>2</sup> The Zhejiang Key Laboratory of Intelligent Education Technology and Application

\* zjnuczz@zjnu.edu.cn

**Abstract:** As a crucial 21st-century competency, computational thinking (CT) has gained prominence in K-12 education. Addressing the weak CT cultivation and operational overemphasis in junior high schools, this study developed a CT development framework using Mind+ programming and McQueen educational robots through maker education activities. Experimental results from assessment scales and interviews demonstrate significant improvements in students' CT cognition and practical abilities, offering an effective approach for junior high CT education.

**Keywords:** Computational Thinking, Maker Education, Mind+, McQueen Educational Robot

## 1. Introduction

The intelligent era has fundamentally transformed human cognition and lifestyles, making computational thinking (CT) cultivation a crucial educational objective (Wing, 2006). CT encompasses five core competencies: **abstraction**, **decomposition**, **algorithms**, **generalization**, and **evaluation** (Selby & Woppard, 2013). Currently, the cultivation of CT has been incorporated into technology courses around the world (e.g., Ministry of Education of China, 2020). While maker education integrated with CT shows potential for enhancing innovation and collaboration, practical implementation often prioritizes technical operations over cognitive development ("doing over thinking"). Therefore, this study integrates constructivism theory into maker education activities cultivated with CT, constructs a framework of maker education activities oriented to the cultivation of CT, and analyzes the effects of students' CT improvement from various aspects using scales, interviews, and students' works (Uzumcu & Bay, 2021). The analysis aims to answer the following research questions:

- (1) What is the overall effect of the maker education model for cultivating CT on improving students' CT ability?
- (2) Under the premise of using this teaching model, what are the different impacts on the five core aspects of students' CT ability?

## 2. Literature Review

### 2.1. The cultivation of CT

Since Wing's seminal definition of computational thinking (CT) in 2006, research has expanded across disciplines. The Computer Science Teachers Association (CSTA) established K-12 computer science standards in 2011, driving CT integration into STEM fields. Recent studies include Martin et al.'s (2024) 3C model for teaching coding and CT skills and Sun et al.'s (2023) educational gaming framework employing problem decomposition and process modeling.

### 2.2. Incorporating CT in Maker Education

Studies have developed multiple integration models. Li et al. (2024) positioned CT as an innovative thinking conduit for STEM-maker education synthesis, while Jiang et al. (2024) emphasized design thinking's catalytic role in value-driven maker projects. Both approaches employ project-based learning to cultivate collaborative problem-solving competencies, demonstrating pedagogical synergy. However, existing models predominantly prioritize practical implementation over theoretical underpinnings. Grounded in constructivism theory, this study establishes a CT-oriented

maker pedagogy framework and empirically evaluates its multidimensional impacts on students' CT development in authentic classroom settings.

### 3. Research design

#### 3.1. Theory model construction

Aligning with constructivism's dynamic epistemology (Fosnot & Perry, 1996), this study develops the **Ideation-Programming-Feedback** (IPF) model that systematically couples CT cultivation with maker pedagogy through three iterative stages. In the Ideation stage, learners define complex project problems and split them into sub-problems, developing abstraction and decomposition skills. In the Programming stage, learners collaboratively plan, draw flowcharts, and build prototypes, enhancing abstraction, algorithmic thinking, and evaluation abilities. In the Feedback stage, learners present their work, summarize functions, and optimize based on feedback, improving evaluation and generalization skills.

#### 3.2. Participants

Participants (N=35) were eighth-grade students. According to Piaget's cognitive development theory, they are in the formal operations stage. Students here are curious but impatient, especially towards things they're not interested in. Yet, their abstract thinking is rapidly developing. So, the author thinks middle school - level maker education can enhance learning and CT levels. Also, pre-test showed most students have computer operation skills and have used graphical programming software, meaning they share a similar starting point.

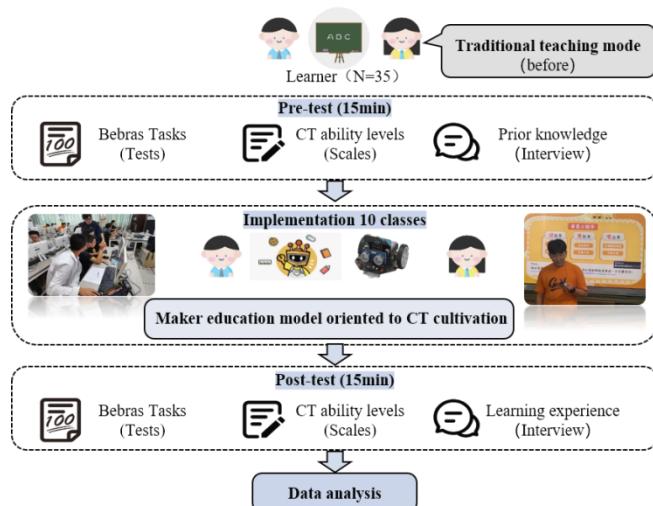


Fig.1 The research framework.

#### 3.3. Study Design and Procedure

This research mainly uses Mind+ software as the basis for the maker course, and uses the McQueen educational robot as the starting point, aiming to cultivate students' CT. Using Mind+ combined with McQueen educational robots for maker teaching has greater advantages in cultivating students' CT than using graphical programming software or open-source hardware alone. At the same time, this study also uses a maker education model for CT cultivation. In the process of solving problems and completing project tasks, students continuously optimize problem-solving solutions and exercise their problem-solving abilities.

This study designed two projects with a total of 10 class periods, each class period was about 40 minutes. The teaching cases selected were from the STEM education series "Python and micro:bit Robot Programming for Middle School Students" published by Tsinghua University Press. On the basis of the teaching materials, the author made certain modifications according to the current situation of the learners and determined the teaching implemented this time. The cases were "My Fancy McQueen" and "Thread-Seeking McQueen"(as shown in *Figure 1*).

#### 3.4. Data collection and analysis

To comprehensively assess students' computational thinking (CT) before instruction, pre-tests using Korkmaz's CTLS scale (12 items across 5 dimensions, 5-point Likert scale) and 5 Bebras Challenge questions (1 point each) were administered. Post-instruction analysis combined CTLS post-tests, performance assessments, and student interviews to evaluate CT development. Data analysis was conducted using SPSS 20.0, examining changes in CT cognition and problem-solving abilities through both quantitative (scale scoring) and qualitative methods.

## **4. Results**

### **4.1. Bebras Tasks**

This study used paired samples t-test to analyze the differences between pre-and post-test. The results showed significant differences ( $p=0.04<0.05$ ) in the pre-and post-tests. The maker education model aimed at cultivating CT ( $MD=3.57$ ,  $SD=1.092$ ) obtained higher scores than the traditional teaching model ( $MD=3.03$ ,  $SD=0.985$ ).

### **4.2. CT ability levels**

In this study, students' scores in five aspects of CT were statistically calculated in SPSS, and a paired sample T test was conducted to detect whether students' CT abilities have been improved. There are significant differences in the four abilities of Abstraction, Decomposition, Generalization, and Evaluation as well as the total score ( $p<0.05$ ), but there is no difference in Algorithmic thinking ( $p=0.16>0.05$ ). Among the five dimensions of CT ability Levels, the differences are most significant in Generalization and Evaluation. At the same time, it can also be seen from the table that the pre-test mean of the total score is smaller than the post-test mean.

### **4.3. Interview**

This study conducted interviews with students to further understand how students play a role in self-exploration and group cooperation, as well as their mastery of relevant knowledge. More than 80% of students think they perform well under the maker education model. More than half of the students believe that under the maker education model for CT training, they can perform excellently in problem-solving and group cooperation.

## **5. Discussion and Conclusion**

This study introduces the cultivation of CT in maker education, takes junior high school students as the research object, and constructs a maker teaching model oriented to the cultivation of CT based on constructivism theory, using Mind+ and McQueen educational robots for teaching practice. After two rounds of teaching practice and analysis of data and materials, the research results are now summarized:

### **(1) The maker education model oriented towards cultivating CT has a positive effect on improving students' CT ability.**

This study used a maker education framework for CT cultivation. Analysis of students' CT pre-tests and post-tests, self-evaluation forms, and interviews revealed that the relevant teaching improved students' CT abilities, with significant gains in Generalization and Evaluation. Interviews suggested that in maker activities, students collect information, experiment, and adjust plans, naturally enhancing their inductive abilities. Maker projects, focused on problem-solving, require students to evaluate different solutions and choose the best approach, developing their evaluation and decision-making skills. Therefore, when designing maker education activities, teachers should guide students to focus on the learning and exploration process, not just the final results. Teachers should also facilitate reflection and sharing of learning experiences at the end of activities to further improve CT skills. However, no significant improvement was found in students' algorithmic thinking. This may be due to their prior exposure to similar programming tasks and a basic grasp of the course algorithms. As a result, students focused on practical operations and overlooked algorithm learning during maker activities. To address this, teachers can provide targeted guidance and algorithm resources to students who need them, encouraging deeper exploration of algorithmic concepts.

### **(2) Maker education can enhance students' attention and interest in information technology courses.**

It can be seen from the interviews that students like this type of maker courses very much. In the traditional information technology teaching process, teachers often only impart basic knowledge and skills to students, resulting in students' lack of enthusiasm and interest in learning. However, if Mind+ is combined with McQueen educational robots, students can acquire corresponding knowledge and skills through creative projects, teamwork, etc., thereby improving their problem-solving abilities and stimulating their interest in information technology courses. Therefore, we can integrate more courses related to maker education into junior high school teaching, or adopt similar teaching methods to effectively promote the overall development of students.

Herein, this study explores the impact of the maker education model on the CT levels of junior high school students. In addition, recommendations for educational practice are provided. It is hoped that the content of this research can help the development of maker education in the future.

### Acknowledgements

This paper is supported by the National Key R&D Program of China under Grant (2022YFC3303600), Zhejiang Provincial Philosophy and Social Sciences Planning Project (24NDJC191YB), and the Zhejiang Key Laboratory of Intelligent Education Technology and Application.

### References

- Fosnot, C. T., & Perry, R. S. (1996). Constructivism: A psychological theory of learning. *Constructivism: Theory, perspectives, and practice*, 2(1), 8-33.
- Jiang, H., Islam, A. A., Gu, X., & Guan, J. (2024). How do thinking styles and STEM attitudes have effects on computational thinking? A structural equation modeling analysis. *Journal of Research in Science Teaching*, 61(3), 645-673.
- Li, Z., & Oon, P. T. (2024). The transfer effect of computational thinking (CT)-STEM: a systematic literature review and meta-analysis. *International Journal of STEM Education*, 11(1), 44.
- Martin, D. A., Curtis, P., & Redmond, P. (2024). Primary school students' perceptions and developed artefacts and language from learning coding and computational thinking using the 3C model. *Journal of Computer Assisted Learning*, 40(4), 1616-1631.
- Selby, C., & Woollard, J. (2013). Computational thinking: the developing definition.
- Sun, L., Guo, Z., & Hu, L. (2023). Educational games promote the development of students' computational thinking: a meta-analytic review. *Interactive Learning Environments*, 31(6), 3476-3490.
- Uzumcu, O., & Bay, E. (2021). The effect of computational thinking skill program design developed according to interest driven creator theory on prospective teachers. *Education and Information Technologies*, 26(1), 565-583.
- Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, 49(3), 33-35.

# Evaluation of ChatGPT's Implementation in Undergraduate Java Programming Tutorials

Zilin Wang<sup>1</sup>, Di Zou<sup>2\*</sup>,  
 Lap-Kei Lee<sup>1</sup>, Haoran Xie<sup>3</sup>, Fu Lee Wang<sup>1</sup>  
<sup>1</sup>Hong Kong Metropolitan University  
<sup>2</sup>The Hong Kong Polytechnic University  
<sup>3</sup>Lingnan University  
 \*daisy.zou@polyu.edu.hk

**Abstract:** This study investigates the impact of integrating ChatGPT into undergraduate Java programming tutorials, focusing on its effects on students' knowledge acquisition, motivation, and self-efficacy. A quasi-experimental design was employed over a seven-week period, involving 53 undergraduate students divided into a ChatGPT-centered learning group (CCLG) and a teacher-centered learning group (TCLG). Both groups completed programming tasks, assessments, and surveys to measure learning outcomes. While the results indicated no statistically significant improvement in programming knowledge, students in the ChatGPT-centered group demonstrated significantly higher motivation and self-efficacy compared to their peers in the traditional instructional setting. These findings suggest that ChatGPT can enhance engagement and confidence in programming education, though its direct impact on knowledge acquisition may be limited within short-term interventions. The study highlights potential challenges, including students' adaptation to AI-assisted learning and limitations in foundational programming skills. Future research should explore strategies for optimizing ChatGPT's implementation, addressing digital literacy gaps, and evaluating its long-term effectiveness in programming education.

**Keywords:** Programming Education, ChatGPT, Educational Technology, Learning Motivation, Self-Efficacy

## 1. Introduction

OpenAI's ChatGPT is an advanced language model based on the GPT (Generative Pre-trained Transformer) framework. By utilizing deep learning algorithms, it can interpret and produce human-like text with remarkable accuracy. Trained on an extensive dataset of internet-based content, ChatGPT demonstrates significant potential in educational applications (Liang et al., 2021). Research findings indicate that ChatGPT can enhance student learning outcomes (Cooper, 2023), offer personalized learning pathways (Ghanizadeh et al., 2015), and support collaborative and communicative learning environments (Hinton & Wagemans, 2023). These insights highlight ChatGPT's value as an educational tool.

Despite its potential, assessing ChatGPT's effectiveness within specific educational contexts remains essential. Certain challenges must be addressed, such as concerns over its generalizability and lack of subject-specific expertise risks of misinformation limited human interaction capabilities (Watters & Lemanski, 2023) and ethical issues like its potential misuse for academic dishonesty (Alshurafat et al., 2024). As ChatGPT reshapes traditional educational practices, it becomes increasingly important to explore strategies for optimizing its use in real-world settings.

To contribute to this discussion, this study investigated the impact of ChatGPT in a Java programming tutorial over seven weeks. A total of 53 undergraduate students were divided into two groups: one engaged in ChatGPT-centred learning (CCLG) and the other in teacher-centred instruction (TCLG). Participants completed tutorials, knowledge assessments, and surveys to evaluate their learning achievements, motivation, and self-efficacy.

## 2. Programming Education

Programming education is essential in higher education, equipping students with computational thinking skills vital for the AI era (Botirovich et al., 2020). However, mastering programming requires complex cognitive abilities, including syntax selection, debugging, problem-solving, and communication (Coşkunserçe, 2023). Traditional teaching methods often emphasize rote learning, leading to low motivation and self-efficacy among students (Botirovich et al., 2020; Konecki, 2014). Teacher-centered learning offers a structured approach by guiding students through coding concepts with direct instruction, enhancing engagement (Elen et al., 2007). However, its effectiveness is limited by the lack of immediate, personalized feedback, particularly during debugging.

To address this, this study explores the use of ChatGPT as a programming assistant. By providing real-time error detection and tailored feedback, ChatGPT can support students in overcoming coding challenges, fostering a more engaging learning experience.

This study investigates:

- (1) Does a ChatGPT-centered approach improve students' knowledge, motivation, and self-efficacy compared to traditional teaching?
- (2) Does prior Java proficiency significantly impact on learning outcomes, motivation, and self-efficacy?

### 3. ChatGPT-based Java Programming Tutorial

CCLG in this study utilized ChatGPT as a central tool to assist students in completing programming tasks, while the teacher's role was primarily to deliver learning objectives and monitor the class. Figure 1 reports the situation in which the student was using the university ChatGPT web portal to complete the programming task.

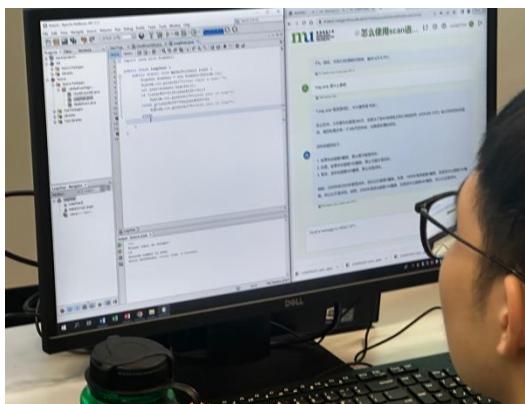


Fig.1 ChatGPT-centered programming learning

Each tutorial followed a structured format consisting of three phases. In the first phase, the teacher introduced the background and context of the programming tasks and explained the objectives in detail. For example, the teacher might assign tasks such as: "Create a game that allows the user to guess an integer number until their answer is correct" or "Create a function that allows the user to determine their grade point when they input a score." While the teacher introduced the tasks, students quickly listed the programming syntax they would need to use, such as "for loop," "if statement," and "switch statement." This helped students focus on the relevant concepts and tools required to complete the tasks.

In the second phase, students worked on the coding tasks with the support of ChatGPT. They were encouraged to use ChatGPT to translate task instructions into code, define programming syntax they were unfamiliar with, and debug errors in their code. For instance, a student working on the grade-point function task might consult ChatGPT to understand how to implement conditional logic to map score ranges to grade points. During this stage, the teacher moved around the classroom to provide technical support as needed, ensuring that any unresolved issues could be addressed promptly.

The final phase involved a reflective activity designed to deepen students' understanding of both the programming process and their problem-solving strategies. After reviewing the students' completed tasks, the teacher posed reflective

questions, such as: “What problems did you encounter, and how did you overcome them?” “Can you explain your problem-solving logic?” and “Did you use the same solutions as ChatGPT? If not, which solution do you think is better, and why?” These questions encouraged students to critically evaluate their own approaches and compare them to the solutions suggested by ChatGPT, fostering metacognitive skills and a deeper understanding of programming concepts.

#### **4. Methodology**

##### **4.1. Participants**

Our study initially aimed to recruit 60 sophomore students from a university in Hong Kong. By the end of the experiment, 53 students had completed all the learning activities and tests and were thus included in the final sample. The participants, aged between 18 and 26, were all enrolled in a Java Programming Fundamentals course. According to their responses in the biographical survey, the students had one to two years of programming education experience. However, most had minimal exposure to technology-enhanced learning, AI-based learning, or prompt training.

The participants were randomly assigned to two groups: a control group (TCLG, N = 27) and an experimental group (CCLG, N = 26).

##### **4.2. Experimental Procedures**

In week one, participants completed a 70-minute pre-test assessing their Java knowledge and a 20-minute pre-questionnaire on learning motivation and self-efficacy.

From weeks 2 to 9, students attended weekly two-hour programming tutorials in computer labs. Both groups received identical learning materials but followed different instructional methods. The control group engaged in a teacher-centred approach, where the instructor guided students through tasks, provided explanations, and addressed common coding issues. In contrast, the experimental group utilized a ChatGPT-based learning method.

In week 10, all students took a post-test to assess their learning outcomes and completed post-questionnaires.

##### **4.3. Instruments**

This study utilized a pair of programming knowledge tests and questionnaires to evaluate learning outcomes and changes in motivation and self-efficacy. The pre- and post-programming knowledge tests consisted of four questions, with a total maximum score of 50. These included short-answer questions, multiple-choice questions, and code-writing tasks.

The pre- and post-questionnaires were identical and comprised two sections: motivation and self-efficacy. Questions one to six assessed motivation, using items adapted from Wang and Chen (2010) framework on the influence of game strategies on learning motivation. For example, “When I have the opportunity, I choose course assignments that I can learn from even if they don’t guarantee a good grade.” Responses were rated on a five-point Likert scale (1 = strongly disagree, 5 = strongly agree), with higher scores indicating stronger intrinsic and extrinsic motivation.

Questions seven to fourteen measured self-efficacy, defined as an individual’s belief in their ability to perform tasks or achieve goals (Garcia & Pintrich, 2023). This section consisted of eight items adapted from Pintrich’s (1991) self-efficacy questionnaire. An example statement is: “I’m confident I can understand the most complex material presented by the instructor in this course.” Responses were also rated on a five-point Likert scale, with higher scores reflecting greater confidence in programming-related tasks.

We initially conducted reliability analyses and the Cronbach’s Alpha values for both 14-item questionnaires, along with the Cronbach’s Alpha Based on Standardized Items, exceed .8 and approach 90%, indicating that the scales possess high internal consistency. This high level of internal consistency points to a significant degree of reliability in the measurements.

##### **4.4. Data Analysis**

At the outset, we conducted a series of Shapiro-Wilk tests to assess the normality of the post-knowledge-test scores and post-questionnaire responses for both groups. The results indicated that only the control group’s post-knowledge-test scores were normally distributed ( $p = .377 > .05$ ). Consequently, we primarily employed

non-parametric tests for data analysis. Firstly, we used Wilcoxon Signed-Rank tests to compare the paired pre- and post-test scores within each group. This allowed us to assess the changes in scores over time for the same participants. Secondly, we employed Mann-Whitney tests to compare the differences between the two groups. Lastly, we utilized Quade's Ranking-Based Covariance Analysis to further confirm the impact of the learning methods on learning performance, while controlling for specific covariates.

## 5. Results

### 5.1. Results of Students' Knowledge Test

In this study, pre-test and post-test scores were collected for both groups. The descriptive statistics for these scores are reported in Table 1. The results of the Wilcoxon signed-rank test indicated no significant differences between the pre-test and post-test scores for either group, with both p-values exceeding 0.05. However, it is noteworthy that the mean score of the experimental group demonstrated a slight increase, whereas the control group exhibited a slight decrease.

Table 1. The descriptive result of students' pre- and post-test scores of Java programming knowledge test

	Pre-test		Post-test	
	Mean	SD	Mean	SD
Control group	37.12	6.09	36.38	6.52
Experimental group	31.31	10.08	31.44	10.95

### 5.2. Results of Students' Motivation

The Wilcoxon signed-rank tests revealed that the experimental group's post-test scores for motivation were significantly higher than their pre-test scores ( $p = .000 < .05$ ), while no significant differences were found in the control group ( $p = .762 > .05$ ). The results of the Mann-Whitney tests further suggested that the experimental group's post-test scores were significantly higher than those of the control group ( $Z = -3.311, p < .001$ ). These findings indicate that, compared to the teacher-centered learning method, the ChatGPT-centered learning method significantly improved students' motivation. To further confirm whether the learning method significantly impacted students' motivation scores after controlling for their pre-knowledge-test scores, we conducted Quade's ranking-based covariance analysis. The results, as shown in Table 2, indicated a significant effect of the group variable on post-test motivation scores,  $F (1, N-2) = 16.541, p < .001$ . This suggests that the experimental group demonstrated significantly higher motivation scores compared to the control group after adjusting for pre-test knowledge scores.

Table 2. The Quade's ranking-based covariance analysis for students' post-motivation scores

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2905.063 <sup>a</sup>	1	2905.063	16.541	< .001***
Intercept	1.034	1	1.034	.006	.939
Learning method	2905.063	1	2905.063	16.541	< .001***
Error	8957.010	51	175.628		
Total	11862.073	53			
Corrected Total	11862.073				

\*\*\* $p < .001$

### 5.3. Results of Students' Self-Efficacy

The Wilcoxon signed-rank tests revealed that the experimental group's post-test scores for motivation were significantly higher than their pre-test scores ( $p = .000 < .05$ ), while no significant differences were found in the control group ( $p = .762 > .05$ ). The results of the Mann-Whitney tests further suggested that the experimental group's post-test scores were significantly higher than those of the control group ( $Z = -3.311, p < .001$ ). These findings indicate that, compared to the teacher-centered learning method, the ChatGPT-centered learning method significantly improved

students' motivation. To further confirm whether the learning method significantly impacted students' motivation scores after controlling for their pre-knowledge-test scores, we conducted Quade's ranking-based covariance analysis. The results, as shown in Table 2, indicated a significant effect of the group variable on post-test motivation scores,  $F(1, N-2) = 16.541$ ,  $p < .001$ . This suggests that the experimental group demonstrated significantly higher motivation scores compared to the control group after adjusting for pre-test knowledge scores.

Table 3. The Quade's ranking -based covariance analysis for students' post-self-efficacy scores

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2853.975a	1	2853.975	16.578	< .001***
Intercept	1.016	1	1.016	.006	.939
Learning method	2853.975	1	2853.975	16.578	< .001***
Error	8780.095	51	172.159		
Total	11634.070	53			
Corrected Total	11634.070	52			

\*\*\* $p < .001$

## 6. Discussion and Conclusion

This study investigated the impact of a ChatGPT-centred learning method on students' knowledge acquisition, motivation, and self-efficacy in Java programming education. While the results revealed a significant improvement in motivation and self-efficacy for students in the experimental group, the impact on knowledge acquisition was minimal.

The insignificant differences in students' pre-test and post-test scores suggest that while ChatGPT enhanced engagement and confidence, it did not substantially improve their coding proficiency within the study period. One possible explanation is the small sample size (Carlson & Schmidt, 1999), which may have limited the ability to detect statistically significant improvements. In studies with small groups, minor variations in performance can be amplified, leading to non-significant findings even when an actual effect exists. Future studies with larger sample sizes could help clarify whether ChatGPT meaningfully enhances programming knowledge acquisition.

Another contributing factor may be students' limited foundational digital skills. A lack of prior digital literacy can hinder learners' ability to navigate programming environments effectively, thereby reducing their ability to leverage ChatGPT optimally (Mokaya et al., 2022). Students with lower self-efficacy in technology-related domains may struggle with programming concepts, leading to minimal measurable improvement in knowledge despite increased motivation. Addressing these gaps through introductory digital literacy training before engaging with AI-based tools may enhance the effectiveness of such interventions.

Additionally, students' familiarity with traditional teacher-centred learning approaches could have influenced the results. Many learners may have been accustomed to direct instruction and guided problem-solving rather than self-directed learning facilitated by ChatGPT. This shift in instructional methods may have required an adaptation period, limiting immediate knowledge gains. Some students may also have held negative attitudes toward ChatGPT, either due to skepticism about AI-generated feedback or discomfort in relying on a non-human instructor. Such resistance could have affected engagement with the tool and subsequently impacted their learning outcomes. Future studies should explore strategies to gradually integrate ChatGPT into programming education, ensuring sufficient guidance and scaffolding to support students in transitioning to AI-assisted learning.

Despite these limitations, the study aligns with prior research indicating that generative AI enhances motivation and self-efficacy (Mun, 2024; Woo et al., 2024). ChatGPT's ability to provide instant, contextually relevant feedback creates a more interactive learning experience, reducing anxiety and fostering a sense of achievement (Mun, 2024). By minimizing students' fear of failure, ChatGPT enables them to engage with programming in a more enjoyable and stress-free manner (Wu et al., 2024). In terms of self-efficacy, students in the experimental group exhibited greater

confidence in their programming abilities, possibly due to the ease of accessing ChatGPT's explanations and debugging assistance. When students can quickly identify errors and receive immediate feedback, they develop a stronger belief in their ability to solve problems independently (Mun, 2024). However, the challenge remains in ensuring that this increased confidence translates into measurable improvements in programming proficiency.

This study also has limitations. The difficulty level of the knowledge tests may have contributed to low scores, making it harder to capture improvements in students' learning outcomes. Future research should consider adjusting the complexity of assessments to better align with students' abilities. Additionally, the absence of qualitative data limits deeper insights into students' perceptions of ChatGPT-assisted learning. Incorporating interviews or focus groups could provide a more comprehensive understanding of how learners interact with AI tools and what challenges they face.

Future research should address these gaps by incorporating larger sample sizes, refining assessment methods, and integrating qualitative research to capture student experiences more holistically. Additionally, exploring long-term effects and cross-disciplinary applications of ChatGPT in education could provide further evidence of its potential as a learning tool. By refining AI-assisted instructional strategies, educators can better leverage ChatGPT to enhance programming education and support diverse learners effectively.

### Acknowledgements

We thank all the people who wrote previous versions of this document.

### References

- Alshurafat, H., Al Shbail, M. O., Hamdan, A., Al-Dmour, A., & Ensour, W. (2024). Factors affecting accounting students' misuse of chatgpt: an application of the fraud triangle theory. *Journal of Financial Reporting and Accounting*, 22(2), 274–288.
- Botirovich, B. D., Ergashevich, T. K., Eshmirzayevna, M. U., Kholboyevich, A. S., Fayzullayevna, K. N., & Mukhiddinovna, A. U. (2020). The importance of teaching algorithms and programming languages in the creation of electronic education resources. *Journal of Critical Reviews*, 7(11), 365–368.
- Carlson, K. D., & Schmidt, F. L. (1999). Impact of experimental design on effect size: Findings from the research literature on training. *Journal of Applied Psychology*, 84(6), 851.
- Cooper, G. (2023). Examining science education in chatgpt: An exploratory study of generative artificial intelligence. *Journal of Science Education and Technology*, 32(3), 444–452.
- Coşkunserçe, O. (2023). Comparing the use of block-based and robot programming in introductory programming education: Effects on perceptions of programming self-efficacy. *Computer Applications in Engineering Education*, 31(5), 1234–1255.
- Elen, J., Clarebout, G., Léonard, R., & Lowyck, J. (2007). Student-centred and teacher-centred learning environments: What students think. *Teaching in Higher Education*, 12(1), 105–117.
- Garcia, T., & Pintrich, P. R. (2023). Regulating motivation and cognition in the classroom: The role of self-schemas and self-regulatory strategies. In *Self-regulation of learning and performance* (pp. 127–153). Routledge.
- Ghanizadeh, A., Razavi, A., & Jahedizadeh, S. (2015). Technology-enhanced language learning (TELL): A review of resources and upshots. *International Letters of Chemistry, Physics and Astronomy*, 54, 73–87.
- Hinton, M., & Wagemans, J. H. M. (2023). How persuasive is AI-generated argumentation? An analysis of the quality of an argumentative text produced by the GPT-3 AI text generator. *Argument & Computation*, Preprint, 1–16. <https://doi.org/10.3233/AAC-210026>
- Konecki, M. (2014). Problems in programming education and means of their improvement. *DAAAM International Scientific Book*, 2014, 459–470.

- 
- Liang, J.-C., Hwang, G.-J., Chen, M.-R. A., & Darmawansah, D. (2021). Roles and research foci of artificial intelligence in language education: an integrated bibliographic analysis and systematic review approach. *Interactive Learning Environments*, 1–27.
- Mokaya, M., Kyallo, F., Vangoitsenhoven, R., & Matthys, C. (2022). Clinical and patient-centered implementation outcomes of mHealth interventions for type 2 diabetes in low-and-middle income countries: a systematic review. *International Journal of Behavioral Nutrition and Physical Activity*, 19(1), 1.
- Mun, I. B. (2024). A study of the impact of ChatGPT self-efficacy on the information seeking behaviors in ChatGPT: the mediating roles of ChatGPT characteristics and utility. *Online Information Review*.
- Pintrich, P. R. (1991). A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ).
- Wang, L., & Chen, M. (2010). The effects of game strategy and preference-matching on flow experience and programming performance in game-based learning. *Innovations in Education and Teaching International*, 47(1), 39–52.
- Watters, C., & Lemanski, M. K. (2023). Universal skepticism of ChatGPT: a review of early literature on chat generative pre-trained transformer. *Frontiers in Big Data*, 6, 1224976.
- Woo, D. J., Wang, D., Guo, K., & Susanto, H. (2024). Teaching EFL students to write with ChatGPT: Students' motivation to learn, cognitive load, and satisfaction with the learning process. *Education and Information Technologies*, 1–28.
- Wu, T.-T., Lee, H.-Y., Li, P.-H., Huang, C.-N., & Huang, Y.-M. (2024). Promoting self-regulation progress and knowledge construction in blended learning via ChatGPT-based learning aid. *Journal of Educational Computing Research*, 61(8), 3–31.

# Students' Perceptions and Usage Patterns of ChatGPT as an Automated Writing Evaluation (AWE) Tool

Linqian Ding<sup>1\*</sup>, Di Zou<sup>2</sup>, Lucas Kohnke<sup>3</sup>

<sup>1,3</sup> The Education University of Hong Kong

<sup>2</sup> The Hong Kong Polytechnic University

\* 15080791126@163.com

**Abstract:** The emergence of Generative AI (GAI) has significantly influenced writing evaluation, with recent literature primarily focusing on comparing GAI feedback or scores with those provided by humans. However, fewer studies have explored students' perspectives, which play a crucial role in the effectiveness of feedback. To address this gap, this study investigates how EFL students interact with ChatGPT as an Automated Writing Evaluation (AWE) tool and their perceptions of its usefulness in improving academic writing. The study involved 47 graduate students at a university in Hong Kong, and data were collected through a questionnaire survey to examine their usage patterns and perceptions of ChatGPT feedback. The findings reveal that students primarily used ChatGPT during the revision stage for grammar and vocabulary enhancement, while its use for higher-order writing tasks was limited. Moreover, students generally viewed ChatGPT as trustworthy, useful, and easy to use, though there were concerns about information security and ethical implications. Despite these challenges, students expressed a willingness to continue using ChatGPT, suggesting its potential as a supplementary tool for academic writing instruction. This study highlights the need to refine GAI tools for better writing support and ethical use.

**Keywords:** Generative AI, ChatGPT, Automated Writing Evaluation (AWE)

## 1. Introduction

The advent of technology has transformed language education significantly, particularly in the field of writing instruction (Wen & Walters, 2022). One notable advancement is the implementation of automated writing evaluation (AWE) tools, which have evolved greatly since their inception (Ding & Zou, 2024). Originally, AWE tools assessed and scored texts, but technological advancements, such as natural language processing and artificial intelligence (AI) have enhanced their ability to provide more accurate, customized feedback (Bai & Hu, 2017). With AWE systems gaining popularity in writing instruction, research has shown that they enhance writing performance across various dimensions and users generally maintain a positive attitude toward these technologies (Barrot, 2021). However, AWE systems have received criticism. Some argue that AWE feedback can be inaccurate, generic, and confusing, and it lacks deep analysis of content and organization (Han & Sari, 2022). With the emergence of generative AI technologies, such as ChatGPT, it is believed they have the potential overcome these limitations (Ding & Zou, 2024).

As an AI-powered chatbot, ChatGPT can mimic human conversation and may revolutionize AWE technologies by providing more effective evaluations and feedback (Mizumoto & Eguchi, 2023). There have been explorations of using ChatGPT as an AWE tool, but there is a lack of comprehensive studies on students' usage patterns and perceptions. This study aims to fill this gap by investigating the student usage patterns and perceptions towards ChatGPT feedback. The study is guided by the following research questions:

RQ1: What usage patterns emerge when students use ChatGPT for writing evaluation?

RQ2: What are students' perceptions of ChatGPT feedback on writing evaluation?

## 2. Methodology

The study involved 47 graduate students majoring in English Education at a university in Hong Kong. The majority of participants were female (n=43), aged between 18-34 years (n=45). All participants were native Chinese speakers with more than 10 years of English learning experience in China from primary to undergraduate education. They were enrolled in the same English Writing course which ran for 16 weeks with one 120-minute session per week.

To address the two research questions, a questionnaire survey was designed to collect data on students' usage patterns and perceptions. The questionnaire included two sections: (1) Usage of ChatGPT as an AWE tool and (2) Perceptions of using ChatGPT as an AWE tool. The first section comprised six multiple-choice questions and the second section consisted of 27 closed-ended items measured on a five-point Likert scale (1 = strongly disagree, 5 = strongly agree) and three multiple-choice questions. The items were adapted from established scales and prior research on AWE feedback (e.g., Davis, 1989; Huang & Renandya, 2020) to probe students' (a) perceived trust in ChatGPT's feedback, (b) perceived usefulness of the feedback, (c) perceived ease of use, (d) perceived value of the feedback for revision and English writing performance, (e) AWE system anxiety, (f) behavioral intention to use ChatGPT as an AWE tool, and (g) beliefs about the ethicality of such usage. The Cronbach's  $\alpha$  coefficients for each were 0.87, 0.94, 0.89, 0.90, 0.82, and 0.84, respectively, indicating good reliability.

This quasi-experimental study was integrated into the course in week 5 with a 60-minute training session on accessing ChatGPT and using it for writing feedback. The session covered uploading files, inputting instructions, and refining feedback through follow-up questions. In week 6, students completed a 60-minute in-class writing task based on an IELTS writing topic. After submitting their first draft, students were instructed to use ChatGPT to receive feedback on their assignment. To facilitate this process, they were provided with sample prompts designed to elicit specific and constructive feedback. Examples were presented below:

“Assume the role of a language teacher. I am writing an IELTS essay. Please provide feedback on my vocabulary/grammar/ structure”; “Explain any mistakes in my writing”; “Provide revised sentence examples/ synonyms”.

The entire interaction process was conducted over 60 minutes during class. Following this, they completed the questionnaire to report how they used the tool and share their perceptions and experiences with it.

### **3. Results**

#### **3.1. Usage pattern of ChatGPT as AWE tool**

The results indicate that a significant portion of the students in the class frequently used ChatGPT to provide writing feedback in their coursework. A majority of the students (n=34) reported using ChatGPT "sometimes" for writing feedback. Conversely, 19.1% of the participants seldom used these tools, with only two students reporting that they "never" used them. The reasons cited by the two students were concerns about potential penalties for using such technology. As for the timing of ChatGPT usage, most students (n=40) used it after completing draft but before submission. Ten students used it during the planning stage, while six used it during drafting. Notably, none used ChatGPT after receiving grades and instructor feedback.

Moreover, students predominantly used ChatGPT to identify and correct grammatical errors or vocabulary mistakes (77.8%), improve vocabulary or sentence structure (75.6%), and provide feedback on the structure, organization, and logic of their writing (64.4%). Over half of the respondents also used ChatGPT to refine entire paragraphs or complete texts,. Less frequently, students used ChatGPT to score their writing, assess originality, prevent plagiarism, or receive suggestions on writing style. One student mentioned using ChatGPT to suggest an outline for their assignment.

Regarding the impact of ChatGPT's feedback on the revision process, most students reported making changes based on the feedback received. Specifically, 35 students made significant or moderate changes to their writing, while eight students made minor changes. In contrast, two students reported that the feedback did not influence their revision process, either because they acknowledged the feedback without implementing changes or they found it unhelpful. In terms of the language used to interact with ChatGPT, the majority of participants (n=29) communicated in English,

while 35.6% used both Chinese and English. Notably, despite their mother tongue being Chinese, none of the participants exclusively used Chinese to interact with ChatGPT.

### **3.2. Perceptions of ChatGPT as AWE tool**

Our findings indicate that students generally perceive ChatGPT as a reliable and useful AWE tool, though concerns remain regarding information security and ethical considerations (see Table 1). Regarding perceived trust, students expressed moderate confidence in ChatGPT's reliability ( $M = 3.51$ ,  $SD = 0.67$ ). However, concerns about information security were evident, with some students remaining cautious about data privacy and potential security risks. Both perceived usefulness ( $M = 3.81$ ,  $SD = 0.79$ ) and perceived ease of use ( $M = 3.70$ ,  $SD = 0.72$ ) were rated positively, which indicates that students found ChatGPT both beneficial for writing improvement and easy to use.

Regarding students' perceived value of the feedback, they found it useful not only for revising their current essays ( $M = 3.76$ ,  $SD = 0.71$ ) but also for improving their long-term English writing skills ( $M = 3.89$ ,  $SD = 0.73$ ). Specifically, most students found ChatGPT feedback particularly helpful for grammar improvement and vocabulary expansion. They also acknowledged its effectiveness in identifying writing problems in future writing practice and enhancing overall writing proficiency. However, their perceived value of the feedback on deeper-level writing aspects, such as organization and content, was relatively lower. These findings align with prior research on AWE tools, which suggests that students value AI-generated feedback primarily for surface-level corrections rather than higher-order writing improvements (Huang & Renandya, 2020). When asked about their intentions to continue using ChatGPT, students showed a moderately positive response ( $M = 3.50$ ,  $SD = 0.72$ ). This indicates that while students acknowledge and value the benefits of ChatGPT, they still have concerns about its limitations.

Table 1. Students' Perceptions of ChatGPT as AWE tool

	Mean	Standard Deviation
Perceived trust	3.51	0.67
Perceived usefulness	3.81	0.79
Perceived ease of use	3.70	0.72
Perceived value for revision	3.76	0.71
Perceived value for English writing performance	3.89	0.73
Anxiety	3.01	0.63
Behavioral intention	3.50	0.72

Regarding students' concerns, they exhibited moderate levels of anxiety when using ChatGPT for writing tasks ( $M = 3.01$ ,  $SD = 0.63$ ). Many students expressed concerns about the possibility of making irreversible mistakes, fearing that errors in their writing might not be accurately identified or appropriately corrected by the tool. Furthermore, some students felt intimidated by the evaluation process itself, as they were uncertain about how ChatGPT generates feedback and whether its suggestions truly reflect their writing proficiency. This underscores the need for explicit training on how to effectively interact with ChatGPT to minimize uncertainty and enhance user confidence.

As for ethical perceptions, students expressed diverse opinions. While only one student considered its use inherently unethical, nearly half of the students regarded it as ethically acceptable under any circumstances ( $n = 20$ ). Beyond this, many students perceived it as cheating in certain situations, 19 students believed it was unethical when explicitly prohibited, while 14 students felt it was unethical when students relied too heavily on its output without making any revisions. Therefore, it is essential for institutions to provide clear guidelines to help them understand what is considered appropriate and responsible use.

## **4. Discussion and conclusion**

This study examined how EFL students use ChatGPT for writing evaluation and their perceptions towards the tool. The findings suggest that students primarily use ChatGPT as a revision tool rather than for initial drafting or brainstorming. Most students rely on it for grammar and vocabulary enhancement, using it after completing their first

drafts but before final submission. However, fewer students engage with ChatGPT for higher-order writing tasks, such as improving argumentation or content organization. This indicates that while students find ChatGPT useful for surface-level corrections, they may perceive its feedback as less effective for deeper revisions, limiting its role in refining critical thinking and coherence in writing.

Students' perceptions of ChatGPT were generally positive, particularly regarding its usefulness and ease of use. However, concerns were raised about information security and ethical considerations. While many students viewed ChatGPT as a helpful writing aid, others expressed uncertainty about its appropriateness in academic settings, especially when its use was restricted or when AI-generated content was integrated without proper modification. Additionally, some students experienced moderate anxiety of using ChatGPT for writing evaluation, particularly regarding the reliability of ChatGPT's feedback and the potential risks of over-reliance on AI-generated suggestions.

In conclusion, this study underscores the need for both pedagogical and technological improvements in the integration of GAI tools in writing evaluation. Instructors should provide explicit training to help students critically engage with ChatGPT feedback, ensuring they use it as a support tool rather than a substitute for independent writing. Institutions must establish clear ethical guidelines to address concerns about academic integrity and responsible GAI usage. From a technological perspective, developers should enhance ChatGPT ability to provide more nuanced feedback on high-order aspects, such as argumentation, structure, and coherence.

## Reference

- Bai, L., & Hu, G. (2017). In the face of fallible AWE feedback: How do students respond? *Educational Psychology*, 37(1), 67–81. <https://doi.org/10.1080/01443410.2016.1223275>
- Barrot, J. S. (2021). Using automated written corrective feedback in the writing classrooms: Effects on L2 writing accuracy. *Computer Assisted Language Learning*, 1–24. <https://doi.org/10.1080/09588221.2021.1936071>
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340.
- Ding, L., & Zou, D. (2024). Automated writing evaluation systems: A systematic review of Grammarly, Pigai, and Criterion with a perspective on future directions in the age of generative artificial intelligence. *Education and Information Technologies*, 1–53. <https://doi.org/10.1007/s10639-023-12402-3>
- Han, T., & Sari, E. (2022). An investigation on the use of automated feedback in Turkish EFL students' writing classes. *Computer Assisted Language Learning*, 1–24. <https://doi.org/10.1080/09588221.2022.2067179>
- Huang, S., & Renandya, W. A. (2020). Exploring the integration of automated feedback among lower-proficiency EFL learners. *Innovation in Language Learning and Teaching*, 14(1), 15–26. <https://doi.org/10.1080/17501229.2018.1471083>
- Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics*, 2(2), 100050. <https://doi.org/10.1016/j.rmal.2023.100050>

# Interactive Effects of Graphic Organizer Strategy and Strategy Preference on Self-Paced Video Learning

Yu Ge <sup>1</sup>, Caixia Liu <sup>2</sup>, Jiumin Yang <sup>1\*</sup>

<sup>1</sup> Faculty in Artificial Intelligence in Education, Central China Normal University

<sup>2</sup> Beijing Bayi School

<sup>\*</sup> yjm@mail.ccnu.edu.cn

**Abstract:** This study explored the mutual influence of GO strategy (watching vs. generating) and strategy preference (with GO preference vs. without GO preference) on the learning outcomes (indicated by learning performance, intrinsic motivation, metacognition and cognitive load) when learning from self-paced videos. The results showed that when generating GO, GO preference increased participants' knowledge retention in both immediate and delayed performance, perceived value of intrinsic motivation and reduced participants' cognitive load. Additionally, when watching GO, GO preference increased participants' perceived value of intrinsic motivation and reduced participants' perceived pressure of intrinsic motivation and cognitive load. Furthermore, GO preference can improve participants' knowledge retention in delayed performance by improving their competence and learning confidence. The results suggested the positive effect of GO preference on the learning outcomes when learning self-paced video with generating and watching a GO.

**Keywords:** instructional videos, graphic organizer, strategy preference

## 1. Introduction

The rise of internet technologies has positioned video learning as a prominent educational tool. However, learners often passively consume video content, limiting opportunities for active knowledge construction—a critical component of meaningful learning (Fiorella & Mayer, 2016; Mohammadhassan et al., 2022). This passive engagement may reduce focus, persistence, and learning efficiency compared to traditional methods (Crook & Schofield, 2017). To address this, researchers emphasize the need for strategies that promote deeper cognitive processing (Yang et al., 2021).

Graphic organizers (GOs) as visual tools like concept maps and flowcharts that enable learners to represent hierarchical relationships and reorganize information spatially (Novak & Gowin, 1984). By structuring knowledge visually, GOs enhance comprehension, retention, and retrieval (Bean et al., 1986). While generating GOs is theorized to foster generative learning through active engagement, debates persist regarding its efficacy versus passively watching pre-made GOs. Proponents argue that generation induces deeper cognitive processing, whereas critics highlight potential cognitive overload risks (Fiorella & Mayer, 2016). Learner differences further complicate this dynamic. Strategy preference is defined as an individual's tendency to adopt specific learning approaches, which can significantly influences learning outcomes (Vos et al., 2011). For GOs, learners with a GO preference (habitual, self-directed use of such tools) may demonstrate superior performance, motivation, and metacognitive regulation due to strategy familiarity (Eielts et al., 2020). Conversely, non-preferring learners might experience cognitive strain when compelled to use GOs.

This study investigates the interactive effects of GO strategy (generating vs. watching) and GO preference on self-paced video learning outcomes, including performance, motivation, cognitive load and metacognition. We hypothesize that GO preference will amplify the benefits of generating GOs, as heightened motivation and metacognition may optimize generative processing without excessive cognitive demands. In contrast, preference may exert minimal influence when learners passively watch GOs. This aligns with prior work suggesting strategy efficacy depends on alignment with learner characteristics (Vos et al., 2011; Yang et al., 2021).

## 2. Method

## 2.1. Participants and design

One hundred and six Chinese undergraduates (aged 18-23, SDage = 1.06; 93 females) from diverse majors (e.g., Pedagogy, Information Engineering) participated voluntarily in a  $2 \times 2$  between-subjects design. Participants were divided into the with GO preference group and the without GO preference group (negative responses to both), then randomly assigned to either watch or generate a GO during video learning (the interface is shown in Figure 1-2). All participants provided informed consent and received 12-15 CNY compensation. This study received approval from the local ethics committee.

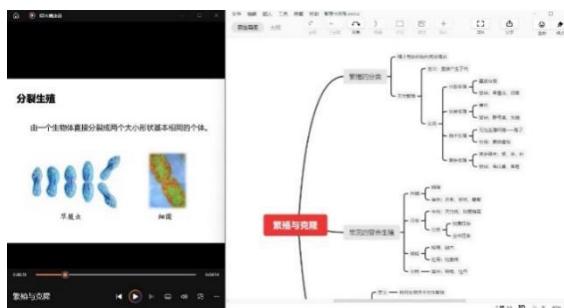


Fig.1 Screenshot of the learning interface of the watching GO group.

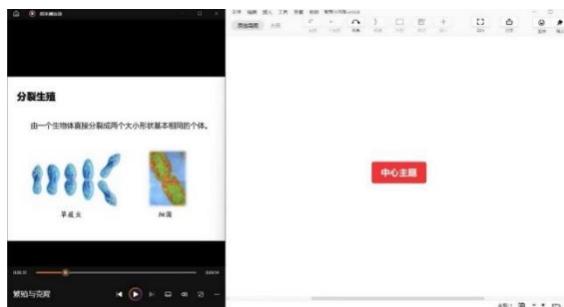


Fig.2 Screenshot of the learning interface of the generating GO group.

## 2.2. Learning material

The instructional video was adapted from the video lecture used by Pi et al. (2019). The topic was Reproduction and Cloning in biology. It was divided into three parts: classification of reproduction, vegetative reproduction and cloning. This video lasted about 5 minutes. Participants were allowed to control the video, but the maximum learning time was 15 minutes.

## 2.3. Measurements

The strategy preference questionnaire assessed participants' inclination toward GOs using two criteria: (1) interest in GOs (1 = "not at all true" to 5 = "very true"; participants rating 4-5 classified as "interested"), and (2) frequency of GO use (0 = "less than once a month" and 1 = "more than once a month"; participants rating 1 classified as "frequent"). The prior knowledge test evaluated baseline understanding with one fill-in-the-blank (1 point) and five multiple-choice questions (10 points; total = 11, Cronbach's  $\alpha = 0.62$ ). The learning performance test measured retention and transfer immediately and one week later. Retention included 17 fill-in-the-blanks (17 points; Cronbach's  $\alpha = 0.70$  and 0.73), while transfer combined seven multiple-choice (14 points), one fill-in-the-blank (1 point), and one short-answer question (5 points; total = 20, Cronbach's  $\alpha = 0.61$  for both). The intrinsic motivation scale comprised 12 items across four subscales (interest, competence, value and pressure) rated on a 7-point Likert scale (1 = "not at all true" to 7 = "very true"). Subscale scores were averaged (pressure items reverse-scored), with acceptable reliability (Cronbach's  $\alpha = 0.75$ -0.86). The cognitive load scale included three 7-point Likert items (total score averaged; Cronbach's  $\alpha = 0.79$ ). The judgment of learning (JOL) scale measured metacognitive confidence via two items: self-rated accuracy (0-100) and predicted test performance. Scores were averaged and Cronbach's  $\alpha = 0.91$ . Metacognition accuracy was derived from the absolute difference between JOL scores and actual performance.

## 2.4. Procedure

The experiment was conducted in a lab and lasted about 40 minutes. Firstly, the researchers introduced the basic information and procedure of the experiment to the participants and led them to sign the informed consent (5 min). Afterward, the participants filled out the demographic questionnaire (e.g., gender, age, and major) and the prior knowledge test (5 minutes). Subsequently, all participants were assigned to one of four groups and asked to watch or generate the GO while watching the instructional video (15 min). After learning, they completed the immediate learning performance test, the intrinsic motivation test and the cognitive load test (15 minutes). A week later, they completed the delayed learning performance test online.

## 3. Results

Data met normality (skewness and kurtosis within -2 to 2) and homogeneity of variance ( $p > .05$ ), justifying ANOVA. Pretests showed no group differences ( $F(3, 102) = 0.43, p = .730, \eta^2 = .013$ ).

Immediate knowledge retention showed a marginal main effect of strategy preference ( $F(1, 102) = 3.64, p = .059, \eta^2 = .034$ ) and an interaction effect ( $F(1, 102) = 4.54, p = .036, \eta^2 = .043$ ). Participants with GO preference outperformed without GO preference after generating GOs ( $MD = 2.39, p = .006$ ). Similarly, delayed knowledge retention showed a main effect of strategy preference ( $F(1, 102) = 6.30, p = .014, \eta^2 = .058$ ) and an interaction effect ( $F(1, 102) = 3.92, p = .050, \eta^2 = .037$ ). Participants with GO preference outperformed without GO preference after generating GOs ( $MD = 2.77, p = .002$ ). But knowledge transfer showed no significant effects ( $p > .050$ ).

Interest showed a main effect of strategy preference ( $F(1, 102) = 148.89, p < .001, \eta^2 = .593$ ) and a main effect of GO strategy ( $F(1, 102) = 12.24, p = .001, \eta^2 = .107$ ). Participants with GO preference and watched GOs both reported higher interest ( $MD = 1.35, p < .001; MD = 0.86, p = .001$ ). Competence only showed a main effect of strategy preference ( $F(1, 102) = 11.33, p = .001, \eta^2 = .100$ ). Participants with GO preference reported higher interest ( $MD = 1.35, p < .001$ ). Value showed a main effect of strategy preference ( $F(1, 102) = 144.81, p < .001, \eta^2 = .587$ ) and a marginal interaction effect ( $F(1, 102) = 3.17, p = .078, \eta^2 = .030$ ). Participants with GO preference reported higher value after both watching and generating GOs ( $MD = 1.40, p < .001; MD = 1.04, p < .001$ ). Pressure showed a main effect of strategy preference ( $F(1, 102) = 14.87, p < .001, \eta^2 = .127$ ) and an interaction effect ( $F(1, 102) = 5.44, p = .022, \eta^2 = .051$ ). Participants with GO preference reported lower pressure after watching GOs ( $MD = 1.30, p < .001$ ).

JOL only showed a main effect of strategy preference ( $F(1, 102) = 18.07, p < .001, \eta^2 = .151$ ) and metacognitive accuracy showed no effects ( $p > .050$ ). Participants with GO preference showed greater confidence ( $MD = 13.06, p < .001$ ).

Cognitive load showed a main effect of strategy preference ( $F(1, 102) = 36.42, p < .001, \eta^2 = .263$ ) and a marginal interaction effect ( $F(1, 102) = 3.49, p = .065, \eta^2 = .033$ ). Participants with GO preference reported lower cognitive load after both watching and generating GOs ( $MD = 1.24, p < .001; MD = 0.65, p = .004$ ).

Mediation analysis (PROCESS v4.0) identified competence ( $a \times b = 0.51, 95\% \text{ CI } [0.0930, 1.1066]$ ) and JOL ( $a \times b = 1.21, 95\% \text{ CI } [0.5397, 1.9974]$ ) as mediators between GO preference and delayed retention, suggesting their roles in enhancing retention through intrinsic motivation and metacognitive confidence.

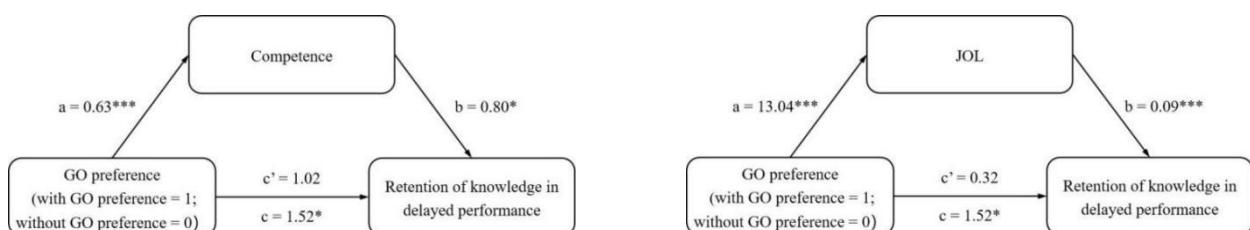


Fig.3 Results of the mediation analysis (\* $p < .05$ ; \*\*\* $p < .001$ ).

## 4. Discussion and Conclusion

This study demonstrates that participants' GO strategy preference significantly enhances learning outcomes in self-paced video instruction. Participants with GO preference showed superior knowledge retention, higher intrinsic motivation, and reduced cognitive load after generating GOs. When watching GOs, participants with GO preference exhibited stronger motivation and lower cognitive demands. Notably, GO preference improved delayed retention through enhanced competence and confidence development. These results may be due to the fact that participants without GO preference likely struggled with GO generation's working memory demands (Wang et al., 2021). While generating GOs remains cognitively challenging even for participants with GO preference (Fiorella & Mayer, 2015), our findings emphasize the motivational benefits of strategy alignment (Hidi & Renninger, 2006). The absence of transfer differences suggests future research should employ more sensitive assessments across task difficulty levels (Kalyuga et al., 2003). These results underscore the pedagogical value of cultivating GO preference and metacognitive awareness (Mazancieux et al., 2023) in self-directed learning environments.

### Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant [62177027].

### References

- Bean, T. W., Singer, H., Sorter, J., & Frazee, C. (1986). The effect of metacognitive instruction in outlining and graphic organizer construction on students' comprehension in a tenth-grade world history class. *Journal of Reading Behavior, 18*(2), 153-169.
- Crook, C., & Schofield, L. (2017). The video lecture. *The Internet and Higher Education, 34*, 56-64.
- Eilts, C., Pouw, W., Ouwehand, K., Van Gog, T., Zwaan, R. A., & Paas, F. (2020). Co-thought gesturing supports more complex problem solving in subjects with lower visual working-memory capacity. *Psychological research, 84*, 502-513.
- Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational psychology review, 28*, 717-741.
- Fiorella, L., & Mayer, R. E. (2015). The relative benefits of learning by teaching and teaching expectancy. *Contemporary Educational Psychology, 43*, 1-7.
- Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational psychologist, 41*(2), 111-127.
- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*(1), 23-31.
- Mazancieux, A., Pereira, M., Faivre, N., Mamassian, P., Moulin, C. J., & Souchay, C. (2023). Towards a common conceptual space for metacognition in perception and memory. *Nature Reviews Psychology, 2*(12), 751-766.
- Mohammadhassan, N., Mitrovic, A., & Neshatian, K. (2022). Investigating the effect of nudges for improving comment quality in active video watching. *Computers & Education, 176*, 104340.
- Vos, N., Van Der Meijden, H., & Denessen, E. (2011). Effects of constructing versus playing an educational game on student motivation and deep learning strategy use. *Computers & education, 56*(1), 127-137.
- Wang, X., Mayer, R. E., Zhou, P., & Lin, L. (2021). Benefits of interactive graphic organizers in online learning: Evidence for generative learning theory. *Journal of Educational Psychology, 113*(5), 1024.
- Yang, J., Zhang, Y., Xu, K., & Pi, Z. (2021). The Effects of Learning Strategies on Video Learning: Imagination, Drawing and Self-explanation Strategies. *e-Education Research, 42*(10):40-47.

# Empowering Socratic Teaching Through Large Language Models: New Opportunities and Challenges in Intelligent Tutoring Systems

Yueru Lang 1\*, Siyang Liu 1, Xiangen Hu 2\*, Shaoying Gong 1

1 School of Psychology, Central China Normal University, Wuhan, China

2 Department of Applied Social Sciences, The Hong Kong Polytechnic University, Hong Kong

\*langyueru\_michelle@163.com; xiangen.hu@polyu.edu.hk

**Abstract:** This paper explores the necessity and potential of empowering Socratic teaching by large language models in intelligent tutoring systems to enable adaptive conversations and facilitate deeper engagement. Using Socratic Playground for Learning as an example, this paper presents the system's architecture and functionality of two components: lesson creation and Socratic interaction. A pilot study was conducted with 30 participants over five days, revealing that SPL significantly improved their performance and was well-accepted by users. In addition, the paper highlights the key opportunities and challenges that need to be addressed for effective implementation.

**Keywords:** Socratic teaching, intelligent tutoring system, large language model, opportunity, challenge

## 1. Introduction

Socratic teaching has long been valued for its educational benefits in cultivating deeper thinking (Elkowitz, 2021). However, applying face-to-face Socratic teaching has been challenging due to the high skill requirement for instructors and the difficulty sustaining meaningful interactions in classroom teaching (Dalim et al., 2022; Hsu et al., 2022). Recent advancements in large language models (LLMs) present transformative opportunities to enhance the abilities of intelligent tutoring systems (ITSs) to conduct Socratic teaching beyond traditional classes.

LLMs can enhance the Socratic teaching by allowing for more dynamic and personalized interactions with lower development costs. However, relying solely on LLMs carries the risk of over-dependence and may hinder critical thinking (Premkumar et al., 2024). Therefore, incorporating domain-specific knowledge models, learner models, and pedagogical strategies within ITSs can refine LLM's responses to ensure they are contextually appropriate and align with learners' needs and learning goals. Based on Socratic principles of encouraging inquiry rather than providing direct answers, large language model-based intelligent tutoring systems (LLM-ITSs) can be steered towards probing questions, challenging assumptions, and clarifying ideas, thereby enhancing cognitive engagement and active learning.

## 2. LLM-ITSs: Exemplified by Socratic Playground for Learning

Inspired by the potential of delivering Socratic teaching by LLM-ITSs, this study introduces Socratic Playground for Learning (SPL) — an ITS powered by ChatGPT. SPL is structured into two main components: lesson creation and Socratic interaction. These two components work together to create a highly personalized, customized, and adaptive learning scenario for learners. Figure 1 shows the system architecture of SPL.

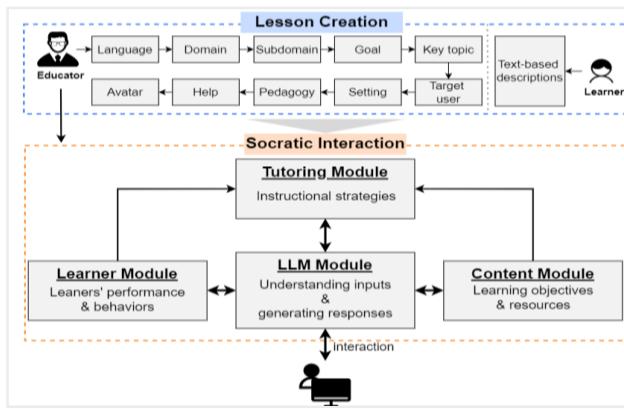


Fig.1 The system architecture of SPL

SPL empowers educators and teachers to customize learning scenarios through its lesson creation component. Educators can set key parameters from a tree-structured framework. Based on the selected parameters, ChatGPT generates a preliminary interactive interface for the lesson. At this stage, educators have the opportunity to further fine-tune the configuration. Besides, SPL goes beyond traditional teacher-driven lesson creation by allowing learners to create a lesson through text-based descriptions.

The Socratic interaction component is where learners actively engage with SPL. SPL initiates the interaction by posing a provoking Socratic question to prompt critical thinking. After receiving learners' inputs, SPL adapts the dialogue that mirrors traditional Socratic teaching to support ongoing engagement and deeper exploration of the topic. SPL can also summarize learners' responses, provide feedback on strengths and weaknesses, give hints when the learner struggles, offer encouragement to avoid frustration and disengagement, and propose follow-up questions that foster a reflective and iterative learning process. If learners need additional support, they can request on-demand help in challenging tasks without feeling overwhelmed.

### 3. A Pilot Study: Examining the capabilities of SPL

This study used a single-group pretest-posttest design to investigate the effects of SPL on learning performance and explore user acceptance toward SPL. A total of 30 Chinese college students were included in the pilot study. All participants were 18-23 years old ( $M = 20.43$ ,  $SD = 0.43$ ).

This study selected 5 typical flaws in experiments from the ARIES program as the learning content (Myers, 2021). The knowledge pretest and posttest were homogeneous transfer tests, consisting of 15 multiple-choice items (15 points) and two case studies (15 points), respectively. The Unified Theory of Acceptance and Use of Technology (UTAUT2) scale developed by Venkatesh et al. (2012) and revised by Zeng (2019) was used to assess the technology acceptance of SPL. Learners were required to score 23 items on the 7-point Likert scale ranging from "strongly disagree" to "strongly agree".

Prior to the learning phase, participants were recruited, signed the informed consent, and completed the prior knowledge test. The learning phase lasted for five days. Participants focused on one specific flaw each day. On average, participants spent 170 minutes learning in SPL. On the day following the learning phase, participants sequentially completed the knowledge posttest and the UTAUT2 scale.

Results of the paired-sample t-test indicated that the knowledge posttest test score was higher than the knowledge pretest score ( $t = 7.57$ ,  $p < 0.001$ ). The proportional learning gain ( $= \frac{\text{posttest ratio} - \text{pretest ratio}}{1 - \text{pretest ratio}}$ ) was calculated at 29.23%. These findings indicate a substantial improvement in learners' knowledge after interacting with SPL.

Results of user acceptance showed that learners generally had favorable attitudes toward SPL, with moderate to high levels of all dimensions of the UTAUT2 scale ( $M > 4.50$ ). Key factors, including performance expectancy ( $r = 0.59$ ), effort expectancy ( $r = 0.52$ ), social influence ( $r = 0.75$ ), hedonic motivation ( $r = 0.77$ ), and habit ( $r = 0.81$ ), were

all significantly and positively associated with intention to use SPL. Learners not only perceived the SPL as beneficial for their learning but also found that using SPL was free of effort, and they were influenced by social environments to adopt this new tool. Learners could derive enjoyment from using SPL, and the significant correlation with habit also supports the potential that SPL could become an ingrained part of lifelong learning. These positive perceptions and strong acceptance highlight the potential of SPL as a valuable tool with sustained engagement.

#### **4. Opportunities and Challenges of LLM-ITSs**

LLM-ITSs can foster the scalability and accessibility of personalized Socratic teaching, have the ability to maintain learning engagement and motivation, enable learning as a service and lifelong learning, promote cross-cultural understanding, and support cultural heritage preservation.

While LLM-ITSs present considerable opportunities to enhance education, their implementation also faces challenges that must be noticed and addressed. Key challenges include technological limitations, user readiness and acceptance, assessment transparency and rationality, ethical concerns, and adaptability to different cultures. In addition, Future research should employ experimental designs with rigorous controls to provide a clearer picture of how LLM-ST influences learning.

#### **References**

- Dalim, S. F., Ishak, A. S., & Hamzah, L. M. (2022). Promoting students' critical thinking through Socratic method: Views and challenges. *Asian Journal of University Education*, 18(4), 1034-1047.  
<https://doi.org/10.24191/ajue.v18i4.20012>
- Elkowitz, D. E. (2021). Socratic questioning to engage learners. In A. Fornari & A. Poznanski (Eds.), *How-to guide for active learning* (pp. 107–119). Springer.
- Hsu, F., Lin, I., Yeh, H., & Chen, N. (2022). Effect of Socratic Reflection Prompts via video-based learning system on elementary school students' critical thinking skills. *Computers & Education*, 183, Article 104497.  
<https://doi.org/10.1016/j.compedu.2022.104497>
- Myers, M. H. (2021). Automatic Detection of a Student's Affective States for Intelligent Teaching Systems. *Brain Sciences*, 11, 331. <https://doi.org/10.3390/brainsci11030331>
- Premkumar, P. P., Yatigammana, M. R. K. N., & Kannangara, S. (2024). Impact of Generative AI on Critical Thinking Skills in Undergraduates: A Systematic Review. *Journal of Desk Research Review and Analysis*, 2(1), 199–215. <https://doi.org/10.4038/jdrra.v2i1.55>
- Venkatesh, T., Thong, J. Y. L., & Xu, X. (2012). Consumer acceptance and Use of information technology: Extending the unified theory of acceptance and Use of technology. *MIS Quarterly*, 36(1), 157.  
<https://doi.org/10.2307/41410412>
- Zeng, J. (2019). A study on user acceptance of mobile-assisted language learning based on UTAUT2 [Chinese]. *Technology Enhanced Foreign Language Education*, (6), 16-24.  
<https://doi.org/10.20139/j.issn.1001-5795.2019.06.003>

# What Makes the Ideal AI Collaborator? Exploring Student and Teacher Perspectives on Roles, Support, and Challenges

Zhou Long<sup>1</sup>, Dehong Luo<sup>1</sup>, Hongli Gao<sup>2</sup>, Sheng Xu<sup>3</sup> and Xiangen Hu<sup>4</sup>(✉)

<sup>1</sup> Huaihua University, Huaihua 418000, China

<sup>2</sup> Xinxiang Medical University, Xinxiang 453003, China

<sup>3</sup> Central China Normal University, Wuhan 430079, China

<sup>3</sup> The Hong Kong Polytechnic University, Hong Kong, China

\* xiangenhu@gmail.com

**Abstract:** The application of artificial intelligence (AI) in collaborative learning situations has much potential to disrupt education, but it is still an understudied field. This study investigated the anticipated roles, benefits, and challenges of AI in group learning as perceived by 15 students and 12 teachers. Students viewed AI as a collaborator in their learning process (to think of ideas, to work on a task, and to engage socially), as well as a tutor who patiently met their needs while completing tasks. Teachers saw AI as a way to augment human instruction, making it more flexible, collaborative, and personalized. Both groups identified challenges, such as lack of curriculum guidance, socio-emotional support, and frustration with the mechanical and emotional limitations of AI. The study reveals both the exciting possibilities and key limitations of using AI for group learning and provides guidance on how to effectively integrate AI into learning environments.

**Keywords:** Artificial Intelligence (AI), Group Learning, AI Collaborator, Teachers' and Students' Expectations

## 1. Introduction

The incorporation of AI in group learning has introduced fresh opportunities for collaborative education, however, its complete capabilities are yet to be fully investigated. Past research has primarily examined the role of AI as an aid to educators, as well as to enrich student learning experiences and promote group dynamics through tools such as chatbots and social robots that foster cognitive and emotional growth (Mavrikis et al., 2021). AI is now often seen as a partner in collaboration alongside individuals in capacities, like teaching support and peer interaction. However, we do not yet know how teachers and students perceive AI in group learning environments. This study bridges the gap by investigating their perceptions of AI in group learning (AIGL) with the goal of identifying effective integration strategies.

AI technology in the field of education is commonly perceived either as a tool or a facilitator for learning enhancement purposes. When viewed as a tool AI is utilized to provide organized information to support students in reaching their objectives (Kovari, 2025). Nonetheless this perspective fails to acknowledge the exchange between students and AI, and the influence of AI on behavior and self-perception. Post humanist ideas like Actor Network Theory (ANT) consider both humans and AI on an equal footing (Latour, 2005). Recent research underscores the capacity of AI to function as an interactive collaborator that can handle information processing and support cognitive development (Holmes, Bialik, & Fadel, 2019). For instance, AI has been demonstrated to aid in generating ideas (Ji, Han & Ko, 2022) self-assessment (Echeverria et al., 2018) and enhancing teamwork dynamics via visualizing information (Han et al., 2021).

AI has the potential to strengthen the connection between students and AI entities. The effectiveness of this relationship, however, depends on how it is put into practice. This study aims to delve deeper into the perceptions of

university students and educators on the role of AI, shed light on the barriers encountered, and provide recommendations to improve teamwork dynamics and overall team performance.

## **2. Research Methods**

### **2.1. Participants**

This study used a purposeful sampling strategy to recruit 27 participants - 15 undergraduate students and 12 university instructors. In order to obtain a representative overview of students regarding AIGL, participants were selected from a range of academic disciplines, performance backgrounds, and attitudes regarding AI. For instructors, a minimum of one year of experience teaching AI was needed. They were involved via embedding AI functionalities (e.g., GPT-based chatbots, AI dashboards, virtual tutors, AI-powered speech/writing assistants) into the instruction (Holmes, Bialik, & Fadel, 2019). The research received ethical approval from the university's Institutional Review Board, and informed consent was obtained from all participants before their participation in the study.

### **2.2. Data Collection and Analysis**

Each participant took in a semi-structured face-to-face interview lasting between 30 and 50 minutes. Data were examined with a reflective theme approach (Braun & Clarke, 2006) to generate themes aligned with the research questions. After multiple readings of the transcripts, three researchers independently coded the significant statements, grouping them into 20 potential themes. Through iterative review and consensus, these were distilled into two themes each related to AI's expected roles, AIGL advantages, and AIGL barriers. Trustworthiness was ensured by the use of respondent validation and moderation (Golafshani, 2003).

## **3. Findings and Discussion**

### **3.1. Participants' Perceptions of AI's Expected Roles in AIGL**

#### **3.1.1. AI as Co-Learning Partners**

Students saw AI as a teammate in learning together. They looked forward to its involvement in group activities. AI was imagined as a source of inspiration, support in generating ideas (S2) and fostering creativity (S9). Despite acknowledging that AI cannot replicate human actions entirely, like expressing emotions or empathy (S3, S15), students saw this distinction as a chance to embrace different viewpoints and outlooks (S15).

Moreover, students believed that AI might function as an ally. It could enhance the social interactions within group study sessions. They compared this phenomenon to teamwork among humans. In human teams, each person's distinct traits enhance collaboration and make it a richer experience. In this scenario, students viewed AI as a digital counterpart with its own special abilities. They separated AI from those of human team members. As an illustration one student observed, "While chatting with peers about challenges we frequently stray from the main subject. I believe having an AI companion could assist in maintaining the focus of the conversation" (S13).

#### **3.1.2. AI as an Instructional Tutor**

Students were looking forward to using AI as a tutor. In their learning process, AI would patiently guide them through tasks. It offers explanations and suggestions when they needed. Students clearly expressed that they wanted AI to guide them to enhance learning experience, like giving them suitable hints (S5) or effective learning strategies (S8).

Teachers envisioned a teaching approach involving both AI and human educators sharing duties and switching roles. For instance, teachers presenting ideas while AI expands on them or guides hands on activities (T4). They recognized AI's expertise in conveying theoretical information. But teachers highlighted the unique ability of human educators to offer emotional and social assistance (T2) nurturing students' intellectual growth and emotional well-being simultaneously (T10).

### **3.2. Participants' Perceived Supports of AIGL**

#### **3.2.1. Improved Task Performance**

Students mentioned that working with AI helped them express their thoughts better. They appreciated how AI could analyze their drafts and offer polished recommendations. Some students especially thanks AI for refining their ideas (S7) and enhancing the overall development process (S11). Although a few students observed that AI's language appeared mechanical compared to human expression preferences (S2, S9, S10), they valued its role in subtly guiding their thought processes towards improvement (S9).

AI was seen as a tool that made tasks easier, by assigning roles and fostering teamwork. A few students noticed that projects took time when using AI because they took into account its recommendations (S9). Individuals who saw AI as an aide often viewed themselves as "leaders" with AI playing a supporting "follower" position (S4, S11).

Additionally, AI has enhanced creativity by providing recommendations (S2) inspiring fresh thoughts (S5) and connecting various ideas (S13). This has empowered learners to delve into methods and create inventive answers.

### **3.2.2. Increased Teaching Flexibility**

Teachers recognized that utilizing AI in the classroom provided chances for stimulating activities. AI creates technology-driven environments. Educators could use these lively environments, like replicating real life situations to tackle intricate issues (T6). Furthermore, teachers saw how AI had the potential to elevate the teaching of subject related material. It enhanced students' comprehension, through exercises on giving presentations (T9), conducting research, and solving problems (T1).

AI also offered flexibility in arranging students into groups. It examined student information to assist teachers in creating groups according to their readiness (T4) passion (T7) or hobbies (T11). This approach reduced the number of teachers required. It allowed personalized support for students with different learning requirements.

A teacher mentioned that AI has the ability to oversee group conversations at once. It spot things like, off topic chats or exceptionally rapid advancements in the discussions. This feature helped lessen the teachers' tasks and enabled them to offer more personalized assistance (T5).

## **3.3. Participants' Perceived Challenges to AIGL**

### **3.3.1. Absence of Systematic Curriculum Design**

A key challenge was the lack of adequate pedagogical support from AI. This hindered its ability to effectively facilitate student learning process. Students wanted AI to automatically care their learning stage, then provide tailored guidance. They called for step-by-step instructions (S2), timely feedback (S5), and detailed material classification (S8). In contrast, excellent students preferred open-ended tasks (S3) and creative brainstorming (S11).

Teachers were also not satisfied with AI's structured course resources. In the current teaching approaches, such as integrating AI into ICT/STEM activities (T11) or interdisciplinary frameworks (T4), these materials often prioritize AI technology over its pedagogical integration. They argued that this limitation might narrow the scope of AIGL and reduce its effectiveness across different disciplines (T4, T9).

### **3.3.2. Insufficient Socio-Emotional Support**

Students complained about AI's limited social and emotional capabilities. They observed that AI rigidly followed basic communication norms, such as starting with "Hi" and ending with "Bye" (S9) or adhering strictly to turn-taking (S4), meanwhile lacked flexibility and humor (S4, S5). This rigidity disrupted the natural social flow of group learning. Unlike human peers, it was hard for AI to express empathy or provide motivational encouragement. However, students perceived these qualities as essential to fostering a supportive and effective collaborative environment.

Teachers had similar concerns. They emphasized that a lack of emotional support from AI could negatively impact students' development. Areas such as empathy (T7) and social communication skills (T3, T9) were specifically mentioned. They warned that excessive interaction with emotionally neutral AI might reduce students' emotional awareness (T2) and impair their ability to form meaningful connections in the real world (T12).

## **4. Conclusion**

This study categorized the expected roles of AIGL, its perceived advantages, and difficulties through interviews with students and teachers. It raises an all-too-familiar question, especially to non-academic audiences: “Should AI replace human teachers?” The findings indicate that AI will not replace human instructors but rather supplement their work. Its function will differ based on the context. Be students sometimes, and learn along with fellow students (Kovari,2025). Other times act as a mentor, offering tailored support beyond the physical classroom (e.g., providing language support for students who are not fluent in English) (Ji, Han & Ko, 2022). These applications are specifically effective in personalized and online learning environments. But, incorporating AI also brings up some ethical issues.

One of the biggest worries is bias in AI-generated content. It can stimulate stereotypes or misinformation if datasets are not adequately managed. To guarantee that AI-assisted learning maintains equity, we must constantly oversee and modify our algorithms. AI systems also gather student information, so data privacy and security are top concerns. Policies and safeguards are needed to build trust and protect data. Another major concern is the long-term effects of AI on students’ independence and critical thinking ability. Though AI can facilitate learning tasks, over-reliance on its suggestions may undermine students’ independent ability to look at problems from multiple perspectives and come up with solutions. A balance must be struck between leveraging AI capabilities and preserving human essential cognitive skills.

### Acknowledgements

We are grateful for the funding support provided by the Research on the Induction and Regulation of Learners' Social Cognitive Conflicts in Virtual Collaboration Situations (Grant No. 24A0561) and the Reform of Teaching Strategies for Social Cognitive Conflict in Online Courses Project (Grant No. HNJG-2021-0923).

### References

- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
- Echeverria, V., Martinez-Maldonado, R., Shum, S. B., Chiluiza, K., Granda, R., & Conati, C. (2018). Exploratory versus explanatory visual learning analytics: Driving teachers' attention through educational data storytelling. *Journal of Learning Analytics*, 5(3), 72–97.
- Golafshani, N. (2003). Understanding reliability and validity in qualitative research. *The Qualitative Report*, 8(4), 597–607.
- Han, J., Kim, K. H., Rhee, W., & Cho, Y.H. (2021). Learning analytics dashboards for adaptive support in face-to-face collaborative argumentation. *Computers & Education*, 163, 104041.
- Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.
- Ji, H., Han, I., & Ko, Y. (2022). A systematic review of conversational AI in language education: Focusing on the collaboration with human teachers. *Journal of Research on Technology in Education*, 55(1), 48–63.
- Kovari, A. (2025). A systematic review of AI-powered collaborative learning in higher education: Trends and outcomes from the last decade. *Social Sciences & Humanities Open*, 11, 101335.
- Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. Oxford university press.
- Mavrikis, M., Cukurova, M., DiMitri, D., Schneider, J., & Drachsler, H. (2021). A short history, emerging challenges and cooperation structures for artificial intelligence in education. *Bildung und Erziehung*, 74(3), 249–263.

# Impact of Using a Smart Online Teaching Tool “Rain Classroom” on Sophomores’ Class Engagement in an Intercultural Communication Course

Guiwu XU<sup>1</sup>, Alex Wing Cheung TSE<sup>2\*</sup>,

Weilan PI<sup>3</sup>

<sup>1,2</sup>The University of Hong Kong

<sup>3</sup>Zhanjiang University of Science and Technology

\* awctse@hku.hk

**Abstract:** This study investigates the impact of Rain Classroom, a smart online teaching tool, on sophomores' engagement in an Intercultural Communication Course (ICC) in China. Using a quantitative embedded quasi-experimental design, the research involved 131 students learning with Rain Classroom and 112 in traditional settings in four weeks. The experimental group utilized the Presentation-Assimilation-Discussion (PAD) teaching model. Pre- and post-tests, analyzed through ANCOVA, revealed that the experimental group exhibited significantly higher cognitive engagement ( $F = 6.34, p < 0.05; \eta^2 = 0.03$ ) compared to the control group. The result suggests that the platform may enhance self-regulated learning and deep learning strategies. However, no significant differences were found in emotional or behavioral engagement between the groups, indicating a need for further research in these areas.

**Keywords:** Rain Classroom, Presentation-Assimilation-Discussion (PAD) Model, Intercultural Communication Course (ICC), Class engagement

## 1. Introduction

The 21st century has put forward new requirements for the teaching and learning of teachers and students, and the importance of educational technology is also clearly indicated in PISA (Petko et al., 2017). UNESCO stresses that the twenty-first-century revolution focuses on the use of educational technology to facilitate the uptake of knowledge by students and their collaborative knowledge-building (Chai et al., 2015). The deep integration of science, technology, and education is one of the future development directions of education. In light of the proliferation of mobile internet and big data, Tsinghua University has introduced a new smart teaching tool called Rain Classroom (Da-Hong et al., 2020). Though this platform is important and widely used, related research is still limited, especially those about students' engagement, reflecting the need for this research. The PAD classroom teaching model is a new teaching model proposed by Prof. Zhang of Fudan University in 2014 (Lv et al., 2024), and it promotes student engagement as an essential aspect that contributes to accomplishing educational goals. Therefore, the following research question is posed in this study: Does Rain Classroom with PAD mode impact sophomores' class engagement in the Intercultural Communication Course?

## 2. Literature Review

### 2.1. Class Engagement

Engaging students in the classroom is essential for effective teaching and learning. Educators can employ various strategies to create a dynamic and interactive learning environment. One key approach is the use of active learning techniques. Fostering collaboration and interaction is another important aspect of class engagement. Offering opportunities for students to make decisions, such as selecting project topics or presentation formats, can foster a sense of ownership and investment in the learning process. Constructive feedback, ongoing support, and a positive learning environment that encourages risk-taking and celebrates student achievements which can further contribute to a more

engaging and enriching classroom experience. Most of the previous studies on classroom engagement were based on classroom organization forms, teachers' teaching methods, or students' learning strategies, and were rarely considered from the perspective of a teaching platform (Kong, 2021; Elmaadaway, 2018). This study is conducted through the Rain Classroom, which echoes the research significance of the relationship between classroom engagement and teaching platforms.

## **2.2. Intercultural Communication Course**

Central to ICC is the understanding of how culture shapes communication styles, perceptions, and behaviors. Students will study cultural dimensions, such as Hofstede's framework and Hall's concept of high-context and low-context cultures, to analyze their impact on verbal and nonverbal communication (Yu et al., 2019). The course addresses the effects of cultural biases and stereotypes on interactions and explores conflict resolution and negotiation techniques in cross-cultural settings.

Additionally, contemporary issues such as globalization, technology, diversity, and ethical considerations in intercultural interactions are discussed. Through lectures, case studies, and experiential activities, students can cultivate a comprehensive understanding of intercultural communication, fostering an appreciation for cultural diversity and enhancing their ability to communicate effectively across cultural boundaries. The nature of ICC requires students' active participation in class, which is also the reason for choosing this course for research.

## **3. Theoretical Framework: Presentation-Assimilation-Discussion (PAD) Mode**

The Presentation-Assimilation-Discussion (PAD) instructional strategy is a structured approach to facilitating effective learning in educational settings (Yu et al., 2019). This three-phase method aims to engage students actively throughout the learning process. By following this three-step process shown, educators can effectively introduce new content, provide opportunities for students to process and apply the information, and engage them in meaningful discussions to solidify their understanding. PAD mode is combined with pre-class preparation and after-class review to form five phases shown in Figure 1. The combination of PAD mode and Rain Class undoubtedly makes the course design in Rain Classroom more orderly, so that teachers can be more confident when using the platform to teach. In this study, Rain Classroom in PAD mode was used as an intervention, as shown in Figure 1.

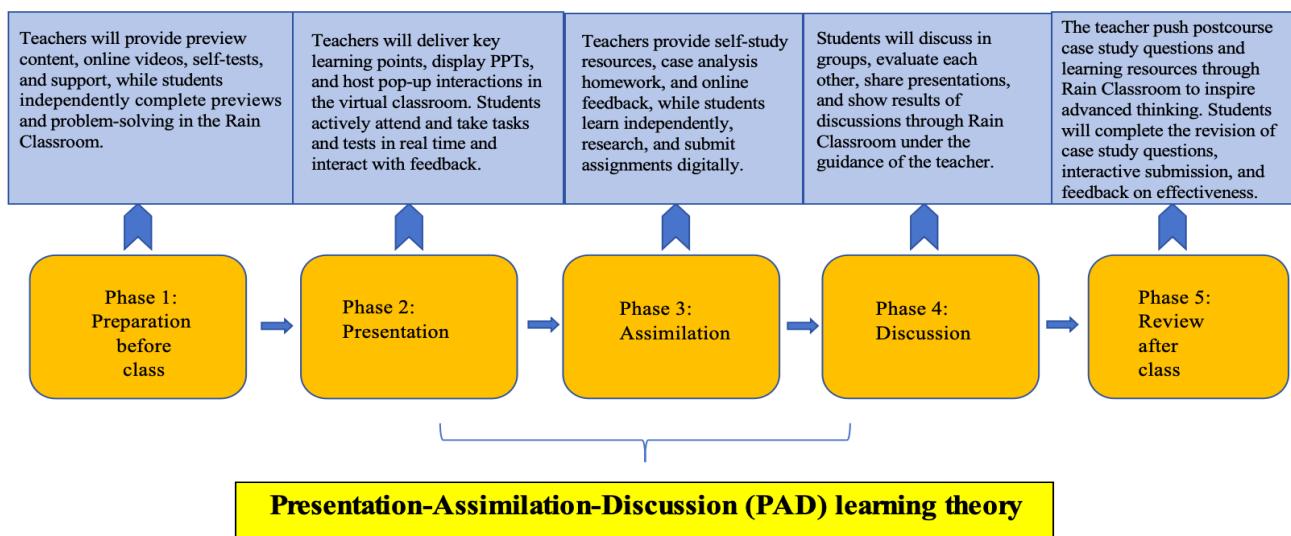


Fig.1 Processes of intervention

## **4. Research Design, Data Collection and Analysis**

This research employed a quasi-experimental design with pretest and posttest measures, involving 243 second-year undergraduate students from a university in Guangdong, China. Students with similar English academic performance were divided into an experimental group ( $n=131$ ) and a control group ( $n=112$ ). Over four weeks, both groups attended

90-minute English classes given by the same teacher. While the control group followed a traditional format using paper-based assignments, the experimental group utilized the PAD model through an online platform for in-class and out-of-class learning.

Pretests and posttests used the class engagement measure (Muir et al., 2022), which has three dimensions: cognitive engagement, emotional engagement, and behavioral engagement. Quantitative data were analyzed using IBM SPSS 29.0. Cronbach's alpha assessed the internal consistency of the Class Engagement Scale based on pretest data. Descriptive analysis and ANCOVA were conducted on pretest and posttest data to evaluate overall student engagement in both the experimental and control groups. Pretest scores were considered as potential confounding factors impacting posttest outcomes. ANCOVA adjusted for these variables, maintaining a 0.95 confidence interval. The dependent variables included total class involvement and its facets, while the independent variable was group assignment. Levene's test verified variance homogeneity between the groups before analysis.

## 5. Results and Findings

The Cronbach's alphas for all dimensions exceeded 0.7 ( $\alpha_{CE} = 0.963$ ,  $\alpha_{EE} = 0.973$ ,  $\alpha_{BE} = 0.955$ ), indicating strong reliability and internal consistency of the scale. An independent samples T-test ( $p=0.09$ ) showed no significant difference in pretest scores between the experimental and control groups before the intervention. Descriptive analysis in Table 1 illustrates a slight improvement in class engagement scores, with the experimental group performing better on average. Levene's test confirmed variance homogeneity across all dimensions (overall engagement:  $F=2.107$ ,  $p > 0.05$ ; cognitive:  $F=0.870$ ,  $p > 0.05$ ; emotional:  $F=1.205$ ,  $p > 0.05$ ; behavioral:  $F=1.059$ ,  $p > 0.05$ ). ANCOVA was conducted to assess differences in posttest scores, revealing significant differences in cognitive engagement ( $F = 6.344$ ,  $p < .005$ ;  $\eta^2 = .026$ ). This suggests that Rain Classroom with PAD mode enhanced cognitive engagement, motivating students in ICC. However, the intervention did not significantly impact emotional or behavioral engagement dimensions.

Table 1. Descriptive Statistics and ANCOVA Results of Class Engagement

Dimension	Pretests		Posttest		ANCOVA		
	Mean	SD	Mean	SD	F	P	$\eta^2$
Experimental: Control							
Overall Engagement	4.123: 3.969	0.689: 0.720	4.253: 4.087	0.663: 0.631	3.636	0.058	0.015
Cognitive Engagement	4.003: 3.859	0.764: 0.799	4.163: 3.915	0.725: 0.734	6.344	< 0.05	0.026
Emotional Engagement	4.139: 3.984	0.720: 0.746	4.274: 4.132	0.669: 0.639	2.516	0.114	0.010
Behavioral Engagement	4.212: 4.041	0.684: 0.746	4.303: 4.172	0.666: 0.666	1.817	0.179	0.008

## 6. Discussion and Conclusion

This study was a pretest-posttest quasi-experimental design to investigate the impact of Rain Classroom, a smart online teaching tool, on class engagement in ICC. To answer the research question, the results show that the use of Rain Classroom with PAD has a certain effect on the participation of English majors in ICC, especially at the level of cognitive engagement, but there is no significant difference in other aspects. Teaching practice helps teachers realize the benefits of using smart educational tools with PAD to make students' learning more motivated because cognitive engagement represents students' willingness to make an effort even when the activities are challenging. As for the result of the significant improvement of students' cognitive participation in the Rain Classroom, teachers can use the platform to arrange pre-reading tasks in advance and formulate students' curriculum learning goals. Meanwhile, a number of online blended learning platforms are available, including [www.icourse163.org](http://www.icourse163.org), [classroom.google.com](http://classroom.google.com), [kahoot.com](http://kahoot.com), and [www.ketangpai.com](http://www.ketangpai.com). These platforms, which include online work correction, performance record and analysis, courseware sharing, and online discussion, are made to make it easier for teachers to engage with their students in the classroom (Gao et al., 2020). This is similar to the effect of Rain Classroom on promoting students' engagement in class. This study has several limitations. First of all, students' class engagement varies in different educational contexts. Future studies are expected to employ different disciplines. Individual factors of teaching and students are taken into

---

account to draw a broader range of conclusions. In addition, since the intervention lasted only 4 weeks, the impact of the intervention on students' course engagement is likely to be quite limited. With longitudinal experiments, further research can be done. Additionally, the participants in this study were mainly sophomore students, and did not attempt to cover all the grades of university students, lacking multiple data to provide a source of triangulation. Data collection relies largely on self-reporting, which can make the results relatively subjective and imprecise. It is possible to conduct more extensive research using a variety of tools, such as leveraging machine learning algorithms or Rain Classroom background data to help consider student classroom engagement. Lastly, combined with qualitative research or comparing rain classroom with other platforms, its advantages and limitations can be more fully assessed.

---

## References

- Chai, C. S., Deng, F., Tsai, P.-S., Koh, J. H. L., & Tsai, C.-C. (2015). Assessing multidimensional students' perceptions of twenty-first-century learning practices. *Asia Pacific Education Review*, 16, 389-398.
- Da-Hong, L., Hong-Yan, L., Wei, L., Guo, J. J., & En-Zhong, L. (2020). Application of flipped classroom based on the Rain Classroom in the teaching of computer-aided landscape design. *Computer Applications in Engineering Education*, 28(2), 357-366.
- Ding, L., Sitthiworachart, J., & Morris, J. (2023). Effect of scaffolding and peer review on learning in a PAD class. *World Journal of English Language*, 13(2).
- Elmaadaway, M. A. N. (2018). The effects of a flipped classroom approach on class engagement and skill performance in a blackboard course. *British Journal of Educational Technology*, 49(3), 479-491.
- Gao, B. W., Jiang, J., & Tang, Y. (2020). The effect of blended learning platform and engagement on students' satisfaction—the case from the tourism management teaching. *Journal of Hospitality, Leisure, Sport & Tourism Education*, 27, 100272.
- Kong, Y. (2021). The role of experiential learning on students' motivation and classroom engagement. *Frontiers in Psychology*, 12, 771272.
- Li, Y., Wen, X., & Liao, X. (2021). Research on the Application of PAD Class Mode in Vocational English Teaching Based on Duifene Platform. *Advances in Applied Sociology*, 11(7), 335-339.
- Lv, H., Tang, L., Luo, G., Meng, M., & Luo, Q. (2024). Rain Classroom and PAD class blended learning mode effectively improves teaching quality in a surgical nursing course. *American Journal of Translational Research*, 16(1), 200.
- Petko, D., Cantieni, A., & Prasse, D. (2017). Perceived quality of educational technology matters: A secondary analysis of students' ICT use, ICT-related attitudes, and PISA 2012 test scores. *Journal of Educational Computing Research*, 54(8), 1070-1091.
- Shen, R., Wang, M., & Pan, X. (2008). Increasing interactivity in blended classrooms through a cutting-edge mobile learning system. *British Journal of Educational Technology*, 39(6), 1073-1086.
- Yue, Y. (2021). Integrating PAD class to college English reading teaching. *Open Access Library Journal*, 8(6), 1-8.
- Yu, S., Ji, H., Lu, Y., Chen, S., Xiong, J., Chi, C., Teliewubai, J., Fan, X., Blacher, J., & Li, J. (2019). Significance of the combination of inter-limb blood pressure differences in the elderly: The Northern Shanghai Study. *The Journal of Clinical Hypertension*, 21(7), 884-892.
- Zhu, H. (2021). Application of Rain classroom in formal classroom learning in the teaching of offshore engineering environment and loads. *Computer Applications in Engineering Education*, 29(3), 603-612.

# AI-Powered Socratic Learning for Psychological Statistics: Enhancing Flipped Classroom Practices with Generative Intelligent Tutoring Systems

Lei Yang<sup>1</sup>, Sheng Xu<sup>2</sup>, Hongli Gao<sup>3</sup>, Zhou Long<sup>4</sup>, Xiangen Hu<sup>5</sup>, Wenhui Xu<sup>1</sup>

<sup>1</sup> Faculty of Education, Henan Normal University

<sup>2</sup> Department of psychology, Central China Normal University & Beijing Jingshi Liyun Education Technology Co., Ltd.

<sup>3</sup> School of Psychology, Xinxiang Medical University

<sup>4</sup> School of Psychology, Huaihua University

<sup>5</sup> Institute for Higher Education Research and Development, Hong Kong Polytechnic University

\*xiangenhu@gmail.com

**Abstract:** This study integrates the Socratic Playground Intelligent Tutoring System (ITS) into a flipped classroom model for a psychological statistics course. The system, based on an AI-supported Learning Management System (LMS) with a Socratic-style large language model, is combined with flipped classroom teaching to explore the impact of adaptive pre-class tasks, personalized grouping, continuous post-class reflection, and various teaching methods on instruction effectiveness. It investigates the optimal timing, methods, and strategies to improve teaching and learning outcomes, with the goal of enhancing classroom interaction, learning atmosphere, student interest, and learning effectiveness, while fostering problem-solving, critical thinking, and autonomous learning skills. The study also provides data-driven insights to refine the system. The process includes pre-class task assignments, in-class personalized grouping, post-class reflection, and evaluation of classroom configuration to assess effectiveness and inform improvements. The results aim to optimize teaching and system design, supporting future implementation and refinement.

**Keywords:** Socratic Playground; Intelligent Tutoring System (ITS); Flipped Classroom; Personalized Learning

## 1. Introduction

In the context of advancing psychological research and the big data era, psychological statistics is a foundational course for psychology majors, offering essential skills for research and critical thinking. However, traditional teaching often struggles to meet students' diverse needs for personalized learning and feedback. To address this, the study explores the integration of generative AI and flipped classrooms under Socratic Playground, utilizing AI-enabled learning management systems and intelligent tutoring systems. These technologies adapt to learners' preferences, reduce cognitive load, and enhance application skills. The research investigates the optimization of teaching strategies through Socratic questioning, adaptive grouping, and real-time feedback, aiming to achieve optimal learning outcomes by matching learner characteristics, knowledge complexity, and pedagogies. Supported by AI algorithms that assess learning behaviors and provide personalized recommendations, the study seeks to enhance the theory-practice connection and improve teaching quality and student outcomes, informed by the works of Hu et al. (2025) and Zhang et al. (2024). In summary, the research questions of this paper are:

Question 1: How can we optimize teaching links to enhance learning effectiveness?

Question 2: Which teaching methods are most effective for adaptive learning at what time and place?

Question 3: How can the support of adaptive technology enhance interaction and students' core competencies?

## 2. Research Methods

A quasi-experimental study uses an AI-enabled flipped classroom LMS (AI-FC-LMS) to explore teaching mode optimization in psychological statistics. Participants are 78 psychology undergraduates. The  $2 \times 3 \times 4$  design compares AI-FC-LMS to traditional flipped classrooms across learning stages (before, during, after) and learning modes (tutoring, vicarious, gamification, teachable agents). Table 1 shows the experimental process.

Table 1. The Specific Experimental Process of AI-FC-LMS

Stage	Research Objectives	Time	Specific Research Content	Research Indicators	Research Methods
	Learner Characteristics		Group by cognitive styles		
Stage 1	Assessment and Experimental Grouping	1 - 2 weeks	(AI - FC - LMS vs. Traditional Flipped Classroom)	Cognitive patterns, learning motivation, statistical anxiety, etc.	Questionnaire survey, theoretical analysis
Stage 2	Implementing Socratic - style Flipped AI Classroom	3 - 15 weeks	Before class: Recommend personalized learning materials During class: Allocate intelligent interaction modes After class: Arrange personalized review	Basic knowledge learning, question generation Interaction mode adaptability Review effectiveness	Theoretical analysis, AI recommendation, AI - adaptive task allocation AI - adaptive task allocation
Stage 3	Integration of Optimal Teaching Strategies	16 - 18 weeks	Multi - dimensional learning effect assessment	Knowledge acquisition, learning motivation, cognitive ability etc.	Structured modeling, learning trajectory analysis

### 3. Expected research findings

This study predicts that AI-FC-LMS will improve the teaching of psychological statistics by personalizing learning through dynamic task adjustment based on student motivation and affective factors. AI will optimize pre-class preparation with intelligent recommendations, enhance classroom interaction via smart grouping and Socratic guidance, and provide personalized post-class feedback to reduce statistical anxiety. Students using AI-FC-LMS are expected to show better knowledge mastery, higher test scores, and increased motivation compared to traditional flipped classrooms. Theoretically, this study advances AI-flipped classroom pedagogy, offering empirical support for educational theories. Practically, it aids teachers in personalized instruction, improves teaching quality, and supports broader educational reforms using AI and flipped classroom models in higher education.

### 4. Innovative Points, Limitations, Conclusion, and Outlook

This study presents an AI-FC-LMS model that integrates artificial intelligence into flipped classrooms to enhance the teaching of psychological statistics. By leveraging the LCC theory and the ICAP framework, the model optimizes three-stage learning processes through personalized tasks, intelligent grouping mechanisms, and instantaneous feedback systems. The AI-driven platform dynamically tailors its teaching strategies based on learners' motivations, cognitive abilities, and preferences, thereby enhancing knowledge acquisition, fostering active participation, and promoting critical thinking skills. Theoretically, this research enriches the pedagogical framework of AI-infused flipped classrooms while providing robust empirical evidence for the efficacy of AI in educational settings. From a practical perspective, it offers actionable teaching strategies to facilitate personalized instruction, improve educational quality, and advance the integration of AI technologies in academic environments. However, the study acknowledges several limitations, including data variability, concerns regarding the generalizability of AI applications, challenges associated with mode switching, the use of small sample sizes, and potential external validity issues. These limitations underscore the need for further validation and refinement to ensure broader applicability of the AI-FC-LMS model across diverse

---

academic disciplines and cultural contexts. Future research should focus on addressing these challenges to maximize the model's potential for transformative impact in education.

## References

- Hu, X., Xu, S., Tong, R., & Graesser, A. (2025). Generative AI in Education: From Foundational Insights to the Socratic Playground for Learning. arXiv:2501.06682v1 [cs.AI].
- Zhang, L., Lin, J., Kuang, Z., Xu, S., & Hu, X. (2024). SPL: A Socratic playground for learning powered by large language model. Electronics, 13(24).

# Neither Schoolgirls nor Schoolboys but STEM Learners: Develop and Evaluate Training Course for Hong Kong Secondary School STEM Teachers to Cultivate Gender Equality in the Learning Environment: Case Study

Sze Lam Choi<sup>1</sup>, Alex Wing Cheung Tse<sup>2</sup>

Faculty of Education, The University of Hong Kong, Hong Kong

<sup>1</sup> u3556532@connect.hku.hk, <sup>2</sup> awctse@hku.hk

**Abstract:** This case study developed and evaluated a 3-week training course with 3 weekly modules pedagogically rooted in Mezirow's transformative learning theory. The course aspired to reshape gender-related teaching concepts and practices of Hong Kong secondary school STEM teachers and eventually cultivate gender equality in 4 STEM teaching aspects that influenced female and male students' expectancy and task value of Eccles's situated expectancy-value theory. Gender equality here did not mean exploiting STEM learning opportunities for male students to put female students at an advantage. It aimed at freeing every STEM learner, regardless of gender, from traditionally stereotypical views on male and female STEM performance and ensuring equal multi-aspect learning access (Balan & Stanciu, 2021). The study purposively sampled 6 male and female first-year STEM teachers at different mainstream co-educational secondary schools to participate and qualitatively leveraged 2-stage semi-structured individual interviews: pre-training interviews on prior gender-related teaching concepts and post-training interviews to gather their voices on 1) whether and how this course reshaped their concepts and 2) practical improvements to support them further (i.e., 2 research questions) where thematic analysis ran. Participant teachers hugely reshaped gender-stereotypical concepts and practices through 3 essential transformative learning phases. It constructed 2 improvements: 1) further transformation of teachers' previous STEM schooling experiences and 2) reachable access to gender-balanced STEM role models (with QR codes).

**Keywords:** Gender equality in STEM learning, Hong Kong secondary school teacher training, Situated expectancy-value theory, Mezirow's transformative learning theory

## 1. Introduction: the need of developing and evaluating gender equality teacher training course

“Boys should innately excel in STEM subjects as they can think logically!” “Girls are good at writing and should pursue art-related careers instead!”. “Masculine STEM” and “feminine arts” teacher gender-stereotypical beliefs have long subsisted in multi-shapes and socially burdened two genders in STEM academic tasks and careers (Eccles & Wigfield, 2020). Maries et al. (2022) revealed that “boys-only” STEM clubs demotivated female students to sustain their STEM strengths. Makarova et al. (2019) echoed that underrepresentation of female STEM role models in secondary school textbooks added anxiety to female students' pursuit of STEM careers. The burden intensified in mainstream secondary education when Grade 9 students chose art-related or STEM-related electives of public examination. Physics topped, and ICT ranked fifth on the male elective choice chart. However, neither topped the six elective choices of female candidates who turned to more art-related fields, such as geography (EOC, 2022). Another turning point situated in Grade 12 when students arranged their university programme preferences that heavily illuminated relevant career options. Male students applying for STEM university programmes to pursue STEM career goals were twofold more than females, who occupied around 70% of humanistic caregiving sectors (EOC, 2022). PISA 2022, moreover, reported the widening gender gap in Hong Kong secondary school mathematics achievement, where female students underperformed male students by fifteen scores more in 4 years (Schleicher, 2023).

Gender-stereotypical teaching concepts seemingly favoured male students. Yet, overwhelmingly high expectations that boys innately excelled in STEM learning lowered the self-efficacy of low-performing male students and pressurized high-achieving ones (Zhang et al., 2022). It urged this study to develop and evaluate a well-polished training course to reshape Hong Kong secondary school STEM teachers' gender-related beliefs and practices to cultivate gender equality in STEM learning. Before so, it should investigate how secondary school STEM teachers' gender-related concepts and practices induced these students' gender struggles in STEM learning.

## **2. Background: from teachers' gender-related concepts to students' achievement motivation**

Eccles's situated expectancy-value theory (hereafter SEVT) summarized two factors for STEM achievement motivation (Eccles & Wigfield, 2020). 1) Expectancy: student beliefs on how well they would accomplish STEM academic tasks. 2) Task value: tetra-dimensional value students derived from STEM academic tasks: intrinsic (how well tasks captivated students), attainment (student perceived significance of STEM task success), utility (usefulness of tasks for student prospective aspirations), and cost (how much inputs students had to devote for STEM academic tasks). In line with Bronfenbrenner's ecological systems theory, teacher socialization, which occupied the microsystem level, primarily and continually influenced student expectancy and task value. Students emotionally internalized and projected gender-associated treatments from STEM teachers onto the world's perceptions of genders in STEM fields (Bronfenbrenner, 1979). Pygmalion effect underlying teacher expectancy and self-fulfilling prophecy contextualized vicious STEM learning cycle: The less rigid the beliefs that acquired potential prioritized over inborn gender-based abilities to strive for STEM academic excellence a teacher held, the less frequent the constructive STEM teaching assets a teacher provided for the student gender group, the more reinforcing the false self-concept that efforts did not help their STEM accomplishment a student clung to, the worse STEM achievement that withheld students decisions to STEM further studies and careers and perpetuated teacher stereotypical gender beliefs a student enacted (Rosenthal, 2002). That said, mainstream teacher training focused on pedagogical knowledge (Makarova et al., 2019). Other cities' designs could not take over due to city-based differences in STEM learning curriculum and foci teaching aspects which teachers should act. This qualitative case study, therefore, aimed to develop and evaluate a 3-week teacher training course to empower Hong Kong secondary school STEM teachers to co-cultivate gender equality. The study addressed two research questions (hereafter RQs): RQ1) Did the teacher training course reshape STEM teacher concepts on cultivating gender equality in the learning environment? If so, how? RQ2) How could the training course be improved to support Hong Kong secondary school STEM teachers in cultivating gender equality in the learning environment?

## **3. Methodology: training course theoretical frameworks and qualitative data**

### ***3.1. Compound theoretical frameworks for 3-week training course content and pedagogy***

The training content contextualized four teaching aspects impacting student expectancy and task value of SEVT, which mainstream secondary school STEM teachers should conceptualize and act to cultivate gender equality (Eccles & Wigfield, 2020). Teacher expectations implied in lesson materials (textbooks and tailor-made notes) showcasing STEM role models of two genders influenced student expectancy (Makarova et al., 2019). Three teaching aspects influenced student task value: 1) elective and career advice (STEM/Art stream), 2) student recruitment to and grouping in STEM workshops, and 3) attributions for STEM academic success and failure (Maries et al., 2022). Mezirow's transformative learning theory framed training pedagogies with three interdependent elemental changes that guided participant teachers to transform in these SEVT-embedded multi-teaching aspects (Spear & Costa, 2018). 1) Psychological change: participant teachers experienced unprecedented emotional ups and downs in the expectancy and task value when female and male students encountered teacher-led gender struggles. 2) Conviction change: participant teachers related them to daily STEM teaching and set their hearts on falsifying relevant gender-related beliefs and practices. 3) Behavioural change: teachers co-constructed new gender-related beliefs and practices in these teaching aspects and felt self-motivated to enact. Contextualizing the elemental changes, each of the three transformative phases in Appendix 1

pedagogically supported one sequential training module. Fleming (2018) unveiled the feasibility of transformative learning over lecture-based instruction to cultivate educational gender equality. STEM teachers acted as active minds-on and hands-on engagers rather than passive receivers who rejected changes because of “valid” deep-seated gender-related misbeliefs and short-term memory to sustain. The course held three 90-minute tailormade weekly modules in April. Appendix 2 described how SEVT and Mezirow’s transformative phase complementarily guided three subsequent module activities.

### **3.2. Qualitative data collection and analysis**

The study purposively sampled 3 female and 3 male first-year STEM teachers to join and evaluate training. As all have organized diversified STEM activities at different Hong Kong co-educational secondary schools and had no prior exposure to gender-related training courses, they could authentically visualize how their transformed practices influenced students’ expectancy and task value at the earliest year of service and extend evaluation of transformative effectiveness to prolongedly unexplored gender-related concepts (Tang et al., 2021). Equal gender proportion removed selection bias, enabling male and female teachers of different prior gender-related STEM schooling experiences to bring refreshing insights. Codes (Female Teachers: FT1, FT2, FT3; Male Teachers: MT1, MT2, MT3) were assigned to preserve anonymity. Individual semi-structured post-training interviews (compared to pre-training interviews) elicited participant teachers’ personalized changes in gender-related beliefs and practices in four SEVT-embedded teaching aspects and how each transformative learning phase facilitated changes to answer RQ1 (Fleming, 2018). They also probed into practical improvements in training content and pedagogies to answer RQ2. Thematic analysis ran on transcribed interview data to deduce theoretically grounded codes.

## **4. Results, findings and discussion**

The following first answered RQ1, which focused on twofold theoretical evaluations: 1) “Did”- evaluating the effectiveness of SEVT-embedded training content on transformed teachers’ gender-related concepts and practices influencing students’ expectancy and task value, and 2) “How”- evaluating the effectiveness of Mezirow’s transformative learning training phase to call forth corresponding transformations.

### **4.1. Transformative gender-related concepts influencing students’ expectancy in STEM learning**

It evaluated effectiveness of SEVT-embedded training content in module 1 to transform gender-related concepts and practices in teaching aspect of lesson materials (textbooks and tailor-made notes) influencing expectancy of SEVT and the corresponding transformative phase “disorienting dilemma and critical assessment” to bring about the transformation.

#### **4.1.1. Effectiveness of SEVT-embedded training content**

Almost all participant teachers were shocked by male-dominating STEM role models in textbook extracts (excluding FT2, who already noticed) (sample theme categorization in Appendix 3). MT2 realized, “It is out of my expectations that many male scientists appear in the science textbooks” and strived for gender-balanced representations masked by prior foci on teaching content. Remarkably, FT3, who previously justified gender equality in STEM textbooks by the scarcity of field-specific female scientists, recognized that gender-unbalanced issues existed in general textbook topics and devised an alternative of exemplifying herself in the STEM field to include female STEM contributions in elevating female students’ expectancy to succeed in STEM. Moreover, all participant teachers revealed their post-training willingness to balance but not overwhelm role models of two genders in STEM lesson notes so that students could mirror themselves in STEM careers. MT1, who priorly selected three male scientists to include in his tailor-made STEM teaching notes, proposed how he would balance the gender representations in the forthcoming STEM teaching chapter, “When I teach about DNA structures, I can introduce students to Franklin’s (a female scientist) photo 51.” Extensively, FT1 put her transformed ideas of gender-balanced notes into practice, recalling, “Towards the end of the academic year, I am trying to incorporate more examples of female scientists into learning worksheets to inform students that aside from male scientists, female students can also contribute to STEM.”

Participant teachers learned why they should change in bid for students' expectancy and STEM achievement motivation. FT1 recognized, "If teacher expectations are without gender bias, male and female students will believe they can really perform well in STEM-related subjects (expectancy), work harder for better STEM results, and solve problems actively." FT2 and MT3 grasped magnifying influence of teacher expectations on secondary school students' expectancy at Erikson's "identity versus roles" stage that the course selectively targeted, "It is essential to manage our expectations on students who are still exploring whether they can do something, no matter STEAM or not." (Erikson, 1950)

#### **4.1.2. Effectiveness of Mezirow's transformative learning pedagogy**

Nearly all participant teachers (excluding MT3) agreed that "triggered emotional struggles" and "self-assessed similarity in teaching" facilitated the abovementioned changes (sample theme categorization in Appendix 4). Given extracts effectively triggered MT1 and MT2 emotional struggles of resonating with female students' "dilemma" of seeing themselves and sustaining expectancy in STEM careers under overwhelming teacher male-dominated lesson materials and motivated them to watch out for textbook gender-imbalance and balance gender representations in lesson notes. FT1 echoed, "Textbook extracts surprised me a lot because I didn't pay much attention to the gender of scientists. Teachers need to amend some teaching notes and add more examples of female scientists," which facilitated her above practices of amending gender-balanced STEM lesson notes. FT3 and FT2 recognized that constrained lesson preparation time induced her slothfulness in recalling and searching for female scientists. "Because of time constraints during lesson preparations, we will follow the textbooks and suggestions from the curriculum." "Critical assessment" with textbook extracts acknowledged the importance of real-life practice on top of on-paper awareness to elevate expectancy.

These sections revealed that expectancy-embedded content in module 1 hugely empowered participant teachers' transformation of gender-related concepts and practices in lesson materials through the first Mezirow learning phase to cultivate gender equality. Some found textbook extracts engaged them unprecedently in students' frustrations to stay in STEM fields with gender-unbalanced representations of STEM role models (anchoring to "disorienting dilemma"). Affective struggles alerted them not to place the same burden on their students and falsify their previous and even inconspicuous real-life teaching concepts and practices (anchoring to "critical assessment") (Fleming, 2018). Some, therefore, became aware of gender-unbalanced role models in mainstream STEM textbooks and strived for more gender-balanced STEM representations in their tailor-made notes of forthcoming chapters. With this transformation, female students, who long-lastingly saw themselves as under-represented in STEM fields, could visualize females' remarkable STEM contributions and revitalize the expectancy of believing their potential contribution to STEM careers. It also helped male students recognize female STEM contributions and detach their gender stereotypes to extend peer-level gender equality in STEM learning (Becker & Nilsson, 2021).

#### **4.2. Transformative gender-related concepts influencing students' task value in STEM learning**

It evaluated the effectiveness of SEVT-embedded training content in module 2 to transform gender-related concepts and practices in three teaching aspects of elective and career advice (STEM/Art stream), student recruitment to and grouping in STEM workshops, and attributions for STEM academic success and failure influencing task value and the transformative phase "acquisition of new knowledge and provisional trying of roles" to bring about the transformation.

##### **4.2.1. Effectiveness of SEVT-embedded training content**

After the training, FT1, MT1, and MT2 transformed from gender-based to ability/interest-based elective and career advice. MT1 recognized his flawed interest-based advice, "Although I previously claimed that students should take what they want for high school subjects, I tried to push boys to choose STEM-related subjects and girls to opt for non-STEM subjects. I should not have done that." Meanwhile, FT2, FT3, and MT3 saw training course reinforcing their practices in giving ability-/interests-based advice to maintain the STEM task value of female and male students. FT2 added, "As long as students like to participate in STEM, they do not have to make a stunning contribution. They can

still play their roles in STEM and practice STEM in daily life.”, respecting students’ interests rather than focusing on student genders to elevate students’, regardless of gender, the utility value of daily STEM tasks.

MT2 and FT1, who previously recruited male students to STEM workshops first, transformed to open opportunities for two genders to receive equal teacher attention and resources at extracurricular STEM workshops. FT1 proposed how she could maintain equality: “Ask students to raise their hands and join the programme instead.” Nearly all (except FT3) participant teachers transformed from gender-based to gender-free roles in heterogeneous grouping during workshops, including rotatory roles (MT3), “For this time, we assign presentation roles to male students. Next time, we ask female students to be responsible for the presentation. We give both the chance to taste different roles in STEM learning.” and performance-based roles (FT2), “the role of proofreading or logical thinking should not be limited to genders, but rather their strengths.” to sustain cost and attainment value of both genders.

FT1 and MT3 transformed from gender-dependent to gender-free attribution for success and failure, while MT1, FT2, and FT3 formed a more rigid one. FT1 saw the importance of regular positive reinforcement to recognize achievement (high attainment)/effort (for low performance) to revitalize STEM task value and sustain achievement motivation. MT3 freed his attribution from gender labels undermining task value, “Two student genders can achieve the same if they receive the same. We should not link gender to success or failure in STEM tasks.”

Participant teachers internalized transformation conducive to elevated task value of both genders. MT2 saw teacher soundless gender practices mattered, “When teachers agree (implied in these teaching aspects) that gender matters most in STEM-related subjects, students will reject to choose STEAM-related subjects.” FT3 added, “Everyone should have chances to engage. That’s what gender equality looks like (aligning to true gender equality the course emphasized).”

#### **4.2.2. Effectiveness of Mezirow’s transformative learning pedagogy**

Participant teachers concurred that “provisional trying of roles” underlying board game facilitated above transformation in task value. FT2 foresaw STEM pursuits at the end of secondary education (game endpoint) that gender mattered more than ability from student perspectives and transformed to assign performance-based roles in STEM workshops. “High-performing male students have a teacher-led shortcut to higher STEM achievement goals. It widens our horizon to probe into gender stereotypes behind this unfairness.” MT3 additionally enlightened teacher role in gender equality through board game results, “If teachers had provided enough chance for both genders, they could have achieved the same goal.” MT1, who recognized he should not push low-performing male students to choose STEM, added that immersive forward and backward movements of each student character, which simulated ups and downs in task value due to different teacher-led gender privileges and challenges, reminded him of overlooking pressure on male students and hence low intrinsic value with over-pronounced masculine stunning STEM achievement. FT1 further exemplified, “A challenge card was about teacher’s blame on a low-performing male student who then had lower motivation to learn and STEM results. It reminds me not to blame their gender to think logically but to concentrate on their interests,” where “trying of roles” explained FT1’s above transformation in praising every student rather than attributing failure to gender.

FT1, FT3, and MT3 acknowledged fake gender equality practices (acquisition of new knowledge) that this training course intentionally added. They recognized that pull-out female STEM programmes, which aimed to promote gender equality, exploited male students’ participation. They moreover saw similarities between gender equality practices and daily STEM teaching. FT1, who acted as a high-ability male student rejected from pull-out female programmes, felt the unfairness of gender-based student recruitment to STEM workshops. FT2 recognized unnoticeable gender-biased teaching practices undermining male students’ task value, “I paid more attention to girls and helped them solve problems. I realize that when I gave such special treatment to girls, I ignored the boys. I do not solve gender inequality.”

The two sections manifested that task value-embedded content in module 2 hugely supported participant teachers’ gender-related transformation in three teaching aspects through the second learning phase to cultivate gender equality. Tailor-made board game anchored closely to “provisional trying of roles” where role-taking along the board game journey helped teachers experience fluctuations in task value of Hong Kong female and male secondary school students

under teacher-led opportunities and challenges to pursue STEM studies and careers (Fleming, 2018). They could project commonness from games (including teacher-led struggles and game results that high-ability male students usually edged out in STEM future) into their daily STEM teaching to raise abovementioned transformation (Fleming, 2018). Participant teachers moreover saw “acquisition of new knowledge” in not only how their gender-related practices influenced students’ task value aside from expectancy but also fake gender equality. Balan and Stanciu (2021) unveiled that worldwide practices of including ‘A’ in STEAM, which intentionally increased female students’ STEM interests, indeed intensified gender inequality, making feminine arts appealing to female students. Board games, therefore, exemplified fake gender inequality (pull-out female STEM programmes) prevalent in Hong Kong, which effectively reminded participant teachers of true gender equality and its applicability to daily practices (balanced FT2 attention on males).

With these transformative processes, transformation to ability-/interest-based teacher elective and career advice fortified female students’ utility value on STEM tasks to transfer knowledge for future career choices and male students’ intrinsic value to retain with genuine enjoyment (Zhang et al., 2022). Its importance amplified in Hong Kong secondary education to prevent post-secondary dislocation of female students aspiring in STEM to arts while male students aspiring in arts to STEM fields (Zhang et al., 2022). Transformation to academic-/interest-based student recruitment to STEM workshops of extensive teacher resources and attention increased the intrinsic, cost, and attainment value of female students who received equal teacher scaffolds (Maries et al., 2022). Transformation to gender-free roles in heterogeneous grouping revitalized female students’ attainment and utility value in seeing their achievement for group success and male students’ intrinsic value to take preferred roles in STEAM tasks (Maries et al., 2022). Transformation to gender-free attribution for success and failure (success of high-performing female students and failure of low-performing males were not due to uncontrollable and unchangeable luck and gender, as per attribution theory) enabled students to see cost value behind devoting effort and attainment value from teacher recognition (Weiner, 1985).

#### **4.3. Effectiveness of last transformative phase-“building self-efficacy”**

All participant teachers saw that "working together added confidence." FT3, who struggled between gender equality and teaching efficiency, saw this course as networking support for her difficulties. FT1 moreover unveiled the usefulness of scenario cards with critical gender-related incidents of Hong Kong Grade 7-12 students (e.g., taking electives, choosing undergraduate programmes) to put transformation into practice, “We had scenario cards and discussed how we could prevent this scenario at very beginning or what we should not do in the future. My groupmates gave lots of useful suggestions.” FT2 exemplified how collaborative discussions reinforced her confidence in believing in students’ uniqueness, “Being a teacher in one school limits knowing the strengths of students. With other participants, we can argue for gender stereotyping that boys may not excel in STEM subjects, and girls can also do very well.” These together showed that collaborative groups in module 3 hugely boosted participant teachers’ self-efficacy to put gender-related concepts and practices influencing expectancy and task value into authentic teaching from collective teachers’ perspectives with scenario cards specific to the Hong Kong context. The uplifted self-efficacy sustained their beliefs, preventions, and solutions in cultivating gender equality in balance with teaching efficiency (Tang et al., 2021).

#### **4.4. Improvements for training content and pedagogy**

Although SEVT-embedded training course content and pedagogy rooted in Mezirow’s transformative learning guided participant teachers to transform majoritively in the abovementioned teaching aspects, an implicit improvement (interpreted from participant interviews) for SEVT-embedded training content and an explicit improvement (direct suggestions from participants) for transformative pedagogy were necessary to add practicality into paperwork theories with participants’ real-life STEM teaching experiences for further empowerment of gender equality (RQ2).

##### **4.4.1. Implicit improvement on SEVT-embedded training content: further transformation of teacher prior STEM schooling experiences**

MT1’s post-training attribution, “male students may naturally have mathematical sense, but female students’ effort can help.” might still undermine task value of female and low-performing male students in seeing themselves at

comparatively inferior STEM starting points. He explained, “When I was in secondary school, boys had higher academic achievement in STEM-related subjects.” Some teachers projected previous self- and peer- gender-associated STEM performance (their gender-associated struggles, the big picture of gender achievement differences, gender ratio in STEM classes) to their students (Tang et al., 2021). MT1, who generalized gender performance in previous STEM classrooms and formulated parallel gender-related expectations on their students, overlooked STEM potential of female students. Misbeliefs could also form from socialization of former teachers. When they were students, their teachers unconsciously developed their gender-based expectancy and task value (Dinh & Nguyen, 2019). When they became STEM teachers, they exerted reinforced gender-related beliefs and practices on their students whose gender-based expectancy and task value needlessly went up and down. When the cycle perpetuated, teacher-led gender struggles formed more complex ties for future cohorts (Dinh & Nguyen, 2019). Training content should have linked them aside from current teaching to teacher-led gender struggles. Participant teachers should discuss guiding questions in Appendix 5 that probed into their previous STEM schooling experiences right after each board game movement in module 2 for substantial gender equality.

#### **4.4.2. Explicit improvement on transformative learning training pedagogy: Reachable access to gender-balanced STEM role models**

FT2 was aware of gender-unbalanced STEM role models in textbooks yet was slothful for balanced representations in her teaching notes due to “time constraints during lesson preparations.” Reachable access to gender-balanced STEM role models would help solve time constraints, encouraging her to put gender-balanced concepts into real-life practices and strengthening the transformative phase of “building self-efficacy.” MT3 saw the potential of accessible technology (QR codes) for extensive gender equality as some teachers in other schools had little exposure to topic-specific contributions of female scientists, “we can share with teachers at other schools. They will not be limited to very famous scientists in new topics.” Becker and Nilsson (2021) echoed twofold reasons for teachers’ gender-unbalanced representation in STEM lesson materials. First was little knowledge about less well-known topic-specific female scientists (like FT3). Second was adhesion to commonly used textbooks, which saved their lesson preparation time (like FT2) but stood in contrast to “building self-efficacy” that devised gender-related practices should be actualized. QR codes for gender-balanced STEM role models of different STEM chapters of Hong Kong curriculum (sample in Appendix 6) therefore leveraged handy technology to save participant teachers’ efforts and time for accessing less well-known scientists, motivate them to put gender-balanced STEM role models into lesson notes and share with their colleagues to cultivate gender equality at teacher networking level (MT3 suggested) and secondarily peer level which falsified expectations of masculine STEM contributions among peers (Becker & Nilsson, 2021).

### **5. Conclusion: prospective gender equality in the future**

In sum, the tailor-made teacher training course largely reshaped gender-stereotypical concepts and practices influencing the expectancy and task value of SEVT of first-year participant STEM teachers serving at Hong Kong co-educational secondary schools through three Mezirow’s transformative learning phases. It constructed two improvements: 1) further transformation of teachers’ previous STEM schooling experiences and 2) reachable access to gender-balanced STEM role models (with QR codes). Despite limited resources, it hopefully brightened future course directions on co-cultivating gender equality at peer levels (i.e., students are under heavy peer pressure to stay in “unexpected” streams), such that Hong Kong students in current and future cohorts can take off socializing gender lens (schoolboys or schoolgirls) to perceive themselves simply as STEM learners and enter aspiring career streams, as the case study title suggested.

### **References**

- Balan, S. M., & Stanciu, C. (2021). Gender stereotypes and steam education. *International Journal of Advanced Studies in Sexology*, 3(2), 120-125. <https://doi.org/10.46388/ijass.2021.13.48>

- Becker, M. L., & Nilsson, M. R. (2021). College chemistry textbooks fail on gender representation. *Journal of Chemical Education*, 98(4), 1146-1151. <https://doi.org/10.1021/acs.jchemed.0c01037>
- Bronfenbrenner, U. (1979). *The ecology of human development*. Harvard University Press.
- Dinh, D. H., & Nguyen, Q. L. (2019). The involvement of gender in STEM training for teachers. *European Journal of Educational Research*, 9(1), 363-373. <https://doi.org/10.12973/eu-jer.9.1.363>
- Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: a developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology*, 61, 1-60. <https://doi.org/10.1016/j.cedpsych.2020.101859>
- Equal Opportunities Commission (EOC). (2022). *Gender equality in Hong Kong*. EOC (Hong Kong).
- Erikson, E. H. (1950). *Childhood and society*. W. W. Norton & Company.
- Fleming, T. (2018). *Critical theory and transformative learning: Mezirow and the theory of transformative learning*. IGI Global.
- Makarova, E., Aeschlimann, B., & Herzog, W. (2019). The gender gap in STEM fields: the impact of the gender stereotype of math and science on secondary students' career aspirations. *Frontiers in Education*, 4, 1-11. <https://doi.org/10.3389/feduc.2019.00060>
- Maries, A., Whitcomb, K., & Singh, C. (2022). Gender inequities throughout STEM. *Journal of College Science Teaching*, 51(3), 1-16. <https://doi.org/10.1080/0047231X.2022.12290557>
- Piaget, J. (1964). Cognitive development in children: development and learning. *Journal of Research in Science Teaching*, 2, 176-186. <https://doi.org/10.1002/tea.3660020306>
- Rosenthal, R. (2002). *The Pygmalion effect and its mediating mechanisms*. Academic Press.
- Schleicher, A. (2023). *PISA 2022 insights and interpretations*. OECD.
- Spear, A.M., & Costa, R.B. (2018). Potential for transformation? Two teacher training programs examined through a critical pedagogy framework. *Teaching and Teacher Education*, 69, 202-209. <https://doi.org/10.1016/j.tate.2017.10.013>
- Tang, H. H., Chun, W. S., Leung, C. H., & Yau, S. H. (2021). *Aspiring and becoming STEM teachers in Hong Kong*. Routledge.
- Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, 92(4), 548–573. <https://doi.org/10.1037/0033-295X.92.4.548>
- Zhang, Q., Chia, H. M., Chen, K. (2022). Examining students' perceptions of STEM subjects and career interests: An exploratory study among secondary students in Hong Kong. *Journal of Technology Education*, 33(2), 4-19. <https://doi.org/10.21061/jte.v33i2.a.1>

## Appendices

<https://drive.google.com/file/d/149-bot4pqI SmNOteb14IAkdDn9b29Syq/view?usp=sharing>

# Exploring EFL Learners' Academic Emotions and Emotion Regulation Strategies in AI-Assisted Collaborative Academic Writing Tasks

Miao Jia<sup>1</sup>, Yuhua Tong<sup>1</sup>, Zeting Yuan<sup>1</sup>, Zitong Liu<sup>1</sup>, Shuting Wang<sup>1</sup>, You Su<sup>1\*</sup>,

<sup>1</sup> Beijing University of Posts and Telecommunications

\* suyou@bupt.edu.cn

**Abstract:** This mixed-methods study investigated the academic emotions and emotion regulation strategies (ERS) of two groups of EFL learners ( $n=8$ ) during a 6-week AI-assisted collaborative academic writing project. Data from questionnaires and semi-structured interviews revealed that learners experienced diverse emotions, with negative emotions—particularly anxiety, confusion, and anger—outweighing positive and mixed emotions. Moreover, co-regulation, task-related regulation, and cognitive change emerged as dominant ERS, highlighting the role of peer interaction and adaptive problem-solving in managing challenges like AI feedback limitations during AI-assisted collaborative academic writing tasks.

**Keywords:** AI-assisted collaborative writing, Academic emotions, Emotion regulation strategies, EFL learners

## 1. Introduction

Collaborative academic writing fosters skill development but challenges for L2 learners, including task complexity and feedback interpretation. AI tools (e.g., ChatGPT and Doubao) mitigate these issues by providing real-time feedback (Barrot, 2023). Academic emotions and ERS significantly influence collaboration (Järvenoja & Järvelä, 2009), yet research on emotion regulation in AI-assisted writing remains limited. Therefore, this study explored the academic emotions and emotion regulation strategies that EFL learners experience in the face of challenges during AI-assisted collaborative academic writing tasks.

## 2. Methods

This study involved 8 students in two groups completing five collaborative academic writing tasks. Data were collected through vignette-based scenarios to analyze academic emotions and regulation strategies, supplemented by semi-structured interviews. Quantitative analysis involved frequency coding of emotional responses, while qualitative thematic analysis identified patterns in strategy use.

## 3. Results

The results showed that negative emotions (68%) slightly outnumbered positive emotions (9%), mixed emotions (20%), and no emotions (3%). Specifically, anxiety, confusion, and anger were the three most prevalent emotions among the fifteen specific emotions studied. Furthermore, through thematic analysis, the study identified a framework of emotion regulation (ER) strategies, consisting of six strategy families: co-regulation (38.68%), task-related regulation (32.08%), cognitive change (13.68%), situation modification (6.13%), response modification (4.25%), and attention deployment (2.38%).

A thematic analysis of interview data from six participants revealed four primary factors inducing negative emotions in academic contexts. Firstly, environmental and technological challenges predominated, particularly regarding AI tool limitations. Learners employed peer negotiation for model adjustments while demonstrating cognitive acceptance of technological constraints. Secondly, time pressure from procrastination and final-year workloads prompted strategic task management and peer negotiation. Thirdly, academic writing difficulties generated affective

barriers (e.g., boredom, stress, confusion) that were mitigated through seeking help from peers and teachers and iterative practice to improve performance and skills. Lastly, group interaction challenges were mediated via mutual empathy cultivation and peer negotiation strategies. Notably, peer negotiation emerged as a cross-thematic coping mechanism, supplemented by context-specific approaches like technological adaptation and emotional regulation techniques.

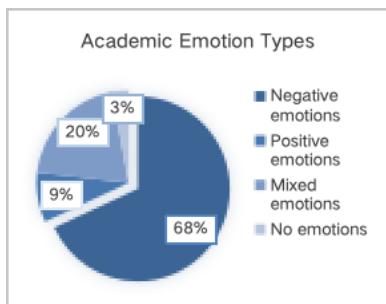


Fig.1 Percentages of academic emotion types

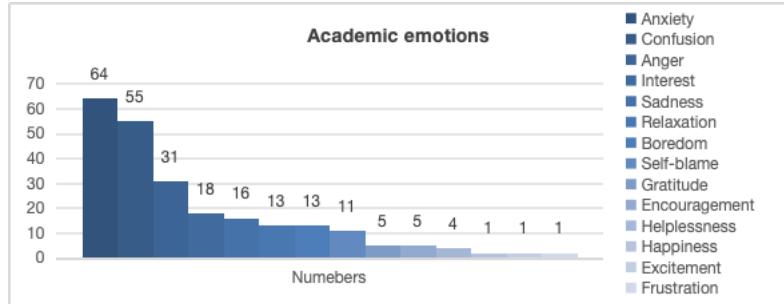


Fig.2 Numbers of specific academic emotions

Table 1. ER strategies framework and their numbers and percentages.

ER strategies	Co-regulation	Task-related Regulation	Cognitive Change	Situation Modification	Response Modification	Attention Deployment	Vague Responses
Numbers.	82	68	29	13	9	6	5
Percentages.	38.68%	32.08%	13.68%	6.13%	4.25%	2.38%	2.36%

#### 4. Discussion and Conclusion

EFL learners experienced significant negative emotions in AI-assisted writing, driven by technological and collaborative challenges. Co-regulation emerged as the most frequently used strategy, followed by task-related regulation and cognitive change. This suggests that students, especially in collaborative environments, benefit from peer interactions and collective problem-solving. This finding echoes previous research that highlights the importance of social regulation strategies in academic settings (Zhang et al., 2021). In conclusion, this study emphasizes the need for further research to explore how individual and group-based emotion regulation strategies can be enhanced in AI-assisted learning environments to enhance emotional resilience and academic writing outcomes in AI-integrated academic environments.

#### Acknowledgements

This study was supported by the Postgraduate Teaching Reform Projects (2023Y037) of Beijing University of Posts and Telecommunications.

#### Reference

- Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, 57, 100745. <https://doi.org/10.1016/j.asw.2023.100745>
- Järvenoja, H., & Järvelä, S. (2009). Emotion control in collaborative learning situations: Do students regulate emotions evoked by social challenges? *British Journal of Educational Psychology*, 79(3), 453–471. <https://doi.org/10.1348/000709909X402811>
- Zhang, Z., Liu, T. & Lee, C.B. (2021). Language learners' enjoyment and emotion regulation in online collaborative learning. *System*, 98, 102478. <https://doi.org/10.1016/j.system.2021.102478>

# ChatGPT Usage Patterns in Essay Writing: A Case Study of Advanced, Intermediate, and Low-Proficiency English Learners

Zitong Liu<sup>1</sup>, Zeting Yuan<sup>1</sup>, Miao Jia<sup>1</sup>, Yuhan Tong<sup>1</sup>, You Su<sup>1\*</sup>

<sup>1</sup> Beijing University of Posts and Telecommunications

\* suyou@bupt.edu.cn

**Abstract:** This study examines how second language learners at different English proficiency levels interact with ChatGPT during argumentative essay writing. By analyzing three university students' writing processes and interviews, the research reveals distinct AI tool usage strategies: high-proficiency learners critically refine language and structure, medium-proficiency students adaptively generate content, and low-proficiency participants heavily rely on AI for argument development and translation. The findings underscore the importance of understanding AI's role in language learning and developing critical AI literacy.

**Keywords:** academic writing, ChatGPT feedback, second language learners, students' usage patterns

## 1. Introduction

ChatGPT shows exceptional advantages in real-time text generalization and revision, which may help solve student essay writing challenges. Despite research on ChatGPT's use in revision (e.g., Fathi & Rahimi, 2024; Koltovskaya et al., 2024), few studies have explored how students' language proficiency influences their interaction with the tool. Therefore, this study investigated ChatGPT's feedback effects on students' essay writing across different language proficiency levels, exploring usage patterns among advanced, intermediate, and low-proficiency L2 learners.

## 2. Method

The study recruited three university students with varying English proficiency levels (IELTS scores ranging from 8.0 to 6.5, majoring in English, History, and E-commerce Engineering and Law) to investigate their ChatGPT usage strategies. The experiment comprised two phases: a 1-hour IELTS argumentative essay writing assisted by ChatGPT and a 40-minute Chinese stimulated recall interview. Data analysis examined students' writing process through video recordings and interviews, revealing AI tool usage strategies across proficiency levels.

## 3. Results

### 3.1. Learner F1: Advanced Proficiency, High-level Self-monitoring, with Critical Evaluation of AI Feedback.

High-proficiency participant demonstrated advanced self-monitoring of her interactions with ChatGPT, while critically evaluating AI's role in supporting academic writing. She did not copy and paste ChatGPT's feedback directly into her writing during the earlier stages. Only after completing the initial draft did she incorporate AI feedback directly into her writing. When asked for the rationale, she explained in the stimulated recall interview, "Directly copying and pasting in the content generalization stage would diminish self-motivation. Paraphrasing based on ChatGPT's feedback could enhance my language skills. However, in the revision stage, copying and pasting is more efficient. (F1-1)"

In the evaluation phase, she critically reflected on ChatGPT's role in academic writing. (F1-2).

F1-2: "The primary function of ChatGPT in academic writing is to assist with grammar and vocabulary improvement. Subsequently, I'm inclined to seek feedback from peers or instructors for further enhancement. Moreover, GPT can't critically assess my arguments, it accepts them without evaluating the internal coherence."

### 3.2. Learner F2: Med-level proficiency, Strategic Prompt Iteration, and ChatGPT-Centered Learning.

The medium-language-proficiency participant frequently refined her prompts according to ChatGPT's responses and primarily relied on ChatGPT's outputs as the main source of material.

Interestingly, when participant F2 did not receive her desired response, she repeatedly revised her prompts and specified her requirements (F2-1). Moreover, participant F2 primarily used the summary generated by ChatGPT as the main framework and arguments, incorporating minor revisions and personal supplementary content (F2-2).

F2-1: *"When I didn't get my desired answer, I first revised my prompt because I believed that ChatGPT had the ability to provide the ideal feedback; I just needed to adjust my prompt. I needed to repeatedly emphasize the need for 'concise' in my prompts and explicitly limit the word count to receive my desired response."*

F2-2: *"Since the summary generated by ChatGPT aligned with my pre-existing viewpoints and the length of it was also acceptable. Therefore, I used it as the main body of my writing."*

### **3.3. Learner M1: Low-proficiency, Heavy Dependence on ChatGPT Feedback Especially in Content Generalization**

The low-proficiency participant showed heavy dependence on ChatGPT feedback. This reliance was so pronounced that he expressed disappointment when unable to incorporate more ChatGPT-generated arguments into his work (M1-1). Furthermore, participant M1 showed a distinct L1-mediated approach when engaging with ChatGPT during his writing. He first composed a Chinese version essay, utilizing L1 prompts to interact with ChatGPT. Finally, he used ChatGPT to translate his completed Chinese essay into English at the end of the writing process. When asked for the reason in using L1, he expressed in the stimulated recall interview, "Writing in English is less efficient for me than using Chinese. With Chinese, I can quickly assess whether the AI-generated content aligns with my views. Additionally, I frequently rely on AI for translation and have great trust in its translation capabilities (M1-2)."

M1-1: *"The adoption rate for arguments should reach 50% or higher. For example, if ChatGPT provides me with four arguments, I should be able to use at least three."*

## **4. Conclusion**

This study provides valuable insights into the differentiated use of ChatGPT by students of varying English proficiency levels. The high-proficiency participant strategically used AI to refine her writing, maintaining a critical stance toward its outputs. The Medium-proficiency participant demonstrated adaptive strategies to maximize the benefits of AI, while the low-proficiency participant relied heavily on AI to overcome L2 challenges as an EFL learner.

## **Acknowledgments**

This study was supported by the Postgraduate Teaching Reform Projects (2023Y037) of Beijing University of Posts and Telecommunications.

## **References**

- Fathi, J., & Rahimi, M. (2024). Utilising artificial intelligence-enhanced writing mediation to develop academic writing skills in EFL learners: a qualitative study. *Computer Assisted Language Learning, 1–40*.
- Koltovskaia, S., Rahmati, P., & Saeli, H. (2024). Graduate students' use of ChatGPT for academic text revision: Behavioral, cognitive, and affective engagement. *Journal of Second Language Writing, 65, Article 101*