

Proceedings of the International CALL Research Conference INCLUSIVE CALL

3 – 5 Oct, 2025

**The University of Queensland, Brisbane,
Australia**

Editors

**Yijen Wang, Paul Moore,
Teresa Hsieh & Hao Tran**

4th Floor, Silverstream House, 45 Fitzroy Street Fitzrovia, London W1T 6EB, United Kingdom
Level 9, 440 Collins Street, Melbourne, Victoria 3000, Australia
2nd Floor Daiya Building, 2-2-15 Hamamatsu-cho, Minato-ku, Tokyo 105-0013, Japan
447 Broadway, 2nd Floor #393, New York NY, 10013, United States
First published 2025 by Castledown Publishers, Melbourne

Information on this title:

www.castledown.com/reference/9781763711624

DOI: 10.29140/9781763711624

Transitions in CALL

© Yijen Wang, Paul Moore, Teresa Hsieh & Hao Tran (Editors) 2025

All rights reserved. This publication is copyright. Subject to statutory exception and to the provisions of relevant collective licencing agreements, no reproduction, transmission, or storage of any part of this publication by any means, electronic, mechanical, photocopying, recording or otherwise may take place without prior written permission from the author.

Typeset by Chennai Publishing Services Pvt. Ltd.

ISBN: 978-1-7637116-2-4 (Paperback)

ISBN: 978-1-7637116-3-1 (Digital)

Castledown Publishers takes no responsibility for the accuracy of URLs for external or third-party internet websites referred to in this publication. No responsibility is taken for the accuracy or appropriateness of information found in any of these websites.

Table of Contents

Using Voice-Enabled AI Chatbot to Improve Academic Presentation Skills	1
<i>Aisha Aslam and Dr. Saqib Aziz</i>	
The Relationship Between Media and Technology Usage and Learning Styles in Language Learning	13
<i>Ph.D. Candidate Elham Azad and Prof. Dr. Mehmet Ali Yavuz</i>	
Multi-Agent Debate in L2 Argumentative Writing: An Exploratory Study of Chinese EFL Learners	21
<i>Tianshu Chu</i>	
Undergraduates' Reflection on Applying ChatGPT in ESP Writing Skills: A Case from Foreign Trade University HCMC	29
<i>Thi My Dung Dang</i>	
L2 Spanish and Indonesian Learners' Engagement with ChatGPT Written Corrective Feedback	35
<i>Ms Alicia Ximena Gazmuri Sanbueza, Ms Franciele Spinelli and Dr David Wijaya</i>	
Enhancing Inclusivity and Equity through a Broader Understanding of Data-Driven Learning in Pre-tertiary Education: A Mixed-Methods Systematic Review.....	45
<i>Shiya Huang and Qing Ma</i>	
AI-Supported Collaborative Learning to Reduce EFL College Students' Anxiety in Story Writing.....	51
<i>Siao-Cing Lai, Yu-Fen Yang, Christine Chifen Tseng and Min-Chuan Tsai</i>	
Exploring Digital Literacy in ESP Tourism Presentations Using Importance-performance Analysis	59
<i>Gayoung Lee</i>	
From Lookup to Output: Exploring Mobile Vocabulary Searches for Self-Expression in Multicultural Japanese Language Learning.....	65
<i>Yuzhe Li and Kei Ishiguro</i>	
The Transformative Role of E-Readers in Inclusive Online Mandarin Education: A Duo-Ethnographic Study.....	71
<i>Chuan Liu and Jing (Vera) Huang</i>	
Using a Custom GPT to Scaffold Corpus-Based Vocabulary Learning for Low-Proficiency Learners	77
<i>Jing Liu and Qing Ma</i>	
Vietnamese Pre-service Language Teachers' Technology-Mediated Task Creation: An Ecological Perspective	85
<i>Paul J. Moore, Giang Hong Nguyen and Quang Vinh Nguyen</i>	

Beyond the Game: The Theoretical Exploration of Gamification in Foreign Language Teaching.....	91
<i>Nguyen Thi La</i>	
Leveraging the Affordances of Artificial Intelligence Chatbots to Create Differentiated Instructions and Cater to Learners' Differences in an English as a Foreign Language Class.....	99
<i>Trang Thi Thuy Nguyen and Phuc Kim Nguyen</i>	
Using Dialogic Multimodal Feedback to Enhance University EFL Students' Writing: Affordances for Language Learning Styles.....	107
<i>Nguyen Thi Thuy Trang, Edwin Creely and Huynh Anh Tuan</i>	
A Large-Scale Mixed-Methods Study of Japanese University Students' Use of ChatGPT for L2 Learning.....	115
<i>Louise Obashi and Suwako Uehara</i>	
Designing Collaborative Resources in Data-Driven Learning to Accommodate Learner Diversity	123
<i>Muhammad Rudy, Emi Emilia, Wawan Gunawan and Destiani</i>	
AI as a Thinking Partner? Exploring International Students' Use of GenAI to Support their Academic English Needs.....	145
<i>Franciele Spinelli</i>	
Instructional Design for Co-Created and Context-Based VR Chinese Learning Environments.....	155
<i>Satoko Sugie</i>	
Understanding the Digital Literacies of English Major Students in Tertiary Education	161
<i>Yong Tan</i>	
Beyond Technical Competencies: A Critical Analysis of Global Research on Language Teacher AI Literacy.....	167
<i>Danping Wang and Qianwei Zhang</i>	
Evaluating Advanced EFL Learners' Writing Performance and Perceptions of Large Language Model Assistance	173
<i>Mengkai Wang</i>	
Artificial Intelligence-Driven Personalized Language Learning: Customizing Content and Feedback to Learners' Needs and Proficiency Levels.....	179
<i>Mengkai Wang and Yifei Gao</i>	
Developing College Students' Creative Thinking and English Story Writing Through AI-supported Collaborative Learning.....	185
<i>Shih-Chiao Wang and Yu-Fen Yang</i>	
Ethical Principles for Generative AI in Research Paper Writing for EFL Graduate Students in China.....	191
<i>Zeng Jianbin and Zhang Xubua</i>	
Using Large Language Models to Enhance Data-driven Learning in Academic Writing	197
<i>Xubua Zhang and Jianbin Zeng</i>	

Using Voice-Enabled AI Chatbot to Improve Academic Presentation Skills

Aisha Aslam¹  and Dr. Saqib Aziz² 

¹The University of Queensland, Australia

²Bahauddin Zakariya University, Pakistan

Abstract

Presenting research is as important as conducting research; requiring novice researchers to practice and review their presentations continuously. Given the limited time in language classrooms, developing impactful presentation skills is difficult. Tools like AI chatbots could help overcome this challenge. While research shows AI tools aid in developing oral proficiency, there is limited empirical evidence on the effectiveness of voice-enabled chatbots for improving academic presentation skills. This study investigates the effectiveness of a voice-enabled chatbot 'Oratio', in enhancing academic presentation skills, mainly the lexicogrammatical and structural features. The study also explores the possibility of Oratio reducing the students' anxiety related to giving academic oral presentations. The participants of this experimental study were 20 Master of Philosophy, first and second year students from the English department at a public university in Pakistan. Pre-intervention and post-intervention presentations were assessed using a rubric designed by the researchers. The experimental group used Oratio and the control group did not use any interventions to support their presentation skills improvement. The Wilcoxon signed rank test showed no significant improvement in the pre-test and post-test results in the control and experimental groups; however, there is a marked improvement reflected by the median scores in the post-intervention performance for all the participants. Using Oratio appears to have helped students build their confidence by improving lexicogrammatical

E-mail: aisha.aslam@uq.edu.au & saqib@bzu.edu.pk

and structural issues in their presentations. Although this pilot study has limitations, it illustrates the potential of using voice-enabled AI chatbots in improving academic oral presentation skills of novice researchers.

Keywords: oral presentation, academic speaking, AI chatbots

Introduction and Background

There is an increasing expectation from researchers to not only disseminate their research with other scientists but also with a wider, non-expert audience (Lamarche et al., 2024; Plo-Alastrué & Corona, 2023). Not being able to naturally communicate well should not stop researchers from learning the craft of delivering engaging presentations (Youknovsky & Bowers, 2024). It is crucial that university students are taught to share their research with a wider audience in an engaging and accessible manner. AI chatbots can play a vital role in providing the support that is needed (Belda-Medina & Calvo-Ferrer, 2022; Belda-Medina & Kokošková, 2023; Labadze et al., 2023; López-Minotta et al., 2025; Pituxcoosuvann et al., 2025).

Research on AI chatbots has shown their potential in developing English language learners' speaking skills. A systematic review conducted by Du and Daniel (2024) focused on the literature (published between 2017 and 2023) on English-speaking skills improvement using AI chatbots, and concluded that using AI chatbots has generally proved to be effective in vocabulary (Chen et al., 2023) and grammar learning (Tai, 2024); confidence building (Zou et al., 2023); alleviating anxiety (Han, 2020; Tai & Chen, 2023); speaking fluency (Wu et al., 2023) and improving learners' motivation. There are a few studies done to date, investigating the improvement of oral presentation skills using AI chatbots. Garcia et al. (2024) studied the effectiveness of AI chatbots on reducing public speaking anxiety and improving self-perceived confidence and reported positive results. Another small-scale, preliminary study attempted to use AI chatbots along with the physiological detection model to help reduce anxiety in students while giving presentations (Kimani & Bickmore, 2019).

However, the research on the effectiveness of AI chatbots has shown certain limitations. For example, the studies have reported challenges in speech recognition leading to lack of accuracy of feedback given by the AI chatbots (Chen et al., 2023; Moussalli & Cardoso, 2016) resulting in lack of user satisfaction (Tai, 2024). The other limitation of AI chatbots is the occasional comprehensibility issues between the chatbots and the learners (Chen et al., 2023; Wu et al., 2020).

The lack of empirical evidence on using chatbots to improve lexicogrammatical and structural features in academic presentations of students in tertiary education has motivated this pilot study. Therefore, the researchers recruited the MPhil students to use a voice-enabled AI chatbot, Oratio, to help them practice their academic presentations. ‘Oratio’ is developed by the researchers with the help of a developer, and its efficacy is tested to answer the following questions:

- RQ1. How effective is Oratio in improving the research students’ lexicogrammatical features, structure, clarity, confidence, and engagement in academic presentations?
- RQ2. Does Oratio help students reduce their anxiety related to delivering academic presentations?

Methodology

The main objective of this study was to assess the effectiveness of a voice-enabled AI chatbot to improve students’ academic presentation skills, therefore, ‘Oratio’ – a voice-enabled chatbot, was developed because most voice-enabled chatbots are behind paywalls. The chatbot in its beta version is developed using Open AI’s GPT-4o-mini model. The current version takes audio input; therefore, the students can practice delivering their presentation in a safe, stress-free manner. Its feedback in this version is text-only, but voiced output can be an added feature in its full version. We decided to go against using the voiced feedback at this stage as the freely available chatbot voices are very robotic. Once funding is secured for this project, a full version of Oratio will have more realistic voiced as well as written feedback. Oratio gives feedback on vocabulary, grammar, structure, clarity, and engagement of the presentation based on the features of TED Talks.

The data were collected as oral presentations delivered by the MPhil English students in two groups: first year and second year students. Each group was divided into two sub-groups: (a). the Experimental group that used the voice-enabled chatbot ‘Oratio’, and (b). the Controlled group that did not receive any instructions or feedback. After the pre-intervention presentations, the latter group was suggested to watch some examples of good presentations on YouTube, such as TED Talks while the Experimental group used Oratio as their intervention. The students in first year presented a topic from a published research paper while the second year students presented their own research as they were already working on their MPhil thesis.

Table 1. Examination with pre- and post-tests across first and second year Mphil students

Groups	N	Median	Z value	p value	r
Pretest 1 st year control group	5	15	-1.76	0.7	-1.795
Post-test 1 st year control group	5	17			
Pre-test 1 st year experimental group	5	17	-1.84	.06	-1.718
Post-test Pretest 1 st year experimental group	5	21			
Pre-test 2 nd year control group	5	15	-.67	.50	-4.719
Post-test 2 nd year control group	5	18			
Pre-test 2 nd year experimental group	5	16	-1.84	.06	-1.718
Post-test 2 nd year experimental group	5	18			

Using the rubric designed by the researchers (see Appendix A), students' presentations were evaluated. The Wilcoxon signed-rank test was performed to analyze the differences in students' presentations, pre- and post-interventions.

Results and Discussion

The results of the Wilcoxon signed rank test did not show any significant improvement in the pre-test and post-test results in control and experimental groups; however, there is a marked improvement reflected by the median scores in the post-intervention performance across all the groups (see Table 1).

Structure of Presentations

Most of the students ($n = 7/10$) showed structural awareness in presentations. It is through the use of sentence structure that clearly indicates the part of the presentation being presented. The following examples from the first and second year experimental group show this. For example, they clearly

mentioned the sections in their presentation, as can be seen in the examples below:

(a) *Then there is an introduction.*

Then there is a difference of the EMI, ICLH and CLIC:

Then there is an English medium paradigm which is a ... (HA Pre-T-1)

The second part *is concerned with the rise of English education.*

Then the next part is *concerned with the key differences. (HA Post-T-1)*

Grammatical Accuracy/Complexity

More than half of the students ($n = 6/10$) used grammatically complex sentences in the experimental group. These students used complex sentence structures with appropriate use of connectors or linking words. In the example below, the first year student used a simple grammatical structure to describe ‘code switching’ in the pre-test, whereas in the post-test, the same participant used a complex grammatical structure to describe ‘code switching’.

(b) *Code switching is a very interesting [...]* (NA Pre-T-1)

So as far as the usage of code switching is concerned, *it is used [...]*

(NA Post-T-1)

Vocabulary Range and Appropriacy

The experimental group, second year, also showed variation in the use of vocabulary post-intervention.

(c) *Today, let's **delve into** the fascinating evolution of dictionaries. I'll*

delineate *how dictionaries have evolved over time, specifically the shift from prescriptive to descriptive approaches in lexicography. (AL Post-T-2)*

Overall Clarity and Impact

Majority of the students in the experimental group ($n = 7/10$) showed improvement in overall clarity in presentations. The following extracts show that in the post-intervention, the student has conveyed the same ideas with more clarity.

- (d) *So to conclude that replication is vital yet underdeveloped in applied linguistics [...]* (MA Pre-T-1)
So replications hold a lot of significance for the applied linguistics [...] (MA Post-T-1)

Audience Engagement and Persuasion

More than half the students in the experimental groups ($n = 6/10$) displayed awareness in engaging the audience. The following example (e) shows that the experimental group student has used the phrase “*as you see*” to engage the audience.

- (e) *We see that the present study... We see the findings that the words like Janak, [...]*as well*.* (AA Pre-T-1)

*For example, we see in the findings that [...] frequency **as you see** in the article* (AA Post-T-1)

Use of Metaphor, Simile, or Analogy

A few students in the experimental group ($n = 4/10$) made a use of analogies to explain the content of their presentation. In example (f) the student has used an analogy to make their presentation interesting.

- (f) *Myself, [name, roll number]. Today I am going to explore the topic, The English [...] Education.* (HS Pre-T-2)
*My name is [...] **Imagine navigating a vast ocean of knowledge and language is a vessel that carries us through.*** (HS Post-T-2)

The findings of the current study show that the students who used Oratio tried to make their presentations more interactive and interesting (see Appendix B for screenshots of students’ interaction with Oratio). The use of Oratio has helped in reducing their anxiety related to presentation. These results are consistent with the findings of the previous research (Askari & Rahim, 2024; Kimani & Bickmore, 2019). This could be explained in the light of Krashen’s (1982) Affective Filter Hypothesis and Sweller’s (2011) Cognitive Load Theory. More practice coupled with structured and personalized feedback may have reduced affective filter and lower the cognitive load for the students. The control group has shown little improvement in the use of vocabulary, structure, and engagement with the audience.

As their presentations lacked overall clarity and impact, and they did not try to engage with the audience, they were less confident and descriptive in their presentation. A possible reason could be the lack of awareness of the various features of academic presentation.

Conclusion

It is important that postgraduate students learn to improve their academic presentations. In the current study, the use of Oratio, a voice-enabled chatbot, has shown some improvement in the structure and confidence of the novice researchers. There are various limitations and challenges that this study faced. The small size of participants, discrepancy in tasks given to first year and second year students, the short time available to the Oratio group participants for practising their presentations before the post-tests, and dealing with natural disaster and unforeseen political tension with the neighbouring country were a few challenges that the researchers had to face during the data collection process. Above all, the AI chatbot, Oratio, was developed by the researchers on a self-finance basis. Due to limited funds, Oratio could not be designed as per the needs of the learners. We plan on using WhisperAI, deepgram and GPT-5 in the full version of Oratio to enable us to add better voiced output and playback along with the text feedback. We would also improve the UX design of Oratio and may develop a companion mobile app.

References

- Askari, M., & Rahim, A. S. (2024). The role of AI in enhancing presentation skills in English language teaching. *International Journal of Business and Applied Social Science*, 10(12), 1–6. <https://doi.org/10.33642/ijbass.v10n12p1>
- Belda-Medina, J., & Calvo-Ferrer, J. R. (2022). Using chatbots as AI conversational partners in language learning. *Applied Sciences*, 12(17), Article 8427. <https://doi.org/10.3390/app12178427>
- Belda-Medina, J., & Kokošková, V. (2023). Integrating chatbots in education: Insights from the Chatbot-Human Interaction Satisfaction Model (CHISM). *International Journal of Educational Technology in Higher Education*, 20(1), Article 62. <https://doi.org/10.1186/s41239-023-00432-3>
- Chen, H. H. J., Yang, C. T. Y., & Lai, K. K. W. (2023). Investigating college EFL learners' perceptions toward the use of Google Assistant for foreign language learning. *Interactive Learning Environments*, 31(3), 1335–1350. <https://doi.org/10.1080/10494820.2020.1833043>

- Du, J., & Daniel, B. K. (2024). Transforming language education: A systematic review of AI-powered chatbots for English as a foreign language speaking practice. *Computers and Education: Artificial Intelligence*, 6, Article 100230. <https://doi.org/https://doi.org/10.1016/j.caeai.2024.100230>
- Garcia, F. V., Jr., Pascual, N. F., Sybingco, M. E., & Ong, E. (2024). Addressing public speaking anxiety with an AI speech coach. In A. Kashiwara et al. (Eds.), *Proceedings of the 32nd International Conference on Computers in Education*. Asia-Pacific Society for Computers in Education. <https://doi.org/10.58459/icce.2024.4955>
- Kimani, E., & Bickmore, T. (2019). Addressing public speaking anxiety in real-time using a virtual public speaking coach and physiological sensors. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents (IVA '19)* (pp. 260–263). Association for Computing Machinery. <https://doi.org/10.1145/3308532.3329409>
- Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Pergamon.
- Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: Systematic literature review. *International Journal of Educational Technology in Higher Education*, 20(1), Article 56. <https://doi.org/10.1186/s41239-023-00426-1>
- Lamarche, V. M., Kung, F. Y. H., Finkel, E. J., Jayawickreme, E., Rattan, A., & Wheatley, T. (2024) How to give great research talks to any audience. *Nature Human Behaviour*, 8, 616–621. <https://doi.org/10.1038/s41562-024-01839-2>
- López-Minotta, K. L., Chiappe, A., & Mella-Norambuena, J. (2025). Implementation of Artificial Intelligence to improve English oral expression. *Multidisciplinary Journal of Educational Research*, 15(1), 43–71. <https://doi.org/10.17583/remie.16188>
- Moussalli, S., & Cardoso, W. (2016). Are commercial ‘personal robots’ ready for language learning? Focus on second language speech. *CALL Communities and Culture—Short Papers from EUROCALL*, 325–329. <https://doi.org/10.14705/rpnet.2016.eurocall2016.583>
- Pituxcoosuvann, M., Tanimura, M., Murakami, Y., & White, J. S. (2025). Enhancing EFL speaking skills with AI-powered word guessing: A comparison of human and AI partners. *Information*, 16(6), Article 427. <https://doi.org/10.3390/info16060427>
- Plo-Alastrué, R., & Corona, I. (Eds.). (2023). *Digital scientific communication: Identity and visibility in research dissemination*. Springer.
- Sweller, J. (2011). Cognitive load theory. In *Psychology of learning and motivation* (Vol. 55, pp. 37–76). Academic Press.

- Tai, T. Y. (2024). Effects of intelligent personal assistants on EFL learners' oral proficiency outside the classroom. *Computer Assisted Language Learning*, 37(5–6), 1281–1310. <https://doi.org/10.1080/09588221.2022.2075013>
- Tai, T. Y., & Chen, H. H. J. (2023). The impact of Google Assistant on adolescent EFL learners' willingness to communicate. *Interactive Learning Environments*, 31(3), 1485–1502. <https://doi.org/10.1080/10494820.2020.1841801>
- Wu, C., Lam, H., Kong, A., & Wong, W. (2023). The design and evaluation of a digital learning-based English chatbot as an online self-learning method. *International Journal of Engineering Business Management*, 15. <https://doi.org/10.1177/18479790231176372>
- Wu, E. H. K., Lin, C. H., Ou, Y. Y., Liu, C. Z., Wang, W. K., & Chao, C. Y. (2020). Advantages and constraints of a hybrid model K-12 E-Learning assistant chatbot. *Ieee Access*, 8, 77788–77801. <https://doi.org/10.1109/ACCESS.2020.2988252>
- Youknovsky, A., & Bowers, J. (2024). The SELL method: Prepare your presentation. In A. Youknovsky & J. Bowers (Eds.), *SELL YOUR RESEARCH: Public Speaking for Scientists* (pp. 5–14). Springer International Publishing. https://doi.org/10.1007/978-3-031-60887-2_2
- Zou, B., Du, Y., Wang, Z., Chen, J., & Zhang, W. (2023). An investigation into Artificial Intelligence speech evaluation programs with automatic feedback for developing EFL learners' speaking skills. *SAGE Open*, 13(3). <https://doi.org/10.1177/2158244023119381>

Appendix A

The Rubric to Evaluate Presentations

Criterion	Level 1 (Limited)	Level 2 (Developing)	Level 3 (Proficient)	Level 4 (Strong)
Structure of Presentation	<i>Lacks clear beginning, body, and end. Ideas are disjointed.</i>	<i>Attempt at structure but transitions are unclear; some logical flow.</i>	<i>Clear structure with recognizable introduction, body, and conclusion. Transitions mostly smooth.</i>	<i>Strong and cohesive structure with well-signposted transitions; smooth flow throughout.</i>
Grammar Accuracy (CEFR A2–B2)	<i>Frequent errors impede understanding (A1–A2 level).</i>	<i>Some basic grammatical control, though errors are noticeable (low B1).</i>	<i>Generally good control of simple and some complex forms; few errors (mid–high B1).</i>	<i>Uses a range of grammatical structures with high accuracy and flexibility (B2+).</i>
Vocabulary Range and Appropriacy (CEFR A2–B2)	<i>Limited vocabulary; repetitive or incorrect word choices; lacks academic register.</i>	<i>Some variety; attempts to use academic vocabulary; occasional misuse.</i>	<i>Appropriate range of vocabulary, including some academic terms; mostly accurate.</i>	<i>Wide lexical range; precise and appropriate use of academic and topic-specific terms.</i>
Confidence in Delivery	<i>Hesitant, reads from notes/ script; avoids eye contact; posture closed.</i>	<i>Some hesitations, occasional eye contact; limited engagement with audience.</i>	<i>Generally confident, maintains eye contact; uses gestures; some natural delivery.</i>	<i>Speaks fluently with clear articulation; confident body language and tone; highly engaging.</i>
Overall Clarity and Impact	<i>Difficult to follow; lacks coherence or persuasiveness.</i>	<i>Somewhat clear but lacks impact or persuasiveness.</i>	<i>Clear and coherent; makes an effort to engage the audience.</i>	<i>Clear, impactful, and engaging; demonstrates persuasive communication.</i>
Audience Engagement and Persuasion	<i>Monotonous or disengaged; lacks connection with the audience.</i>	<i>Occasional attempts to engage the audience.</i>	<i>Generally engaging; uses some rhetorical questions or hooks to persuade or connect.</i>	<i>Highly engaging and persuasive; uses rhetorical questions, 'survey questions' and/or hooks effectively.</i>
Use of Metaphor, Simile, or Analogy	<i>No figurative language used; presentation is literal and dry.</i>	<i>Rare use of metaphor/ simile; not clearly connected to content.</i>	<i>Effective use of metaphor or analogy in parts of the presentation.</i>	<i>Creative and well-integrated use of metaphor/ simile/analogy that enhances understanding and engagement.</i>

Note. The inter-rater reliability of the rubric is tested for both pre-test and post-test evaluations. Pre-test scores was examined using a two-way mixed-effects model of the intraclass correlation coefficient (ICC), assessing consistency between two evaluators' ratings of speaking proficiency. The analysis of pre-test evaluation showed a single-measures ICC of .85, 95% CI [.67, .94], indicating substantial agreement when considering one rater alone. The average-measures ICC was .92, 95% CI [.80, .96], reflecting excellent reliability. Similarly, the analysis of post-test evaluation showed a single-measure ICC of .93, 95% CI [.84, .97], indicating excellent agreement when considering one rater alone. The average-measures ICC was .96, 95% CI [.91, .98], reflecting excellent reliability.

Appendix B

The screenshots from the chatbot- students interaction may be accessed [here](#).

The Relationship Between Media and Technology Usage and Learning Styles in Language Learning

Ph.D. Candidate Elham Azad  and

Prof. Dr. Mehmet Ali Yavuz 

Department of English Language Teaching (ELT), Cyprus
International University, Cyprus

Abstract

This study examines the correlation between Iranian students' learning styles and their media and technology use for language acquisition in Australian higher education. Understanding of how learners' technology use aligns with their personal learning preferences can offer valuable insights for more effective pedagogical design. In this study, the Learning Style Survey (LSS) and the Media and Technology Usage and Attitudes Scale (MTUAS) were used in a quantitative survey to measure dominant learning styles and technology use frequency among 85 participants. The results revealed that approximately 75% of participants preferred visual learning, and the most frequently used tools for language learning were smartphones and text messaging. Correlation analysis revealed that there was a weak association between specific learning styles and particular technologies—for example, visual learners with television viewing and auditory learners with phone calls. These results suggest that, in multicultural contexts, designing Computer-Assisted Language Learning (CALL) environments should prioritize accessibility and prevailing digital habits rather than strict adherence to learning style alignment.

Keywords: Learning styles, computer-assisted language learning (CALL), higher education, Iranian students, technology-enhanced language learning

E-mail: e.azad00@gmail.com & mayavuz@ciu.edu.tr

Introduction

Rapid advances in digital technology have transformed the ways in which languages are taught and learned, expanding opportunities for interaction and personalization in educational settings. Mobile applications, AI-driven platforms, and other digital instruments can enhance learner motivation and autonomy as well as the development of language skills (Lai et al., 2022). Yet the effectiveness of these technologies is not uniform across learners. Differences in how individuals process and retain information—often described as learning styles such as visual, auditory, or kinesthetic—may shape how they select and engage with digital resources. Although recent research has cast doubt upon matching instruction to individual preferences in this way (Emban et al., 2025), it is still valuable for optimizing teaching design to investigate their association with the use of technology.

Despite extensive research on learning styles and tech-mediated language learning, little research has investigated an area's crossing within an international students' context of Iranian international students in Australian higher education—a growing cohort whose learning experiences are shaped by cultural adaptation, varying English proficiency, and differing prior exposure to educational technologies. This study addresses that gap by the investigation of students' learning styles and their digital language learning tool engagement, which provides evidence of how cultural and educational backgrounds may influence technology engagement. Results will be useful in designing culture-responsive EdTech tools and extending theoretical understandings on learner–technology interaction in multicultural environments.

Literature Review

Currently, technology plays an essential role in language learning, offering clear benefits such as stronger vocabulary retention, better fluency, and richer opportunities for collaboration. An earlier study by Lai et al. (2014) revealed that with clear instructional goals, digital tools can enhance all core language skills. In a recent study, Lee and Lee (2024) reported that AI-guided, personalized language learning significantly boosts learners' performance. Zhang et al. (2023) found that chatbot-assisted learning can outperform traditional methods when design and interaction features are well optimized. Mobile-assisted language learning (MALL) has been shown to foster autonomy and engagement by allowing learners to personalize and direct their own study (Nurmala et al., 2023).

Learning style as individual preferences for processing information can influence how learners choose and interact with digital tools. While earlier research (Cohen et al., 2001) suggested that aligning teaching to such preferences could improve learning outcomes, later studies have been more cautious. Newton and Miah (2017) and Emban et al. (2025) both highlighted the limited evidence for strict alignment, though they note that multimodal approaches, which accommodate a variety of preferences, can help maintain motivation and engagement in technologically advanced settings.

Despite the growing interest in both learning styles and technology-assisted language learning, their intersection has received limited attention, particularly in international higher education contexts. Research on adaptive platforms has shown that technology can accommodate diverse preferences (Kukulska-Hulme & Viberg, 2017). However, newer developments of AI-driven personalisation of CALL have seen limited interest in this respect. Examining how Iranian students' learning styles interact with their use of technology can help in designing EdTech that is responsive to culture and further theoretically develop learner–technology interaction across multicultural settings.

Research Questions

The current study seeks to answer the following research questions:

1. What are the most common learning styles among Iranian students enrolled in Australian higher education institutions?
2. How frequently, and for what purposes, do these students use different forms of media and technology for language learning?
3. How do students' learning styles relate to the frequency and type of media and technology use for language learning?

Methodology

The study involved 85 Iranian students from various program, aged 18–40, enrolled in undergraduate and postgraduate programs at Australian universities. This age range reflects the typical university demographic, with mature learners over 40 excluded. Participants were recruited via university mailing lists and Iranian student networks. Ethics approval was granted by Cyprus International University (Approval No. EKK24-25/16/10).

To collect the necessary data, two instruments were employed. The Learning Style Survey (Cohen et al., 2001) identified learners' dominant

preferences. This study focused on the Visual, Auditory, and Kinesthetic (VAK) categories due to their relevance to technology use in language learning although the instrument includes several dimensions. While the VAK model has been debated (Newton & Miah, 2017), it remains pedagogically useful when applied with caution. The other validated instrument—the Media and Technology Usage and Attitudes Scale (MTUAS; Rosen et al., 2013) assessed students' frequency of digital tool use and attitudes. Only subscales relevant to language learning—smartphones, messaging, email, and video streaming—were included. A small pilot study confirmed cultural appropriateness and language clarity.

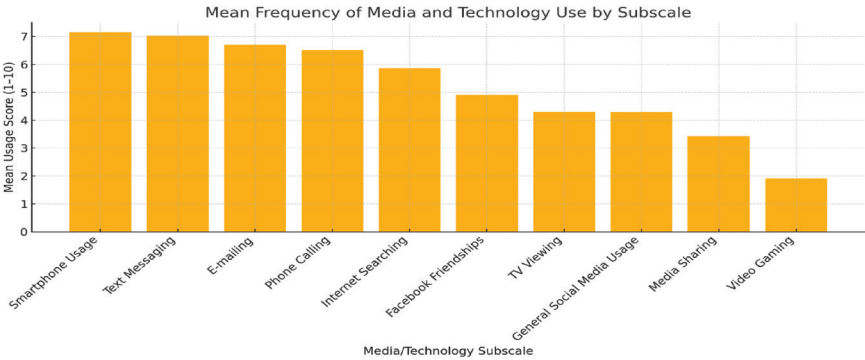
Data collection was conducted online via Google Forms over a three-week period. Eighty-five out of 102 cases were retained. Descriptive statistics, Pearson correlations, and one-way ANOVA were used to examine learning style prevalence and its relationship to technology use.

Result (Hypothetical Overview)

To address the first research question, descriptive statistics were used to summarize the distribution of dominant learning styles. Visual learning was the most prevalent preference (75%) followed by auditory (15%) and kinesthetic (10%). No significant variation was observed across age groups or gender.

In response to the second question, analysis of MTUAS subscales indicated that smartphone use and text messaging were the most frequently used tools for language learning, with over 80% of participants reporting daily use. This was followed by email (58%) and phone calls (42%). In contrast, TV viewing (22%), video gaming (15%), and online friendships (18%) were less common. These patterns suggest a strong preference for mobile-based and text-driven tools in learners' digital routines.

To explore the third research question, Pearson correlation coefficients were calculated to examine associations between learning styles and specific technology use. Visual learners had a weak positive correlation with TV viewing ($r = .24, p < .05$), while auditory learners were moderately associated with phone calling ($r = .31, p < .01$). Kinesthetic learners demonstrated a small negative correlation with text messaging and smartphone use ($r = -.18, p = .07$), indicating a limited alignment between their preferred learning mode and the dominant digital platforms. Overall, these findings indicate that while certain tendencies exist, the relationship between learning styles and technology use is modest.



Discussion and Conclusion

The findings of this study are consistent with recent research showing weak correlations between learning styles and technology use (Sui et al., 2023). In this sample, mobile technology adoption appeared to be driven more by convenience and accessibility than by instructional fit. The dominance of smartphones and messaging tools among Iranian students reflects broader trends such as mobile-first internet use and limited access to computers in Iran (Mokhtarinia et al., 2024).

These findings align with the Technology Acceptance Model (Davis, 1989), in which perceived ease of use and usefulness outweigh individual cognitive style preferences. More recent studies (e.g., Feng et al., 2025) also highlight that habitual behaviors and available infrastructure serve as stronger predictors of technology adoption than cognitive styles. Weak correlations in this domain may also reflect the limitations inherent in self-reported learning style data (Newton & Salvi, 2020). Considering cultural adaptation, language barriers, and limited resources, inclusive CALL environments should emphasis mobile-friendly, multimodal resources that allow flexible engagement instead of rigid learning style alignment.

This study indicates that everyday digital habits, particularly mobile use, can outweigh learning style preferences among Australian-educated Iranians. The findings challenge assumptions behind style-based EdTech and support learner-focused, technology-integrated curricula grounded in actual use. In practice, this means prioritizing mobile-friendly platforms, using multimodal resources, and accounting for infrastructure in both local and global contexts. Further research should explore diverse cultural groups and track changes in technology preferences over time.

References

- Cohen, A. D., Oxford, R., & Chi, J. (2001). *Learning style survey: Assessing your own learning styles*. Center for Advanced Research on Language Acquisition, University of Minnesota.
- 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>
- Emban, E. M., Pactos, E. M. C., Pelare, G. B., & Nanud, M. N. D. (2025). Students' learning styles and academic performance. *United International Journal for Research & Technology*, 6(3), 57–65. Retrieved from <https://uijrt.com/paper/students-learning-styles-and-academic-performance>
- Feng, J., Yu, B., Tan, W. H., Dai, Z., & Li, Z. (2025). Key factors influencing educational technology adoption in higher education: A systematic review. *PLOS Digital Health*, 4(4), e0000764. <https://doi.org/10.1371/journal.pdig.0000764>
- Kukulska-Hulme, A., & Viberg, O. (2017). Mobile collaborative language learning: State of the art. *British Journal of Educational Technology*, 49(2), 207–218. <https://doi.org/10.1111/bjet.12580>
- Lai, C., Yang, J., & Gu, M. (2022). Learning strategies in self-directed language learning using mobile technology in higher education: A systematic scoping review. *Education and Information Technologies*, 27(18), 7749–7780. <https://doi.org/10.1007/s10639-022-10945-5>
- Lai, C. L., & Hwang, G. J. (2014). Effects of mobile learning time on students' conception of collaboration, communication, complex problem-solving, meta-cognitive awareness, and creativity. *International Journal of Mobile Learning and Organisation*, 8(3–4), 276–291. <https://doi.org/10.1504/IJMLO.2014.067029>
- Lee, H., & Lee, J. H. (2024). *The effects of AI-guided individualized language learning: A meta-analysis*. *Language Learning & Technology*, 28(2), 134–162. <https://hdl.handle.net/10125/73575>
- Mokhtarinia, H. R., Heydari Torkamani, M., Farmani, N., & Gabel, C. P. (2024). Smartphone addiction prevalence, patterns of use, and experienced musculoskeletal discomfort during the COVID-19 pandemic in a general Iranian population. *BMC Public Health*, 24, 161. <https://doi.org/10.1186/s12889-024-17654-3>
- Newton, P. M., & Miah, M. (2017). Evidence-based higher education — Is the learning styles 'myth' important? *Frontiers in Psychology*, 8, Article 444. <https://doi.org/10.3389/fpsyg.2017.00444>

- Newton, P. M., & Salvi, A. (2020). How common is belief in the learning styles neuromyth, and does it matter? A pragmatic systematic review. *Frontiers in Education*, 5, 602451. <https://doi.org/10.3389/feduc.2020.602451>
- Nurmala, I., Irianto, S., Franchisca, S., Amsa, H., & Susanti, R. (2023). Technology-enhanced language learning: A meta-analysis study on English language teaching tools. *Journal on Education*, 6(1), 2188–2195. <https://doi.org/10.31004/joe.v6i1.3221>
- Rosen, L. D., Whaling, K., Carrier, M. A., Cheever, N. A., & Rokkum, J. (2013). The media and technology usage and attitudes scale: An empirical investigation. *Computers in Human Behavior*, 29(6), 2501–2511. <https://doi.org/10.1016/j.chb.2013.06.006>
- Sui, Y., Xu, H., & Zheng, L. (2023). *Exploring factors influencing self-regulated learning in technology-enhanced environments: An extended technology acceptance model*. *Education and Information Technologies*, 28, 8453–8473
- Zhang, S., Shan, C., Lee, J. S. Y., Che, S., & Kim, J. H. (2023). Effect of chatbot-assisted language learning: A meta-analysis. *Education and Information Technologies*, 28(11), 15223–15243. <https://doi.org/10.1007/s10639-023-11805-6>

Multi-Agent Debate in L2 Argumentative Writing: An Exploratory Study of Chinese EFL Learners

Tianshu Chu 

Capital Normal University, China

Abstract

Argumentative writing poses significant challenges for second-language (L2) learners. Generative AI (GenAI) offers pedagogical support, but existing tools typically act as single assistants and fail to simulate the dynamic, multi-perspective dialogue essential for robust argumentation. We introduce a Multi-Agent Debate (MAD) framework in which multiple AI personas engage in structured debates to foster divergent thinking and expose learners to varied viewpoints. In a one-group pretest–posttest exploratory study ($N = 22$ Chinese EFL undergraduates), we investigated MAD’s impact on argumentative writing and learner perceptions. Preliminary results indicate that MAD enhanced argument structure and quality, promoted metacognitive awareness, cultivated critical AI skepticism, and reduced anxiety when addressing controversial topics. These findings suggest MAD is a promising approach for improving both L2 argumentative skills and GenAI literacy.

Keywords: Generative AI, Multi-Agent Debate (MAD), Chinese EFL learners, argumentative writing

Background

Effective L2 argumentative writing requires not only constructing logical claims but also anticipating and refuting counterarguments. This is particularly challenging in Chinese English as a Foreign Language (EFL) contexts,

E-mail: tianschu@cnu.edu.cn

where traditional pedagogies may neglect the dialogical processes crucial for rebuttal skills. While in-class debating can be useful (e.g., El Majidi et al., 2021), it often presents hurdles such as linguistic barriers and performance anxiety.

Generative AI (GenAI) offers a new opportunity to support learners. Prior research suggests that GenAI can enhance argument structure (Pratama et al., 2025) and support claim generation (Guo et al., 2023). However, most existing tools serve as “reactive” assistants, responding solely to user prompts and predefined criteria (Su et al., 2023), thus limiting opportunities for learners to engage in the multi-sided debate necessary for sophisticated argumentation. For instance, Guo et al. (2023) found no significant improvement in students’ ability to formulate counterarguments or rebuttals when using a chatbot for argumentation, underscoring the need for more interactive, dialogical approaches.

To address this gap, this study proposes a Multi-Agent Debate (MAD) framework that draws on research in computational argumentation (e.g., Du et al., 2024; Hu et al., 2025; Liang et al., 2024). MAD leverages multiple AI agents to simulate debates (Du et al., 2024; Hu et al., 2025; Liang et al., 2024), which counteracts the “Degeneration-of-Thought” phenomenon in single-agent large language models (LLMs) (Liang et al., 2024) and creates a richer, dialogical environment for L2 writing development.

Research Questions

1. To what extent does engagement in a Multi-Agent Debate (MAD) framework improve Chinese EFL learners’ argumentative writing skills?
2. How do learners perceive the role of the MAD framework in their argumentative writing process?

Methodology

Participants

Twenty-two undergraduate English majors (CEFR B2–C1) at a mainland Chinese university participated in the study as part of their foundational writing course.

The Multi-Agent Debate (MAD)

The MAD in this study was implemented using widely available LLMs such as DeepSeek-R1, GPT-4o, and Grok 3. Students were free to select the model they preferred and interacted with it through standard web-based chat

interfaces. Rather than relying on a pre-built educational platform, the debate environment was created through carefully designed prompts within a single model session. The design draws on existing MAD frameworks in computational argumentation research (e.g., Du et al., 2024; Hu et al., 2025; Liang et al., 2024). In particular, the inclusion of persona creation follows Hu et al. (2025), who argue that assigning distinct, high-level beliefs and viewpoints to each debating agent ensures the diversity of perspectives. This prompting process corresponds with Step 3 and Step 4 in Table 1. For reference, sample prompts are included in the Appendix.

For the current study, the overall MAD workflow consists of five steps (see Table 1). Steps 1 and 2 involved preparatory reading and stance-taking without AI involvement. In Step 3, students prompted the LLM to generate debater personas with names, roles, and core stances, and could revise them based on their own knowledge. In Step 4, they instructed the LLM to run a multi-round debate between these personas, generating claims, evidence, counterarguments, and rebuttals. In Step 5, students reflected on the debate using guiding questions and revised their original arguments.

Procedure

A one-group pretest-posttest design was conducted over five sessions. In the initial session, students received instruction and wrote a pretest essay on the topic “*Should subway systems implement mandatory security checks?*” Over the next three intervention sessions, students engaged with the MAD (see Table 1 for the full workflow), exploring a new topic each session. In the final session, students composed a posttest essay on the topic “*Should universities require AI checks on term papers?*”. For both the pretest and posttest, students conducted 15 minutes of online research followed by 40 minutes of offline writing. Ten participants were interviewed after the posttest.

Data Analysis

We analyze three data sources:

- a. **Pre-/Posttest essays:** Scored using rubrics (adapted from El Majidi, 2020; Guo et al., 2023) for structural complexity (based on the Toulmin model), argument quality, and language.
- b. **Reflection logs:** Thematically analyzed to identify recurring patterns in student engagement with the MAD system.
- c. **Interview transcripts:** Used to validate log findings and explore learner perceptions in depth.

Table 1. The Multi-Agent Debate (MAD) Workflow

Steps	Student-Directed AI Generation	Student Learning Actions
1. Read & Understand	– (No AI action in this step. Students access human-selected materials.)	Students read a pre-selected opinion article and are given a related debate topic. <u>Example:</u> Article – “Want Kids to Stay in School? Let Them Keep Their Phones” (Sixth Tone). Debate topic – “Dpo the benefits of cellphone bans in rural Chinese schools outweigh the drawbacks?”
2. Take a Position	– (No AI action)	Students write a short paragraph stating their initial opinion on the topic. <u>Example:</u> “I believe the benefits of banning cellphones in rural Chinese schools do outweigh the drawbacks. Without phones engage more in class, sports, and friendship...”
3. Create personas	Students prompt the LLM to generate at least two debater personas with names, roles, and core stances.	Students receive the generated personas and can compare them to any they imagined themselves. <u>Example personas:</u> Persona A: Zhang Wei, Rural Middle School Principal, “Structured learning environments free from digital distractions are essential for academic focus and discipline in under-resourced schools.” Persona B: Chen Lihua, Education Policy Researcher, “Preparing students for a digital future requires responsible access to technology, not blanket bans that widen the digital divide.” ...
4. Simulate Debate with LLM	Students instruct the LLM to run a multi-round debate between the personas.	Students read the AI-generated debate transcript, identifying claims, evidence, counterarguments, and rebuttals. They annotate surprising or strong arguments, and note weak or repetitive points.
5. Reflect & Expand	– (No AI action)	Student reflect on the debate using reflection questions, then revise or write their own argumentative essay. <u>Sample reflection questions:</u> What were the 2–3 most surprising or strong arguments generated by the AI? What challenges did you face in analyzing the AI debate? How did this debate change your thinking or writing?

Representative Results

Although the full data analysis is ongoing, preliminary results have revealed several notable trends. Posttest essays showed marked improvement in structural complexity over the predominantly one-sided pretest essays. Most posttest essays incorporated opposing perspectives, shifting from simply acknowledging counterarguments to involving substantive rebuttals. While overall linguistic complexity saw minimal change, there was a marked increase in phrases signalling dialogical argumentation, such as “Conversely,” “While proponents argue,” and “This viewpoint, however, fails to consider...”.

Reflection logs and interviews revealed three key themes in learner perceptions. First, students reported enhanced metacognition; the debates helped them identify “blind spots” and logical fallacies in their own reasoning. As one participant noted, “The debate lets me see different points in a critical way and new angles can be added.” Second, learners developed critical AI skepticism. They actively evaluated the AI-generated content by fact-checking claims, flagging “unvalidated data,” and noting “clichéd phrasing” or “stereotypical responses” from personas. Finally, the framework’s low-stakes nature reduced writing anxiety. It created a safe space for learners to explore contentious topics. Several participants noted that the simulated debate help them “rehearse the points” and “boost confident in expressing views”.

Discussion

This study’s preliminary evidence suggests the MAD framework significantly enhances L2 argumentative writing. A key finding is that MAD appears to promote critical GenAI literacy (Li, 2025). Rather than passively accepting AI output, students actively interrogated it. This suggests MAD can function as a “cognitive mirror,” prompting learners to scrutinize both AI-generated arguments and their own, which is a crucial skill for responsible AI integration in pedagogy.

Additionally, MAD appears to lower the affective filter for engaging with controversial topics. The low-stakes simulated environment allowed students to “rehearse” arguments without the social pressure of live debate. This is valuable in cultural contexts where direct contention may be discouraged, offering a scaffold for students to build confidence for authentic civic discourse.

Limitations and Future Directions

The study’s conclusions are preliminary and should be interpreted with caution. Key limitations also include the small sample size ($N = 22$), the

pre-experimental design that lacks a control group, and the short intervention period. These factors limit the generalizability of the findings and restrict causal claims.

Despite these limitations, this exploratory study highlights that the MAD framework holds significant potential. It uniquely addresses the need for multi-perspective engagement critical for developing rebuttal skills, fosters the metacognitive evaluation of AI, and provides a safer pathway for Chinese students to engage with complex social issues. Future research should employ rigorous quasi-experimental designs to validate and extend these findings.

References

- Du, Y., Li, S., Torralba, A., Tenenbaum, J. B., & Mordatch, I. (2024). Improving factuality and reasoning in language models through multiagent debate. In *Proceedings of the 41st International Conference on Machine Learning (ICML '24)*, 235, Article 467, 11733–11763. <https://doi.org/10.48550/arXiv.2305.14325>
- El Majidi, A., de Graaff, R., & Janssen, D. (2020). Debate as L2 pedagogy: The effects of debating on writing development in secondary education. *Modern Language Journal*, 104(4), 804–821. <https://doi.org/10.1111/modl.12673>
- El Majidi, A., Janssen, D., & Graaff, R. de. (2021). The effects of in-class debates on argumentation skills in second language education. *System*, 101, Article 102576. <https://doi.org/10.1016/j.system.2021.102576>
- Guo, K., Zhong, Y., Li, D., & Chu, S. K. W. (2023). Effects of chatbot-assisted in-class debates on students' argumentation skills and task motivation. *Computers & Education*, 203, Article 104862. <https://doi.org/10.1016/j.compedu.2023.104862>
- Hu, Z., Chan, H. P., Li, J., & Yin, Y. (2025). Debate-to-write: A persona-driven multi-agent framework for diverse argument generation. In *Proceedings of the 31st International Conference on Computational Linguistics* (pp. 4689–4703). Association for Computational Linguistics. <https://doi.org/10.48550/arXiv.2406.19643>
- Li, S. (2025). Generative AI and second language writing. *Digital Studies in Language and Literature*, 2(1), 122–152. <https://doi.org/10.1515/dsll-2025-0007>
- Liang, T., He, Z., Jiao, W., Wang, X., Wang, Y., Wang, R., Yang, Y., Shi, S., & Tu, Z. (2024). Encouraging divergent thinking in large language models through multi-agent debate. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* (pp. 17889–17904).

- Association for Computational Linguistics. <https://doi.org/10.18653/v1/2024.emnlp-main.992>
- Pratama, R. D., Widiati, U., & Hakim, L. N. (2025). Effects of human-AI collaborative writing strategy on EFL students' argumentative writing skills. *Computer Assisted Language Learning*, 1–32. <https://doi.org/10.1080/09588221.2025.2503900>
- Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing*, 57, Article 100752. <https://doi.org/10.1016/j.asw.2023.100752>

Appendix

Sample User Prompts for Persona Creation and Debate Simulation

User prompt for persona creation:

You are designing a debate on the following topic: “Do the benefits of cellphone bans in rural Chinese schools outweighs the drawbacks?” Create three personas who are going to debate on the topic. For each persona: Provide a name and concrete role of the persona. State their worldview or values in one sentence. Output exactly in this format:

Persona A: [name], [role], [worldview].

Persona B: ...

User prompt for debate simulation:

Now please simulate a debate on the topic above. Three debater personas have been created, each with distinct roles, backgrounds, and reasons for opposing the proposition.

Debate Rules:

There are five rounds. Each round follows the order: Persona A → Persona B → Persona C. Each persona must respond in 3–4 sentences per turn. Language must be respectful, clear, and logically structured. Avoid repetition of points. Persuasive strategies such as examples, analogies, and rhetorical questions are encouraged.

Round Structure: Opening Statements: Each persona presents their main claim and two supporting reasons. First Rebuttal: Each persona challenges one or more points made by other personas, identifying weaknesses or offering counterexamples. Second Rebuttal: Each persona deepens their challenge by addressing counter-challenges or introducing new evidence. Synthesis Round: Each persona clarifies their position in light of earlier rounds, acknowledging valid points from others while reinforcing their own stance. Closing Statements: Each persona summarizes why their position is stronger than others and appeals to the audience’s logic or values.

Output Format: Clearly label each round and each persona’s turn (e.g., “Round 1 – Rural middle school principal”).

Keep responses concise and relevant to the proposition. Maintain coherence across rounds so that arguments build on previous exchanges.

Undergraduates' Reflection on Applying ChatGPT in ESP Writing Skills: A Case from Foreign Trade University HCMC

Thi My Dung Dang 

Foreign Trade University, Ho Chi Minh City Campus, Vietnam

Abstract

Artificial intelligence (AI)-driven automated feedback is fast becoming a norm. ChatGPT can provide meaningful writing samples, adjust the difficulty level of texts matching learners' proficiency level and facilitate guided writing. ChatGPT is capable of rapidly generating texts which are not only appropriate but are also error-free in terms of language. This study investigates Business students' reflection on applying ChatGPT in ESP writing skills. The study involved fifty-two undergraduates in Business who were taking an English language course during one semester at Foreign Trade University (FTU) – HCMC, a public university of Vietnam. Mixed-methods design was used to investigate students' beliefs in and perspectives on the effectiveness of ChatGPT application in English Teaching and Learning. Two study tools were employed: a questionnaire and a semi-structured interview. Comprising ten items, the questionnaire measured undergraduate learners' perceptions of using ChatGPT for improving their writing skills and at the same time focusing on higher-order thinking skills as they are related to effective writing. The semi-structured interview, on the other hand, consisted of the pros and cons of utilizing ChatGPT-based essay writing versus self-dependent essay writing. The results revealed most respondents confirmed ChatGPT's effectiveness in saving time and enhancing their language accuracy. However, the progress they could make in acquiring writing skills and increasing higher-order

E-mail: dangthimydung.cs2@ftu.edu.vn & dtd1ocd@bolton.ac.uk

thinking abilities could not be seen significant. The research paper also made some implications and recommendations for language educators in Vietnamese higher institutions.

Keywords: ChatGPT, English for specific purposes, thinking skills, Vietnamese higher educational institutions

Introduction

In Vietnam's higher education system, non-English-major students, such as those in Business, study English as a compulsory subject tailored to domain-specific needs, unlike English majors who focus primarily on language studies. This distinction shapes ESP instruction, which is critical for preparing Vietnamese business undergraduates for global workplaces. The integration of AI tools like ChatGPT into Computer-Assisted Language Learning (CALL) offers potential to enhance ESP writing but raises concerns about creativity and critical thinking. Despite growing global research, ChatGPT's impact in Vietnam's ESP context remains underexplored, particularly for non-English majors facing unique linguistic and cultural barriers (Nguyen, 2023). This study addresses three research questions (RQs):

1. What are Vietnamese business undergraduates' perceptions of ChatGPT in ESP writing?
2. How do ChatGPT-assisted and self-dependent writing compare in accuracy, creativity, and critical thinking?
3. What pedagogical strategies optimize ChatGPT integration while mitigating risks like plagiarism?

Literature Review

AI tools have revolutionized CALL by offering adaptive feedback and personalized learning (Stockwell, 2015). ChatGPT supports ESP writing with error-free texts and tailored suggestions (Baskara & Mukarto, 2023), creating conducive environments for Business English (Kim et al., 2023). However, concerns about plagiarism and reduced creativity persist (Cotton et al., 2024; Al-Alami, 2024). Self-dependent writing fosters critical thinking (Yamashita, 2021), and AI should complement, not replace, it (Zhao, 2022). In Vietnam, ESP learners face challenges due to domain-specific language demands and limited exposure to authentic English contexts (Nguyen, 2023). Recent meta-analyses confirm AI improves writing accuracy but less so higher-order thinking (Zang & Wang,

2025). This study fills a gap by examining ChatGPT's role in Vietnam's ESP context, where cultural and educational constraints necessitate tailored pedagogical approaches.

Methodology

Participants

Fifty-two (52) business undergraduates aged 18–22 (balanced gender) at FTU HCMC participated, majoring in Business Administration, Banking and Finance, and Accounting. All were enrolled in a standard ESP English course during the 2024–2025 academic year, focusing on business-specific communication (e.g., sales contracts, business reports, emails).

Research Design

A 10-item, five-point Likert scale questionnaire adapted from Al-Alami (2024) was used to assess perceptions of ChatGPT's impact on ESP writing skills such as vocabulary, coherence, critical thinking, and authenticity. Sample items included: "ChatGPT helps improve my vocabulary specific to business writing" and "Using ChatGPT encourages me to think critically about my writing." Semi-structured interviews with 16 volunteers explored detailed experiences comparing ChatGPT-assisted and self-dependent writing, focusing on language accuracy, creativity, ethical concerns, and task authenticity.

Procedure

Over one semester, students completed weekly ESP writing assignments in class. For each task, participants produced two drafts: one independently, and one revised using ChatGPT feedback targeting grammar, vocabulary, and structure. Students finalized all texts themselves. Questionnaires were completed post-intervention in class, with items explained to ensure clarity. Interviews were conducted individually (approximately 4 minutes each). Ethical protocols ensured voluntary participation, anonymity, and consent.

Data Analysis

Quantitative questionnaire data were analyzed as percentages of agreement to gauge perceptions. Qualitative interview data were thematically coded using NVivo, focusing on themes like efficiency, language accuracy, creativity, plagiarism risk, and critical thinking development. Percentage summaries were used only for quantitative data, while qualitative insights provide depth and nuance.

Results

Advantages of ChatGPT and Self-dependent Writing

- 60% of students agreed that self-dependent writing improved their overall writing proficiency through practice and reflection.
- 50% of respondents reported enhanced critical thinking and problem-solving skills from self-dependent writing.
- 45% of them found self-dependent writing enabled more authentic and personalized expression.
- For ChatGPT, 60% valued the time-saving benefits in ESP writing tasks.
- 55% noted improved grammatical accuracy via ChatGPT feedback.
- 25% appreciated the ease of task completion facilitated by AI assistance.
- Smaller portions of learners (15% and 10%) highlighted better-structured responses and incorporation of business-specific vocabulary, respectively.

Disadvantages of ChatGPT and Self-dependent Writing

- 33% of students expressed concerns that ChatGPT encourages passive learning habits.
- 27% of them felt their authentic writing voice was diminished with AI reliance.
- 22% of respondents raised plagiarism concerns, highlighting a need for institutional policies and teacher guidance.
- Self-dependent writing was criticized by 65% for its time-consuming nature.
- 25% of learners mentioned recurring linguistic inaccuracies when writing independently.

Thematic Insights

Interviewees preferred a hybrid approach: initial drafts with ChatGPT support, followed by independent revision to nurture creativity and critical thinking. Significant emphasis was placed on teacher mediation, clear guidelines on AI use, and fostering student responsibility to avoid ethical pitfalls.

Discussions

Addressing research question 1, 60% of students valued ChatGPT's time-saving and 55% its accuracy, but only 30% saw writing proficiency gains, suggesting limited impact on skill development. For research question 2, ChatGPT improved accuracy but hindered creativity (27%) and critical thinking (35%),

while self-dependent writing fostered authentic expression (45%) and thinking skills (50%). This aligns with Al-Alami (2024) and Yamashita (2021), emphasizing balanced pedagogy. For research question 3, a hybrid approach integrating ChatGPT with independent revision mitigates plagiarism risks (22%) and supports authentic learning. In Vietnam's ESP context, cultural emphasis on rote learning (Nguyen, 2023) underscores the need for teacher training and institutional policies to mediate AI use effectively. Limitations of this study include the small, business-specific sample and self-reported data. Future research should use objective measures and explore diverse ESP domains.

Conclusion

ChatGPT serves as a valuable aid for improving ESP writing efficiency and error reduction among Vietnamese business undergraduates. However, to foster originality, critical thinking, and authentic language use, it must be integrated within blended instructional frameworks that emphasize independent revision, ethical awareness, and teacher guidance. Institutional policies and teacher training are essential to maximize the pedagogical benefits while avoiding misuse. Further research is needed to optimize AI-supported ESP teaching across varied disciplines and contexts.

References

- Al-Alami, S. (2024). EFL learners' attitudes towards utilizing ChatGPT for acquiring writing skills in higher education: A case study of computing students. *Journal of Language Teaching and Research*, 15(4), 1029–1038.
- Baskara, F. X. R., & Mukarto, F. X. (2023). Exploring the implications of ChatGPT for language learning in higher education. *Indonesian Journal of Language Teaching and Applied Linguistics*, 7(2), 343–358.
- Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61(2).
- Kim, S., Shim, J., & Shim, J. (2023). A study on the utilization of OpenAI ChatGPT as a second language learning tool. *Journal of Multimedia Information System*, 10(1), 79–88.
- Nguyen, T. H. (2023). Challenges in ESP instruction for non-English majors in Vietnam. *Vietnam Journal of Education*, 7(3), 45–52.
- Stockwell, G. (2015). Mobile-assisted language learning: Concepts, contexts, and challenges. *The Modern Language Journal*, 99(3), 578–594.

- Yamashita, T. (2021). Corrective feedback in computer-mediated collaborative writing and revision contributions. *Language Learning & Technology*, 25(2), 75–93.
- Zhang, J., & Wang, J. (2025). Student perceptions of hybrid feedback: Using Gen-AI to enhance engagement with EAP writing feedback. *The JALT CALL Journal*, 21(2), 1–27.
- Zhao, X. (2022). Leveraging artificial intelligence (AI) technology for English writing: Introducing Wordtune as a digital writing assistant for EFL writers. *RELC Journal*, 54(3).

L2 Spanish and Indonesian Learners' Engagement with ChatGPT Written Corrective Feedback

Ms Alicia Ximena Gazmuri Sanhueza¹ ,
Ms Franciele Spinelli¹  and Dr David Wijaya² 

¹The University of Queensland, Australia

²The University of Sydney, Australia

Abstract

Written corrective feedback (WCF) plays a crucial role in second language (L2) writing by helping learners address local (grammar) and global (content) errors (Crosthwaite et al., 2021). The rise of Generative Artificial Intelligence (GenAI) has increased interest in its role in WCF; Tools such as ChatGPT can provide personalised feedback by explaining grammar, defining words and translating between languages (Kohnke et al., 2023). However, studies have focused exclusively on English (Lin & Crosthwaite, 2024), limiting the linguistic inclusivity in the research. In addition, despite the recognised importance of learner engagement for the effectiveness of WCF, only a few studies (Koltovskaia et al., 2024; Zhang & Hyland, 2022) have examined learners' engagement with WCF; however, little is known about the follow-up questions students may ask regarding the feedback they receive.

This study applies a multidimensional framework of learner engagement (Zhang & Hyland, 2018) to analyse how 13 intermediate L2 learners of Spanish and Indonesian engaged cognitively, behaviorally and attitudinally with ChatGPT's feedback. Participants wrote a short L2 essay on their reasons for learning a L2 and then used a pre-prepared prompt to seek WCF from ChatGPT. A semi-structured interview followed this. Students' drafts, revisions and follow-up questions were coded for behavioural and cognitive engagement. The interview data were analysed thematically for cognitive and attitudinal engagement.

E-mail: a.gazmuri sanhueza@uq.edu.au, f.spinelli@uq.edu.au & david.wijaya@sydney.edu.au

Spanish and Indonesian students noticed most WCF at local and global levels, but relied more on ChatGPT for the former. While participants found ChatGPT's feedback mostly useful, they remained critical of its relevance and accuracy. These findings highlight that while ChatGPT could offer accessible WCF, its effectiveness depends on learners' ability to critically evaluate it. The study shows the opportunities and challenges of integrating GenAI for WCF, broadening the understanding of how diverse linguistic communities engage with technology for L2 writing.

Keywords: Written Corrective Feedback, Cognitive engagement, Behavioural Engagement, Attitudinal Engagement, GenAI, ChatGPT, Languages other than English

Introduction

Written corrective feedback (WCF) refers to corrections or comments on writing to improve global (e.g., content, organisation) and local (e.g., grammar, vocabulary) linguistic issues (Lin & Crosthwaite, 2024). Although widely implemented in L2 classrooms, its effectiveness remains debatable, with research indicating learners' engagement as a key mediating role (Papi et al., 2024).

Generative AI (GenAI) tools (e.g., ChatGPT) have reshaped WCF in L2 learning by offering specific, actionable, and personalised insights (Guo & Wang, 2024) and allowing learners to ask follow-up questions (Kohnke et al., 2023). However, GenAI's feedback may also be inaccurate or redundant (Lin & Crosthwaite, 2024; Yoon et al., 2023), particularly for languages other than English (LOTEs) (Kohnke et al., 2023), which may impact different aspects of learner engagement (e.g., reducing trust, being more cautious and/or critical about the feedback, etc). This highlights the need to critically evaluate the role of GenAI in providing WCF for L2 learners of LOTEs.

To investigate learner engagement with WCF, Zhang and Hyland (2018) used a tridimensional framework comprising behavioural (revision actions and time), cognitive (attention and revision strategies) and attitudinal (attitudes towards WCF) engagement. Building on this framework, Zhang and Hyland (2018) emphasise that engagement with WCF varies considerably among learners, shaping the extent and quality of their writing revisions. Studies applying this framework to GenAI-generated WCF (Koltovskaia et al., 2024; Zhang & Hyland, 2022) have been conducted primarily in English, limiting the linguistic inclusivity. Moreover, while the studies showed that learners critically evaluated GenAI's feedback and prioritised local over global-level corrections,

these lacked insight into how learners engage with GenAI feedback through follow-up questions.

This study addresses these gaps by examining how L2 learners of Spanish and Indonesian, two widely spoken and taught LOTES but under-researched, engage with ChatGPT's WCF. Employing the tri-dimensional engagement framework, the study is guided by the following research question:

How do Spanish and Indonesian L2 learners engage behaviourally, cognitively and attitudinally with ChatGPT's WCF as they revise essays written in their L2?

Methodology

Following ethical approval (ID: 2024/HE001234), thirteen undergraduate L2 students from two major Australian universities ($n = 7$ Spanish, $n = 6$ Indonesian) participated in the study. Based on university placement, their L2 proficiency is B1 ($n = 4$) or B2 ($n = 9$) (Council of Europe, 2001). The research design and instruments for data collection are summarised in Figure 1.

The data was analysed in two stages. First, similar to Koltovskaia et. al (2024), students' writing tasks (before and after the revisions) and ChatGPT's WCF were analysed to explore participants' behavioural engagement. Using a spreadsheet, the researchers counted and categorised:

- ChatGPT's WCF type (i.e., global or local)
- Whether students accepted or rejected the WCF
- How they revised their essay (e.g., copying from ChatGPT, retyping)
- Length spent on revisions
- Follow-up questions after receiving the WCF

Second, following Saldaña's (2016) guidelines for thematic analysis, the interview transcripts were analysed for cognitive and attitudinal engagement.

Findings

Participants showed different behavioural engagement with ChatGPT's WCF. Out of the 95 WCF instances received by Spanish learners (64% local, 36% global), 56% were accepted, with a higher acceptance of local feedback (77%). Indonesian learners received 66 WCF instances, mostly language-focused (80%), with a high acceptance of 89%. All participants reviewed the WCF

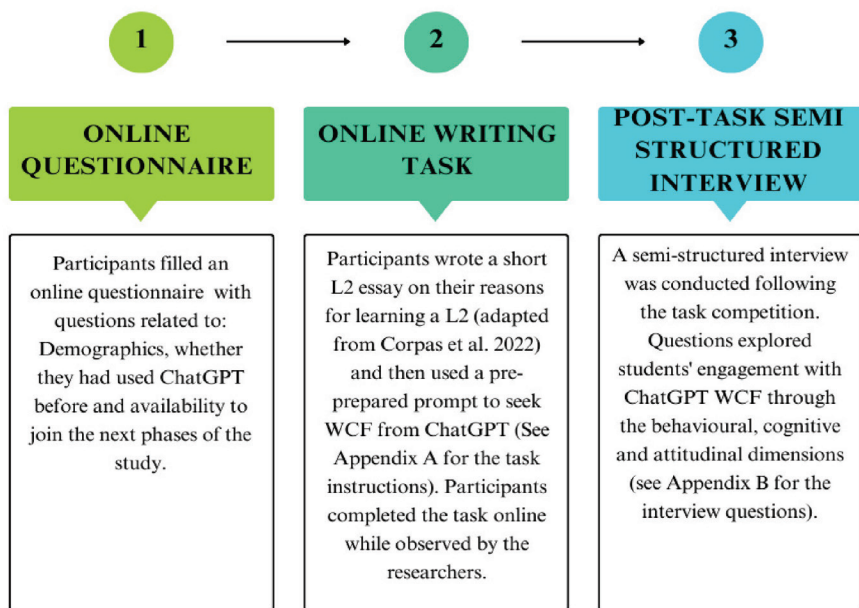


Figure 1. Research design and instruments for data collection

before revising their essays. While all Spanish learners typed the suggestions into their revisions, some Indonesian learners translated the WCF into English and then copied and pasted ChatGPT's suggestions into their revisions. On average, Spanish learners spent 15 minutes revising their essays, compared to 21 minutes for Indonesian learners. All Indonesian learners asked follow-up questions, but only 2 Spanish learners did the same.

In the interviews, participants explained how they engaged cognitively with ChatGPT's WCF. They explained that they accepted and incorporated most of ChatGPT's WCF into their writing, especially regarding grammar, vocabulary and punctuation. Similarly, WCF on global aspects, such as writing a clear introduction, were widely accepted to enhance the essay's coherence. However, some grammar-related feedback was rejected because students acknowledged that the suggested structures were beyond their current knowledge, thus preferring to use more familiar forms. Students also approached WCF about content (e.g., adding more examples) selectively, as they believed they were unnecessary or exceeded the scope of a short essay (see Appendix C for sample quotes related to cognitive engagement).

Attitudinally, participants valued the detailed WCF ChatGPT provided, explaining that it helped them notice aspects to improve the accuracy and coherence of their essays. Those who asked follow-up questions appreciated

being able to do so without fear of judgment. However, they noted that the WFC was at times overly wordy, poorly organised, generic, misaligned with their L2 proficiency, and occasionally contained errors. Despite the challenges, all students plan to continue using ChatGPT for language support but would appreciate clearer university policies and guidelines (see Appendix C for sample quotes related to attitudinal engagement).

Discussion and Conclusion

Using the tri-dimensional framework of learner engagement (Zhang & Hyland, 2018), the study explored how L2 learners of Spanish and Indonesian engaged with ChatGPT-generated WFC. Overall, the findings partially align with those of Koltovskaia et al. (2024), who demonstrated that students engaged actively with ChatGPT's feedback by accepting and rejecting comments. Behaviourally, participants actively revised their text using ChatGPT suggestions and asked follow-up questions, with a primary focus on language accuracy (Zhang & Hyland, 2022), as they perceived them as more relevant than global-related comments. While time spent on revisions may indicate a high level of engagement (Koltovskaia et al., 2024), our results showed that remaining on task revising from feedback is an indicator of sustained engagement.

Cognitively, participants noticed and understood the feedback, an important predictor of revision quality (Papi et al., 2024). All students compared ChatGPT's suggestions with their original text, deciding to accept or reject the WFC based on their language styles and proficiency. Local-level feedback was generally prioritised, helping them notice correct forms, while global suggestions were often dismissed as unclear or unnecessary. Indonesian learners asked more follow-up questions to fill knowledge gaps, whereas Spanish learners chose not to ask when feedback was clear, revealing a deeper cognitive engagement in their learning.

Attitudinally, participants expressed high satisfaction with ChatGPT's suggestions regarding L2 accuracy, coherence, and depth, as well as a reduction in anxiety when asking follow-up questions (Lo et al., 2024). However, concerns about inaccuracies (Steiss et al., 2024), processing lengthy feedback (Guo & Wang, 2024), and prompting effectively (Lin & Crosthwaite, 2024), reflect a critical stance and active engagement. These responses highlight the potential and challenges of integrating GenAI for WFC practices for LOTEs.

Despite the small sample size and the use of a structured prompt, which may not reflect real-world uses of GenAI, these findings underscore the importance of critically engaging in writing revisions following feedback. To maximise the effectiveness of integrating GenAI tools, training teachers and

students to use detailed prompts tailored to the specific goal of the writing may facilitate deeper engagement (Liang et al., 2023; Crosthwaite et al., 2024) as these are more effective than generic ones (Koltovskaia et al., 2024; Yoon et al., 2023). Teachers should encourage learners to critically evaluate ChatGPT's feedback, viewing it as a writing assistant rather than a definitive authority (Teng, 2024). Further research could investigate learner engagement with GenAI-generated WCF across proficiency levels, the use of learner-designed prompts independently, and expand the scope to other LOTEs.

Acknowledgement

ChatGPT, Claude, Gemini and Copilot were used to edit this article for clarity, and Grammarly was consulted for grammatical and lexical errors.

References

- Corpas. J., Garmendia. A., Sánchez., N & Soriano. C. (2022). *Aula Internacional 4*. Campus Difusión.
- Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press. <https://rm.coe.int/1680459f97>
- Crosthwaite, P., Smala, S. and Spinelli, F. (2024). Prompting for pedagogy? Australian F-10 teachers' generative AI prompting use cases. *The Australian Educational Researcher*, 1–31. <https://doi.org/10.1007/s13384-024-00787-0>
- Guo, K., & Wang, D. (2024). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-023-12146-0>
- Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, 54(2), 537–550. <https://doi.org/10.1177/00336882231162868>
- 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, 101130. <https://doi.org/10.1016/j.jslw.2024.101130>
- Liang, W., Zhang, Y., Cao, H., Wang, B., Ding, D., Yang, X., Vodrahalli, K., He, S., Smith, D., Yin, Y., McFarland, D., & Zou, J. (2023). Can large language models provide useful feedback on research papers? A large-scale empirical analysis. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2310.01783>

- Lin, S., & Crosthwaite, P. (2024). The grass is not always greener: Teacher vs. GPT-assisted written corrective feedback. *System*, 127, 103529. <https://doi.org/10.1016/j.system.2024.103529>
- Lo, C. K., Hew, K. F., & Jong, M. S. Y. (2024). The influence of ChatGPT on student engagement: A systematic review and future research agenda. *Computers & Education*, 219, Article 105100. <https://doi.org/10.1016/j.compedu.2024.105100>
- Papi, M., Abdi Tabari, M., & Sato, M. (2024). The importance of seeking feedback for benefiting from feedback: A case of second language writing. *The Modern Language Journal (Boulder, Colo.)*, 108(2), 489–512. <https://doi.org/10.1111/modl.12923>
- Saldaña, J. (2016). *The coding manual for qualitative researchers* (Third edition.). SAGE.
- Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., Moon, Y., Tseng, W., Warschauer, M., & Olson, C. B. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. *Learning and Instruction*, 91, Article 101894. <https://doi.org/10.1016/j.learninstruc.2024.101894>
- Teng, M.F. (2024). "ChatGPT is the companion, not enemies": EFL learners' perceptions and experiences in using ChatGPT for feedback in writing "ChatGPT is the companion, not enemies": EFL learners' perceptions and experiences in using ChatGPT for feedback in writing. *Computers and Education. Artificial Intelligence*, 7, <https://doi.org/10.1016/j.caeai.2024.100270>
- Wang, X., & Wang, S. (2024). Exploring Chinese EFL learners' engagement with large language models: A self-determination theory perspective. *Learning and Motivation*, 87, Article 102014. <https://doi.org/10.1016/j.lmot.2024.102014>
- Yoon, S. Y., Miszoglad, E., & Pierce, L. R. (2023). Evaluation of ChatGPT feedback on ELL writers' coherence and cohesion. *arXiv preprint*, arXiv:2310.06505 <http://arxiv.org/abs/2310.06505>
- Zhang, Z., & Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing*, 36, 90–102. <https://doi.org/10.1016/j.asw.2018.02.004>
- Zhang, Z., & Hyland, K. (2022). Fostering student engagement with feedback: An integrated approach. *Assessing Writing*, 51, Article 100586. <https://doi.org/10.1016/j.asw.2021.100586>

Appendices

Appendix A

Online writing task instructions

1. In the space below, write an essay of at least 150 words in Spanish OR Indonesian explaining why you chose to study a foreign language at university, including your personal motivations, specific experiences or influences and the anticipated future benefits. Do not use dictionaries or machine translators, such as Google Translate or DeepL. You have 25 minutes to complete the task.

2. Use ChatGPT to revise your text, pasting the following prompt:
“You are a Spanish OR Indonesian language teacher. I am a second language learner of Spanish OR Indonesian. I have written this essay and would like you to provide specific, actionable feedback on its content, organisation, language and mechanics in Spanish. If needed, provide examples of how I could improve the essay. Do not provide a revised version of the essay.” Once you have received the feedback from ChatGPT, you may like to ask for clarification of the feedback. Please revise your essay accordingly in the space below.

Note: Unless specifically told otherwise by your Course Coordinator and the Course Electronic Profile (ECP), the use of Generative Artificial Intelligence technologies, such as ChatGPT, to complete any university assessment is strictly prohibited and may constitute student misconduct under the Student Code of Conduct.

Appendix B

Interview questions

1. In which aspect did ChatGPT provide you with feedback, e.g., vocabulary, grammar, formality, etc?
2. Do you think the suggestions provided by ChatGPT are correct? If so, how? If not, why?
3. Did you find the suggestions from ChatGPT useful? If so, how? If not, why?
4. Which suggestions provided by ChatGPT did you decide to accept? Why?
5. Which suggestions provided by ChatGPT did you decide to reject? Why?
6. Were there any suggestions provided by ChatGPT that you did not know whether to accept or reject? If so, which one?
7. What is the most difficult part of using ChatGPT as a feedback tool?
8. What is the most beneficial part of using ChatGPT as a feedback tool?
9. How is it different from the teacher's feedback?
10. What other questions did you have to ask ChatGPT to complete/revise your essay? And how useful was the feedback from those questions?
11. Do you think your revised essay version is better than the original? How? In what ways?
12. Would you use ChatGPT again in future written language practices? Why? Why not?
13. Do you think teachers should guide students on how to use ChatGPT for feedback in language classes? Why? Why not? If so, how?

Appendix C

Participants' Sample Quotes

1. Cognitive engagement

At the end of the first paragraph, I had written “be desarrollado”. But then it said, to change it to “desarrollé”. I think between those two past tenses, I can be confused about when I should use one or the other. So then, it was like that past tense and not the other one. (Isabella, Spanish learner)

However, I did think that it was important to include a paragraph, just because having a really big chunk of information isn't digestible. So then I split it up like suggested by ChatGPT. (Mawar, Indonesian learner)

2. Attitudinal engagement

I was expecting it [feedback] to be more generic. But it gave a lot of examples from the essay on where I could improve (Taylor, Spanish learner).

A positive way that it's different is that it's immediately responsive in any situation. If you're at home, you can get immediate feedback, which is really helpful, especially for weekly tasks. And it is able to highlight the reasons behind the changes it makes, which makes it more helpful. [...] there are certain contexts when you don't really have the time, or, you know, the place to be able to converse with a teacher. So a tool like this is still helpful, even if it is slightly less ideal in certain aspects, I still think that it's valuable. (Nuri, Indonesian learner)

The very first thing I didn't really understand was ‘you've mentioned that you studied Spanish for 6 years’, and it just told me what I had done. It wasn't helpful. (Luke, Spanish learner)

Enhancing Inclusivity and Equity through a Broader Understanding of Data-Driven Learning in Pre-tertiary Education: A Mixed-Methods Systematic Review

Shiya Huang  and Qing Ma 

Department of Linguistics and Modern Language Studies, The Education University of Hong Kong, Hong Kong SAR, China

Abstract

Data-driven learning (DDL) is a significant approach in computer-assisted language learning (CALL), but its adoption in pre-tertiary education is limited, raising concerns about education equity and inclusivity. This study reviews 39 empirical studies published between 1994 and 2023 to assess DDL's effectiveness, efficiency, learner perceptions. Quantitative findings show a large within-group effect size ($d = 1.77$) and a medium-to-large between group effect size ($d = 0.88$). Qualitative findings reveal that most learners (74%) report positive views. This review confirms DDL's suitability for non-university settings and offers strategies to enhance accessibility, address equity barriers, and promote inclusive practices, ensuring that all pre-tertiary learners, regardless of diverse background or resources, can benefit from DDL.

Keywords: Data-driven learning, language education, pre-tertiary, mixed-method review, inclusivity, equity

Introduction

Over the past decades, DDL has gained increasing attention in language education research (Boulton & Vyatkina, 2021). However, studies on DDL's

E-mail: s1151957@s.edu.hk & maqing@edu.hk

use in pre-tertiary education is still limited (Boulton & Cobb, 2017; Boulton & Vyatkina, 2021; Pérez-Paredes, 2022), highlighting gaps in understanding its accessibility and suitability for younger learners with diverse language skills and learning contexts. Additionally, previous reviews often overlook learners' experiences and practical implementation challenges, such as disparities in technology access that exacerbate equity issues in CALL. This study addresses these gaps by systematically analyzing 39 empirical studies (1994–2023) on DDL at the pre-tertiary level, assessing learning outcomes and learner feedback to promote accessible and adaptable DDL practices that foster inclusivity across varied educational settings.

In line with the aim of this study, the following research questions were formulated to guide the quantitative and qualitative synthesis respectively:

RQ1: To what extent is the implementation of DDL effective and efficient in facilitating language learning and teaching at the pre-tertiary level?

RQ2: How do pre-tertiary students perceive the effectiveness of corpus-based learning?

Methodology

Three databases (Web of Science, the Education Resources Information Centre [ERIC], and Google Scholar) and reference digging (Plonsky & Brown, 2015) were used to search for eligible studies. After selection based on the inclusion and exclusion criteria (Table 1), 39 studies were included in this synthesis.

To answer RQ1, we quantitatively synthesized 24 studies by calculating unbiased Cohen's d for between-group (effectiveness) and within-group (efficiency) studies, following methods from previous DDL meta-analyses (Boulton & Cobb, 2017; Mizumoto & Chujo, 2015; Ueno & Takeuchi, 2023). Outliers were managed through winsorizing method and capped the effect size at 3.0 to enhance the robustness of our results. The 24 studies yielded 40 separate effect sizes. To answer RQ2, we qualitatively synthesized 28 studies using grounded theory (GT), analyzing data on students' perceptions, attitudes, and experiences from observations, interviews, open-ended questionnaires, and other qualitative methods. Thornberg and Charmaz (2014) described GT as the simultaneous collection and analysis of data, blending induction and abduction to form categories. Following Charmaz (2006), we conducted initial coding, focused coding, and axial coding to connect categories and subcategories coherently.

Table 1. Inclusion/exclusion criteria

Inclusion criteria	Exclusion criteria
Studies published between 1994 and 2023	Studies not published between 1994 and 2023
Published empirical articles from journals or books	Not Published empirical articles from journals or books
Text written in English	Text not written in English
Related to corpus technology, DDL, and learning or teaching language	Not related to corpus technology, DDL, and learning or teaching language
Targeting at pre-tertiary level (Articles focusing on non-university teacher education or corpus-based materials development are considered if including pre-tertiary students' involvement)	Not targeting at pre-tertiary level

Findings

Quantitative synthesis

The aggregated effect size for all the within-group studies is 1.77, 95% CI [1.10, 2.43] ($p < 0.01$, $I^2 = 91\%$), and the aggregated effect size for all the between-group studies is 0.88, 95% CI [0.46, 1.29] ($p < 0.01$, $I^2 = 87\%$). Plonsky and Oswald (2014) highlight the need to contextualize Cohen’s d within specific academic disciplines and propose adjusted benchmarks for second language research: small ($d = 0.40$ for between-group, $d = 0.60$ for within-group), medium ($d = 0.70$ for between-group, $d = 1.00$ for within-group), and large ($d = 1.00$ for between-group, $d = 1.40$ for within-group) effects. According to these benchmarks, our within-group effect size is extremely large, while the between-group effect size is medium to large.

Compared to previous studies, within-group effect sizes range from 0.9 (Mizumoto & Chujo, 2015) to 1.50 (Boulton & Cobb, 2017), with Liu and Ma (2025) reporting 1.43. The present study’s within-group effect size ($d = 1.77$) is the highest, indicating greater efficiency. Between-group effect sizes range from 0.71 (Liu & Ma, 2025) to 0.95 (Ngo & Chen, 2024), with our study’s $d = 0.88$ falling in the middle, suggesting moderate effectiveness consistent with prior findings.

Qualitative synthesis

This qualitative synthesis analyzed 433 initial codes from student perceptions of DDL at the pre-tertiary level, revealing a predominance of positive (74%, 319 codes) over negative (26%, 114 codes) feedback. The perceptions were organized into eight categories and 30 sub-categories, with positive categories including “positive feelings,” reflecting learners’ excitement and enjoyment from DDL’s novelty and game-like designs; “efficacy,” highlighting its usefulness in simplifying complex concepts and improving skills; “autonomy,” showing increased motivation and independent learning through collaboration; “cognitive enhancement,” indicating improved language awareness and meta-cognitive skills via concordancing; and “combination with other learning resources,” noting benefits from integrating other tools such as dictionaries. Negative categories included “challenges,” addressing skills gaps and technical issues; “negative feelings,” capturing overwhelm and monotony among some learners; and “habits,” revealing a preference for traditional methods due to familiarity and test-focused learning styles.

DDL is valued by students for its innovative and enjoyable approach, enhancing learning effects, simplifying complex concepts, and boosting student autonomy and cognitive skills. However, it faces challenges including a lack of skills, language difficulties, insufficient teacher support, technical issues, and some students’ preference for traditional methods, leading to stress and reduced engagement. These perceptions underscore DDL’s potential to promote inclusivity and equity through positive learner engagement, while also revealing areas where technology access and support are essential to mitigate negative experiences.

Discussion

The quantitative results confirm DDL’s effectiveness and efficiency in pre-tertiary language education, showing its potential to improve language learning while supporting equity. Qualitatively, overall positive student attitudes reinforce these results. However, negative feedback highlights challenges such as limited DDL skills, negative feelings, and a preference for teacher-centered instruction and deductive approaches (e.g., using dictionaries or grammar notes). These challenges often stem from inequities in technology access, a key concern in inclusive CALL, where disparities in digital resources can hinder participation for underrepresented or resource-limited learners. To address these concerns inclusively, practical strategies identified through qualitative synthesis include pre-editing DDL materials, teacher supervision, multi-modal

resources, diverse activities, and scaffolding prompts. These approaches not only enhance accessibility but also promote equity by accommodating diverse learner needs, such as varying levels of technological proficiency and socioeconomic backgrounds, thereby aligning DDL with inclusive CALL principles. Ultimately, these strategies aim to bridge research and practice, ensuring equitable opportunities for all learners to benefit from DDL.

Conclusion

Despite the successful implementation of DDL, its adoption has yet to become mainstream and has remains limited to certain contexts. It necessitates a period of adjustment and dedicated effort to fully integrate DDL into existing practices. To promote greater educational equity and inclusivity, future research should aim to broaden its scope and support a wider range of educations in adopting and implementing DDL across diverse educational settings.

References

- Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, 67(2), 348–393. <https://doi.org/10.1111/lang.12224>
- Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions over time. *Language Learning & Technology*, 25(3), 66–89.
- Charmaz, K. (2006). *Constructing Grounded Theory*. SAGE.
- Liu, J., & Ma, Q. (2025). Examining corpus-based language pedagogy (CBLP) practices in data-driven learning (DDL) for low-proficiency L2 English learners: A meta-analysis. *Educational Technology & Society*, 28(2).
- Mizumoto, A., & Chujo, K. (2015). A meta-analysis of data-driven learning approach in the Japanese EFL classroom. *English Corpus Studies*, 22, 1–18.
- Ngo, T. T.-N., & Chen, H. H.-J. (2024). The effectiveness of corpus use in ESL/EFL writing: A meta-analysis. *Language Teaching Research*, 00(0), 1–27. <https://doi.org/10.1177/13621688241260183>
- Pérez-Paredes, P. (2022). A systematic review of the uses and spread of corpora and data-driven learning in CALL research during 2011–2015. *Computer Assisted Language Learning*, 35(1–2), 36–61. <https://doi.org/10.1080/09588221.2019.1667832>
- Plonsky, L., & Brown, D. (2015). Domain definition and search techniques in meta-analyses of L2 research (Or why 18 meta-analyses of feedback have

- different results). *Second Language Research*, 31(2), 267–278. <https://doi.org/10.1177/0267658314536436>
- Thornberg, R., & Charmaz, K. (2014). Grounded theory and theoretical coding. In U. Flick (Ed.), *The SAGE handbook of qualitative data analysis*. SAGE.
- Ueno, S., & Takeuchi, O. (2023). Effective corpus use in second language learning: A meta-analytic approach. *Applied Corpus Linguistics*, 3(3), 100076. <https://doi.org/10.1016/j.acorp.2023.100076>

AI-Supported Collaborative Learning to Reduce EFL College Students' Anxiety in Story Writing

Siao-Cing Lai¹ , Yu-Fen Yang² , Christine Chifen Tseng³ and Min-Chuan Tsai⁴

¹Institute of Education, National Yang Ming Chiao Tung University, Taiwan

²Department of Applied Foreign Languages, National Yunlin University of Science and Technology, Taiwan

³Department of Applied English, National Taichung University of Science and Technology, Taiwan

⁴Mount Dawu College, National Pingtung University, Taiwan

Abstract

In traditional collaborative learning, English as a Foreign Language (EFL) students with similar levels of foreign language writing anxiety (FLWA) frequently showed limited knowledge organization and teamwork effectiveness. This study proposed AI-supported collaborative learning to reduce EFL college students' FLWA in story writing. The 36 students participated, with one class receiving AI-supported collaborative learning (the experimental group) and the other receiving traditional collaborative learning (the control group). Data were from pre- and post-FLWA scales and story-writing tests. Results revealed that the experimental group reported significantly lower FLWA and higher story writing proficiency than the control group. Different from traditional collaborative learning, AI-supported collaborative learning provided automated feedback (story chain, grammar check, word choice, and evaluation) that helped the students overcome challenges (e.g., idea generation, storyline organization, and lexical usage). This AI support, with positive feedback

E-mail: cing0615@gmail.com, yangy@yuntech.edu.tw,
chifen@nutc.edu.tw & mtsai0913@mail.nptu.edu.tw

among peers, initiated positive emotional responses. Under the broaden-and-build theory, these positive emotions broadened the students' thoughts, built knowledge, and ultimately reduced FLWA and improved English story writing.

Keywords: AI-supported collaborative learning, broaden-and-build theory, college students, English story writing, foreign language writing anxiety

Study Background

English story writing competency is crucial for college students in their future careers, as stories are widely used in marketing to attract customers' attention (Ahmad et al., 2020). Story writing refers to narrating fictional or non-fictional events, requiring various elements (e.g., characterization, plot organization, and dialogue) (Vaezi & Rezaei, 2018). Given its complexity, many English as a Foreign Language (EFL) college students frequently experience foreign language writing anxiety (FLWA), defined as negative emotions triggered in writing contexts (Horwitz, 2017; Lai & Yang, 2024). FLWA is commonly provoked by fears of negative evaluations, grammatical concerns, and low self-confidence, which hinders writing development (Prasetyaningrum et al., 2021).

The broaden-and-build theory (Fredrickson, 2001), a key model of positive psychology, suggests the role of positive emotions in counteracting negative emotions. Positive emotions, defined as positive emotional responses to events, are signals initiating a cascade of optimal behaviors. That is, positive emotions broaden thoughts and actions, build personal knowledge, and ultimately undo negative emotions (Fredrickson, 2001). In educational settings, positive emotions are shaped by learning environments, and collaborative learning is encouraged, as social interactions can be predictors of happiness (Volet et al., 2019). However, the benefits of collaborative learning may be hindered, particularly when anxious students work together because they tend to demonstrate limited idea organization and generation (Ye & Zhou, 2022).

Generative AI can serve as knowledgeable scaffolding to help anxious students collaboratively address challenges by analyzing learning processes and providing automated feedback (Shadiev & Feng, 2024). Nevertheless, limited research has examined the effects of AI-supported collaborative learning on emotions and outcomes (Yan, 2024), and further investigation is required (Jin et al., 2021). Thus, this study investigated the effects of AI-supported collaborative learning on EFL college students' FLWA and story writing proficiency. The following research questions were addressed:

1. What were the effects of AI-supported collaborative learning on students' FLWA?

2. What were the effects of AI-supported collaborative learning on students' story writing proficiency?

Methods

This 14-week study (three hours per week) involved 36 EFL students (CEFR B1), aged 20–22, from an elective story writing course at a university in central Taiwan. Two classes were randomly assigned to the experimental and control groups. Data were collected through the FLWA scale and story-writing test in weeks 1 and 14. The FLWA scale (Cronbach's $\alpha = .88$), adapted from Cheng (2004), incorporated seven items for somatic anxiety (physiological reactions), seven items for avoidance behavior (writing avoidance), and eight items for cognitive anxiety (mental perceptions). The story-writing test, adapted from Lai (2011), consisted of four pictures for story writing. Before instructions, both groups shared similar levels of FLWA ($Z = -1.928$, $p = .054$) and writing proficiency ($Z = -.749$, $p = .454$). Instructions started from weeks 2 to 13. Both groups first received five-week lectures on story components. Then, the experimental group received AI-supported collaborative learning, while the control group received traditional collaborative learning for the last seven weeks.

AI-supported collaborative learning (see Figure 1) was designed based on the broaden-and-build theory (Fredrickson, 2001). The in-house story writing system features four functionalities: Story Chain (idea generation), Grammar Check (detection of grammatical errors), Word Choice (recommendations for vocabulary precision), and Story Evaluation (positive feedback on five story components). It also allows for peer review. The four instructional stages are described below.

- **Stage 1 (3 weeks):** Students co-wrote English stories. For example, G3 drafted,
"... Difference between the little boy and his sisters, wasn't only these, when they grow up, his sister wants to move out of the countryside ..."
During drafting, they received AI support. For example, Grammar Check provided instant feedback: *"Clarify what 'these' refers to. Ensure tense consistency: 'wants to' should be 'wanted'...."*
- **Stage 2 (1 week):** Students conducted peer reviews to exchange positive feedback based on three criteria: specific (areas of strength), understandable (actions for improvement), and achievable (goals). For example, *"Narrations were clear [specific]. You can try to add more dialogue [understandable] to make the story more vivid! [achievable]"* (S13).

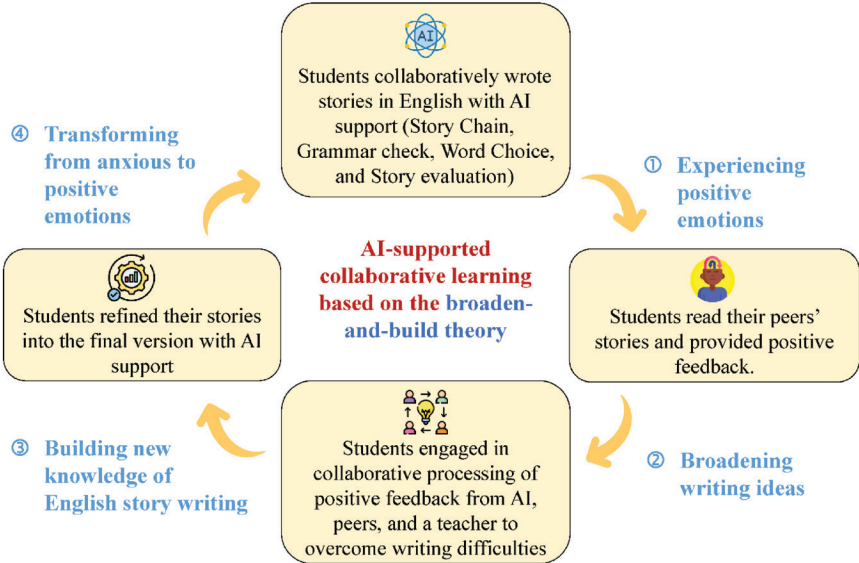


Figure 1. AI-Supported collaborative learning.

- **Stage 3 (1 week):** Students co-processed positive feedback from AI, peers, and the teacher, as S6 noted, “Peers’ comments enhance my confidence. Besides, AI helps us generate ideas and check grammar, reducing my fear of making errors during writing.”
- **Stage 4 (2 weeks):** Students refined their stories based on feedback. For example, G3 improved their story by adding dialogues and correcting grammatical errors:
“...These were not the only differences between the children. When they grew up, the girl wanted to move out of the countryside. ... ‘You are just a girl. You are just dreaming for something that you are not deserved!’ the father shouted.
...”

The control group followed the same four-stage sequence and educational content. However, they wrote and revised stories through group discussions without AI and positive feedback support.

Results

The Wilcoxon signed-rank test was used due to non-normally distributed data and a small sample size ($n < 30$) to ensure statistical robustness (Orcan, 2020). Table 1 shows the experimental group greatly reduced FLWA and

Table 1. College students’ FLWA and story writing proficiency before and after instructions

Groups	Factors	n	Pre-test		Past-test		Wilcoxon Signed-rank Test		
			M	SD	M	SD	Z	p	Cohen's d
Exp.	Somatic anxiety	16	21.19	6.23	14.13	5.68	-3.07	.002**	.77
	Avoidance Behaviour	16	20.81	5.39	16.19	4.97	-3.05	.002**	.76
	Cognitive anxiety	16	28.00	7.01	20.00	5.38	-3.52	<.001***	.88
	Overall FLWA	16	70.00	15.86	50.31	12.50	-3.41	.001**	.85
	Overall story writing proficiency	16	12.13	1.41	15.56	2.37	-3.31	.001**	.83
Con.	Somatic anxiety	20	21.25	4.83	20.70	4.91	-.60	.552	.13
	Avoidance Behaviour	20	19.25	2.22	18.85	3.54	-.49	.623	.11
	Cognitive anxiety	20	24.15	2.46	24.35	2.13	-.40	.688	.09
	Overall FLWA	20	64.65	6.82	63.90	6.59	-1.22	.221	.27
	Overall story writing proficiency	20	11.85	1.53	12.35	2.30	-1.32	.189	.29

Note. Exp. = experimental group; Con. = control group **p* < .05, ***p* < .01, ****p* < .001 (two-tailed test).

improved writing proficiency, with moderate (*d* ≥ .5) and large effect sizes (*d* ≥ .8) observed. Contrarily, the control group showed no significant change. Moreover, the experimental group reported significantly lower FLWA (*Z* = -3.236, *p* = .001*) and higher writing proficiency (*Z* = -3.757, *p* < .001***) than the control group.

Discussion and Conclusion

In AI-supported collaborative learning, the students received support with idea generation, grammar checking, and word suggestions. This AI support reduced the students’ fear of making mistakes, further facilitating positive emotions, such as confidence. Positive feedback also lessened their fears of negative social evaluation. These findings aligned with the broaden-and-build theory (Fredrickson, 2001), highlighting positive emotions could broaden ideas, build knowledge, and transform negative into positive emotional states. Contrarily, the control group lacked AI and positive feedback support, faced emotional challenges, and remained fearful of social judgment. This study

suggests that instructors incorporate AI tools and positive feedback to create supportive collaborative environments that reduce FLWA and enhance story writing proficiency.

References

- Ahmad, A., Samiullah, M., & Ahmed, G. (2020). Development of story writing skills through Communicative Approach at secondary level in Pakistan. *Global Regional Review, 1*, 143–150. [https://doi.org/10.31703/grr.2020\(V-I\).18](https://doi.org/10.31703/grr.2020(V-I).18)
- Cheng, Y.-S. (2004). A measure of second language writing anxiety: Scale development and preliminary validation. *Journal of Second Language Writing, 13*(4), 313–335. <https://doi.org/10.1016/j.jslw.2004.07.001>
- Fredrickson, B. L. (2001). The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. *American Psychologist, 56*(3), 218–226. <https://doi.org/10.1037/0003-066X.56.3.218>
- Horwitz, E. K. (2017). On the misreading of Horwitz, Horwitz, and Cope (1986) and the need to balance anxiety research and the experiences of anxious language learners. In C. Gkonou, M. Daubney, & J.-M. Dewaele (Eds.), *New insights into language anxiety: Theory, research and educational implications* (pp. 31–47). <https://doi.org/10.21832/9781783097722-004>
- Jin, Y., Dewaele, J.-M., & MacIntyre, P. D. (2021). Reducing anxiety in the foreign language classroom: A positive psychology approach. *System, 101*, Article 102604. <https://doi.org/10.1016/j.system.2021.102604>
- Lai, P. (2011). *A guide to better writing skills* (連環圖作文寫作指引). Taiwan: Ivy.
- Lai, S.-C., & Yang, Y.-F. (2024). Alleviating college students' writing anxiety through online collaboration to improve story writing. In Y. Wang & M. S. C. Claros (Eds.), *Proceedings of the International CALL Research Conference* (pp. 115–120). Castledown.
- Orcan, F. (2020). Parametric or non-parametric: Skewness to test normality for mean comparison. *International Journal of Assessment Tools in Education, 7*(2), 255–265. <https://doi.org/10.21449/ijate.656077>
- Prasetyaningrum, A., Nazri, A., & Asrobi, M. (2021). A study of learners' writing anxiety in EFL context. *Jo-ELT (Journal of English Language Teaching) Fakultas Pendidikan Bahasa & Seni Prodi Pendidikan Bahasa Inggris IKIP, 8*(1), 19–31. <https://doi.org/10.33394/jo-elt.v8i1.3707>
- Shadiev, R., & Feng, Y. (2024). Using automated corrective feedback tools in language learning: A review study. *Interactive Learning Environments, 32*(6), 2538–2566. <https://doi.org/10.1080/10494820.2022.2153145>

- Vaezi, M., & Rezaei, S. (2018). Development of a rubric for evaluating creative writing: A multi-phase research. *New Writing*, 16(3), Article 303317. <https://doi.org/10.1080/14790726.2018.1520894>
- 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. <https://doi.org/10.1080/03075079.2019.1665314>
- Yan, D. (2024). Comparing individual vs. collaborative processing of ChatGPT-generated feedback: Effects on L2 writing task improvement and learning. *Language Learning and Technology*, 28(1), 1–19. <https://hdl.handle.net/10125/73597>
- Ye, J.-m., & Zhou, J. (2022). Exploring the relationship between learning sentiments and cognitive processing in online collaborative learning: A network analytic approach. *The Internet and Higher Education*, 55, Article 100875. <https://doi.org/10.1016/j.iheduc.2022.100875>

Exploring Digital Literacy in ESP Tourism Presentations Using Importance-performance Analysis

Gayoung Lee 

Ewha Womans University, Seoul, Republic of Korea

Abstract

This study investigates the development of digital literacy in ESP tourism education using Importance-Performance Analysis (IPA). A pre-post survey and semi-structured interviews were conducted after the course with 32 undergraduates enrolled in a Digital Tourism English course. Results revealed significant gaps in self-regulation and information literacy, while AI and multimodal literacies improved. The IPA matrix identified clear instructional priorities, highlighting the need for systematic support. The study suggests integrating Bloom's Digital Taxonomy to enhance critical thinking and proposes pedagogical implications for tourism English instruction.

Keywords: ESP, digital literacy, tourism English, importance-performance analysis, Bloom's digital taxonomy

Background

The global tourism industry's digital transformation has reshaped how services are delivered and how clients interact. Kraus et al. (2022) demonstrate that these changes improve efficiency and enhance customer experience. Consequently, tourism professionals are required to manage booking systems, curate content for destination websites, and respond to customer needs through digital platforms.

In response, educational institutions are responding by integrating digital competencies into Tourism English instruction. Chang (2024) emphasizes that proficiency in English and digital literacy have become essential competencies,

E-mail: lifelikegreenery@hotmail.com

with technology playing a crucial role in optimizing language learning and professional performance in specialized ESP education.

Digital literacy, as defined by Gilster (1997), involves understanding and using information in multiple formats from various digital sources. Contemporary frameworks identify key components including information, search, ethical, critical, and expanded digital literacy (Morgan et al., 2022; Pegrum et al., 2022). While digital literacy enhances classroom engagement (Morgan et al., 2022; Kirovska-Simjanoska, 2022), students often lack critical evaluation skills. In Tourism English, digital literacy positively impacts destination promotion and intercultural communication (Ritonga, 2023; Dabić, 2017).

However, few studies have applied pedagogical frameworks to Tourism English digital literacy. While Bloom's Taxonomy has proven effective in enhancing critical thinking in ESP contexts (Vavilina, 2020; Bilonozhko, 2023), research examining the gap between perceived importance and actual performance of digital literacy in tourism English contexts remains limited. This study examines how students perceive and develop digital literacy competencies within ESP tourism education, using Importance-Performance Analysis to identify areas where students need additional support. The study addresses three research questions.

1. To what extent do importance and performance ratings of digital literacy skills differ between pre- and post-tests?
2. What patterns emerge in the importance-performance relationships of digital literacy components using IPA in pre- and post-test data?
3. How do students perceive different digital literacy components in Tourism English instruction?

Methodology

This mixed-methods study employed a pre-post survey design with 32 undergraduate students enrolled in a Digital Tourism English course. The course comprised two main components: conversation practice and theoretical study. Students were required to explore professional tourism terminology while examining official tourism websites, developing both language skills and digital literacy.

The instructional design was based on previous research demonstrating that systematic application of Bloom's Digital Taxonomy enhances critical thinking and reduces student anxiety (Vavilina, 2020; Bilonozhko, 2023). Previous research shows that students can use digital tools effectively but

often lack critical evaluation skills (Morgan et al., 2022). A structured presentation development process was therefore implemented to address this issue. Accordingly, students progressed through six cognitive levels (Remember to Create) from basic information retrieval to critical evaluation and multi-modal presentation creation. Reflective learning logs helped students track their source selection and develop critical evaluation skills, while students were required to use English-language websites for information literacy practice.

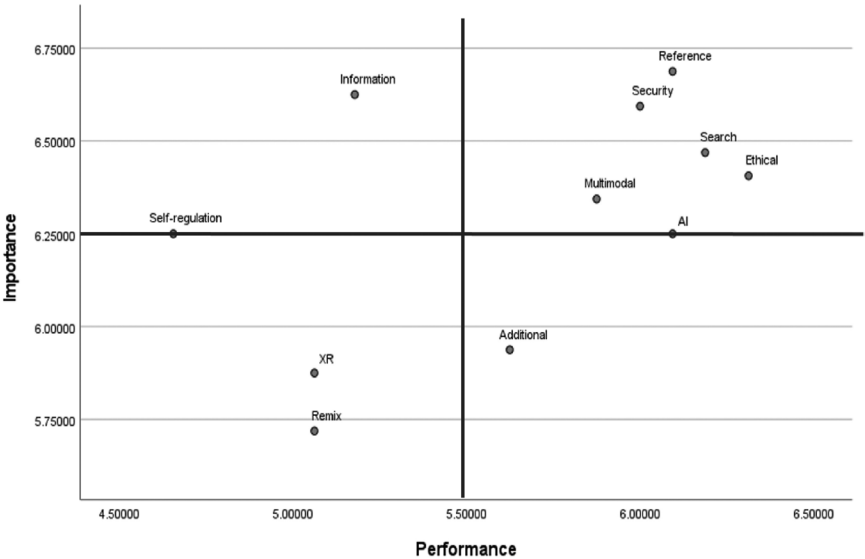
Pre- and post-surveys measured students' perceived importance and performance across 11 digital literacy attributes using 7-point Likert scales (Pegrum et al., 2022). Quantitative data were analyzed using paired t-tests and Importance-Performance Analysis (IPA). The IPA created a four-quadrant matrix by plotting importance against performance to identify priority areas (Martilla & James, 1977). Semi-structured interviews were conducted with all participants and analyzed using Miles and Huberman's (1994) three-phase process.

Representative Results

Research Question 1 examined differences between importance and performance ratings across pre- and post-tests. Self-regulation demonstrated the most substantial gaps in both measurements (pre-test: $d = 1.656$, $p < .001$; post-test: $d = 1.594$, $p < .001$). Information literacy gaps increased from pre-test ($d = 0.75$, $p = 0.032$) to post-test ($d = 0.906$, $p < .001$). Conversely, AI literacy progressed from no significant gap in pre-test ($p = 0.914$) to ratings being aligned in the post-test. Multimodal literacy also improved, with the initial gap ($d = 0.469$, $p = 0.033$) narrowing substantially.

Research Question 2 explored IPA matrix patterns. The post-test analysis positioned Search, Ethical, Reference, and Security literacies in the "Keep up the good work" quadrant (high importance, high performance). Self-regulation and Information literacy occupied the "Concentrate here" quadrant (high importance, low performance). XR and Remix literacies fell into the "Low priority" quadrant, while AI and Multimodal literacies moved into the "Possible overkill" quadrant.

Research Question 3 investigated student perceptions through qualitative analysis. Participants expressed enthusiasm for XR's immersive potential, noting the appeal of being able to "travel without leaving our seats" (N9), while voicing concerns that "actual travel will decrease" (N15, N25, N27). Regarding AI literacy, students appreciated practical benefits but raised ethical concerns about "small business owners" (N1, N2, N3).



(Reproduced from Lee & Lee, 2024)

Figure 1. Post-test IPA chart for 11 parameters

Conclusion and Discussion

The findings highlight the significant role of digital literacy in ESP tourism education. The persistent self-regulation gaps, despite successful alignment in AI and multimodal literacies, indicate that while digital natives recognize competency importance, they struggle with practical implementation (Spante et al., 2018). These results emphasize the need for systematic self-regulation programs in ESP tourism education.

The IPA results provide clear educational priorities. Self-regulation’s placement in the “Concentrate here” quadrant reveals the need for immediate intervention, aligning with Kirovska-Simjanoska’s (2022) findings on student focus challenges in digital environments. Search and ethical literacy positioned in “Keep up the good work” demonstrate adequate capabilities, while XR’s low priority reflects current limited tourism applications.

Qualitative analysis revealed student recognition of XR’s immersive value alongside concerns about replacing actual travel experiences. Regarding AI literacy, students acknowledged convenience while raising ethical concerns about corporate manipulation. These findings suggest balanced technology integration emphasizing both technical competency and ethical judgment.

This study demonstrates how Bloom's Digital Taxonomy provides a structured framework for ESP tourism education. However, the single-university sample limits generalizability. Future research should include multiple institutions and diverse departments to expand digital literacy frameworks across technological competencies.

References

- Bilonozhko, N. (2023). Use of Bloom's taxonomy in the ESP final exam preparation as the way to avoid the feelings of anxiety in the war period. *European Humanities Studies State and Society*, 1, 24–40.
- Chang, H. J. (2024). ESP instruction design based on digital literacy for tourism English learners. *Journal of English Teaching through Movies and Media*, 25(3), 60–71.
- Dabić, T. (2017). Useful online resources and educational tools for teaching English for tourism and hospitality. *Zbornik Radova Univerziteta Sinergija*, 1(2), 1–6.
- Gilster, P. (1997). Digital literacy. Wiley.
- Kirovska-Simjanoska, D. (2022). Digital literacy as a tool for fostering engagement and motivation in online/hybrid ESP classes. *Journal of Teaching English for Specific and Academic Purposes*, 10(3), 533–546.
- Kraus, S., Durst, S., Ferreira, J., Veiga, P., Kailer, N., & Weinmann, A. (2022). Digital transformation in business and management research: An overview of the current status quo. *International Journal of Information Management*, 63(4), 1–18.
- Lee, G., & Lee, H. (2024). Importance-performance analysis of digital literacy in ESP tourism English presentation activities among Korean university students. *Journal of Language Sciences*, 31(4), 241–273.
- Martilla, J. A., & James, J. C. (1977). Importance-performance analysis. *Journal of Marketing*, 41(1), 77–79.
- Miles, M. B., & Huberman, A. M. (1994). Qualitative data analysis: An expanded sourcebook. Sage.
- Morgan, A., Sibson, R., & Jackson, D. (2022). Digital demand and digital deficit: Conceptualising digital literacy and gauging proficiency among higher education students. *Journal of Higher Education Policy and Management*, 44(3), 258–275.
- Pegrum, M., Hockly, N., & Dudeney, G. (2022). Digital literacies. Routledge.
- Ritonga, A. K. (2023). An analysis of English for tourism, digital literacy, and business success to a tourist destination promotion. *Journal Iqra*, 8(2), 325–339.

- Spante, M., Hashemi, S. S., Lundin, M., & Algers, A. (2018). Digital competence and digital literacy in higher education research: Systematic review of concept use. *Cogent Education*, 5(1), 1519143.
- Vavilina, S. H. (2020). Using Bloom's revised taxonomy to develop digital literacies in the ESP classroom. *Information Technologies and Learning Tools*, 79(5), 168–183.

From Lookup to Output: Exploring Mobile Vocabulary Searches for Self-Expression in Multicultural Japanese Language Learning

Yuzhe Li^{*1}  and Kei Ishiguro² 

¹University of Macau, Department of Japanese, Macau SAR, China

²National Institute for Japanese Language and Linguistics, Tokyo, Japan

Abstract

Mobile dictionaries and translation tools provide learners with instant and extensive access to vocabulary, but their role in supporting searches for language production remains underexplored. This study examines output-oriented vocabulary searches, which may aid self-expression, without directly measuring learning outcomes. Using smartphone screen recordings, we collected 3,552 searches from 80 Japanese language learners in China, South Korea, and Vietnam. Each search was coded by purpose (comprehension or output) and depth (quick or in-depth). Only 14.2% were output-oriented, and most were quick and superficial, with little attention to nuances or contextual meaning. The findings reveal a paradox in mobile-assisted vocabulary learning: while access is widespread and inclusive, depth is limited. The study highlights the need for purposeful tool design and pedagogical scaffolding to foster more productive vocabulary engagement.

Keywords: Vocabulary lookup, smartphones, screen recording function, productive vocabulary

*Corresponding Author E-mail: yuzheli@um.edu.mo & ishigurokei@ninja.ac.jp

Introduction

Vocabulary lookup is the foundation of language learning, and mobile technology has made this process much more efficient. Devices such as dictionary apps in smartphones, and search engines provide low-cost, easily accessible, and instant ways to look up the meanings and usage of words.

While these tools can effectively support vocabulary learning focused on comprehension (Wang & Stockwell 2024; Lai & Wang 2025), their role in improving learners' productive vocabulary is less clear. In this study, the focus is on vocabulary search behavior rather than on the direct assessment of learning outcomes. And productive vocabulary refers to the ability to actively select and use words and expressions to convey intended meanings in speaking and writing, which we define as the self-expression purpose. Lai and Wang (2025) emphasize that the purpose of media use has a significant influence on vocabulary learning. Also, differences in learners' cultural backgrounds, educational systems, and information and communication technology (ICT) environments may also influence their vocabulary search behavior.

This study aims to examine how learners in different educational and cultural contexts use mobile tools to search for vocabulary for the purpose of self-expression, and how such search behaviors vary across contexts. It also considers the pedagogical implications for designing instructional strategies that help learners move from simple words lookup to meaningful and accurate use of vocabulary in mobile-assisted language learning.

Vocabulary Lookup and Language Output

Vocabulary lookup primarily serves two purposes: comprehension (input side) and language production (output side). Prior studies have shown that dictionary tools effectively support input-oriented vocabulary search, such as memorizing words, reading and extracting vocabulary, looking up definitions, and even creating personal dictionaries (Chen, 2010; Levy & Steel, 2015). However, output-oriented search, which supports learners in finding and applying words for speaking or writing, requires more complex processing and often deeper engagement with usage contexts (Laufer & Nation 1995; Dell et al., 1999). However, search that supports expressive abilities faces greater challenges in practice due to individual differences in linguistic competence, digital literacy, and metacognitive strategies.

To explore the above situation and issues from some basic perspectives, this study proposes the following research questions:

- RQ1: To what degree do Japanese language learners use mobile tools for output-oriented vocabulary search?
- RQ2: What differences exist in the behaviors of learners from different educational and cultural backgrounds (e.g., China, South Korea, Vietnam) when using mobile tools for output-oriented vocabulary search?

Methods

Participants

The study participants included 80 Japanese language learners from China ($n = 48$), South Korea ($n = 22$), and Vietnam ($n = 10$), regions that differ significantly in terms of cultural background, education systems, and ICT infrastructure. The learners' Japanese proficiency levels ranged from beginner to upper-intermediate, but all had the habit and basic skills of using mobile devices to look up words.

Data Collection and Analysis

Participants used the screen recording function of their smartphones to document their vocabulary search behavior over approximately one week, resulting in a total of 3,552 search instances. Of these, 3,427 searches were conducted via mobile tools, categorized as follows: Dictionary apps and online dictionaries: 2,486 searches, educational websites (e.g., vocabulary lists, Q&A sites): 131 searches, general search engines: 233 searches, machine translation tools (e.g., Google Translate): 141 searches, electronic dictionaries (non-smartphone): 469 searches, various combinations of mobile tools: 92 searches. Additionally, participants recorded the vocabulary items searched, the reasons and context for the search, the purpose (understanding or output), and their self-assessment in supplementary text records.

Although the number of participants from the three countries was unbalanced, the total number of searches recorded was relatively similar: China (1,583 searches), South Korea (824 searches) and Vietnam (1,020 searches). This to some extent reduced the impact of participant imbalance on cross-national comparisons.

Based on the above data, we coded each search behavior using the following two-level criteria. First level: Purpose (input-oriented vs. output-oriented); Second level: In output-oriented contexts, conducting "deep searches" (Li 2024) for specific scenarios. Specifically, this involves an in-depth analysis of usage contexts and nuances, including collocations, example sentences, idioms, common contexts, comparisons with the native language, and exploration of

hidden meanings and related concepts in Japanese culture, as well as efforts to complete and improve self-expression. The data was further analyzed through comparisons with different countries, tool types, and search objectives to identify trends and differences.

Ethical Considerations

The use of screen recording raises significant ethical issues regarding the protection of personal privacy. All data was obtained with informed consent, allowing students to withdraw their videos at any time and ensuring that this would not affect their grades. NINJAL ensures the security of uploaded data and deletes any content that accidentally captures personal information. All public videos are reviewed by participants.

Findings and Discussion

In response to RQ1, the data showed that out of 3,427 vocabulary searches recorded, only 488 (14.2%) were output-oriented, while the remaining 2,770 vocabulary searches were input-oriented (the remainder included cases that were unclear). This suggests that Japanese learners frequently prioritized simplicity and speed, completing quick translation lookups without additional investigation of usage or context.

In response to RQ2, significant cross-regional differences emerged. Chinese learners conducted the most output-oriented searches (324 out of 1,583 vocabulary searches, or 20.5%), followed by Korean learners (153 out of 824 vocabulary searches, or 18.6%). In comparison, only 11 (1.1%) of the 1,020 queries conducted by Vietnamese learners were output oriented. These differences may reflect differences in curriculum focus, pedagogical practices, and ICT familiarity. Chinese and Korean learners may use their familiarity with Chinese characters to directly infer the meaning of words from their orthography, which may affect their search objectives and strategies. In some cases, these learners only search for pronunciations (e.g., to confirm on-yomi or kun-yomi) or to verify the approximate meaning of Chinese characters they already know, which differs from Vietnamese learners, who typically need to search for character forms, meanings, and usage.

The above findings suggest that while certain tools (such as search engines and machine translation) can provide rich contextual data, learners rarely utilize them to support self-expression. Most output-oriented searches focus on vocabulary usage, collocation, and idioms, but few involve in-depth searches considering discourse nuances or contextual sensitivity. It is important to note that this study examines search behavior as an observable activity and

does not measure whether actual vocabulary learning or retention took place. Objectively, we have the technological conditions to support self-expression, but we do not see the results. As mentioned by Wang and Stockwell (2024), mobile tools can provide inclusive access, but without pedagogical guidance or training, they cannot promote equitable or transformative language learning outcomes.

Limitations and Future Studies

This study has several limitations, including an imbalance in participant numbers across the three countries, which means that the overall results may reflect the behavior of Chinese participants more strongly and may be less representative of learners from Vietnam. Individual differences in digital literacy, language proficiency, and learning motivation were not fully controlled. Despite these constraints, the study provides evidence about output-oriented vocabulary searches in mobile learning contexts. Future research should explore how to support more equitable and effective output-oriented vocabulary learning during the searching phase to meet the diverse needs of learners.

Acknowledgments

This study is part of the NINJAL collaborative research project “A Survey on the Use of online Dictionary Resources by Japanese Language Learners” and was supported by JSPS KAKENHI Grant Number JP21K18375.

References

- Chen, Y. (2010). Dictionary use and EFL learning: A contrastive study of pocket electronic dictionaries and paper dictionaries. *International Journal of Lexicography*, 23(3), 275–306. <https://doi.org/10.1093/ijl/ecq013>
- Dell, G. S., Chang, F., & Griffin, Z. M. (1999). Connectionist models of language production: Lexical access and grammatical encoding. *Cognitive Science*, 23(4), 517–542. https://doi.org/10.1207/s15516709cog2304_6
- Lai, C., & Wang, Q. (2025). Online informal learning of English and receptive vocabulary knowledge: Purpose matters. *ReCALL*, 37(1), 62–78. <https://doi.org/10.1017/S095834402400017X>
- Lai, C., & Wang, Q. (2025). Profiles of media use purpose in informal digital learning of English and their association with vocabulary knowledge.

- Computer Assisted Language Learning*, 1–31. <https://doi.org/10.1080/09588221.2025.2501705>
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16(3), 307–322. <https://doi.org/10.1093/applin/16.3.307>
- Levy, M., & Steel, C. (2015). Language learner perspectives on the functionality and use of electronic language dictionaries. *ReCALL*, 27(2), 177–196. <https://doi.org/10.1017/S095834401400038X>
- Li, Y. (2024). The depth of vocabulary search on smartphones among Japanese language learners in China. *Proceedings of the International CALL Research Conference*, 2024, 149–154. <https://doi.org/10.29140/9780648184485-23>
- Wang, Y., & Stockwell, G. (2024). Training to use machine translation for vocabulary learning. In Teng, M. F., Kukulska-Hulme, A., & Wu, J. G. (Eds.), *Theory and practice in vocabulary research in digital environments* (pp. 132–153). Routledge.

The Transformative Role of E-Readers in Inclusive Online Mandarin Education: A Duo-Ethnographic Study

Chuan Liu¹  and Jing (Vera) Huang²

¹Western University, Canada

²Vancouver Mandarin School, Canada

Abstract

Many Mandarin programs in Canada rely on prescriptive, textbook-driven curricula that may overlook the diverse learning needs and backgrounds of plurilingual students. This research seeks to challenge and reimagine this traditional curriculum by exploring how Mandarin educators can integrate E-readers as alternative materials to address students' emerging learning needs. In the post-pandemic era, online language education has become the new normal. While extensive insights have emerged from online English language teaching regarding its challenges, benefits, and implications, the specific context of online Mandarin education remains insufficiently explored, particularly in North America. Given the growing demand for Mandarin language learning in North America, this study engaged two Mandarin teachers in Canada in a 12-week series of online sessions utilizing E-readers (i.e., iChineseReader) as core teaching materials. Employing a duo-ethnographic approach, this research collected data: video-recorded weekly critical dialogues, reflective journals, and lesson plans as pedagogical artifacts. Narrative inquiry guided data analysis, focusing on the teachers' evolving ideologies surrounding E-reader integration in online Mandarin classrooms. This study presents key stories in terms of: (1) E-readers as a support for ethics-oriented curricula; (2) E-readers as mediators of learner autonomy; and (3) E-readers as tools for plurilingual, multimodal, and identity-affirming learning. This research invites educators to creatively use E-readers to enhance inclusivity, foster learner agency, and recognize students' linguistic and semiotic resources in online Mandarin learning.

E-mail: cliu783@uwo.ca & verahuang@vancouvermandarinschool.com

Keywords: Online Mandarin education, E-readers, inclusive education, learner autonomy, duo-ethnography

Many Mandarin programs in Canada rely on prescriptive, textbook-driven curricula, often neglecting the diverse learning needs of plurilingual students (Du, 2017; Jin & Liu, 2023). With online language teaching becoming increasingly common in the post-pandemic era (Amitabh, 2020), Mandarin educators face the pressing challenge of creating inclusive and responsive online learning environments. Although numerous research endeavors address online English language education (e.g., Liu & Bangou, 2022), research insights on online Mandarin education in North America remain limited (Pan et al., 2023). This study aims to address this gap by examining how the integration of E-readers in online Mandarin classrooms fosters inclusive language teaching practices. Specifically, this study explores: How do Mandarin teachers perceive the role of E-readers in enhancing learner autonomy and addressing the diverse needs of plurilingual students?

Theoretical Framework

This study employs the curriculum paradigm articulated by Heydon and Wang (2006) to analyze Mandarin curricula for plurilingual learners. A prescriptive curriculum paradigm refers to a fixed scope of content, teacher-directed instruction, and standardized expectations for all students. Accordingly, teachers primarily deliver content, and students are expected to follow predetermined materials and goals, often neglecting individual needs and interests. In contrast, an ethics-oriented curriculum paradigm recognizes students' diverse subjectivities and identities. Teachers adapt instruction to students' emerging needs and invite them to co-construct learning goals, thereby promoting educational equity and fostering learner autonomy by drawing on students' linguistic and cultural resources (Cope & Kalantzis, 2023). In this study, teachers engaged in critical dialogues, continuously reflecting on the role of E-readers in promoting inclusive curricula that better address linguistic diversity and learner autonomy.

Methodology

Employing a duo-ethnographic approach (Norris & Sawyer, 2012), this research engaged two Mandarin teachers in Canada in a 12-week series of online sessions. Rather than contrasting cases, the study included teachers

working with different learner populations to facilitate critical dialogue on curriculum practices across varied educational contexts. Teacher V holds a Master of Arts in education and has a decade-long Mandarin teaching experience in diverse contexts. Teacher C is a PhD student in education and has more than 15 years of language teaching experience. Both teachers utilized E-readers (i.e., iChineseReader, n.d.) as core teaching materials for their online classes. The E-Readers provide a wide selection of graded Mandarin lessons on real-life topics and include digital affordances, such as vocabulary translation, grammar notes, audio support, and interactive exercises. V focuses on adult language learners, while C focuses on young language learners. They conducted weekly online critical dialogues that served as reflective and transformative spaces after their teaching. Their dialogues centered around how the use of E-readers will enhance language learning for their students, such as how they planned and delivered their lessons and how students reacted. Data collection included video recordings of weekly online discussions, reflective journals documenting teachers' experiences, and lesson plans as pedagogical artifacts. Narrative inquiry (Clandinin & Connelly, 2004) guided the thematic analysis and data presentation, focusing on how curriculum negotiation unfolded (Kalantzis & Cope, 2010) regarding the integration of E-readers in online Mandarin classrooms.

Findings

Three significant themes emerged from this study, illustrating transformative shifts in curriculum practices with the integration of E-readers in online Mandarin classrooms: (1) E-readers as a support for ethics-oriented curricula; (2) E-readers as mediators of learner autonomy; and (3) E-readers as tools for plurilingual, multimodal, and identity-affirming learning.

Theme 1

E-readers enabled a shift toward ethics-oriented curriculum practices by allowing students to select reading materials that matched their language proficiency, interests, and backgrounds. For instance, C observed that two boys in his class were becoming increasingly disengaged with Mandarin learning. Initially, he attributed their reluctance as personal shortcomings: "It's their problem, not mine." However, through critical dialogue with V, C recognized his curriculum's efficiency-oriented nature had marginalized students' unique needs. He critically acknowledged, "It's my fault not to help them gain confidence." By integrating E-readers, C was able to offer a range of texts at different levels and on diverse topics, empowering students to choose readings that resonated with

their own experiences and abilities. As a result, students actively selected reading materials, formulated questions, and collaboratively established learning goals and assessments, thus foregrounding ethical curriculum making.

Theme 2

The integration of E-readers fundamentally changed both instructional planning and classroom dynamics by promoting learner autonomy. For example, V, who initially relied on grammar- and vocabulary-centric textbook materials, shifted to offering students greater autonomy by enabling them to choose texts based on their thematic interests. The built-in features enabled students to navigate texts independently, seek clarification as needed, and engage with content at their own pace. V observed substantial improvements in student confidence and engagement, remarking, “Students with different backgrounds became more confident...they find the things they want to learn as active learners themselves.” By leveraging the affordances of E-readers, teachers were able to foster greater student agency, decrease instructional pressures, and create a more responsive and dynamic learning environment.

Theme 3

Through E-readers, Mandarin classes evolved from traditional monolingual instruction to a more inclusive plurilingual and multimodal curriculum. V and C encouraged students to utilize E-reader features such as translation tools to access content in their home languages, which supported their comprehension and affirmed their linguistic repertoires. E-readers also enabled students to express meaning through a variety of modes, such as personalized picture-books and visual stories, using E-reader images as prompts for multilingual storytelling, and sharing reflections across languages. C observed, “They fluidly shift between languages and modes to make meaning...” demonstrating how E-readers facilitated the development of students’ multilingual identities and diverse semiotic resources.

These findings underscore the transformative impact of E-readers in fostering culturally responsive, inclusive, and empowering Mandarin teaching practices.

Discussion

This study explored how two Mandarin educators developed a critical friendship to examine the role of E-readers in supporting more inclusive language learning and moving beyond textbook-based instruction in online classrooms. The findings demonstrate that, when intentionally implemented, E-readers

served as powerful tools for fostering student choice, learner autonomy, and plurilingual meaning-making. By enabling students to select readings that matched their interests and language levels, access content in their home languages, and engage with multimodal resources, E-readers facilitated a shift toward more equitable and responsive teaching practices. Overall, this research provides valuable guidance for Mandarin educators aiming to leverage digital tools like E-readers to create equitable, inclusive, and plurilingual online language learning environments.

Acknowledgements

The authors would like to acknowledge the **Canadian Association of Second Language Teachers (CASLT)**, whose generous grants supported this research.

References

- Amitabh, U. (2020, August 31). How technology will transform learning in the COVID-19 era. *World Economic Forum*. <https://www.weforum.org/agenda/2020/08/how-edtech-will-transform-learning-in-the-covid-19-era/>.
- Clandinin, D. J., & Connelly, F. M. (2004). *Narrative inquiry: Experience and story in qualitative research*. John Wiley & Sons.
- Cope, B., & Kalantzis, M. (2023). Towards education justice: The multiliteracies project revisited. In *Multiliteracies in International Educational Contexts* (pp. 1–33). Routledge.
- Du, X. (2017). Rethink about heritage language learning: A case study of children's Mandarin Chinese learning at a community language school in Ontario, Canada. *Language and Literacy*, 19(1), 4–20. <https://doi.org/10.20360/G2KP45>
- Heydon, R. M., & Wang, P. (2006). Curricular ethics in early childhood education programming: A challenge to the Ontario kindergarten program. *McGill Journal of Education/Revue des Sciences de L'éducation de McGill*, 41(1).
- iChineseReader. (n.d.). *iChineseReader*. Retrieved August 5, 2025, from <https://ichinesereader.com/>
- Jin, J., & Liu, Y. (2023). Towards a critical translanguaging biliteracy pedagogy: The 'aha moment' stories of two Mandarin Chinese teachers in Canada. *Literacy*, 57(2), 171–184. <https://doi.org/10.1111/lit.12323>

- Kalantzis, M., & Cope, B. (2010). The teacher as designer: Pedagogy in the new media age. *E-learning and Digital Media*, 7(3), 200–222. <https://doi-org.proxy1.lib.uwo.ca/10.2304/elea.2010.7.3.200>
- Liu, C., & Bangou, F. (2022). The influence of a group of Chinese EFL teachers' beliefs on lesson planning with video-based synchronous computer-mediated communication: A qualitative multiple case study. *Technology in Language Teaching and Learning*, 4(2), 41–61. <https://doi.org/10.29140/tltl.v4n2.780>
- Norris, J., & Sawyer, R. D. (2012). *Duoethnography: Dialogic methods for social, health, and educational research*. Left Coast Press.
- Pan, L., Sun, D., Zou, Y., Cao, Y., Zhang, J., & Li, F. (2023). Psycho-linguistic and educational challenges in teaching Chinese (Mandarin) language: Voices from none-Chinese teachers of Mandarin language. *BMC Psychology*, 11(1), 390. <https://doi.org/10.1186/s40359-023-01432-8>

Using a Custom GPT to Scaffold Corpus-Based Vocabulary Learning for Low-Proficiency Learners

Jing Liu^{1,2} and Qing Ma^{1*} 

¹Department of Linguistics and Modern Language Studies, The Education University of Hong Kong, Hong Kong SAR, China

²School of International Education, Chengdu Polytechnic, Chengdu, China

Abstract

A corpus-based approach is often considered unsuitable for low-proficiency learners, who frequently experience cognitive overload in such learning contexts. To help these learners manage the cognitive demands of corpus-based vocabulary learning, this study designed a custom GPT configured with six scaffolding methods to address three key challenges commonly faced by low-proficiency learners, and integrated it into the framework of corpus-based language pedagogy (CBLP). Unlike teacher-led scaffolding, which may not allow simultaneous support for multiple learners due to inevitable physical constraints, the custom GPT delivers on-demand, contingent support through learner-initiated requests. This mixed-methods study compared two within-group conditions: CBLP with and without custom GPT scaffolding, involving twenty low-proficiency L2 English learners at a vocational college in China. Data sources included vocabulary tests, GPT chat logs, and semi-structured interviews. Results showed that GPT-scaffolded CBLP significantly improved vocabulary knowledge, with the most substantial gains in use. Content analysis of GPT chats revealed that the custom GPT delivered six scaffolding methods in ways that responded contingently to the specific challenges learners faced. Thematic analysis of interview transcripts further indicated that GPT scaffolding demonstrated the characteristic of contingency. Integrating quantitative and qualitative findings, we found that GPT scaffolding appeared

*Corresponding Author E-mail: maqing@eduhk.hk

to have enhanced the effectiveness of CBLP. The present study contributes to the limited but much-needed knowledge of how to integrate contingent scaffolding, enabled by generative AI chatbots, into CBLP lesson designs aimed at enhancing vocabulary learning for low-proficiency learners.

Keywords: Corpus-based language pedagogy (CBLP), GPT scaffolding, contingency, vocabulary knowledge, low-proficiency learners

Background

Despite extensive evidence supporting the effectiveness of corpus-based vocabulary learning (e.g., Karras, 2016; Kaur & Hegelheimer, 2005), a corpus-based approach is often considered unsuitable for low-proficiency learners because heavy cognitive demands may exceed their linguistic competence (Lee et al., 2020). Given this limitation, our study adopted the Corpus-based Language Pedagogy (CBLP) framework (Ma et al., 2022), which emphasizes designing instructional scaffolds to facilitate learners' engagement with corpus-based tasks. However, current corpus instructional practices rely mainly on teacher-led guidance (e.g., Smart, 2014) or pre-designed materials (e.g., Braun, 2005). Although valuable, these supports often fall short in providing contingent scaffolding, meaning moment-by-moment assistance tailored to learners' evolving needs (van de Pol et al., 2010). Drawing on scaffolding theory, we investigated the use of a custom GPT for delivering on-demand assistance during CBLP tasks. Furthermore, prior research on corpus-based vocabulary learning has typically focused on overall vocabulary gains, and few studies have examined its differential effects across individual dimensions. To address this gap, the present study evaluated the impact of GPT-scaffolded CBLP on three dimensions of vocabulary knowledge: form, meaning, and use (Nation, 2022).

Methodology

Designing GPT Scaffolding

We created a custom version of ChatGPT and configured it to implement six scaffolding means as outlined by van de Pol et al. (2010). Aligned with corpus-based activities such as concordance comprehension and inductive discovery, the system delivered three types of support: (1) linguistic support that clarified unfamiliar language items; (2) metalinguistic support that explicated language structures and their functions; and (3) cognitive support that guided learners' lexical inference and pattern detection. Unlike teacher-led scaffolding, which is constrained by the teacher's ability to attend to multiple learners

simultaneously in the classroom setting, generative AI tools like the custom GPT offers a promising opportunity to provide on-demand, contingent support in real time (Baidoo-Anu & Ansah, 2023). These three types of support were specifically designed to address the three typical challenges low-proficiency learners face in corpus-based tasks (Liu & Ma, 2025), with the aim of sustaining cognitive engagement with concordance lines and facilitating inductive language discovery.

Participants

The study recruited 20 first-year L2 English learners from a vocational college in China using convenience sampling. All participants scored below 90 out of 150 on the English section of National College Entrance Exam, a criterion used to indicate low proficiency in previous research (Cheng & Liu, 2022) and had a vocabulary size below the 4,000 word-family threshold for intermediate proficiency (Nation & Meara, 2010).

Task Design

In this study, we sampled concordance lines from COCA and SKELL, identifying collocations using a minimum MI score of 3 (Hunston, 2002) to ensure lexical authenticity. Given the potential variability in difficulty across concordance lines (Boulton, 2009), we conducted a lexical coverage analysis to confirm that at least 90% of the text (excluding target words) met the threshold for comprehensibility (Hu & Nation, 2000).

Experiment Design

This study adopted a mixed-methods, repeated-measures design to evaluate the effectiveness of GPT scaffolding in corpus-based vocabulary learning. The intervention included two conditions: one without GPT support and one with GPT support. Participants took part in six 45-minute sessions, with three sessions conducted under each condition. Both conditions followed the same instructional sequence, based on the four-step CBLP lesson design model (Ma et al., 2022), with the only difference being the presence or absence of GPT scaffolding. In the GPT-supported condition, learners used pre-designed prompts to elicit support.

Data Collection and Analysis

A vocabulary pretest and posttest were administered for each intervention condition to assess learning gains. Additionally, semi-structured interviews were conducted at the end of the study to gather qualitative insights into participants' experiences with GPT scaffolding. GPT chat logs were automatically

archived during each session for subsequent qualitative analysis. To evaluate vocabulary knowledge gains across the three dimensions both within and between conditions, permutation tests were used to assess statistical significance, given the small sample size and non-normal distribution of the data. Wilcoxon signed-rank tests were used to estimate effect sizes for significant results. Monte Carlo simulations were conducted to assess the statistical power of the findings. Qualitative content analysis was conducted to identify patterns in GPT scaffolding strategies. In addition, thematic analysis of interview data was carried out to uncover key features of GPT scaffolding from the learner perspective.

Results

Within each CBLP condition, a decreasing trend in effect sizes was observed across the three vocabulary dimensions. In the condition without GPT scaffolding, the effect sizes were use ($z = 3.59, p < .001, r = 0.57$), form ($z = 3.22, p = .001, r = 0.51$), and meaning ($z = 2.15, p = .032, r = 0.34$). Similarly, in the GPT-supported condition, the effect sizes were use ($z = 3.54, p < .001, r = 0.56$), form ($z = 3.48, p < .001, r = 0.55$), and meaning ($z = 2.67, p = .007, r = 0.42$). For between-condition comparisons, the use dimension showed significant improvement in favor of the GPT-supported condition ($z = 2.44, p = .015, r = 0.39$). Effect sizes were interpreted following Plonsky and Oswald's (2014) benchmarks: $r = 0.25$ (small), $r = 0.40$ (medium), and $r = 0.60$ (large). No significant differences were found for form or meaning. Power analyses indicated that, except for meaning, there was sufficient power to detect significant effects for use and form within each condition, as well as between-condition differences in use. In terms of GPT scaffolding strategies, we observed that GPT adapted its responses based on the type of challenge learners faced. Thematic analysis of interview data showed that learners had an overall positive perception of the GPT scaffolding, frequently highlighting its support delivered in a responsive and tailored manner (e.g., "GPT can immediately identify where I'm struggling and provide timely guidance." [S17]).

Discussion

The improvement in the use dimension under the GPT-supported condition may be explained by the nature of corpus-based learning, which offers repeated, contextualized exposure to authentic language, aligning with usage-based theories (Ellis, 2012). GPT scaffolding likely reinforced this effect by

directing learners' attention to recurring patterns and providing targeted guidance, thereby activating input enhancement (Sharwood Smith, 1993), promoting noticing (Schmidt, 2001), and increasing involvement load (Laufer and Hulstijn, 2001). As for the form dimension, both conditions showed notable gains, but no significant between-condition differences emerged. This is likely because both groups repeatedly viewed concordance lines with bolded target word forms, which enhanced saliency and triggered implicit learning. Such input flooding (Szczudarski & Carter, 2016) may have already provided sufficient form-focused input, rendering the impact of GPT on this dimension less pronounced. The meaning dimension showed the smallest gains, probably because corpus tasks promote contextual inference that may divert attention from consolidating precise definitions (Tsai, 2019). By integrating quantitative and qualitative findings, the study revealed clear convergence. The custom GPT employed scaffolding strategies that offered contingent support aligned with learners' levels of understanding through learner-initiated interactions. This contingency appears to have contributed to the effectiveness of CBLP, particularly in supporting vocabulary use.

References

- Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI*, 7(1), 52–62. <https://doi.org/10.2139/ssrn.4337484>
- Boulton, A. (2009). Testing the limits of data-driven learning: Language proficiency and training. *ReCALL*, 21(1), 37–54. <https://doi.org/10.1017/S0958344009000068>
- Braun, S. (2005). From pedagogically relevant corpora to authentic language learning contents. *ReCALL*, 17(1), 47–64. <https://doi.org/10.1017/S0958344005000510>
- Cheng, X., & Liu, Y. (2022). Student engagement with teacher written feedback: Insights from low-proficiency and high-proficiency L2 learners. *System*, 109, 102880. <https://doi.org/10.1016/j.system.2022.102880>
- Ellis, N. C. (2012). Frequency effects. In P. Robinson (Ed.), *The Routledge encyclopedia of second language acquisition* (pp. 260–265). Routledge.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139524773>
- Karras, J. N. (2016). The effects of data-driven learning upon vocabulary acquisition for secondary international school students in Vietnam. *ReCALL*, 28(2), 166–186. <https://doi.org/10.1017/S0958344015000154>

- Kaur, J., & Hegelheimer, V. (2005). ESL students' use of concordance in the transfer of academic word knowledge: An exploratory study. *Computer Assisted Language Learning*, 18(4), 287–310. <https://doi.org/10.1080/09588220500280412>
- Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22(1), 1–26. <https://doi.org/10.1093/applin/22.1.1>
- Lee, H., Warschauer, M., & Lee, J. H. (2020). Toward the establishment of a data-driven learning model: Role of learner factors in corpus-based second language vocabulary learning. *The Modern Language Journal*, 104(2), 345–362. <https://doi.org/10.1111/modl.12634>
- Liu, J., & Ma, Q. (2025). Examining corpus-based language pedagogy (CBLP) practices in data-driven learning (DDL) for low-proficiency L2 English learners: A meta-analysis. *Educational Technology & Society*, 28(2), 53–76. [https://doi.org/10.30191/ETS.202504_28\(2\).RP04](https://doi.org/10.30191/ETS.202504_28(2).RP04)
- Ma, Q., Tang, J., & Lin, S. (2022). The development of corpus-based language pedagogy for TESOL teachers: A two-step training approach facilitated by online collaboration. *Computer Assisted Language Learning*, 35(9), 2731–2760. <https://doi.org/10.1080/09588221.2021.1895225>
- Nation, P. (2022). *Learning vocabulary in another language* (3rd ed.). Cambridge University Press.
- Nation, P., & Meara, P. (2010). Vocabulary. In N. Schmitt (Ed.), *An introduction to applied linguistics* (pp. 34–50). Hodder Education.
- Plonsky, L., & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, 64(4), 878–912. <https://doi.org/10.1111/lang.12079>
- Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge University Press.
- Sharwood Smith, M. (1993). Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition*, 15(2), 165–179. <https://doi.org/10.1017/S0272263100011943>
- Smart, J. (2014). The role of guided induction in paper-based data-driven learning. *ReCALL*, 26(2), 184–201. <https://doi.org/10.1017/S0958344014000081>
- Szudarski, P., & Carter, R. (2016). The role of input flood and input enhancement in EFL learners' acquisition of collocations. *International Journal of Applied Linguistics*, 26(2), 245–265. <https://doi.org/10.1111/ijal.12092>
- Tsai, K.-J. (2019). Corpora and dictionaries as learning aids: Inductive versus deductive approaches to constructing vocabulary knowledge. *Computer*

Assisted Language Learning, 32(8), 805–826. <https://doi.org/10.1080/09588221.2018.1527366>

- van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in Teacher–Student Interaction: A Decade of Research. *Educational Psychology Review*, 22(3), 271–296. <https://doi.org/10.1007/s10648-010-9127-6>

Vietnamese Pre-service Language Teachers' Technology-Mediated Task Creation: An Ecological Perspective

Paul J. Moore¹ , Giang Hong Nguyen²  and Quang Vinh Nguyen² 

¹The University of Queensland, Australia

²Hanoi University, Vietnam

Abstract

As an emerging context in CALL literature, Vietnam provides the international research community striking insights into attempts at CALL integration and its links to national educational development, from macro-level policy development to meso-and micro-level implementation. This qualitative study explores the major assessment task of a course “Integrating ICT in Language Teaching and Learning,” as well as lesson-plan data from a group of four pre-service teachers, to provide insights into the influence of macro-level policy on those who are really in charge of its implementation—trainee language teachers. Analysis of lesson plans and semi-structured interviews revealed themes of growth in technological capabilities, technology overload, simplification of the relationship between theory and practice, and external influences on pedagogy. We draw on these findings to argue for a holistic, contextualised, ecological approach to teacher training in CALL.

Keywords: Technology mediation, lesson planning, teacher training, Vietnam

Introduction

As with many countries across Asia, the Vietnamese government, through its National Foreign Languages Project (Government of Vietnam, 2017), is investing heavily in improving English language learning across all educational sectors, with a specific focus on the sustainable integration of technology (Ministry of Education & Training, 2014).

E-mail: p.moore5@uq.edu.au, giangnh@hanu.edu.vn & vinhnq@hanu.edu.vn

While much capital has been expended on policies at the macro-level, their aim is to influence the micro-level of classroom interaction (Moore et al., 2023). If policy and standards implementation is wholly a top-down endeavour, it risks irrelevance and ineffectiveness (Hubbard, 2023). What is needed is evidence from micro-level implementation as to the relevance and effectiveness of the government's investment (Nguyen et al., 2025).

In the context of an undergraduate course on “Integrating ICT in Language Teaching and Learning,” this paper explores the perspectives and practices of a group of four pre-service EFL teachers in their collaborative development of a technology-mediated lesson plan. We draw an ecological approach to SLA (van Lier, 2004) and CALL integration (Gruba, 2021) to investigate the contribution of a micro-level lesson-planning task to teachers' developing capabilities in technology integration.

Methodology

Participants

The study triangulates three sets of qualitative data from four undergraduate trainee English language teachers in Vietnam. Data were collected from an existing course; as such, the students who volunteered for the study represented a convenience sample (Cohen et al., 2018).

Data Collection and Analysis

The three sets of qualitative data used in the study were: (i) a collaboratively designed 90-minute lesson plan; (ii) a 900-word justification of the design; and (iii) individual interviews where participants were asked questions about theoretical, technological and pedagogical decisions made.

Content analysis (Cohen et al., 2018) was drawn on to analyse lesson plan documents (14 pages) and transcribed interviews (152 minutes). The aim was to provide insights into how learners perceived their own technology-mediated lesson planning capabilities, and how these perceptions triangulated with their actual lesson plans.

Results

Four major themes arose from the analysis: growth in technological capabilities; technology overload; the relationship between theory and practice; and external influences on pedagogy.

Theme 1: Growth in Technological Capabilities

All learners reported perceptions of growth in their technology-mediated lesson planning capabilities. These were represented as developments in the use of available technologies, awareness of theoretical and pedagogical approaches to integrating technologies (e.g., “access to a wider range of strategies” [P3], and feelings of improved *efficiency* [P1], *confidence* [P2], *flexibility* [P3], and *competence* [P4].

Theme 2: Technology Overload

The learners created an integrated skills (vocabulary-listening-speaking) lesson aimed at primary school children on the topic of “[my] favourite movie.” The content of the lesson, and its structure (warm-up, presentation, practice, and production), were based on the chosen coursebook (Bosiak & Kim, 2009).

Figure 1 provides an overview of the tools chosen.

As can be seen above, the learning management system for the class was Google Classroom, while the online lesson was to be conducted via Zoom. In addition to the functionality of Zoom (chat, breakout rooms, drawing tools), six external tools were chosen.

Noting the large number of tools used by the group, P2 stated “I realized that we applied a lot of online tools and apps in our lesson plans.” P4 provided an explanation for the group’s strong focus on (somewhat redundant) technological tools at the expense of a strong pedagogical approach: “I think we are learning how to integrate technologies into teaching ... so *I have to apply as many tools as possible* to be familiar with these tools” (italics added).

Theme 3: Relationship between Theory and Practice

The participants’ justification of their lesson plan revealed a simplistic direct relationship between theory and practice.

In their interviews all four identified discrete theories taught in the course, including behaviourism, cognitivism, constructivism, as well as theories specific to technology (“multimedia learning effects” [P4]) and culture (collectivism [P2]). They also noted that such theories were taught separately from practice. For example, P1 stated “in some lessons I only learned theory ... no application.”

In the written lesson plan justification, the participants explained the use of online quiz tools (e.g., *classkick.com*) as follows: “the teacher uses the stimulus and responses method ... based on the Behaviorist method.”

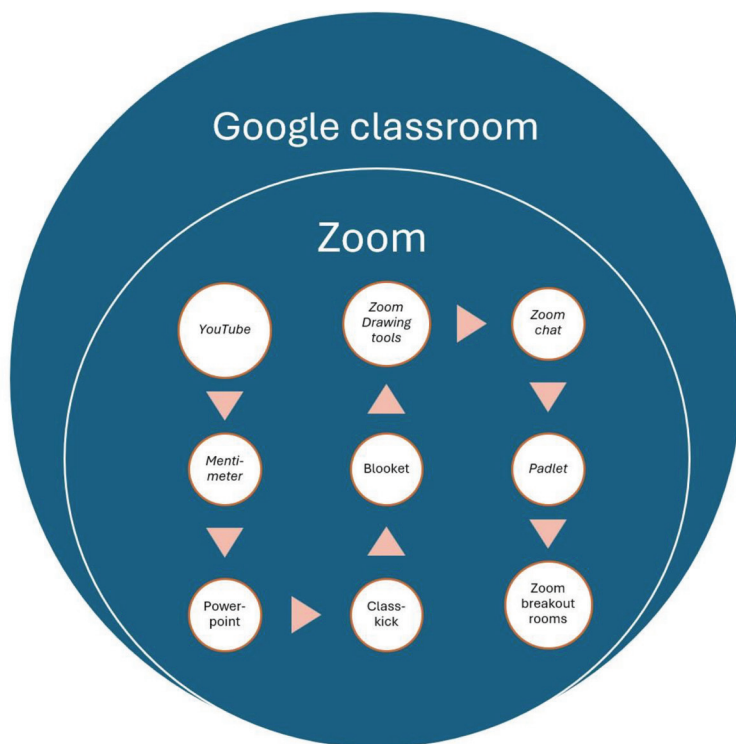


Figure 1. Technological tools used in the lesson plan

Theme 4: External Influences on Pedagogy

Three key influences on the group lesson planning project included: the CALL course text (Bower, 2017), which provided theoretical and pedagogical information for the participants' justification; the chosen language coursebook (Bosiak & Kim, 2009), which determined the language teaching approach and materials; and prior teacher training experience, including micro-teaching, which P1 explained as "traditional" (unsupported by technology), and others explained as less *interesting*, *interactive* and *collaborative* than what they felt could be achieved with technology-integrated learning.

Discussion and Conclusion

The analysis above has provided insight into pre-service Vietnamese language teachers' perspectives and capabilities with regard to technological integration. This research represents an effort to highlight micro-level implementation

of national policy, to inform the development of further pedagogical and policy activity at the meso- (institutional) and macro- (national) levels. The tenuous link between participants' perceptions and practices (Theme 1), and their strong focus on technologies at the expense of pedagogy (Theme 2) are common features of research into CALL professional development within Vietnam and beyond (Moore et al., 2023).

Similarly, the somewhat superficial link between existing materials—specifically professional development texts and language learning coursebooks—and the participants' lesson plan and justification (Theme 3), highlights the necessary role of a critical, evidence-based, approach to existing theory, research and pedagogy (Levy & Moore, 2018), as well as a principled approach to integrating pedagogical approaches and technologies (Bax, 2003).

Finally, the fact that participants' approaches were influenced by their prior experience and other contextual data (Theme 4) provides evidence for the necessity of a holistic, contextualised, and ecological approach to research into the developing technological capabilities of pre-service language teachers.

References

- Bax, S. (2003) CALL – past, present and future. *System*, 3(1), 13–28. [https://doi.org/10.1016/S0346-251X\(02\)00071-4](https://doi.org/10.1016/S0346-251X(02)00071-4)
- Bosiak, I., & Kim, V. (2009) *Step by step listening 3*. e-future. <https://www.efuture-elt.com/books/view.do?bookkey=491>
- Bower, M. (2017). *Design of technology-enhanced learning: Integrating research and practice*. Emerald.
- Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). Routledge.
- Government of Vietnam. (2017). *Decision No. 2080/QĐ-TTg on approval of adjustment and addition in foreign languages project in the national education, Period 2017–2025*. Government of Vietnam. <https://thuvienphapluat.vn/van-ban/Giao-duc/Quyét-dinh-2080-QĐ-TTg-2017-bo-sung-De-an-day-hoc-ngoai-ngu-trong-he-thong-giao-duc-quoc-dan-370658.aspx>
- Gruba, P. (2021, December 10). *Leading the integration of technology* [Keynote Speech]. Technology Standards in Foreign Language Teacher Education in Asia Symposium, Brisbane & Hanoi.
- Hubbard, P. (2023). Contextualizing and adapting teacher education and professional development. In D. Tafazoli & M. Picard (Eds.), *Handbook of CALL teacher education and professional development: Voices from under-represented contexts* (pp. 3–14). Springer Nature.

- Levy, M. & Moore, P. J. (2017). 'Language and technology': Theory and practice, options and issues in computer-assisted language learning. In Son, J.-B. and Windeatt, S. (Eds.), *Language teacher education and technology: Approaches and practices*. Bloomsbury.
- Ministry of Education and Training. (2014). *Official letter 792/BGDĐT-NGCBQLGD-Guidelines for the implementation of the basic requirements for general English teachers' competencies*. MOET. <https://thuvienphapluat.vn/cong-van/Lao-dong-Tien-luong/Cong-van-792-BGDĐT-NGCBQLGD-nam-2014-Yeu-cau-co-ban-nang-luc-giao-vien-tieng-Anh-pho-thong-221908.aspx>
- Moore, P. J., Nguyen, G. H., & Nguyen, Q. V. (2023). Professional standards in CALL teacher training in Vietnam: Towards an ecological approach to CALL integration. In D. Tafazoli & M. Picard (Eds.), *Handbook of CALL teacher education and professional development: Voices from under-represented contexts* (pp. 463–481). Springer Nature.
- Nguyen, Q. V., Moore, P. J., & Nguyen, G. H. (2025). Making the case for micro-level leadership in CALL integration in Vietnam. In L. McCallum, & D. Tafazoli (Eds.), *Computer-assisted Language Learning in the Global South* (pp. 147–163). Routledge.
- van Lier, L. (2004). *The ecology and semiotics of language learning: A sociocultural perspective*. Kluwer Academic Publishers.

Beyond the Game: The Theoretical Exploration of Gamification in Foreign Language Teaching

Nguyen Thi La 

National Academy of Education Management (NAEM),
Vietnam.

Abstract

Increasingly common in education, gamification is the use of game-like features in non-game environments to foster involvement. Particularly for underperforming students, this paper investigates how gamification increases motivation, lowers anxiety, and promotes inclusivity in foreign language education. Through a qualitative analysis of literature from 2010 to 2024, the findings show that gamified activities can enhance learner engagement, foster self-efficacy, and create supportive learning environments. Nonetheless, challenges such as over-reliance on extrinsic rewards, digital divide, and cognitive overload highlight the need for careful design. The paper underscores the educational potential of gamification while calling for further research to ensure sustainable, learner-centered implementation.

Keywords: gamification, foreign language teaching, motivation, inclusivity, self-efficacy, language anxiety

Background of the Study

In the digital era, a fascinating educational tool has been the inclusion of game-design components into non-game settings. Gamification, as Deterding et al. (2011) define, aims to use the motivational affordances of games to boost user involvement in daily tasks, including education. Gamification offers a

E-mail: tobe31082@gmail.com

promising complement to conventional approaches in foreign language education, where students may experience demotivation, performance anxiety, and memory challenges (Pinto et al., 2021). Though traditional classroom teaching may be effective for high-achieving students, it is no longer capable of fulfilling the varied needs of learners, especially underperforming students (Hu, 2024). Often struggling with grammar acquisition, vocabulary retention, and low confidence, these students hinder their academic progress (Thuan, 2023).

While gamification in language learning has been widely explored, several important questions remain. First, there is a limited understanding of how gamification works in practice to boost motivation and reduce language anxiety, particularly for learners who struggle with confidence and engagement. Second, the potential of personalized game elements to strengthen self-efficacy and sustain active participation in the classroom has not been thoroughly examined. Third, while gamification is often praised for its engaging nature, its role in fostering inclusivity, ensuring that learners with different styles, needs, and proficiency levels have equal opportunities to participate and succeed, has yet to be explored in depth. By focusing on these areas, this study aims to provide a more precise and more practical understanding that carefully considered use of gamified language learning changes the classroom into a vibrant, learner-centered environment.

Methodology

This paper employs a qualitative literature review methodology, synthesizing findings from peer-reviewed journal articles, case studies, and meta-analyses published between 2010 and 2024 from reputable academic databases such as Google Scholar, Scopus, and Web of Science.

The 2010–2024 timeframe was chosen because gamification began to attract substantial scholarly attention in education around 2010. The selected studies were located in digital education, applied linguistics, and educational psychology databases, using key terms such as “gamification and language learning”, “gamified learning and foreign language teaching”, “game-based learning and EFL/ESL”, and “gamification and motivation/anxiety/self-efficacy”.

The inclusion criteria required that studies focus on gamification in foreign language teaching and learning, address psychological factors, including motivation, language anxiety, self-efficacy, or pedagogical considerations, and be peer-reviewed. Studies were excluded if they focused solely on non-educational gaming, lacked scholarly analysis, or were not written in English or Vietnamese.

Findings and Discussion

Gamification – A Catalyst for Motivation in Language Learning

Language development depends on motivation, influencing students' tenacity and work in acquiring new language tools. Deci and Ryan's (2000) Self-Determination Theory holds that motivation develops when students feel autonomy, competence, and relatedness. Gamified systems are particularly well qualified to satisfy these criteria. For example, while customizable avatars and peer competition encourage independence and connection, points and badges strengthen competence (Kao et al., 2018). From a teaching perspective, educators can strategically integrate these mechanics into lesson plans by setting incremental goals, using leaderboards to foster friendly competition, or designing thematic quests that align with curriculum objectives. In some digital platforms like Duolingo and Memrise, which show how gamification increases motivation, students do brief, engaging activities, get immediate feedback, and earn virtual reward components that promote a feeling of accomplishment and development (Dichev et al., 2015; Munday, 2016). For underperforming students who might find it challenging to remain motivated through conventional drills or textbook exercises, teachers can redesign tasks into game-like challenges to sustain engagement throughout the learning process.

Addressing Language Learning Anxiety through Game-Based Strategies

Language anxiety is a barrier to language learning, especially in speaking and writing (Horwitz et al., 1986). Students' fears of being judged or making mistakes often lead them to avoid practicing, stalling their fluency development. Learners can practice trying out different words and techniques without the risk of being ridiculed in gamified learning environments, which are much more supportive. Several researchers have pointed towards the existence of a positive correlation between gamification and reduced performance anxiety due to the repeated exposure, safe opportunities provided, and positive feedback (Judkins, 2018; Reinhardt & Sykes, 2014; Vallejo Balduque, 2018). Therefore, instructors can implement role-play simulations, timed word challenges, or collaborative storytelling games where learners contribute without fear of public embarrassment. Teachers also play a crucial role in framing errors as part of the learning journey, using gamified feedback loops to normalize risk-taking in language use.

Enhancing Self-Efficacy through Personalized Gamified Learning

Personalization is another advantage of gamified learning. Several platforms incorporate algorithms to modify content difficulty based on the learner's level, making sure that the content is appropriately challenging. Self-adaptive learning supports the self-efficacy belief of one's ability to succeed, which is particularly important for low-performing students (Locke, 1997). For example, in a traditional classroom setting, feeling disheartened and detached from the learning process is very common for learners who are unable to apply their skills and knowledge. In a gamified system, however, every little win, whether it is getting past a level or getting new content, enables teachers to celebrate incremental progress, thereby reinforcing their students' feelings of mastery. As noted by Laksanasut (2025), the continuous feedback and goal-oriented structure characteristic of the gamified environment build learners' confidence over time, which encourages greater persistence and resilience.

Promoting Inclusive and Engaging Learning Environments

In addition, inclusivity is fostered through countless avenues of achievement. Students from diverse learning paradigms, such as visual, auditory, and kinesthetic, can take advantage of the gamified tasks as noted by Yan et al. (2022). A student who identifies as a "Visual learner" can do image-based vocabulary games, while a student who identifies as an "Auditory learner" can do listening-based challenges. Furthermore, teachers can create mixed-modality gamified tasks, facilitate team-based competitions to encourage collaboration, and assign rotating roles to ensure every student contributes. This equity framework mitigates the performance divide while ensuring that all students have access to equal opportunities for learning, especially students who need the most support in traditional educational environments.

Challenges and Considerations

Despite its many pedagogical benefits, gamification has some concerns in foreign language education. One major issue is the reliance on extrinsic motivation, which could lower the learners' intrinsic motivation, causing disengagement with content at a lower level (Corpus & Wormington, 2014). In addition, implementing gamified elements, including lesson planning, instructional technology, and teaching expertise, is time-consuming (Baalsrud Hauge et al., 2015). This may be professionally unmanageable for some teachers who

lack training or institutional support. Digital inequality also emerges wherein unequal access to technology can unfairly disadvantage some learners, particularly those from underserved backgrounds (Tan et al., 2023).

Additionally, mismanaged gamified activities may lead to cognitive overload or distract from learning goals to unwanted tasks. Moreover, prior cultural attitudes, along with personal preferences towards learning, are likely to affect the degree of support and success, as some students may deem a game-based approach as trivial or unwarranted in academic settings. Therefore, effective gamification should result from careful planning, guided implementation, and clear justification to maintain credibility while balancing universality and diversity.

Conclusion

Gamification holds significant promise for revolutionizing foreign language education. By enhancing motivation, reducing anxiety, and personalizing learning experiences, gamified approaches create a more inclusive and engaging environment for all learners, especially those who struggle in traditional classrooms. As supported by the literature, gamification can bridge the gap between underperforming and high-achieving students by making language learning more accessible, enjoyable, and effective.

Educators and researchers should continue to explore and refine gamified methods, ensuring they serve diverse student populations. With thoughtful design and implementation, gamification can go beyond the game to truly revolutionize the way we teach and learn languages.

This article is based solely on English- and Vietnamese-language sources and lacks empirical data. As a theoretical review, its findings may not fully capture the nuances of different cultural or institutional contexts. To provide a more comprehensive and context-sensitive understanding of gamification in language education, future research should integrate empirical studies and multilingual sources.

References

- Baalsrud Hauge, J. M., Stănescu, I. A., Carvalho, M. B., Stefan, A., Banica, M., & Lim, T. (2015, August). Integrating gamification in mechanical engineering systems to support knowledge processes. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* (Vol. 57052, p. V01BT02A015). American Society of Mechanical Engineers.

- Corpus, J. H., & Wormington, S. V. (2014). Profiles of intrinsic and extrinsic motivations in elementary school: A longitudinal analysis. *The Journal of Experimental Education*, 82(4), 480–501. <https://doi.org/10.1080/00220973.2013.876225>
- Deci, E. L., & Ryan, R. M. (2000). The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01
- Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining “gamification”. In *Proceedings of the 15th International Academic MindTrek Conference* (pp. 9–15). ACM. <https://doi.org/10.1145/2181037.2181040>
- Dichev, C., Dicheva, D., Angelova, G., & Agre, G. (2015). From gamification to gameful design and gameful experience in learning. *Cybernetics and Information Technologies*, 14(4), 80–100. <https://doi.org/10.1515/cait-2014-0007>
- Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, 70(2), 125–132. <https://doi.org/10.1111/j.1540-4781.1986.tb05256.x>
- Hu, J. (2024). The challenge of traditional teaching approach: A study on the path to improve classroom teaching effectiveness based on secondary school students’ psychology. *Lecture Notes in Education Psychology and Public Media*, 50(1), 213–219. <https://doi.org/10.54254/2753-7048/50/20240945>
- Judkins, B. N. (2018). *The Effect of Gamification on Anxiety and Motivation in Algebra* (Doctoral dissertation, Northcentral University).
- Kao, D., & Harrell, D. F. (2018, April). The effects of badges and avatar identification on play and making in educational games. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1–19).
- Laksanasut, S. (2025). Gamification in ESL/EFL education: Transforming language learning and teaching through play. *TESOL and Technology Studies*, 6(1), 16–29. <https://doi.org/10.48185/tts.v6i1.1562>
- Locke, E. A. (1997). Self-efficacy: The exercise of control. *Personnel Psychology*, 50(3), 801.
- Munday, P. (2016). The case for using Duolingo as part of the language classroom experience. *RIED: Revista Iberoamericana de Educación a Distancia*, 19(1), 83–101. <https://doi.org/10.5944/ried.19.1.14581>
- Pinto, R. D., Peixoto, B., Melo, M., Cabral, L., & Bessa, M. (2021). Foreign language learning gamification using virtual reality—A systematic review of empirical research. *Education Sciences*, 11(5), 222. <https://doi.org/10.3390/educsci11050222>

- Reinhardt, J., & Sykes, J. (2014). Special issue commentary: Digital game and play activity in L2 teaching and learning. *Language Learning & Technology*, 18(2), 2–8. <http://lt.msu.edu/issues/june2014/commentary.pdf>
- Tan, A. R., Kilinc, M., Yildiz, M., & Balta, M. (2023). Equity and inclusion in remote education: Unearthing disparities, interrogating strategies. *SSD Journal*, 8(39), 21–27. <https://doi.org/10.31567/ssd.1012>
- Thuan, P. K. (2023). Reviewing the significance of practice in learning English as a second language. *Journal of Knowledge Learning and Science Technology*, 2(2), 62–67. <https://doi.org/10.60087/jklst.vol2.n2.p.67>
- Vallejo Balduque, B. (2018). Gamification and the affective aspects EFL students: Effects on anxiety and motivation. *Language Education and Emotions*.
- Yan, H., Zhang, H., Su, S., Lam, J. F., & Wei, X. (2022). Exploring the online gamified learning intentions of college students: A technology-learning behavior acceptance model. *Applied Sciences*, 12(24), Article 12966. <https://doi.org/10.3390/app122412966>

Leveraging the Affordances of Artificial Intelligence Chatbots to Create Differentiated Instructions and Cater to Learners' Differences in an English as a Foreign Language Class

Trang Thi Thuy Nguyen¹  and Phuc Kim Nguyen² 

¹University of Languages and International Studies (ULIS),
Vietnam National University (VNU), Hanoi, Vietnam

²Faculty of Foreign Languages, Hanoi University of Science and
Technology, Hanoi, Vietnam

Abstract

Given the rapid evolution of artificial intelligence (AI), a growing body of literature has investigated the potential benefits it can bring for education in general and language learning in particular. Nonetheless, little research has looked into how to leverage the affordances of AI chatbots to create differentiated instruction (DI) and cater to English as a foreign language (EFL) learners' differences. Thus, we conducted this action research to shed light on (1) learners' perceptions about the affordances of AI chatbots as a tool for differentiated instruction, (2) perceived challenges of integrating AI chatbots in EFL classrooms for differentiated learning, and (3) stakeholders' suggestions to leverage AI chatbots to implement differentiated instruction effectively. This study follows a mixed-method approach, with three main data collection instruments, namely a survey of the students, a follow-up interview with the students, and the teachers' reflective journals. Thematic analysis was used to analyze data from the interviews and reflective journals, while quantitative analysis was used for the surveys. Data analysis reveals the students' and teachers' generally positive perceptions about a variety of affordances provided by AI chatbots, such as quick, personalized feedback, and multimodal input. However, different students perceive some of the affordances differently,

E-mail: ntttrang.vnu@ulis.vnu.edu.vn & phuc.nguyenkim@hust.edu.vn

either as opportunities or difficulties for learning, depending on their learning styles, and most importantly, AI readiness level. Some specific challenges are unearthed, together with the corresponding solutions.

Keywords: AI chatbots, differentiated instructions, learners' differences, affordances of AI

Background

Despite the potential of artificial intelligence (AI) chatbots for education (Alazemi, 2024), little is known about English as a foreign language (EFL) teachers' and learners' perspectives on how AI can be leveraged to support differentiated instruction (DI), which involves adapting the content, process, product and environment to match students' readiness, interests and learning profiles (Tomlinson, 2017).

According to Kamalov et al. (2023) and Roe et al. (2025), AI tools have the potential to help teachers adapt content and activities in real time, support students of different academic backgrounds, tier tasks and help EFL teachers design materials at different difficulty levels.

Hence, through the lens of Tim et al.'s (2017) Technology Affordances Framework (with three stages: affordance perception, actualization and outcomes), this study investigates the use of AI to support content differentiation in EFL. The research questions are:

1. *What are the learners' and teachers' perceptions regarding the affordances of AI chatbots as a tool for DI?*
2. *What are learners' and teachers' perceived challenges of integrating AI chatbots in EFL classrooms for differentiated learning?*
3. *What are learners' and teachers' suggestions for leveraging AI chatbots to implement DI effectively?*

Research Methodology

This action research adopted a mixed-method design, allowing for robust data triangulation through three main data collection instruments, namely a survey of the students, a follow-up interview with the students, and the teachers' reflective journals. Tim et al.'s (2017) Technology Affordances Framework was used to design items in the survey and interviews.

The study was conducted in an EFL class with 15 students (B1-B2 level) at a language center in Vietnam. The two major AI chatbots used for content differentiation in the intervention phase are ChatGPT and NotebookLM.

Thematic analysis was conducted with qualitative data from the interviews and teaching journals, whereas descriptive statistics were used for quantitative data from the survey.

Findings and Discussion

Perceived Affordances of AI for DI

Data from the teaching journals reveals the tremendous potential benefits of AI for DI. These affordances involve supporting teachers in modifying and adjusting learning materials to learners' needs, designing learning activities to suit students at different levels, and creating multimodal learning materials that not only match but also stretch students' learning styles and increase flexibility (Gregersen & MacIntyre, 2014). For instance, teachers could use AI chatbots to create mindmaps, an audio summary of the lesson (via Notebook LM), vocabulary worksheets or supplementary learning materials (ChatGPT) to activate multiple senses, and support students at lower levels in the class.

The student survey and interview data also illuminate students' perceived importance of the various affordances offered by AI to facilitate DI.

Table 1 indicates that the most important perceived affordance is supporting mixed-level classes (highest mean: 4.86). Discussion functionality ranks second (mean: 4.71; standard deviation (SD): 0.71; small variation). The affordance of creating AI-assisted multimodal input (text, audio, images, etc.) which can match and stretch students' learning styles is also perceived to be essential (mean: 4.0) but with very wide variation (SD: 1.31), suggesting the influence of students varied learning style preferences on their perceptions of the importance of this multimodal affordance.

Data from the interview also confirms that multimodal input designed with the aid of AI can "cater to various learning styles, thus fostering engagement and knowledge retention" (Student 5).

Table 2 illustrates how students act upon the affordances of AI for DI. The affordance that the students often take full advantage of the most is the opportunity to discuss different questions regarding EFL lessons with AI chatbots regularly to understand the knowledge more deeply, with the highest mean (5.71), and the lowest variation (SD: 0.45).

The interview data shed more light on students' positive evaluation of this discussion functionality. As Student 6 stated, "Students who are shy or afraid to ask questions in class can feel more confident when learning with a chatbot. It's like having a personal tutor who is always ready to help." This reflects the possible influence of collectivist culture in Vietnamese educational contexts,

Table 1. Perceived importance of AI affordances for DI

Opportunities	Minimum	Maximum	Median	Mean	Standard Deviation
Opportunities to access quick, personalized AI-assisted feedback for speaking/ writing	3.0	5.0	4.0	3.86	0.64
Opportunities to learn with AI-assisted multimodal input (text, audio, images, etc.) which can match and stretch students' learning styles	1.0	5.0	4.0	4.0	1.31
Opportunities to study with learning materials designed for students at different levels in mixed-level classes	3.0	6.0	5.0	4.86	0.99
Opportunities to discuss different questions regarding EFL lessons with AI chatbots to understand the knowledge more deeply	4.0	6.0	5.0	4.71	0.7

where students may feel anxious about questioning teachers in class, being afraid of making mistakes in front of others (Bao & Nguyen, 2020). AI chatbots can become a comfortable learning environment to help students alleviate anxiety and help them feel free to ask questions.

Interestingly, the students perceive and act upon AI affordances for DI in different ways. The interview data shows that some students recognize various learning opportunities, but others may only recognize a limited number of affordances. Moreover, most students said they made use of AI for personalized learning, but others who are more used to using textbooks tend to prefer learning in traditional ways with textbooks. This may be due to the discrepancy in students' AI readiness (Li & Liang, 2025), including their knowledge of AI affordances and their willingness and creativity in interacting with AI tools for individual learning purposes.

Table 2. How students act upon the affordances of AI for DI

Opportunities	Minimum	Maximum	Median	Mean	Standard Deviation
I take full advantage of opportunities to access quick, personalized AI-assisted feedback for speaking/ writing by practising speaking and writing with the help of AI regularly.	4.0	5.0	5.0	4.71	0.45
I take full advantage of opportunities to learn with AI-assisted multimodal input (text, audio, images, etc.) by regularly using the AI-assisted multimodal learning materials to foster my memory and comprehension.	2.0	6.0	4.0	4.29	1.16
I take full advantage of opportunities to study with learning materials designed to suit my language proficiency by regularly using the activities suggested for my level.	4.0	6.0	5.0	5.29	0.7
I take full advantage of opportunities to discuss different questions regarding EFL lessons with AI chatbots regularly to understand the knowledge more deeply.	5.0	6.0	6.0	5.71	0.45

Potential challenges in using AI for DI

For the teachers, three major challenges emerged in the implementation process. The first challenge is the **requirements for profound AI-Technological Pedagogical Content Knowledge (AI-TPACK)** (Ning et al., 2024),

especially the pedagogical knowledge to select and adapt the materials suggested by AI. In the teaching journals, the teachers' reflections emphasize that even with the support of multiple AI tools that assist teachers in designing DI materials, if teachers do not have the necessary pedagogical knowledge to critically select and adapt those materials for learning, the application may still face various hindrances. Teachers may even feel even confused with the overwhelming amount of data and suggestions from AI. The second challenge is the **requirement to have an in-depth understanding of the students' level and their needs**. For instance, the teaching journals reveal that if teachers design multimodal supplementary materials for students to foster their understanding and memory, it is crucial to take students' levels and individual needs into account. If the AI-generated audio summary is too fast for students to hear, for instance, the DI activity may still fail. Finally, **time constraints** are of no less importance. It is quite time-consuming to design differentiated learning materials, even with the support of AI.

For the students, several difficulties were mentioned in the survey and interviews. Notably, 100% of the students perceived overdependence on AI as a common risk of using AI in learning and DI. Moreover, the risks of inaccurate or misleading AI-generated information, and of students being deskilled when just copying answers without sharpening their skills, also attracted the concern of the majority of the students.

Suggestions for Best Practices of AI-assisted DI

The teachers' reflections unveil several suggestions for best practices. First, teachers should start with a good textbook and then develop supplementary DI materials, rather than building materials from scratch to save time and ensure quality. Second, it is fundamental for teachers to equip themselves with profound AI-TPACK, especially pedagogical knowledge, since AI cannot replace teachers' thinking. Teachers should be the final ones who decide on the content or process of DI lessons. Third, teachers are advised to identify their students' specific needs to create a baseline for designing appropriate AI-assisted learning materials. Finally, to save time, teachers should be selective instead of adapting the whole lesson; it is better to design DI components for specific important sections of a lesson.

Students also suggested that teachers should use AI chatbots as virtual teaching assistants that deliver customized learning materials, immediate responses to student-led inquiry and personalized feedback, allowing educators to focus on higher-level instruction while ensuring every student receives appropriate support.

Conclusion

Overall, both teachers and students perceived various affordances that AI chatbots bring to facilitate DI in EFL, including designing different learning materials and personalizing feedback. However, the effective application of AI affordances in DI depends heavily on teachers' sufficient AI-TPACK and students' AI readiness. Notable suggestions involve teachers having a clear understanding of the students' needs, determining the important content of each lesson to design DI practically and instructing students to actively utilize AI chatbots as tutoring tools.

As the study focuses solely on DI regarding content, such aspects as the process and product have not been addressed adequately. Future research should explore how AI tools can support DI in learning processes and products in EFL.

References

- Alazemi, A. F. T. (2024). Formative assessment in artificial integrated instruction: delving into the effects on reading comprehension progress, online academic enjoyment, personal best goals, and academic mindfulness. *Language Testing in Asia*, 14, 44. <https://doi.org/10.1186/s40468-024-00319-8>
- Bao, D., & Nguyen, T.-M. (2020). How silence facilitates verbal participation. *English Language Teaching Educational Journal*, 3(3), 3. <https://doi.org/10.12928/eltej.v3i3.3004>
- Gregersen, T., & MacIntyre, P. D. (2014). *Capitalizing on language learners' individuality: From premise to practice*. Multilingual Matters & Channel View Publications. <https://doi.org/10.2307/jj.26931978>
- Kamalov, F., Calong, D. S. & Gurrib, I. (2023). *New era of artificial intelligence in education: towards a sustainable multifaceted revolution*. arXiv.org. <https://arxiv.org/abs/2305.18303>
- Li, N., & Liang, Y. (2025). Teachers' AI readiness in Chinese as a foreign language education: Scale development and validation. *System*, 129, 103597. <https://doi.org/10.1016/j.system.2025.103597>
- Ning, Y., Zhang, C., Xu, B., Zhou, Y., & Wijaya, T. T. (2024). Teachers' AI-TPACK: Exploring the relationship between knowledge elements. *Sustainability*, 16(3), 3. <https://doi.org/10.3390/su16030978>
- Roe, J., Perkins, M. & Furze, L. (2025). *From assessment to practice: Implementing the ALIAS framework in EFL teaching and learning*. arXiv.org. <https://doi.org/10.48550/arXiv.2501.00964>

- Tim, Y., Pan, S. L., Bahri, S. & Fauzi, A. (2017). Digitally enabled affordances for community-driven environmental movement in rural Malaysia. *Information Systems Journal*, 28(1), 48-75. <https://doi.org/10.1111/isj.12140>
- Tomlinson, C. A. (2017). *How to differentiate instruction in academically diverse classrooms*. ASCD.

Using Dialogic Multimodal Feedback to Enhance University EFL Students' Writing: Affordances for Language Learning Styles

Nguyen Thi Thuy Trang¹ , Edwin Creely²  and
Huynh Anh Tuan¹ 

¹University of Languages and International Studies (ULIS),
Vietnam National University (VNU), Hanoi, Vietnam

²School of Curriculum, Teaching and Inclusive Education,
Faculty of Education, Monash University, Australia

Abstract

This study examines the implementation of dialogic multimodal feedback (DMF) in English as a Foreign Language (EFL) writing instruction at a Vietnamese university. Using a mixed-methods approach, the research investigated stakeholder perceptions of DMF affordances and their impact on students with diverse learning styles. Results revealed that DMF offers technical, pedagogical, and social affordances that enhance writing development across various learning preferences. While students and teachers showed positive attitudes toward DMF compared to conventional text feedback, the study highlighted the importance of individual differences in perceiving and engaging with feedback affordances. These findings suggest that DMF can effectively support inclusive writing pedagogy when implemented thoughtfully.

Keywords: Dialogic multimodal feedback, EFL writing, affordances, learning styles

E-mail: ntttrang.vnu@ulis.vnu.edu.vn, huynhanhtuan@vnu.edu.vn &
edwin.creely@monash.edu

Background

English as a Foreign Language (EFL) writing instruction faces ongoing challenges in providing effective feedback that accommodates diverse learner needs and preferences. Conventional text feedback, while valuable, often fails to engage students meaningfully or address different learning styles comprehensively (possibly due to the potential lack of clarity (Penn & Brown, 2022) and unimodal approach).

Despite growing interest in video feedback (Pachuashvili, 2024), most implementations remain monologic, treating students as passive recipients rather than active participants in knowledge construction (Mahoney et al., 2019). This limitation has created a research gap for the emergence of dialogic multimodal feedback (DMF) that combines multiple information sources while enabling student-teacher dialogue and discussion.

Additionally, limited research exists on how students with different learning styles perceive and benefit from multimodal feedback approaches.

The current study addresses these gaps by investigating DMF implementation in Vietnamese EFL contexts, where cultural factors such as Confucian ideals may influence student participation in feedback processes.

The research aims to understand both the affordances (learning opportunities [Williams & Mercer, 2016]) that DMF provides and how learners with varying learning styles experience and benefit from this feedback modality through the lens of Affordance Theory (Gibson, 1979) Specifically, the study aims to answer the following research questions:

- *What perceived affordances does DMF offer for supporting EFL university students' writing development across various learning styles?*
- *How do learners with different learning styles perceive the impacts of DMF on their writing skills?*

Methodology

This mixed-methods action research study employed a pragmatism paradigm to investigate DMF implementation in an EFL writing course at a Vietnamese university. The study involved 39 EFL students enrolled in a writing-focused English course during one semester. The research design nested quantitative survey data within a larger qualitative framework to provide comprehensive insights into participant experiences and perceptions.

Loom was chosen as the main platform for recording and delivering DMF. The lecturer recorded multimodal video feedback in Loom, sent the video links

to students so that the students could interact with the feedback by directly commenting on the videos, and receiving lecture responses in Loom platform.

The primary data collection instruments included student surveys measuring perceptions of DMF affordances and effectiveness, semi-structured interviews exploring students' experiences in depth, and reflective teaching journals maintained by the instructor throughout the implementation period.

Qualitative data from interviews and teaching journals underwent thematic analysis to identify key patterns and themes, while quantitative survey data received descriptive statistical analysis. This triangulated approach ensured a robust investigation of the research questions from multiple perspectives and data sources.

Results

Perceived Affordances of DMF

The students' survey revealed a comprehensive range of perceived important affordances that students and teachers attributed to DMF implementation (Table 1). These affordances clustered into three main categories: technical, pedagogical, and social opportunities for learning enhancement.

Technical affordances included replay functionality and transcript availability that enabled personalized learning experiences. Students particularly valued the ability to watch feedback videos multiple times at their preferred pace and location, with scores indicating high importance (Mean: 4.97; SD: 1.02).

Pedagogical affordances encompassed the multimodal nature of feedback that could accommodate different learning style preferences through varied information sources including audio, visual, and textual elements. Interviewees reported that this combination helped avoid misunderstandings common with brief text-only feedback, where unclear symbols or minimal comments often left learners confused about instructors' intentions.

Social affordances emerged as equally significant (Mean: 4.83; SD: 1.04), particularly the discussion function that enabled students to ask clarifying questions and engage in dialogue with instructors. This dialogic feature received high importance ratings from students and was consistently emphasized during interviews as transformative for their learning experience.

Learning Style Accommodation and Flexibility

Analysis revealed that over 93% of students agreed that DMF provided learning opportunities that either matched their preferred learning styles or stretched

Table 1. Perceived importance of the affordances of DMF

Perceived Affordances of DMF	Proportion						Other Descriptive Statistics	
	Not Provided by IVF	Provided, But Not Important	Provided & Slightly Important	Provided & Important	Provided & Very Important	Provided & Extremely Important	Mean (Out of 6)	SD
1. Opportunities to absorb knowledge more deeply and easily from multimodal resources (with the aid of multiple types of information: audio, text, gestures, etc.)	0.00%	0.00%	16.67%	40.00%	36.67%	6.67%	4.33	0.83
2. Learning opportunities that match learners' preferred language learning styles (with the aid of multiple types of information: audio, text, gestures, etc.)	0.00%	6.67%	6.67%	26.67%	46.67%	13.33%	4.53	1.02
3. Learning opportunities that stretch learners' language learning styles (with the aid of multiple types of information: audio, text, gestures, etc.), hence improving learning flexibility	0.00%	3.33%	16.67%	33.33%	26.67%	20.00%	4.43	1.09
4. Opportunities to improve memory retention by combining visual and auditory stimuli.	0.00%	3.33%	13.33%	30.00%	36.67%	16.67%	4.5	1.02
5. Opportunities to make the learning process more interactive and engaging compared to traditional written feedback	0.00%	0.00%	16.67%	26.67%	30.00%	26.67%	4.67	1.04
6. Opportunities to ask questions and have discussions regarding lectures' comments to clarify information	0.00%	3.33%	10.00%	13.33%	46.67%	26.67%	4.83	1.04
7. Opportunities to connect with lectures emotionally (via reaction to the feedback video or engagement with teachers' voice, tone, or non-verbal language, etc.)	0.00%	3.33%	20.00%	36.67%	23.33%	16.67%	4.3	1.07
8. Opportunities to watch the feedback video as many times as possible, at any speed and space	0.00%	0.00%	10.00%	23.33%	26.67%	40.00%	4.97	1.02
9. Opportunities to use the script if needed	0.00%	6.67%	16.67%	23.33%	26.67%	26.67%	4.5	1.23
10. Opportunities to reflect and make revision on the first draft of the writing	0.00%	0.00%	16.67%	23.33%	30.00%	30.00%	4.73	1.06
11. More time for reflection before having questions compared to face-to-face oral feedback	0.00%	3.33%	10.00%	20.00%	40.00%	26.67%	4.77	1.05

their capabilities to improve learning flexibility (Gregersen & MacIntyre, 2014). Audio-visual learners specifically highlighted the significance of combining multiple information modalities to support their comprehension and retention.

Cultural and Emotional Considerations

An unexpected finding concerned the emotional support that DMF provided, particularly relevant in the Vietnamese cultural context. Students reported reduced anxiety compared to face-to-face feedback situations, attributing this to the private nature of online interactions that allowed them to process information calmly and ask questions without fear of public embarrassment.

This finding reflects the influence of collectivist culture and the lingering impacts of Confucian ideals on Vietnamese students. Despite recent cultural shifts towards student-centredness, open dialogue, and progressive changes in student-teacher power relations (Nguyen et al., 2025), to some Vietnamese students, questioning teachers publicly may be perceived as disrespectful (Bao & Nguyen, 2020). DMF provided a comfortable platform for active engagement while maintaining cultural sensitivity and encouraging student agency in the feedback process.

Comparative Effectiveness and Preferences

Figure 1 below shows that students strongly favoured DMF over conventional approaches, with 80% of participants selecting it as their preferred feedback modality. Comparative analysis revealed that students perceived DMF as more detailed, user-friendly, and effective than traditional text feedback, with particularly strong ratings for accommodating diverse learning style preferences.

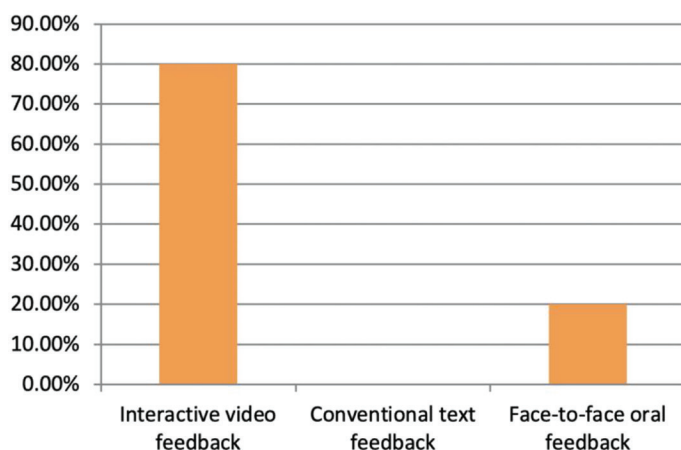


Figure 1. The most preferred feedback modality.

Student interview data and teaching journal provided additional insights into these preferences, highlighting DMF's capacity for personalized learning experiences, enhanced comprehension through multimodal information sources, and improved retention compared to single-mode feedback approaches. Students also appreciated the self-regulated learning opportunities that replay functionality provided.

Interestingly, teaching journals revealed that despite consistent preferences for DMF, dissimilar students may perceive and act upon the available affordances in dissimilar ways, showing different levels of engagement, and attributing to different perceived impacts on writing skill.

Discussion and Conclusion

This study demonstrates that DMF offers significant affordances for EFL writing instruction, particularly in accommodating diverse learning styles and fostering inclusive pedagogical practices. The technical, pedagogical, and social affordances identified suggest that DMF can enhance traditional feedback approaches while addressing limitations of monologic, text-only methods.

The strong student preference for DMF, coupled with positive teacher perceptions, indicates potential for broader implementation in EFL contexts. However, the study also reveals important considerations regarding individual differences in affordance perception and engagement. Some students demonstrated higher levels of interaction with DMF features, while others engaged more passively, suggesting that successful implementation requires careful attention to promoting active participation.

Cultural factors emerged as particularly significant in this Vietnamese context, where DMF provided mechanisms for overcoming traditional barriers to student-teacher interaction. The emotional support affordances of DMF may be especially valuable in collectivist educational environments where public questioning of authority may be culturally discouraged (Bao & Nguyen, 2020).

Future research should explore long-term effects on writing development and how varied cultural contexts influence DMF effectiveness and implementation strategies.

Acknowledgements

This research was funded by VNU University of Languages and International Studies (VNU-ULIS) in the project No. N.25.15

References

- Bao, D., & Nguyen, T.-M. (2020). How silence facilitates verbal participation. *English Language Teaching Educational Journal*, 3(3), Article 3. <https://doi.org/10.12928/eltej.v3i3.3004>
- Gregersen, T., & MacIntyre, P. D. (2014). *Capitalizing on language learners' individuality: From premise to practice*. Multilingual Matters & Channel View Publications. <https://doi.org/10.2307/jj.26931978>
- Mahoney, P., Macfarlane, Susie, & Ajjawi, R. (2019). A qualitative synthesis of video feedback in higher education. *Teaching in Higher Education*, 24(2), 157–179. <https://doi.org/10.1080/13562517.2018.1471457>
- Nguyen, N. T., Vo, L. N. H., & Do, T. V. H. (2025). *Online EFL Classrooms: How Teacher Power Affects Attitudes Toward Learning?* 8(1).
- Penn, S., & Brown, N. (2022). Is Screencast Feedback better than text feedback for student learning in higher education?: A systematic review. *Ubiquitous Learning: An International Journal*, 15(2), 1–18. <https://doi.org/10.18848/1835-9795/CGP/v15i02/1-18>
- Steen-Utheim, A., & Wittek, A. L. (2017). Dialogic feedback and potentialities for student learning. *Learning, Culture and Social Interaction*, 15, 18–30. <https://doi.org/10.1016/j.lcsi.2017.06.002>
- Williams, M., & Mercer, S. (2016). *Exploring psychology in language learning and teaching*. Oxford University Press.

A Large-Scale Mixed-Methods Study of Japanese University Students' Use of ChatGPT for L2 Learning

Louise Ohashi¹  and Suwako Uehara² 

¹Gakushuin University, Japan

²The University of Electro-Communications, Japan

Abstract

This mixed-methods paper reports on the adoption of ChatGPT for L2 learning at university level in Japan. It presents data from an online questionnaire conducted with 2,521 Japanese undergraduate students who studied English ($n = 2,519$) and 17 other L2s ($n = 1,881$). Quantitative data revealed that approximately half had used ChatGPT for language learning, mainly to assist with writing, vocabulary, and grammar. Thematic coding of written usage accounts provided deeper insights into how ChatGPT was used and the range of learning goals it supported. Although many students used it for basic support, such as corrective feedback and translation, some reported more experimental uses. Most written accounts indicated facilitative use of ChatGPT, but some responses pointed to breaches in academic integrity. While not generalisable, the large-scale dataset offers noteworthy insights into the practices of Japanese university students and allows for recommendations that provide a valuable starting point for further exploration in contexts within Japan and abroad.

Keywords: ChatGPT, language education, AI-assisted language learning, Japanese higher education, student use

Background to the Study

Since the release of ChatGPT in November 2023, numerous studies have reported on its potential impact on L2 education, highlighting benefits such

E-mail: ohashijalt@gmail.com & uehara.suwako@uec.ac.jp

as personalised learning, and problems such cheating (Alm & Ohashi, 2024; Dizon et al., 2025; Stroud, 2024). As it can both support and undermine learning, it is crucial to understand how it is being used. This short paper presents a large-scale study of Japanese university students' use of ChatGPT for L2 learning, drawing data from a wider study that explores students' usage and perceptions. The research questions are as follows:

- RQ1. What proportion of Japanese university students use ChatGPT for L2 learning? Do more students use it for classwork, homework, or self-study?
- RQ2. How do Japanese university students use ChatGPT for L2 learning?

Methodology

This mixed-methods study employed a convergent parallel design (Creswell & Plano Clark, 2018) through a questionnaire designed by the first author¹ that gathered quantitative and qualitative data. The questionnaire was distributed online in the participants' L1 (Japanese) in July 2024. Participants were gathered through snowball sampling, principally by distributing it to teachers and asking them to share it with students. While this method limits generalisability due to the potential of sampling bias, it is “a recognized and viable method of recruiting study participants not easily accessible or known to the researcher” (Leighton et al., 2021, p. 38) that was employed to gather large-scale data. The questionnaire was administered anonymously to promote candid responses for improved reliability.

All 2,521 participants were Japanese (1,346 female, 1,125 male, 16 non-binary, 34 not provided) undergraduate students who had taken at least one L2 course at university-level in Japan. In this context, English is generally required regardless of the course of study, and additional L2 credits are a part of many degrees. Almost all participants studied English (2,519) and 1,881 studied other languages, with 1,879 participants studying more than one language. Participants per language were as follows: Chinese (637), Korean (468), French (299), Spanish (280), German (261), Italian (29), Russian (18), Portuguese (6), Indonesian (5), Arabic (5), Latin (4), Thai (4), and Greek/Ancient Greek (3), with one each for Ainu, Czech, Hebrew, and Malay. Due to the brevity of this paper, languages other than English are grouped together as “other L2.”

¹ Louise Ohashi, JSPS KAKENHI Grant Number JP24K16135.

RQ1 was addressed through descriptive analysis of quantitative data (Creswell & Plano Clark, 2018). RQ2 drew on quantitative data from 12 uses adopted from Alm and Ohashi (2024) that focus on language skills and tasks, as well as qualitative thematic analysis (Saldaña, 2021) of written usage accounts. Each account was coded by the authors to capture use behaviours, then pattern coding was employed to group the codes into broader themes. Discrepancies were resolved through discussion. Coding was iterative, with multiple sessions of independent and joint coding to develop the themes and ensure reliability. Representative quotes were translated into English by the two English-Japanese bilingual authors and carefully reviewed to ensure the original meaning of the text was preserved. All quotes were identified with participant IDs (S1–S2521).

Results and Discussion

RQ1. What proportion of Japanese university students use ChatGPT for L2 learning? Do more students use it for classwork, homework, or self-study?

Half of the participants (50.5%) reported using ChatGPT for learning at least one language. Usage was slightly more common for English learners (43.2%) than other L2 learners (40.4%). Classwork, defined as in-class activities, ranked highest for English (31.4%), followed by homework set by the teacher (27.8%), then self-study (26.6%). For other L2s, all percentages were lower, with homework most common (24.5%), then self-study (23.5%) and classwork (22.9%). The limited level of adoption for language learning can be partially explained by low ChatGPT adoption rates more generally, as 29.6% had never used ChatGPT before. While usage for L2 learning is relatively low, comparison with data collected from 521 Japanese university students nine months prior to this study found approximately 25% had used ChatGPT for English learning (Dizon et al., 2025), making the rise to 43.2% in the present study noteworthy.

RQ2. How do Japanese university students use ChatGPT for L2 learning?

Quantitative data revealed that more English learners than other L2 learners used ChatGPT for every category investigated (Table 1). The highest ranked use was writing support for English (30.4%) and vocabulary support (19.6%) for other L2s. This is possibly related to the Japanese educational context, where there are typically differences in prior learning (English from primary school; other L2s from university) and course requirements (English writing courses; grammar/vocabulary tests for other L2s). Notwithstanding this,

Table 1. ChatGPT use for L2 learning (rank order by language group)

English		Other L2s	
Writing	30.4%	Vocabulary	19.6%
Grammar	26.1%	Grammar	16.5%
Vocabulary	25.6%	Writing	15.7%
Feedback/Assessment	18.7%	Reading	11.9%
Reading	17.7%	Feedback/Assessment	9.7%
Conversation Partner	16.0%	Quiz/Test	8.8%
Quiz/Test	11.9%	Conversation Partner	8.7%
Speaking	11.3%	Study Planner	7.7%
Study Planner	9.4%	Speaking	6.7%
Listening	8.9%	Listening	6.7%
L2 Games	7.1%	Virtual Tutor	5.6%
Virtual Tutor	6.6%	L2 Games	5.5%

writing, grammar, and vocabulary support ranked highest in both groups, while L2 games and virtual tutor ranked lowest.

Qualitative data provided further insights through 3,991 codes marked across 33,258 characters of predominantly Japanese text² collected from the written accounts. Approximately 60% (1,539) were from non-users or too vague (e.g., S2187: “For homework”) so excluded from further analysis, with all others thematically coded, and multi-coded when multiple themes appeared within a single account. Over 20 themes emerged, including subthemes, with the most common ones reported in Table 2.

Table 2 highlights the leading role ChatGPT played in individualised feedback, particularly for checking written work (S246, S2078) and providing translations (S1399, S1436). It also provides a more nuanced understanding of usage patterns; i.e., the findings demonstrate different levels of engagement with ChatGPT, from basic use that simply replaced dictionaries, translators, and grammar checkers (Feedback S246, Vocabulary S421, Grammar S2153) to more experimental use for creating personalised study tools and learning experiences (Feedback S2078, Vocabulary S1064, Grammar S2460).

Questions over academic integrity were raised in approximately 2% of the accounts, either through obvious cheating (S639: “When I had to write an essay in English for class, I had ChatGPT write the entire text.”) or suspected misuse (S834: “I used it to speed through my homework.”). However, it is

² Approximately 13,300–16,600 words based on character-to-word ratios of 2 to 2.5 characters per word.

Table 2. Most common themes from students' written accounts of ChatGPT use for L2 learning

Theme (Freq.)	Representative quotes (English Translations)
Feedback on own work (294)	S246: "I had ChatGPT correct sentences that I wrote." S2078: "I asked ChatGPT for corrections on English sentences I wrote, with explanations of the changes and the meanings of some vocabulary."
Translation (219)	S1399: "I used ChatGPT to translate English text into Japanese." S1436: "I used ChatGPT to confirm whether what I was trying to say was correct by asking it to translate from Japanese to English and vice versa."
Vocabulary (199)	Subtheme 1: Look up vocabulary (113) S421: "I have used ChatGPT to look up words I don't know." S842: "Especially for English, I used ChatGPT to find synonyms and antonyms to help build my vocabulary." Subtheme 2: Other vocabulary support (73) S928: "I used ChatGPT to generate example sentences for unfamiliar words." S1064: "I asked ChatGPT to make quizzes to help me memorise unfamiliar words from a video I liked." Subtheme 3: Unspecified vocabulary (13) S2360: "Vocabulary study (Chinese)."
Grammar (186)	Subtheme 1: Grammar correction (105) S2153: "I used ChatGPT to check for grammatical errors when completing English writing assignments." S2460: "I practiced Spanish grammar by chatting with ChatGPT and had it point out and correct my mistakes during the conversation." Subtheme 2: Other grammar support (43) S297: "I asked ChatGPT about grammar when trying to understand difficult English sentences." S147: "I asked ChatGPT to explain parts of German grammar I didn't understand. For English, I used ChatGPT to make grammar quizzes while studying on my own, which helped me understand better." Subtheme 3: Unspecified grammar (38) S1366: "English grammar."
Conversation practice (97)	S181: "I used voice mode and had ChatGPT act as a speaking partner to practice English." S8: "I chatted with ChatGPT in Korean because I wanted to improve my communication skills."
Paraphrase or summarise (72)	S108: "I used ChatGPT to look up ways to rephrase my English journal entries." S2014: "When I wanted to get the gist of long English text, I used ChatGPT to summarise it."

Note. Freq. = Frequency of occurrence.

crucial to note that task context was not always explicit, so 2% is likely to be an underestimate. These findings provide some reassurance that the majority of participants reported adhering to ethical standards. Nevertheless, concerns remain over academic integrity and how to guide students towards fair and optimal use.

Conclusion

The convergent parallel study design brought together quantitative and quantitative data, providing multifaceted insights into students' use of ChatGPT for language learning. Overall, the findings showed adoption by only half of the participants, with use for basic functions such as grammar correction, vocabulary checking, and translation more common than engagement for deeper learning. Drawing on these findings, continuous open dialogue between educators and students is recommended: first to better understand the extent of students' knowledge and usage of ChatGPT; second, to raise awareness of how to use ChatGPT for personalised learning support at both basic and experimental levels; and third, to guide students away from academic misconduct. In practical terms, this would involve teachers setting aside course time for discussions and guided exploration of ChatGPT. While not generalisable, the findings offer valuable insights into the Japanese context, and the recommendations could be a starting point for further explorations in educational contexts in Japan and abroad.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP24K16135, allocated to Louise Ohashi.

References

- Alm, A., & Ohashi, L. (2024). A worldwide study on language educators' initial response to ChatGPT. *Technology in Language Teaching & Learning*, 6(1), 1–23. <https://doi.org/10.29140/tl.v6n1.1141>
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). Sage.
- Dizon, G., Gold, J., & Barnes, R. (2025). ChatGPT for self-regulated language learning: University English as a foreign language students' practices and perceptions. *Digital Applied Linguistics*, 3(102510), 3–17. <https://doi.org/10.29140/dal.v3.102510>

- Leighton, K., Kardong-Edgren, S., Schneidereith, T., & Foisy-Doll, C. (2021). Using social media and snowball sampling as an alternative recruitment strategy for research. *Clinical Simulation in Nursing*, 55, 37–42. <https://doi.org/10.1016/j.ecns.2021.03.006>
- Saldaña, J. (2021). *The coding manual for qualitative researchers*. Sage.
- Stroud, R. (2024). Student-directed language learning with ChatGPT: Preferences from within Japanese university classrooms. In Y. Wang & M. S. Cárdenas Claros (Eds.), *Proceedings of the XXIIInd International CALL Research Conference, 2024* (pp. 267–274). Castledown. <https://doi.org/10.29140/9780648184485-40>

Designing Collaborative Resources in Data-Driven Learning to Accommodate Learner Diversity

**Muhammad Rudy¹ , Emi Emilia¹ , Wawan Gunawan¹ 
and Destiani² **

¹ English Language Education, Universitas Pendidikan Indonesia, Indonesia

² Universitas Pendidikan Indonesia, Indonesia and Universitas Lampung, Indonesia

Abstract

Data-Driven Learning (DDL) usually relies on existing corpora like British National Corpus (BNC) and Corpus of Contemporary American English (COCA), but these may not align with students' needs when writing an abstract. To address this issue, designing and implementing a tailored corpus could be a solution. This study explores students' perceptions during the collaborative creation of an abstract corpus (RQ1) and its strengths and weaknesses in an academic writing process (RQ2). This research was conducted as a case study involving 80 out of 94 final-year engineering students who completed the task of building a corpus of abstracts to serve as a writing reference. Despite varying interests, the participants collaborated to compile a corpus of relevant abstracts. They collaboratively collected, annotated, and analyzed the corpus. To support corpus development and the DDL process, Google Drive was used as online storage, and AntConc software was used as an analysis tool. The data were collected through interviews and video recordings. The video recordings were used to capture the DDL process starting from preparation to corpus utilization. The retrospective interviews with five randomly selected participants were conducted to explore participants' perceptions. The findings suggest that collaborative corpus development enhances students'

E-mail: muhammadrudy@upi.edu, emi.emilia@upi.edu, wagoen@upi.edu & destiani@fkip.unila.ac.id

understanding of abstract writing genres. Different research interests did not hinder the corpus development and utilization. An observable weakness was the corpus tool installation problem. In conclusion, the collaborative corpus development fostered inclusivity and deeper involvement, highlighting the potential of DDL to meet diverse students' needs.

Keywords: Collaborative, corpus development, data-driven learning, interdisciplinary projects

Introduction

Interdisciplinary projects have been long implemented in engineering higher education, including in telecommunication engineering, to bring solutions in real life (Rahman et al., 2025; van den Beemt et al., 2020). As a result, the final projects in telecommunication engineering education, known as Capstone Projects, have been applied in many sectors like smart, safe, and sustainable road infrastructure, agriculture, education, etc. (Rahman et al., 2025; Yao et al., 2023). However, the critics of the interdisciplinary projects emerged with the question of a learning program that should accommodate various specific projects that have specific genres and needs (Markauskaite et al., 2024). The obstacles in the English as Foreign Language (EFL) context regarding English for academic writing are that the lecturer must accommodate various specific needs related to topics (Yang, 2022). This is a challenge in order to facilitate all students' various needs.

DDL has become an influential approach in language learning, particularly for developing writing competence through corpus-based exploration (Johns, 1991; Pérez-Paredes & Boulton, 2025). Traditionally, DDL relies on large, pre-compiled corpora such as British Academic Writing English (BAWE) used by Crosthwaite et al. (2023) and Michigan Corpus of Upper-Level Student Papers (MICUSP) used by Karpenko-Seccombe (2024) or even built-in corpora in online analysis tools like SketchEngine used by Crosthwaite and Steeples (2022). Those corpora provide learners with access to authentic language input and usage patterns (Boulton & Cobb, 2017). However, generic corpora often fall short in meeting the specialized needs of learners engaged in discipline-specific writing (Egan, 2019; Flowerdew & Petrić, 2024). For instance, in this study context, engineering students writing abstracts must adhere to distinct rhetorical conventions and terminological expectations that are rarely represented in general corpora. Prior research has acknowledged this issue, noting that learners struggle when linguistic models do not align with their field-specific tasks (Flowerdew, 2015).

One proposed solution is the development of a specialized corpus, which offers both pedagogical and disciplinary relevance (Charles, 2012; Esfandiari & Allaf-Akbary, 2024; Sfeir, 2025). Collaborative corpus building, where students collect, annotate, and analyze texts, provides opportunities to engage more deeply with language while simultaneously constructing a useful writing resource. Charles (2012) and Esfandiari and Allaf-Akbary (2024) argued that such learner-led initiatives support genre awareness and foster autonomy. When students become co-constructors of linguistic resources, the process itself, like annotating and data cleaning, can enhance their genre awareness and engagement (Flowerdew & Petrić, 2024; Pérez-Paredes, & Boulton, 2025; Sfeir, 2025) in ways that static corpora cannot.

In this context, the study investigates how collaborative corpus development can support abstract writing among 80 final-year engineering students who conducted interdisciplinary research projects. It is different from the study conducted by Charles (2012), where the 47 multidisciplinary students work individually to build a corpus for personal use. It is also not similar to the study conducted by Sfeir (2025), where the 11 multidisciplinary students worked either individually or collectively to build their corpora. Thus, this study tried to investigate the students' perception (RQ1) and explore the strengths and weaknesses (RQ2) of the collaborative corpus building in abstract writing.

Methodology

Research Design

This study employed a qualitative case study design (Kumar, 2011) to explore how collaborative corpus supports abstract writing. The population consisted of 94 final-year engineering students enrolled in a scientific writing course at a private university in West Java. The study was conducted in the odd semester in 2024, taking three out of sixteen meetings of a compulsory course, namely Capstone Project. The timeline of DDL implementation can be seen in Figure 1, which is elaborated in the following subsections.

Participants and Corpus Development

94 students from the sixth and seventh semesters were grouped by the Telecommunications Engineering Department of the university. They took a 4-credit Capstone Course where the focus was writing a capstone report. The course lasted for 16 meetings in a semester. As part of the course completion and report submission, abstract writing training is conducted in the last three meetings. The participants were grouped based on their research interests

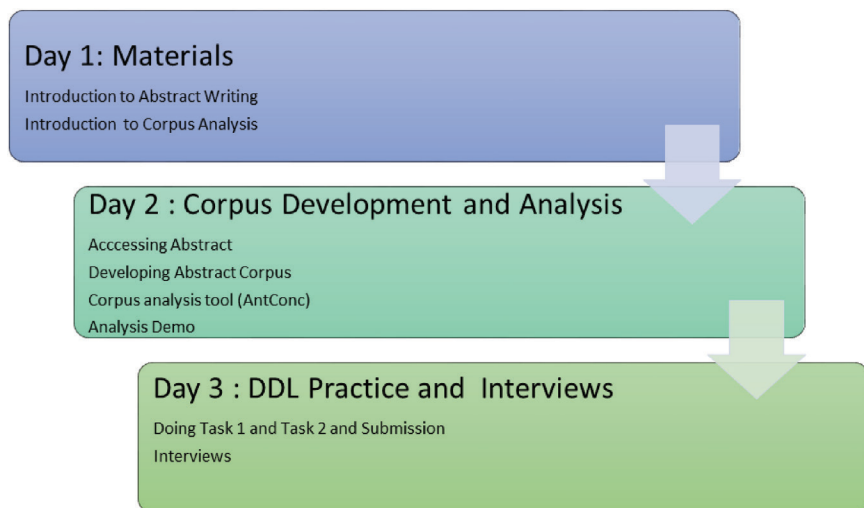


Figure 1. Teaching procedure

(see Appendix A), varying from machine learning to anxiety detection. Each group consisted of two to four people.

The development of the abstract corpus was completed by 80 students from 28 project groups. Each project group contributed at least two abstracts, which were compiled and stored using a folder in Google Drive. Each member of the group was instructed to collect abstracts from journals relevant to their capstone projects. The corpus was not treated with data cleaning, known as dirty corpus (Charles, 2012), to reduce unwanted elements in the text like equations (Sfeir, 2025). The students were instructed to annotate the abstracts manually for identifying abstract sections such as Introduction, Methods, Results, and Discussions (IMRD). This activity aimed to raise their awareness of genre structure and prepare them for corpus analysis.

Teaching Procedure

The DDL teaching procedure shown in Figure 1 was conducted in the thirteenth meeting. The first meeting was conducted to ask consent from the participants and give an explanation of the study. The materials explained on this day were an introduction to abstract writing and corpus analysis. The second meeting was conducted to train corpus development and analysis demo.

Once the corpus was ready, in the last meeting, the students used AntConc (Anthony, 2024) to analyze lexical and phrasal patterns based on abstract

sections. Specific attention was given to word lists and frequently used expressions—particularly phrases following IMRD sections (Swales, 1990). The students worked on answering two questions: “What are ten important words for your abstract?” and “What five writing expressions will you use to start writing the abstract introduction, method, result, and discussion sections?” The participants submitted their answers together with the justification for why they chose those ten words and five phrases. The answers were collected in the Google Form.

Instruments and Analysis

In order to address the two research questions, students’ perception and strengths and weaknesses, this study employed documentation and an interview. The documentation included two submitted tasks, photographs, and videos. The submitted tasks were scrutinized to identify reasons the participants chose words and phrases. The photos and videos captured group discussions and peer correction during the teaching process.

To gain deeper insight into students’ experiences, the interviewers announced the recruitment of interviewees at the end of the abstract writing program. An open-ended interview was conducted with five self-selected participants or volunteers (Brinkmann & Kvale, 2018). The interview focused on students’ perceptions of the collaborative process, corpus usability, and the challenges they encountered. The member checking process was also done by referring to the volunteers’ responses upon the two submitted tasks in the interview process.

The analysis stages were conducted by following suggestions from Kumar (2011, p 278): theme identification, coding, and classification. These three stages were conducted by the second author. The theme identification was through reading the Google form response about selected words and phrases as well as the justifications. The justifications tended to fall on two streams which are, genre awareness and engagement/ownership (Flowerdew & Petrić, 2024; Pérez-Paredes, & Boulton, 2025; Sfeir, 2025). The code was given for the reasons or the causes of choosing the words or phrases. The coded response then was classified into themes (Braun & Clarke, 2006). Peer checking towards the themes and codes was conducted by the third and fourth authors. The coding results can be seen in Table 1 and Table 2.

Similar to the task documents, the interviews were transcribed, and then the second author coded and identified themes. The third and fourth authors conducted peer checking on the transcription, codes, and themes. The improper transcription and code differences were discussed to reach agreement for correction.

Table 1. Distribution of task 1 justifications

No	Code	Number	Percentage %
1	Capstone Project Relevance	28	24.35
2	Abstract Structure	19	16.52
3	Linguistic Function	19	16.52
4	Field Relevance	15	13.04
5	Frequency	12	10.43
6	Academic Needs	10	8.70
7	Conceptual Representation	8	6.96
8	Ease of Understanding	4	3.48
	TOTAL	115	100

Table 2. Distribution of task 2 justifications

No	Code	Number	Percentage %
1	Conceptual representation	47	41.59
2	Abstract structure	26	23.01
3	Ease of understanding	13	11.50
4	Frequency	10	8.85
5	Academic needs	9	7.96
6	Linguistic function	3	2.65
7	Field relevance	3	2.65
8	Capstone Project relevance	2	1.77
	TOTAL	113	100

Results and Discussions

Task 1 showed that most students chose the ten important words that related to their capstone project, with 28 responses (24.35%). For example, participant number 73 showed in Figure 2 that the participant chose ten important words: *fiber, network, IoT, learning, optical, machine, detection, communication, monitoring, and networks*. He said those words were chosen because they were relevant to his Capstone Project. He said, “because these words are common vocabulary in telecommunications engineering, which are often used in capstone design topics in this department.”

In contrast, Task 2 was dominated by Conceptual Representation Code with 41.59%. In choosing five phrases to begin writing each section of IMRD,

participant 33 chose five phrases: *In this paper*, *To overcome the problem*, *This study compares*, *Based on these problems*, and *The method used* as shown by Figure 3. He said, “These five phrases were chosen because they are relevant to conveying research results effectively, both to state findings, prove hypotheses, and describe the impact of research in a professional manner.”

Regarding the RQs in this study, the answers are presented in the following subsections. The perceptions and plan to use the corpus for learning are elaborated. The strengths and weaknesses of corpus-based language learning

	E	F	G
1	app	10 kata penting	Mengapa 10 kata tersebut penting?
73	516521	fiber,network,iot,learning ,optical,machine,detection,communication,monitoring,networks	karena kata kata tersebut merupakan kosa kata umum di lingkungan teknik telekomunikasi yang sering digunakan dalam suatu topik capstone design di jurusan ini. with : karena ini kata hubung yang sangat berperan system : karena ini menjelaskan secara singkat method : langkah dalam penjelasan tugas yang performance : tugas yang sudah dikerjakan past technology : technology pada umumnya inti sari
74	773100	with: svstem; method; per	

Figure 2. Task 1 reasons coded with capstone project relevance.

	E	F	G	H
1		Di bagian mana Anda mengerjakan?	5 Frasa penting	Mengapa 5 frasa tersebut penting?
32	32514	Results	2. The findings of 3. Simulation result show 4. Based on 5. In this	Frasa-frasa tersebut merupakan kalimat pembuka untuk menjelaskan hasil dari penelitian
33	3255	Method	In this paper,to overcome the problem, this study compares, based on these problems, the method used	Karena 5 frasa tersebut menunjukkan bagian method pada abstrak paper atau jurnal suatu proyek dengan spesifik The Methodology: karena memberikan gambaran umum tentang pendekatan yang digunakan dalam penelitian.

Figure 3. Task 2 reasons coded with conceptual representation code.

in the study context highlight genre awareness and engagement. Lastly, the participants reveal a negligible weakness during DDL in academic writing class.

Student Perceptions of Collaborative Corpus (RQ1)

The perception of the participants after completing DDL in the abstract writing program was positive. The corpus-based language learning was relatively easy to follow and it was useful and helpful for their writing process. The five interviewees agreed that the process was easy as can be seen in the excerpts (the complete version in Appendix B).

- Participant 1: Ais
It's easy because, the user interface is also very pleasant
- Participant 2: Aan
It's really easy to find frequently occurring words.
- Participant 3: Kar
Interviewer: So, there are benefits, right? Then, were there any difficulties while learning?
Interviewee: Not yet.
Interviewer: No difficulties yet. So, it's easy.
- Participant 4: Hai
It's easy, sir, because it's very easy to search for words, and that's what the corpus showed. It makes things easier for us.
- Participant 5: Nur
I think I understand because the corpus tool is also quite easy to understand.

All four interviewees explicitly stated that DDL in the writing process was easy. Only one interviewee, Participant 3, did not mention it. However, Participant 3 said during the process he did not find any difficulties, which implied it was easy. Furthermore, the signs of positive perception on DDL for abstract writing intensified as the participants claimed the DDL was useful and helpful. These claims have been proven by the following excerpts:

- Participant 1: Ais
It's quite helpful, so I can better understand the content of abstracts, and also what abstract writing patterns are like.
- Participant 2: Aan
Yes, it is useful, sir. Well, it seems like, it's easier to read without having to click one by one first, we'll read what the results are like, sir.

Participant 4: Hai

I'll probably use it, because it was very helpful.

Participant 5: Nur

It's very useful, especially when searching for words and their frequency.

In addition, the positive perception was also stated by the interviewees. They realized that they acquired new knowledge after the DDL practice in writing abstract. The excerpts are:

Participant 3: Kar

My impression is that I have gained new knowledge for writing

Participant 4: Hai

It broadened my knowledge and helped me learn something new.

The participants' positive perception towards DDL for writing abstracts is because it is easy to practice, useful, helpful, and knowledge-enriching. This encouraging finding is in line with the study conducted by Charles (2012) that her students were enthusiastic about using DIY corpus for writing because it was useful to improve their writing. Crosthwaite and Steeples (2022), though they used large pre-compiled corpora for assisting secondary students in writing, also have found that the students expressed positive perceptions of DDL.

Positive perception was also found by Spivey (2023) when he studied DDL implementation with elementary students. He scaffolded the students by using two types of DDL, paper-based and computer-based. He said that the perception possibly changes over time, and students preferred paper-based DDL after facing difficulties practicing computer-based. The participants in the current study seems to have a consistent perception since they are from telecommunication engineering, where working with computers is part of their daily routine.

In line with a recent study conducted by Esfandiari and Allaf-Akbary (2024), the perception of students (83.7%) taught with hands-on DDL for identifying interactional metadiscourse in EFL writing was positive. DDL motivated the students to realize interactional metadiscourse since it activated autonomy and self-discovery. It is because they were given a chance to develop a corpus and analyze it by using AntConc software, similar to this study. This indicated that a specialized corpus built collaboratively for DDL brings positive perception among participants.

Besides positive perception, collaborative DIY DDL brought strengths and weaknesses discussed in the following subsection.

Strengths and Weaknesses in Writing Practice (RQ2)

Genre Awareness and Engagement

From Tasks 1 and 2, the coded response directed us to draw two streams; Genre Awareness and Engagement. Table 3 informs that Genre Awareness in Task 1 is derived from abstract structure, linguistic function, and academic needs. It occurs 48 times (41.74%) across the participants’ responses.

In addition, the Task 2 presented in Table 4 revealed that the Genre Awareness, similar with the first task, is indicated by the abstract structure, linguistic function, and academic needs. It occurs 38 times (33.63%). The reasonings stated by the participants after choosing ten words and five phrases

Table 3. Theme and code from task 1

No	Themes	Codes	Number	Percentages %
1	Engagement	Capstone Project Relevance, Concept Representation, Field Relationship, Frequency, Ease of Understanding	67	58.26
2	Genre Awareness	Abstract Structure, Linguistic Function, Academic Needs	48	41.74
	Total		115	100

Table 4. Theme and code from task 2

No	Themes	Codes	Number	Percentages %
1	Engagement	Capstone Project Relevance, Concept Representation, Field Relationship, Frequency, Ease of Understanding	75	66.37
2	Genre Awareness	Abstract Structure, Linguistic Function, Academic Needs	38	33.63
	Total		113	100

brought to the conclusion that the collaborative DIY corpus in the DDL enabled them to realize the pattern and genre in abstract.

Across interviews, participants expressed increased genre awareness, particularly in identifying the structure and function of academic abstracts. Three participants noted:

Participant 1: Ais

It has been quite helpful, so I can better understand **the content of abstracts** and also **the patterns of how abstracts** are written.

Participant 2: Nur

From there, **several phrases were displayed**, and then when writing an abstract, we get an idea like, Oh, **the opening words** when writing, especially in English, should be something like this.

Participant 3: Aan

There was no difficulty at all. Because when I looked at the papers that were collected, I thought, Oh, so in these papers, there must be an **abstract, introduction, and results**. Then, within the results, there must be **analysis and discussion**.

Those three participants reflected a heightened sensitivity to abstract writing convention. This hands-on engagement with genre conventions appears to have bridged the gap between theoretical knowledge and practical application.

Another significant finding was the enhanced engagement that resulted from students' sense of ownership over the corpus-building process. Rather than relying on pre-existing materials, students curated texts that reflected their research interests, fostering a sense of personal investment. Data indicated that the students reported feeling more joyful in this task compared to traditional writing tasks (as stated by Participant 2 and 3 above). The collaborative nature of the project—along with its relevance to students' ongoing thesis work—made the activity both meaningful and contextually grounded, reinforcing previous findings that learner-led DDL promotes autonomy and engagement.

Table 3 and 4 have shown that during Task 1 and 2, the students' engagement was seen in the codes like Capstone Project Relevance, Concept Representation, Field Relationship, Frequency, and Ease of Understanding. The response, like in Figure 2, indicated that the participants were strongly bound to their individual capstone projects, although during the DDL implementation, they together built a single corpus.

The interviews conducted with the five participants brought evidence of engagement by using a collaborative DIY corpus. The participants exhibited three out of five engagement codes: Capstone Project Relevant, Concept Representation, and Frequency. Below are the excerpts demonstrating the reasons why they chose a word or a phrase during finishing Task 1 and 2.

- Participant 1: Ais [Frequency] (Commenting about Task 2)
 Interviewer: You chose *the results, the experiment, and the model*. Why did you choose them?
 Interviewee: Because they have a higher frequency, there are several words that are suitable.
- Participant 2: Aan [Capstone Project Relevant] (Commenting about Task 1)
 Interviewer: Why did you choose *performance*?
 Interviewee: Well, because in our results, we display *performance*, that relates to *accuracy* and *precision*, sir. In the paper we wrote, that's *performance*.
 Interviewer: This is closely related to the paper your group wrote? So, your research results show *performance*.
 Interviewee: Yes, that's right.
- Participant 3: Kar [Concept Representation] (Commenting about Task 1)
 Interviewer: You chose *the, and, of, are, results, keyword, research, title, and detection*. Why are the words *the, and, and of* important?
 Interviewee: Because usually *the* is at the beginning, sir. If you don't use *the*. It feels like something is missing in the abstract.
- Participant 4: Hai [Frequency] (Commenting about Task 1)
 Interviewer: Why did you choose those words?
 Interviewee: They appear a lot, sir.
- Participant 5: Nur [Concept Representation] (Commenting about Task 2)
 Interviewer: Why did you choose the phrase *the purpose*?
 Interviewee: Why did I choose that? Because when we create research, there must be a purpose. It must be explained in the abstract. If there is a *the purpose*, there must be a goal for producing that abstract.

A stronger indication of engagement in completing the task during DDL is shown by videos. A Video 1 and Video 2 captured that the students were busy with their laptops to complete Task 1 and 2. Little discussions among the

participants were captured. Even though in the video they were discussing in their group, each of the members submitted his personal answers based on his own preferences. Their discussion was about the stages of submitting the task in order to avoid mistakes.

After compiling the findings from coded responses, videos, and interviews, engagement is relevant to the studies conducted by Charles (2012) and Esfandiari and Allaf-Akbary (2024). Both previous studies discussed the utilization of DIY corpus for writing teaching. Charles (2012) let her students use an individual corpus to help them writing. Esfandiari and Allaf-Akbary (2024) used a corpus created from 20 research articles from two reputable journals to be used in DDL practice to assess interactional metadiscourse. Both pieces of research agreed that DIY corpus improved learners' engagement during the DDL. The difference in the recent study is that the corpus was built collaboratively among interdisciplinary participants as shown by Task 1 and 2, video recording, and interviews.

The strengths of DDL by using a collaborative DIY corpus in this study were facilitating genre awareness and fostering learning engagement during the learning process. The collaborative DIY corpus in this study successfully accommodated the inclusivity of participants' research areas by providing various options of words and phrases to choose from, yet the participants kept focusing on their topic of interest. The corpus reflected a wide range of topics, providing diverse language use samples and discipline-relevant strategies. This variety ensured that each student could find examples closely aligned with their abstract writing needs.

Drawback

Behind the strengths of using a collaborative DIY corpus in DDL previously mentioned, a notable challenge also emerged. First, some students stated difficulty in installing the corpus analysis tool, AntConc, from the website due to internet connection crowding (this part was not recorded). During the DDL practice, all participants accessed the Anthony Laurence website at the same time. The download process was taking time which hindered the smoothness of learning because after the download process the participants were required to install it onto their personal laptop. The DIY corpus by using a similar analysis tool, may face this problem. The experience of this occurred in the interview with Participant 4.

Participant 4: Hai

The only **issue we faced was maybe just during installation.**

The internet connection problem experienced by the participants seems not to relate directly to the practice of DDL with a collaborative corpus. This is probably due to the participants being from a telecommunication engineering major where computer software and program are their core competency. Moreover, positive perception of the participants shown previously confirms that there was no significant difficulty during DDL practice.

Conclusions

This study examined how collaborative corpus development supports abstract writing among telecommunication engineering students within a DDL framework. The participants had a positive perception of the DDL practice. DDL seems easy, useful, and helpful in assisting with the writing of an abstract. This study also indicates that the genre awareness and engagement of the participants were enriched. The participants were involved in curating and analyzing a corpus relevant to their academic interests, enhancing genre awareness and engagement. The process of DDL with a collaborative DIY corpus encouraged participants in understanding disciplinary writing conventions through student-selected texts. The weakness of a collaborative corpus in DDL was unclear due to familiarity bias for telecommunication engineering students.

The limitation of this study was the short time of the implementation of the collaborative corpus in DDL. This becomes a gap to fill by the future research. Although the participants have interdisciplinary topics with different needs, they are from a telecommunications engineering major, which is familiar with the tool, so that they did not face any difficulties on DDL. Future research should involve more diverse participants to strengthen the inclusivity of DDL.

Acknowledgement

We would like to thank the editor and reviewers for the invaluable suggestions and criticisms that helped improve this manuscript.

Funding

This work was supported by Indonesian Education Scholarship (Beasiswa Pendidikan Indonesia/**BPI**), Center for Higher Education Funding and Assessment (Pusat Pembiayaan dan Asesmen Pendidikan Tinggi/**PPAPT**) Kemdiktisaintek, and Indonesian Endowment Fund for Education (Lembaga Pengelola Dana Pendidikan/**LPDP**)

References

- Anthony, L. (2024). AntConc (Version 4.3.1) [Computer Software]. Waseda University. <https://www.laurenceanthony.net/software/AntConc>
- Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, 67(2), 348–393. <https://doi.org/10.1111/lang.12224>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Brinkmann, S., & Kvale, S. (2018). *Doing Interviews*: SAGE Publications. (The Sage Qualitative Research Kit).
- Charles, M. (2012). Proper nouns: Pedagogical tools for teaching academic writing. *Journal of English for Academic Purposes*, 11(1), 54–66. <https://doi.org/10.1016/j.jeap.2011.11.005>
- Crosthwaite, P., Luciana, & Wijaya, D. (2023). Exploring language teachers' lesson planning for corpus-based language teaching: a focus on developing TPACK for corpora and DDL. *Computer Assisted Language Learning*, 36(7), 1392–1420. <https://doi.org/10.1080/09588221.2021.1995001>
- Crosthwaite, P., & Steeples, B. (2022). Data-driven learning with younger learners: exploring corpus-assisted development of the passive voice for science writing with female secondary school students. *Computer Assisted Language Learning*, 37(5–6), 1166–1197. <https://doi.org/10.1080/09588221.2022.2068615>
- Egan, T. (2019). Non-representativeness in corpora: Perils, pitfalls and challenges. *CogniTertes*. Article 1772. <https://doi.org/10.4000/cognitertes.1772>
- Esfandiari, R., & Allaf-Akbary, O. (2024). Assessing interactional metadiscourse in EFL writing through intelligent data-driven learning: the Microsoft Copilot in the spotlight. *Language Testing in Asia*, 14(1). <https://doi.org/10.1186/s40468-024-00326-9>
- Flowerdew, L. (2015). Corpus-based research and pedagogy in EAP: From lexis to genre. *Language Teaching*, 48(2), 153–174. <https://doi.org/10.1017/S0261444814000019>
- Flowerdew, L., & Petrić, B. (2024). A critical review of corpus-based pedagogic perspectives on thesis writing: Specificity revisited. *English for Specific Purposes*, 76, 1–13. <https://doi.org/10.1016/j.esp.2024.05.003>
- Johns, T. (1991). Should you be persuaded: Two examples of data-driven learning. *English Language Research Journal*, 4, 1–16.
- Karpenko-Secombe, T. (2024). *Practical use of corpora in teaching argumentation. In Corpora for language learning* (pp. 111–130). Routledge. <http://dx.doi.org/10.4324/9781003413301-9>

- Kumar, R. (2011). *Research Methodology: a step-by-step guide for beginners*. SAGE Publications Ltd.
- Lee, D., & Swales, J. M. (2006). A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes*, 25(1), 56–75. <https://doi.org/10.1016/j.esp.2005.02.010>
- Lin, M. H. (2021). Effects of Data-Driven Learning on College Students of Different Grammar Proficiencies: A Preliminary Empirical Assessment in EFL Classes. *SAGE Open*, 11(3). <https://doi.org/10.1177/21582440211029936>
- Markauskaite, L., Schwarz, B., Damşa, C., & Muukkonen, H. (2024). Beyond disciplinary engagement: Researching the ecologies of interdisciplinary learning. *Journal of the Learning Sciences*, 33(2), 213–241. <https://doi.org/10.1080/10508406.2024.2354151>
- Pérez-Paredes, P., & Boulton, A. (2025). *Data-driven Learning in and out of the Language Classroom*. Cambridge University Press. <https://doi.org/10.1017/9781009511384>
- Rahman, M. ., Sarwar, H. ., Tin, T. T. ., & Kader, M. A. . (2025). A systematic literature review on final year capstone project in an engineering curriculum. *Edelweiss Applied Science and Technology*, 9(6), 1408–1440. <https://doi.org/10.55214/25768484.v9i6.8146>
- Sfeir, M. (2025). Do-it-yourself corpora and discipline-specific writing: A focus on the benefits of corpus building and cleaning (in the age of large language models). *English for Specific Purposes*, 79, 70–86. <https://doi.org/10.1016/j.esp.2025.04.002>.
- Spivey, M. (2023). Data-driven learning and young learners: Perceptions and attitudes of Japanese elementary school students toward EFL corpus-based activities. *JALT CALL Journal*, 19(3), 394–416. <https://doi.org/10.29140/jaltcall.v19n3.1014>
- van den Beemt, A., MacLeod, M., & van der Veen, J. (2020). *Interdisciplinarity in tomorrow's engineering education*. In J. van der Veen, N. van Hattum-Janssen, H.-M. Jarvinen, T. de Laet, & I. Ten Dam (Eds.), *Engaging Engineering Education: Book of Abstracts SEFI 48th Annual Conference* (pp. 61-74). European Society for Engineering Education (SEFI). <https://www.sefi.be/wp-content/uploads/2020/11/Proceedings-2020-web.pdf>
- Yang, L. (2022). Focus and Interaction in Writing Conferences for EFL Writers. *SAGE Open*, 12(1). <https://doi.org/10.1177/21582440211058200>
- Yao, H., Xu, Z., Hou, Y., Dong, Q., Liu, P., Ye, Z., ... & Wang, D. (2023). Advanced industrial informatics towards smart, safe and sustainable roads: A state of the art. *Journal of traffic and transportation engineering* (English edition), 10(2), 143-158. <https://doi.org/10.1016/j.jtte.2023.02.001>

Appendix A

Students Research Interests

No	Topics	Group Member (Persons)
1	SDN Based 5g Core Network Emulator System with Graphical Interface for 5g Core Network Design	3
2	Observation of the Effect of Anxiety Conditions on the Shift in the Working Frequency of EEG Signals	4
3	Machine Learning Based Multi Disturbance Prediction Modeling for Disturbance Handling in Dense Wavelength Division Multiplexing (DWDM) Longhaul	3
4	Multi Disturbance Prediction Modeling Based on Machine Learning for Handling Disturbances in Optical Network Cables	3
5	Design of Earthquake Detection System Based on Internet of Things	4
6	Resource Allocation in Device to Device Communication Using SCMA Model for 5G Network	3
7	Implementation of Open-Source Cloud-Based WiFi Controller to Improve User Experience in an Open Library Case Study	4
8	Resource Block Allocation for Hetnet (Heterogeneous Network) in 4G AND 5G	4
9	Resource Allocation in Hybrid Interweave-Underlay Scheme in Cognitive Radio Network	3
10	WIH (Monitoring of Water Quality and Automatic Feeding of Tilapia Fish Ponds Based on IoT)	3
11	HRS (Sentiment Analysis of Tourist Attractions in Bandung)	3
12	UMY (Development of Mobile Applications Based on Machine Learning and Internet of Things for Environmental Monitoring Maggot Cultivation as Organic Waste Decomposers	4
13	HRS (Sentiment Analysis of Tourist Attractions)	2
14	Implementation of Internet of Things (IoT) in Aquaponic System for Automatic Monitoring and Controlling of Fish Pond and Hydroponic Maintenance	4

(Continued)

No	Topics	Group Member (Persons)
15	Use of Artificial Intelligence in Audio Watermarking to Improve Information Hiding Performance	4
16	Estimation of Bit Error Rate in Free Space Optic Using Machine Learning	4
17	FlexiWAN-Based uCPE Device as SD-WAN Vendor Lock-In Solution	2
18	Design and Realization of RT RW NET In Girimukti Village	3
19	Trading-dong 2.0: Forex Expert Advisor using Strategy Based on Price Action and Order Block Under The Control of Artificial Intelligence	4
20	Design of Hybrid Access Network for Unifiber Provider by Asianet From Ciwastra Point of Presence to Batununggal Area Bandung	4
21	Automatization of Diagnosis and Claims Process Drugs for Optimization of BPJS Claims by Hermina Hospital Arcamanik	3
22	Optimization of Student Counseling Services at a University Through Digital Applications with NLP Technology	3
23	Improving Network Service Quality Through Ad Blocking and DNS Filtering	3
24	Gender Identification Based on Bite Marks	4
25	Design of Repeated Bending and Twist Test Tools for Patchcord and Precon Fiber Optic Cables at a Test House	4
26	Temperature and Humidity Monitoring and Control System for Incubator for Tempe Fermentation Process Based on IOT Technology	3
27	Design and Analysis of Single Frequency Network on WHOOSH KCIC Between Padalarang Station and Tegalluar Summarecon Station Lines	3
28	Android Application Readers of Javanese Script In Javanese Manuscript and Literary Works	3
		94

Appendix B

Excerpt of Perception

Participant 1: Ais

- Interviewer: I am conducting research related to the learning activities that have been carried out. I have a few questions. First, how do you think learning using the corpus was?
- Interviewee: in my opinion, because this is my first time too. I know... what's the name of this website, because this software... ...it's quite helpful, so I can better understand the content of abstracts, and also what abstract writing patterns are like.
- Interviewer: More or less, there are benefits, right?. In your opinion, was it difficult or easy to learn using this tool?
- Interviewee: Easy.
- Interviewer: Easy?
- Interviewee: It's easy because, the user interface is also very pleasant, sir.
- Interviewer: Okay, Then, after that, are you going to use AntConc or another corpus tool to help you write abstracts?
- Interviewee: I'll try using it again.
- Interviewer: So, will you use it in the future?
- Interviewee: Yes.

Participant 2: Aan

- Interviewer: you've already learned how to use corpus. What's your impression?
- Interviewee: Corpus is that application, right? It's really easy to find frequently occurring words, sir. Because we just input all of them (abstract files). Then we can search for the results, or the frequently occurring words, or how many files are included (Range). Previously, we usually searched for keywords in the abstracts. Previously, to find the most frequently occurring ones (words) was difficult, sir. So, with this application, it's easier, sir.
- Interviewer: So, do you want to continue using this method?
- Interviewee: If I want to write another paper or journal, maybe I will. But I don't know yet, I don't have any plans to do that (write a paper) again, sir.

- Interviewer: So, do you think it's useful?
- Interviewee: Yes, it is useful, sir. Well, it seems like, it's easier to read without having to click one by one first, we'll read what the results are like, sir.
- Interviewer: KWIC interface?.
- Interviewee: yes, that's easier, sir

Participant 3: Kar

- Interviewer: Well, this is related to writing the abstract class that you have been involved; what are your impressions?
- Interviewee: My impression is that I have gained new knowledge for writing.
- Interviewer: Oh, new knowledge. New knowledge in what way?
- Interviewee: Previously, I was confused about writing. But the software makes it easier.
- Interviewer: So, there are benefits, right? Then, were there any difficulties while learning?
- Interviewee: Not yet.
- Interviewer: No difficulties yet. So, it's an easy.

Participant 4: Hai

- Interviewer: According to you, is learning like that, using a corpus tool, difficult or easy?
- Interviewee: It's easy, sir, because it's very easy to search for words, and that's what the corpus showed. It makes things easier for us, sir.
- Interviewer: Were there any difficulties?
- Interviewee: No, sir. The only problem was the installation process, right?
- Interviewer: Oh yes. Was the learning helpful?
- Interviewee: Very helpful, sir. It broadened my knowledge and helped me learn something new.
- Interviewer: Okay, okay. Will you use corpus again when writing abstracts?
- Interviewee: For me, I'll probably use it, because it was very helpful.

Participant 5: Nur

- Interviewer: Okay, you've attended the class, especially the one about writing abstracts. This can be done using a corpus. Do you understand or master the steps taught previously?
- Interviewee: I think I understand because the corpus tool is also quite easy to understand.

- Interviewer: Oh, it's quite easy. Also, there weren't any difficulties. So, do you think this kind of learning, using a corpus, is useful?
- Interviewee: Of course, it's very useful, especially when searching for words and their frequency. Using regular tools takes a long time, but using a corpus is fast, right?
- Interviewer: Oh, right, okay. And also... Do you think what you've learned will be useful for writing abstracts?
- Interviewee: Yes, it's useful.
- Interviewer: What kind of example?
- Interviewee: For example, when creating phrases, for example, we searched for phrases using the corpus (tool). From there, there are several phrases that have been displayed, and then to make the abstract we have an idea for, oh, the initial words in writing the abstract, especially in English, are like this.

AI as a Thinking Partner? Exploring International Students' Use of GenAI to Support their Academic English Needs

Franciele Spinelli 

The University of Queensland, Australia

Abstract

Despite meeting university English proficiency requirements, international students who speak English as a second language often encounter difficulties with academic writing, reading, presentations, class participation and oral comprehension. Generative Artificial Intelligence (GenAI) tools such as ChatGPT can be particularly useful in addressing the academic language challenges of this student cohort, serving as a 'thinking partner' and helping them fulfil the expectations to communicate proficiently in academic English. However, despite GenAI's potential and growing research in education, little is known about how international students have used this technology for academic reading, writing, listening and speaking. In addition, while self-regulated learning (SRL) is pivotal for academic success, no research has examined this cohort's use of GenAI through this lens - an emerging theoretical perspective within SLA and L2 studies. This mixed-methods study aims to explore how international undergraduate students with varying English proficiency levels at an Australian university utilise GenAI for academic language support. Results from questionnaires and semi-structured interviews show that the participants frequently use GenAI to support their academic reading and writing, primarily due to their low confidence in their English skills and drive to meet academic expectations. Less frequent uses include preparing presentations and summarising audio content. While participants employ GenAI to monitor and complete academic tasks (SRL monitoring

E-mail: f.spinelli@uq.edu.au

and control), they use it less for planning and reflection (SRL forethought and reaction/reflection). These findings suggest that international students need more scaffolded support and clearer policies on GenAI use to better self-regulate their academic language needs and maximise their capabilities and academic success. This study contributes to discussions on inclusive CALL, highlighting both opportunities and challenges in integrating GenAI into higher education.

Keywords: Generative Artificial Intelligence, International Students, Academic English Needs, Self-regulated Learning

Background

The growing influx of international students reshaped higher education (HE), particularly in anglophone countries such as Australia (Cho et al., 2021). Yet, many English as an additional language (EAL) international students continue to experience academic language challenges despite meeting English proficiency requirements (Clark & Yu, 2021). Challenges span academic reading, writing, listening and speaking, often due to limited discipline and academic vocabulary (Crosthwaite & Cheung, 2019), unfamiliar academic conventions (Kukatlapalli et al., 2020) and barriers to understanding educators and peers (Park, 2022).

Generative Artificial Intelligence (GenAI) tools such as ChatGPT appear to have broadened and democratised language support for EAL international students. While not originally designed for education, these tools may aid in interpreting assignment tasks in multiple languages (Sterling et al., 2025) and providing feedback similar to interacting with peers or educators (Teng, 2024). In many cases, GenAI became a ‘thinking partner’, helping international students fulfil academic English expectations (Tseng & Warschauer, 2023).

However, research suggests that the use of GenAI for continuous learning and language development depends on self-regulated learning (SRL) (Mirriahi et al., 2025). SRL is considered pivotal for academic success, involving students’ ability to plan (strategising for tasks), monitor (implementing strategies and sustaining effort) and reflect on their learning (evaluating the effectiveness of strategies) (Zimmerman, 2000). Yet, scant research explores EAL international students’ use of GenAI through this theoretical lens, with existing studies focusing on EFL contexts (Teng, 2024) and centred on GenAI’s role in L2 writing (Sterling et al., 2025). This study addresses these gaps by investigating the role of GenAI in supporting the academic English needs of EAL international students in HE, guided by the following research questions:

- RQ1. How do EAL international students use GenAI to address their academic English reading, writing, listening and speaking needs?
- RQ2. What motivates EAL international students to use GenAI to support their academic English needs?

Methodology

Following institutional ethics approval (ref: 2024/HE001084), 170 international students (Female = 109, Male = 60, Undisclosed = 1), aged between 18 to 46 years old and from 17 countries, participated. Participants were enrolled in undergraduate or master's by coursework programs across 6 disciplines at an Australian university.

A mixed-methods design was adopted to answer the RQs on how and why students use GenAI to support their academic English needs. An online Qualtrics survey collected demographic information and measured:

1. GenAI use for academic reading, writing, listening and speaking (to answer RQ1), using 5-point Likert scale questions based on prior research on academic English difficulties (Rahman & Watanobe, 2023); and
2. SRL-guided strategies (to answer RQ1) and motivations for GenAI use (to answer RQ2), using 5-point Likert scale items adapted from the SRL-O questionnaire for online learning (Broadbent et al., 2023).

Survey data were analysed using descriptive statistics to explore general patterns (see Appendix A for the descriptive and internal consistency reliability statistics).

Follow-up semi-structured interviews were conducted via Zoom (approximately 45 minutes each) and explored students' GenAI uses and motivations qualitatively to triangulate the survey responses and provide more in-depth insights for each RQ. Zoom-generated transcripts were edited for accuracy against the recordings (Crosthwaite et al., 2025), and then analysed thematically (Saldaña, 2021).

Results

RQ1 focused on understanding how international students used GenAI. Survey responses (Figure 1) showed that participants used GenAI for all four academic English skills, most frequently for reading and writing. Interviews revealed that GenAI helped students understand complex vocabulary and sentence structures when reading. For writing, it supported appropriate tone,

grammar and lexical choices. Speaking and listening were perceived as more collaborative and less challenging, though students used GenAI to summarise or translate lectures, particularly when unfamiliar English varieties were spoken.

Survey responses (Figure 2) further demonstrated how students used GenAI by indicating that SRL strategies related to Monitoring and Reflection were more common than Planning. Interviews revealed that students monitored their learning by generating original ideas rather than over-relying on GenAI.

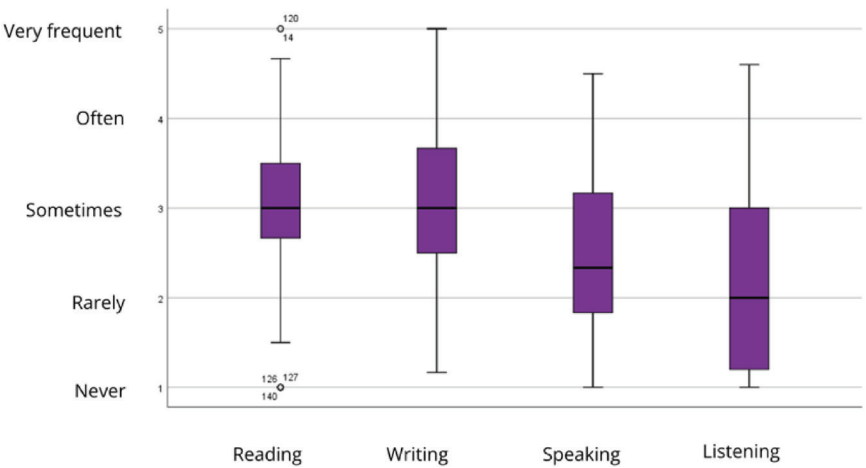


Figure 1. Box plots of academic language needs & GenAI use
<https://figshare.com/s/1d2becfac8099a623cbc>

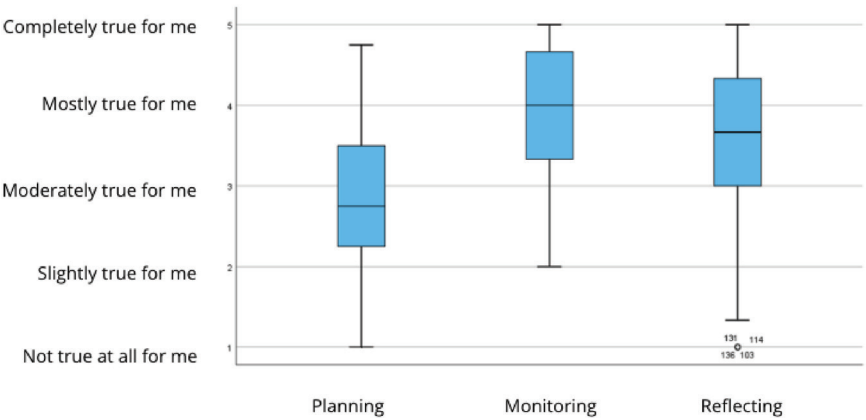


Figure 2. Box Plots of SRL-guided Strategies for Using GenAI
<https://figshare.com/s/1d2becfac8099a623cbc>

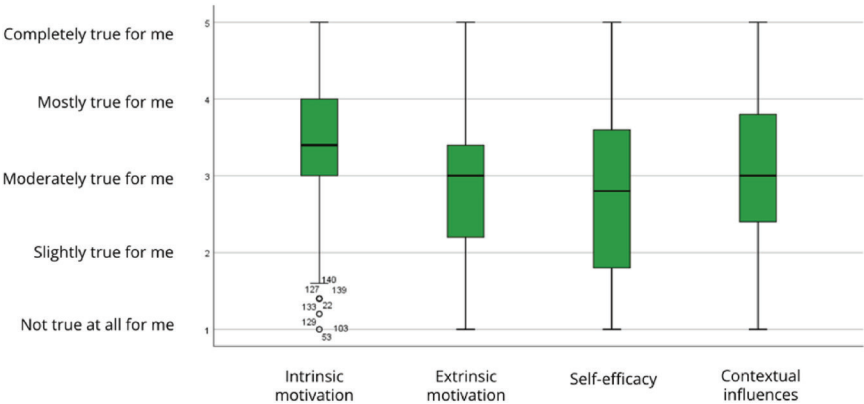


Figure 3. Box Plots of SRL-guided Motivations for Using GenAI
<https://figshare.com/s/1d2becfac8099a623cbc>

For reflection, some critically assessed GenAI language suggestions and reflected on their impact on learning, while others modified GenAI outputs primarily to meet academic integrity requirements. Planning included using GenAI to outline study plans and manage workloads, though some admitted to using it for summaries to expedite tasks.

RQ2 focused on understanding international students’ motivations for using GenAI. Survey responses showed that their reasons spanned across Intrinsic Motivation, Extrinsic Motivation, Self-Efficacy and Contextual Influences (Figure 3), with Intrinsic Motivation as the strongest driver. Students described enjoying the technology and viewing it as essential for improving academic English. Peer and tutor recommendations (Extrinsic Motivation, Contextual Influences) also shaped use. Low self-efficacy, such as fear of language errors affecting grades, further motivated engagement with GenAI.

Discussion

Findings related to RQ1 highlight GenAI’s greater use for reading and writing, reflecting its utility for text-based tasks and still-developing voice features (Vo & Nguyen, 2024). Students’ perception of oral tasks as less demanding and more interactive may also explain lower GenAI use for speaking and listening, consistent with research showing that international students apply various strategies to overcome communication barriers (Park, 2022). Nonetheless, students noted GenAI’s support for pre- and post-listening/speaking tasks, suggesting its growing potential beyond writing-focused applications.

Monitoring was the most reported SRL strategy, consistent with research on technology-enhanced language learning, where students monitored their learning during tasks but engaged less in planning or reflection (Yang et al., 2023). EAL learners may favour GenAI for immediate understanding and feedback (Tseng & Warschauer, 2023) over long-term learning strategies such as planning and reflection, especially if they lack the language proficiency to articulate goals or evaluate GenAI language suggestions (Teng, 2025). The findings suggest their underdeveloped SRL skills, which are further showcased when some described their reflection phase as simply modifying GenAI outputs to avoid academic misconduct rather than engaging in critical evaluation (Tran & Crosthwaite, 2025).

Findings for RQ2 show that all four SRL motivation dimensions influenced participants' use of GenAI. This suggests that students' use of the technology to support their academic English needs is shaped by the interplay between their personal attributes and environmental conditions (Wang, 2024). However, participants reported intrinsic motivation as the strongest motivator, indicating that EAL students who enjoy learning are more likely to value GenAI and use it meaningfully (Mirriahi et al., 2025).

The study offers important implications. While GenAI helps address academic language challenges for EAL students, their engagement with SRL was inconsistent. Echoing previous research (Sterling et al., 2025; Wang, 2024), the findings highlight the need to develop SRL skills, not only technical AI know-how. Despite limitations, including reliance on self-reporting and a potentially engaged sample, the study advances understanding of SRL-guided GenAI use in HE and contributes to inclusive CALL research by identifying opportunities and challenges in supporting EAL learners.

Acknowledgement

ChatGPT 4o was used to edit this article for clarity, and Grammarly was consulted for grammatical and lexical errors.

References

- Broadbent, J., Panadero, E., Lodge, J. M., & Fuller-Tyszkiewicz, M. (2023). The self-regulation for learning online (SRL-O) questionnaire. *Metacognition and Learning*, 18(1), 135–163. <https://doi.org/10.1007/s11409-022-09319-6>
- Cho, H. J., Levesque-Bristol, C., & Mike, Y. (2021). International students' self-determined motivation, beliefs about classroom assessment, learning

- strategies, and academic adjustment in higher education. *Higher Education*, 81(6), 1215–1235. <https://doi.org/10.1007/s10734-020-00608-0>
- Clark, T., & Yu, G. (2021). Beyond the IELTS test: Chinese and Japanese postgraduate UK experiences. *International Journal of Bilingual Education and Bilingualism*, 24(10), 1512–1530. <https://doi.org/10.1080/13670050.2020.1829538>
- Crosthwaite, P., Smala, S., & Spinelli, F. (2025). Prompting for pedagogy? Australian F-10 teachers' generative AI prompting use cases. *The Australian Educational Researcher*, 52(3), 1795–1825. <https://doi.org/10.1007/s13384-024-00787-0>
- Kukatlapalli, J., Doyle, S., & Bandyopadhyay, S. (2020). An investigation into the English language experiences of Indian international students studying in New Zealand universities. *Higher Education Research & Development*, 39(3), 485–499. <https://doi.org/10.1080/07294360.2019.1685940>
- Mirriahi, N., Marrone, R., Barthakur, A., Gabriel, F., Colton, J., Yeung, T. N., Arthur, P., & Kovanovic, V. (2025). The relationship between students' self-regulated learning skills and technology acceptance of GenAI. *Australasian Journal of Educational Technology*. <https://doi.org/10.14742/ajet.10006>
- Park, E. (2022). Coping with diverse accented English: Comprehension strategies deployed by East Asian international students at an Australian university. *International Education Journal: Comparative Perspectives*, 21(1), 1–19. <https://openjournals.library.sydney.edu.au/IEJ/article/view/15287>
- Saldaña, J. (2021). *The coding manual for qualitative researchers* / Johnny Saldaña (4E ed.). SAGE Publishing.
- Sterling, L., Ye, C., Ying, H., & Chen, Z. (2025). Finding your voice: Using Generative AI to help international students improve their writing. *Information*, 16(4), Article 4. <https://doi.org/10.3390/info16040289>
- Teng, M. F. (2024). “ChatGPT is the companion, not enemies”: EFL learners' perceptions and experiences in using ChatGPT for feedback in writing. *Computers and Education: Artificial Intelligence*, 7, 100270. <https://doi.org/10.1016/j.caeai.2024.100270>
- Teng, M. F. (2025). Metacognitive awareness and EFL learners' perceptions and experiences in utilising ChatGPT for writing feedback. *European Journal of Education*, 60(1), e12811. <https://doi.org/10.1111/ejed.12811>
- Tran, H., & Crosthwaite, P. (2025). *Students' self-determination in using machine translation and Generative AI tools for English for academic purposes* (SSRN Scholarly Paper No. 5256503). Social Science Research Network. <https://doi.org/10.2139/ssrn.5256503>

- Tseng, W., & Warschauer, M. (2023). AI-writing tools in education: If you can't beat them, join them. *Journal of China Computer-Assisted Language Learning*, 3(2), 258–262. <https://doi.org/10.1515/jccall-2023-0008>
- Vo, T. K. A., & Nguyen, H. (2024). Generative artificial intelligence and ChatGPT in language learning: EFL students' perceptions of technology acceptance. *Journal of University Teaching and Learning Practice*, 21(06), Article 06. <https://doi.org/10.53761/fr1rkj58>
- Wang, Y. (2024). Cognitive and sociocultural dynamics of self-regulated use of machine translation and generative AI tools in academic EFL writing. *System*, 126, 103505. <https://doi.org/10.1016/j.system.2024.103505>
- Yang, Y., Wen, Y., & Song, Y. (2023). A systematic review of technology-enhanced self-regulated language learning. *Educational Technology & Society*, 26(1), 31–44. <https://www.jstor.org/stable/48707965>
- Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner, *Handbook of Self-Regulation* (pp. 13–39). Academic Press. <https://doi.org/10.1016/B978-012109890-2/50031-7>

Appendices

Appendix A

Descriptive and Internal Consistency Reliability Statistics for Survey Items

<https://figshare.com/s/e926c2edf99635881e66>

Instructional Design for Co-Created and Context-Based VR Chinese Learning Environments

Satoko Sugie 

Hokkai-Gakuen University, Japan

Abstract

The integration of VR and foreign language education offers new opportunities for contextual, interactive, and co-created learning (Durak & Cankaya, 2022). This study investigates how VR enhances co-creative CFL (Chinese as a Foreign Language) at a Japanese university, where students begin with limited language proficiency. The research focuses on the communicative potential of multimodal resources in learner-centered collaboration. Prior research has shown that virtual spaces can increase learner motivation and identity engagement even among struggling learners of English (Chen & Kent, 2019). However, their application in non-English foreign language remains underexplored. We developed VR spaces using “Spatial” through collaboration between students and the teacher. Scenarios were designed to reflect learners’ cultural interests and daily experiences. Classroom sessions incorporated VR-based tasks emphasizing spatial interaction. Data was collected through participant observation and open-ended surveys with 40 students. The results revealed that learners actively co-constructed meaning beyond spoken language, utilizing avatars, shared visual references, and spatial positioning. Motivation increased notably when scenarios aligned with learners’ lived experiences. These findings align with multimodal design approaches in foreign language education, highlighting the importance of integrating linguistic and non-linguistic resources. However, technical barriers such as device access and interface complexity were noted. This study suggests that VR-supported environments enable learners to integrate language with multiple communicative resources and build context-rich, authentic experiences. Such environments shift the focus from traditional four-skill acquisition toward contextually optimized communication.

E-mail: s_sugie@hgu.jp

Keywords: VR, metaverse, CFL (Chinese as a Foreign Language), instructional design, context-based learning, action research

Introduction

Traditional CFL instruction in Japan often emphasizes vocabulary and grammar drills over authentic communication, making it difficult for learners to perceive relevance or purpose in their studies. With the development of immersive technologies such as VR and the metaverse, new opportunities have emerged to reframe CFL learning environments. These tools allow learners to engage in spatial and embodied interaction, creating contexts where language can be experienced and constructed meaningfully (Sugie, 2025). This study explores how VR-supported task-based instruction can address these pedagogical gaps by fostering learner agency, contextual relevance, and multimodal communication.

This action research examines the following research questions:

- RQ1: How do CFL learners perceive the value of VR-based tasks in terms of meaningful learning?
- RQ2: What kinds of learning experiences, needs, and expectations arise through immersive task participation?

Literature Review

This study refers to the ARCS model of motivational design (Keller, 1983) and Self-Determination Theory, which supports autonomous learning (Deci & Ryan, 2009), as its analytical framework. The ARCS model is employed to organize the motivational elements of the VR tasks (Attention, Relevance, Confidence, Satisfaction). Self-Determination Theory is used to confirm how the fulfillment of learners' psychological needs for "Autonomy, Competence, and Relatedness" enhances their intrinsic motivation. From the perspectives of two theories, we explore the learning effects from multiple angles. The analysis of the course evaluation responses was structured around the core constructs of these two theories, using their key elements as analytical dimensions.

Methodology

The study involved 40 first-year students enrolled in a beginner-level CFL course in Fall 2024 at a private Japanese university. Most participants were

humanities majors with limited prior exposure to Chinese. Classes were held weekly for 90 minutes, over 15 sessions.

The instructional design was based on a blended learning course design, integrating both traditional and technology-enhanced components:

1. Textbook-based instruction for vocabulary and grammar, following a standard syllabus;
2. Pronunciation practice via LMS tools using automatic speech recognition to promote self-monitoring;
3. Worksheet-based activities focused on output, such as sentence construction and Q&A practice; and
4. Immersive VR tasks aligned with grammatical objectives, allowing students to apply their learning in simulated social scenarios.

Three original VR tasks were developed collaboratively by the teacher and students using platforms such as Spatial (for VR spaces), Tinkercad and Blender (for 3D modeling), and Monster Mash (for simple interactive objects). These tasks were as follows:

1. Party Preparation: Planning and shopping in Chinese.
2. World Travel, Photo Shooting, and Memory Presentation: Forming travel teams, choosing destinations, taking photos of memorable moments, and reporting travel experiences in Chinese.
3. Room Tidying: Giving and following spatial directions using directional complements.

Each task was aligned with specific grammar points such as “了” (completed action), “要” and “想” (desire and intention), and 趋向补语 (directional complements). Pre-task worksheets and post-task reflections were used to reinforce understanding.

The core of this instructional design is its process of co-creation by the teacher and students. Students are not mere consumers of resources provided by the teacher. Instead, the VR space creation and class activities are built upon student contributions, including their ideas, research, data collection, and added materials. Before the implementation of the classes, some students participated in this project as an extracurricular activity. Students were involved in the following phases:

1. Proposing the theme of “World Travel”: This included devising example sentences and dialogues for countries and regions they wished to visit, using previously learned vocabulary.

2. Conceptualizing and sourcing 3D assets: They searched for typical symbols or iconic landmarks for their travel destinations and collected the corresponding 3D assets.
3. Photographing the travel experience: They took photos to record their travel experiences during the VR trip.
4. Completing the gallery: The final step was to complete a gallery space by displaying their travel photos.

Findings

Learner responses collected through post-task and end-of-semester surveys indicated strong interest and positive emotional engagement with VR-based learning. Over 90% of respondents reported higher motivation and enjoyment. Most learners also agreed that VR tasks deepened their understanding of textbook content.

Qualitative responses emphasized that VR offered “realistic” and “fun” contexts for applying Chinese, lowering anxiety and encouraging active participation. Learners also valued being able to “move and speak naturally” with peers in a virtual space. However, several noted technical issues—including lag, device limitations, and operational difficulties—as barriers to smooth participation.

In terms of instructional needs, many students still desired face-to-face support for pronunciation correction and grammatical guidance. This points to the complementary role of human instruction even in tech-enhanced settings.

Discussion

VR facilitated active, student-centered learning by providing immersive and playful environments where learners could practice Chinese meaningfully. The findings support the relevance of ARCS model (Keller, 1983) and self-determination theory (Deci & Ryan, 2009) in explaining increased motivation.



Figure 1. VR spaces of world travel, photo shooting, and memory presentation.

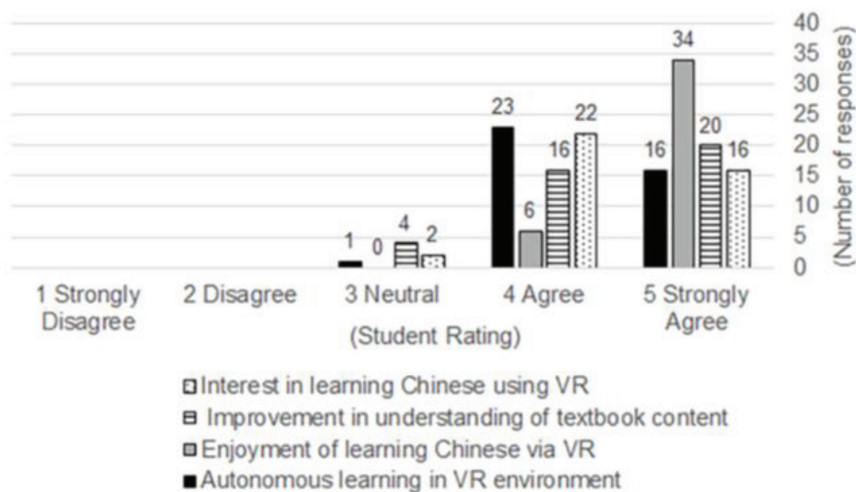


Figure 2. Learners' feedback regarding VR tasks.

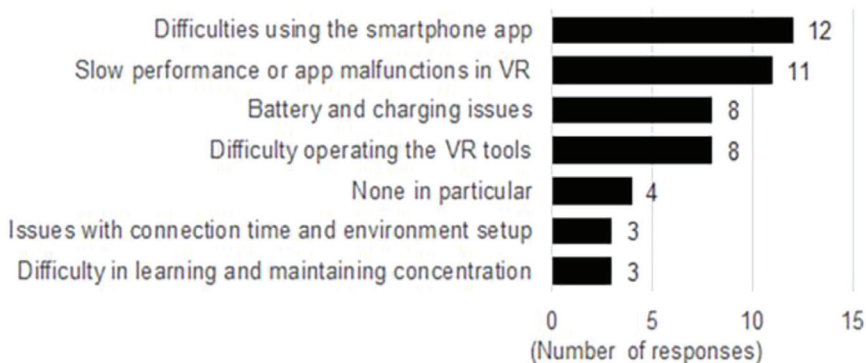


Figure 3. Technical issues and barriers related to VR-based learning.

Three key implications emerged:

1. Task Authenticity: Linking VR tasks to textbook grammar enabled students to apply language meaningfully, turning abstract rules into real-world use.
2. Hybrid Design: A blend of VR and face-to-face interaction proved essential. While VR stimulated engagement, learners still required in-person correction and support.
3. Technical Readiness: Infrastructure and digital literacy remain challenges. Instructors must scaffold VR participation and offer alternatives for learners struggling with devices.

Conclusion

This study demonstrates that VR-enhanced instruction can transform CFL learning into an immersive, co-creative process that fosters learner autonomy and engagement. However, effective implementation depends on balanced instructional design, technological support, and the strategic integration of face-to-face feedback. Future research should explore AI-based feedback systems, multimodal assessment methods, and the scalability of VR-enhanced instruction across diverse educational settings.

References

- Chen, J., & Kent, S. (2019). Task engagement, learner motivation, and avatar identities of struggling English language learners in the 3D virtual world. *System*, 87, 102168. <https://doi.org/10.1016/j.system.2019.102168>
- Deci, E. L., & Ryan, R. M. (2009). Self-determination theory: A macrotheory of human motivation, development, and health. *Canadian Psychology/Psychologie Canadienne*, 49(3), 182–185.
- Durak, G., & Cankaya, S. (Eds.). (2022). *Shaping the future of online learning: Education in the metaverse*. IGI Global. <https://doi.org/10.4018/978-1-6684-6513-4>
- Keller, J. M. (1983). Motivational design of instruction. In C. M. Reigeluth (Eds.), *Instructional design theories and models: An overview of their current status* (pp. 383–433). Lawrence Erlbaum Associates.
- Sugie, S. (2025). Design of multimodal foreign language education. In Proceedings of Language Education Expo 2025. [『マルチモーダルな外国語教育のデザイン』『言語教育エキスポ 2025 予稿集+』]

Understanding the Digital Literacies of English Major Students in Tertiary Education

Yong Tan*

Faculty of Education, University of Canterbury, New Zealand

Abstract

Due to the infusion of technology in education, second language acquisition is transforming enormously. While studies on technology-assisted language learning abound, research on the perceptions of digital literacies for those majoring in languages, especially English, remains scarce. This study investigated relationships between self-perceived digital literacies (knowledge, skills, attitudes), English reading comprehension, and demographics (academic years, geographic areas). 227 Chinese-background English major students from two universities completed a digital literacies questionnaire, a digital literacies test, and an English reading test. Results showed moderate-to-high self-perceived digital knowledge and skills, but no correlation was found between self-perceptions and actual digital competency. A positive correlation emerged between actual digital literacies and reading comprehension ($r = .219$, $p < .001$). Urban participants demonstrated higher self-perceptions and better scores in some skills. A notable perception-performance gap highlights potential overconfidence arising from daily digital practices versus guided academic application.

Keywords: Digital literacies, English as a foreign language, English major students, higher education, digital competency

Introduction

Digital technologies have been significantly integrated with everyday life, as well as English as a second language education. Conventional literacy skills, such as reading and writing, are no longer adequate to define a literate individual in

*Corresponding Author E-mail: yesqq55@163.com

the digital era (Miranda et al., 2018; OECD, 2021). Being digitally literate is not only beneficial for students' academic performance, but also better prepares them for a 21st-century workforce in a digital world (Khan et al., 2022; Martzoukou et al., 2022). Many countries and organizations stressed the pivotal position of digital literacy in Education. UNESCO underscores the application of information and communication technology (ICT) to improve the quality of education (UNESCO, 2022). Similarly, the European Union sets enhancing digital skills and competences of educators and learners as one of the priorities for digital transformation in education (European Commission, 2023). China also proposed the *China's Education Modernization 2035 Plan* and the *New Generation Artificial Intelligence Development Plan* to promote the digital transformation. As emphasized above, it is pivotal for developing countries to mitigate the widening digital divide through improving digital literacies.

Studies about integrating digital technologies into second or foreign language education have been widely conducted in the past decades. The introduction of technologies reshaped language acquisition both in pedagogy and classroom practice (Albadri et al., 2025). However, studies showed inconsistent results regarding digital literacy levels and students' backgrounds (Pinto et al., 2021; Thuy Nguyen & Habók, 2021). Particularly, few studies focus on whether language major students' digital literacies are in line with the availability of digital tools, since some studies suggest that students lack sufficient digital competence while being capable of accessing various digital tools. Therefore, this study aims to understand the digital literacies of students majoring in the English language through their self-perception and to explore the association between their backgrounds and digital literacies. As part of a doctoral study, two research questions were addressed in the main study here:

1. What are the self-perceived digital literacies of Chinese-background English major students?
2. What are the relationships regarding self-perceived digital literacies, actual digital competence and reading comprehension performance of Chinese-background English major students?

Methodology

A total of 227 English major students from two universities in China were included in this study. Students from year one and year three were recruited. The instruments employed included a self-perceived digital literacies questionnaire, a digital literacies test, and an English comprehension test. All

instruments were distributed to one grade at a time via a web-based platform, with the time-consuming and concentration-demanding measure commencing first, and then followed by the questionnaire and digital literacies test. After screening and cleaning, data were analysed by the statistical software package, SPSS version 29.

Results and Discussion

The descriptive analysis of the questionnaire showed that students tended to perceive their digital literacies knowledge as 'Medium' to 'Good'. Although a considerable variation was found in digital literacies skills, most students perceived confidence in skills in areas related to information searching and evaluation. As for the attitudes towards digital literacies in English reading and learning, most students showed a significant positive attitude towards engaging with digital technologies and tools, but varied by the type of technology or tools, and the specific learning context.

By applying the independent samples T-test, statistically significant differences were found in the theoretical knowledge (digital devices, operating system, networks, data safety; all $p < .05$), and skills in keyboard shortcuts and office software ($p < .001$), where third-year students reported having a higher level. However, no significant differences were found in attitudes. These results suggest a consistent positive attitude towards using digital technologies and tools in English reading and learning. A minimal difference between the two academic years was found from actual digital literacies test scores ($p = .080$).

The one-way analysis of variance (ANOVA) employed revealed no significant variance in students' perceptions of knowledge and attitudes towards digital literacies, except those urban students demonstrated stronger self-perceived digital literacies skills in handling software malfunctions ($p = .014$), searching ($p = .013$) and evaluating information ($p = .011$), reading from screens ($p = .005$) and accessing learning resources ($p < .001$). Notably, self-perceived digital literacies showed no significant correlation with the actual digital literacies and reading comprehension. However, a moderate correlation between the actual digital literacies test and reading comprehension test was observed ($r = .219$, $p < .001$).

The preliminary findings of this study are partly inconsistent with Monteiro and Leite's results (2021), which found no difference between students' academic year and digital literacies skills. However, the results of the current study indicate nuanced differences considering the factor of academic year. The misalignment of self-perception and actual performance of digital literacies, indicates that students often overestimate or underestimate their

digital competence, which confirms previous studies (e.g., Kruger & Dunning, 1999; Ng, 2012; Son et al., 2017). The significant positive correlation observed in the present study corroborates with studies exploring the impact of digital technologies on reading literacy, which emphasise that students' reading proficiency improved by using ICT tools in academic work (Lin et al., 2024; Xiao & Hu, 2019). To conclude, the findings of this study suggest that the latent discrepancy of digital literacies exists among English major students, which is easily neglected as they are 'digital natives' and majoring in social science (Prensky, 2001). This study highlights the importance of embedding digital literacies in the curriculum of language majors at the tertiary level and the continuum of future learning. Further insights into students' perceptions of digital literacies for language learning will be explored in the qualitative session of this study.

References

- Albadri, M. A. S., Dashtestani, R., & S. Yahya Hejazi. (2025). Investigating the digital literacy levels of Iraqi EFL university students: Challenges, opportunities and the status quo. *Wasit Journal for Human Sciences*, 21(2), 1099–1072. <https://doi.org/10.31185/wjfh.Vol21.Iss2.917>
- European Commission. (2023, November 23). *Digital Education Action Plan (2021-2027)—European Education Area*. <https://education.ec.europa.eu/focus-topics/digital-education/action-plan>
- Khan, N., Sarwar, A., Chen, T. B., & Khan, S. (2022). Connecting digital literacy in higher education to the 21st century workforce. *Knowledge Management & E-Learning: An International Journal*, 46–61. <https://doi.org/10.34105/j.kmel.2022.14.004>
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121–1134. <https://doi.org/10.1037/0022-3514.77.6.1121>
- Lin, L., King, R. B., Fu, L., & Leung, S. O. (2024). Information and communication technology engagement and digital reading: How metacognitive strategies impact their relationship. *British Journal of Educational Technology*, 55(1), 277–296. <https://doi.org/10.1111/bjet.13355>
- Martzoukou, K., Kostagiolas, P., Lavranos, C., Lauterbach, T., & Fulton, C. (2022). A study of university law students' self-perceived digital competences. *Journal of Librarianship and Information Science*, 54(4), 751–769. <https://doi.org/10.1177/09610006211048004>
- Miranda, P., Isaias, P., & Pifano, S. (2018). Digital literacy in higher education: A survey on students' self-assessment. In P. Zaphiris & A. Ioannou

- (Eds.), *Learning and collaboration technologies. Learning and teaching* (Vol. 10925, pp. 71–87). Springer International Publishing. https://doi.org/10.1007/978-3-319-91152-6_6
- Monteiro, A., & Leite, C. (2021). Digital literacies in higher education: Skills, uses, opportunities and obstacles to digital transformation. *Revista de Educación a Distancia (RED)*, 21(65). <https://doi.org/10.6018/red.438721>
- Ng, W. (2012). Can we teach digital natives digital literacy? *Computers & Education*, 59(3), 1065–1078. <https://doi.org/10.1016/j.compedu.2012.04.016>
- OECD. (2021). *21st-century readers: Developing literacy skills in a digital world*. OECD Publishing. <https://doi.org/10.1787/a83d84cb-en>
- Pinto, M., Sales, D., Fernández-Pascual, R., & Caballero-Mariscal, D. (2021). Exploring social sciences students' perceptions on information literacy and the use of mobile technologies in higher education. *College & Research Libraries*, 82(5), 662–682. <https://doi.org/10.5860/crl.82.5.662>
- Prensky, M. (2001). Digital natives, digital immigrants. *On the Horizon*, 9(5). <https://doi.org/10.1108/10748120110424816>
- Son, J.-B., Park, S.-S., & Park, M. (2017). Digital literacy of language learners in two different contexts. *The JALT CALL Journal*, 13(2), 77–96. <https://doi.org/10.29140/jaltcall.v13n2.208>
- Thuy Nguyen, L. A., & Habók, A. (2021). A study on Vietnamese undergraduates' level of digital skills and the frequency of using digital tools in the EFL context. *International Conference on Computer Systems and Technologies '21*, 171–175. <https://doi.org/10.1145/3472410.3472425>
- UNESCO. (2022). *Guidelines for ICT in education policies and masterplans*. <https://unesdoc.unesco.org/ark:/48223/pf0000380926>
- Xiao, Y., & Hu, J. (2019). The moderation examination of ICT use on the association between Chinese mainland students' socioeconomic status and reading achievement. *International Journal of Emerging Technologies in Learning (ijET)*, 14(15), 107. <https://doi.org/10.3991/ijet.v14i15.10494>

Beyond Technical Competencies: A Critical Analysis of Global Research on Language Teacher AI Literacy

Danping Wang¹  and Qianwei Zhang¹ 

¹University of Auckland, New Zealand

Introduction

Teacher AI literacy is a critical enabler of educational transformation, influencing how technology is interpreted, adapted, and evaluated through human judgment in the classroom. As defined in the UNESCO AI Competency Framework for Teachers (Miao & Cukurova, 2024), teacher AI literacy refers to the knowledge, skills, attitudes, and values that allow educators to critically understand, use ethically, and integrate AI in ways that strengthen human agency and support inclusive education. In recent years, there has been a surge in AI literacy frameworks and guidelines issued by educational organisations worldwide (Zhang et al., 2025). Yet, these initiatives often frame teacher AI literacy primarily in technical competencies, underemphasising its pedagogical, ethical, and societal dimensions.

China's ambitious national policies to embed AI throughout its education system (Roberts et al., 2021), from primary schools to universities, are already reshaping pedagogical models, digital platforms, and language proficiency standards. These changes carry global implications for the teaching of Chinese as a Foreign Language (CFL). For CFL teachers teaching Chinese outside of China, comprehensive AI literacy is increasingly essential to align with evolving policy priorities and innovation agendas set in the home country (Yang et al., 2023). Yet research assessing teachers' readiness and capabilities remains limited (Li & Liang, 2025). The gap between rapid, policy-driven adoption and the absence of robust evaluative frameworks underscores the urgent need for deeper investigation.

E-mail: danping.wang@auckland.ac.nz & qzha455@aucklanduni.ac.nz

Method

To address these gaps, this study presents a systematic review of 79 peer-reviewed articles published between January 2020 and December 2024. It offers a critical analysis of current research focuses and suggests new directions for reconceptualising teacher AI literacy.

Two major academic databases, Scopus and Web of Science, were used for the search. The search terms were “artificial intelligence literacy” OR “AI literacy” AND “teacher” OR “AI readiness”. Articles containing these keywords in their title or abstract were initially screened. To minimise selection bias, a defined set of inclusion and exclusion criteria was applied. The inclusion criteria were as follows: (1) written in English; (2) open access; (3) full-text availability; (4) classified as research articles; and (5) published between 2020 and 2024. The exclusion criteria are: (1) systematic reviews; (2) brief publications such as commentaries, editorials, or workshop papers, which often lack sufficient detail or represent subjective viewpoints; and (3) articles that mentioned AI literacy but did not specifically address teachers’ AI literacy, such as those focusing on other forms of literacy (e.g., data or digital literacy) or on students’ AI literacy. A thematic analysis was then conducted to identify key research trends, conceptual gaps, and implications for teacher AI literacy in the context of language education.

Findings

Four major findings emerged from this review, highlighting the need for a more critical and context-sensitive understanding of teacher AI literacy, particularly for language educators.

First, most studies remain narrowly focused on why, what, and how to use AI tools ($n = 21$). The reviewed literature is concentrated in East Asia (51%), particularly China, followed by North America (15%) and Europe (8%), with large-scale quantitative surveys (64%) dominating the field. These surveys often measure general teacher acceptance of AI or assess digital competencies, even though many studies describe AI literacy as a holistic concept ($n = 12$). Across the reviewed work, AI knowledge, AI use, and AI ethics emerge as three recurring dimensions in AI literacy frameworks. While this body of research contributes to foundational understanding, it frequently overlooks subject-specific needs, especially the pedagogical and epistemological complexities unique to language education.

Secondly, existing AI literacy research is often confined to single cases in university contexts ($n = 19$) or short-term online programmes ($n = 6$). Influencing

factors identified include age, educational background, training experience, and personal attitudes toward technology. These studies are generally limited in duration and broad in scope, most often functioning as one-off tool-driven experiments in a particular type of language classes such as writing (Li et al., 2025). Although language teachers often express positive attitudes toward AI, their ability to integrate it meaningfully and sustainably into classroom practice remains limited, as effective integration requires more than digital skills and an open mindset.

Thirdly, although conceptualisations of AI literacy are plentiful, few studies are supported by reliable, validated assessment tools. Most rely on self-reported Likert-scale instruments ($n = 40$), which lack cross-cultural robustness and long-term applicability (Younis, 2025). Many are further constrained by small sample sizes, limited theoretical grounding, and minimal consideration of ethical or cultural adaptability. As a result, we found that existing body of research on teacher AI literacy remains largely exploratory.

Finally, within the set review period, no studies ($n = 0$) were found that specifically address CFL teaching, with most instead focusing on English language teaching. Tools and frameworks tailored for CFL teachers remain scarce (Li & Liang, 2025). AI is often treated as a generic technological innovation, with little recognition of the linguistic and cultural specificities essential to CFL instruction, such as character recognition, tonal pronunciation, and culturally sensitive content generation. While calls to strengthen CFL teacher AI literacy are increasing, practical and contextually grounded strategies remain underdeveloped.

Taken together, these findings indicate that while global research has laid some groundwork for understanding teacher AI literacy, there is a pressing need to move beyond its current boundaries.

Discussion

The findings of this review reveal persistent challenges in advancing AI literacy among language teachers, particularly within Chinese language education. Theoretical frameworks remain underdeveloped, existing assessment tools lack subject specificity, and classroom practices continue to lag behind technological developments. Moreover, current approaches frequently overlook the sociocultural and linguistic dimensions essential to meaningful language teaching both within and outside of China.

One methodological limitation of this review is that the search was conducted using only English-language terms, which have excluded relevant studies published in Chinese or other languages. Addressing this limitation

in future research would provide a more complete and representative picture of the field.

Going forward, there is a pressing need to develop AI literacy frameworks that are both linguistically and culturally grounded. Such frameworks must recognise teachers' complex pedagogical work, respond to their specific teaching realities, and go beyond surface-level engagement with tools. They should be embedded in a broader, more nuanced understanding of teacher AI literacy, one that foregrounds human agency, critical pedagogy, and the safeguarding of high-quality, personally meaningful, community-oriented language teaching in the age of AI.

Conclusion

As AI becomes embedded in education systems worldwide, there is growing consensus that teacher AI literacy must be reconceptualised. It cannot be reduced to mastering tools or passively accepting technological change. Instead, it must centre on critical understanding, pedagogical agency, and ethical responsibility, with the broader aim of driving transformative and sustainable change in education. For language teachers, this means developing AI literacies that are as nuanced, adaptive, and human as the work they undertake. Without such a shift, AI risks being absorbed as yet another digital tool, rather than fulfilling its potential as a catalyst for profound, lasting educational transformation.

References

- Li, N., & Liang, Y. (2025). Teachers' AI readiness in Chinese as a foreign language education: Scale development and validation. *System*, 103597 <https://doi.org/10.1016/j.system.2025.103597>
- Li, B., Tan, Y. L., Wang, C., & Lowell, V. (2025). Two years of innovation: A systematic review of empirical generative AI research in language learning and teaching. *Computers and Education: Artificial Intelligence*. <https://doi.org/10.1016/j.caeai.2025.100445>
- Miao, F., & Cukurova, M. (2024). *AI competency framework for teachers*. UNESCO Digital Library. <https://www.unesco.org/en/articles/ai-competency-framework-teachers>
- Roberts, H., Cows, J., Morley, J., Taddeo, M., Wang, V., Floridi, L. (2021). The Chinese approach to artificial intelligence: An analysis of policy, ethics, and regulation. In: Floridi, L. (eds) *Ethics, governance, and policies in artificial intelligence*. Philosophical studies series, vol 144. Springer, Cham. https://doi.org/10.1007/978-3-030-81907-1_5

- Yang, T., Xu, T., & LI, Q. (2023). Construction, demonstration and optimization of the intelligent literacy indicator system for international Chinese-language teachers. *Journal of Yunnan Normal University Teaching and Research on Chinese as a Foreign Language Edition*, 21(3), 41–51. <https://doi.org/10.16802/j.cnki.ynsddw.2023.03.014>
- Younis, B. (2025). The artificial intelligence literacy (AIL) scale for teachers: A tool for enhancing AI education. *Journal of Digital Learning in Teacher Education*, 41(1), 37–56. <https://doi.org/10.1080/21532974.2024.2441682>
- Zhang, K., Li, L., & Yu, H. C. (2025). Analysis on concept of AI literacy according and future education to the result of systematic review. *Journal of Lifestyle and SDGs Review*, 1, e03293. <https://doi.org/10.47172/2965-730X.SDGsReview.v5.n01.pe03293>

Evaluating Advanced EFL Learners' Writing Performance and Perceptions of Large Language Model Assistance

Mengkai Wang* 

Beijing Foreign Studies University, China

Abstract

Since the release of Large Language Models (LLMs), they have been widely used in language teaching and acquisition. Previous studies have explored LLMs' assistance in improving language performance, but their impact on advanced English Foreign Language learners' writing remains unclear. This study conducts an experiment based on their IELTS or TOEFL scores. After a pre-test to exclude participants who could not follow instructions, they were randomly divided into two groups. Participants in the experiment group were instructed to complete the essay with guidance but not to directly generate it. Following the task, participants completed a motivation and perception questionnaire. Surprisingly, results show that LLMs did not significantly improve writing performance in advanced EFL learners. However, the questionnaire revealed strong positive perceptions regarding LLM utility and motivation. Additionally, it appears that the current level of LLMs matches the writing capabilities of advanced learners. This study explores how LLMs affect EFL performance and learner perceptions, offering insights into their potential assistance.

Keywords: EFL learners, writing performance, LLMs, learner perception, motivation

Introduction

The rapid integration of large language models (LLMs) such as ChatGPT, into educational contexts has generated widespread interest regarding their utility

*Corresponding Author E-mail: tayloricy@bfsu.edu.cn

for second language acquisition. However, the effectiveness of LLM assistance for advanced English as a Foreign Language (EFL) learners remains relatively underexplored. In line with the conference theme of “Inclusive CALL”, ensuring that educational technologies are truly inclusive requires understanding their utility across the entire spectrum of learner proficiency. Advanced learners, a group often overlooked in the design of support tools, represent a critical demographic for this investigation.

Therefore, this study addresses this research gap through a dual-focus approach. It primarily examines whether ChatGPT-assisted writing significantly improves the objective writing performance of advanced Chinese EFL learners. Concurrently, it explores learners’ subjective perceptions of LLM utility and motivation, aiming to bridge the potential gap between perceived usefulness and measurable performance outcomes.

Literature Review

Previous research predominantly focuses on intermediate-level learners, suggesting notable benefits from LLM assistance in grammar accuracy, lexical development, and overall writing proficiency (Barrot, 2023; Song & Song, 2023). These learners often find LLM-generated feedback and suggestions highly valuable for overcoming specific linguistic hurdles.

However, the context for advanced learners is theoretically different. Characterized by high proficiency levels, they typically operate near the upper bounds of their Zone of Proximal Development (ZPD), as conceptualized by sociocultural theory (Vygotsky, 1978). In this framework, learning is most effective when external support aids a learner in tasks slightly beyond their independent capability. For advanced learners who already possess strong autonomous writing skills, the scaffolding provided by current LLMs might offer only limited benefits (Liu et al., 2023). Consequently, they might derive less significant cognitive advantages from LLMs compared to their less proficient counterparts. This highlights a potential bifurcation in LLM impact: while cognitive gains may be minimal for this group, the affective benefits, such as increased motivation or reduced writing anxiety, could still be substantial.

Methodology

A randomized controlled design was employed, involving 50 advanced Chinese EFL learners recruited from university English programs. All participants met the advanced proficiency criteria (IELTS score ≥ 7.5 or TOEFL score ≥ 105).

Prior to the task, all participants in the experimental group received a 15-minute standardized training session. This session demonstrated how to use ChatGPT for brainstorming, outlining, and receiving feedback on grammar and vocabulary, while explicitly instructing them not to use it for generating complete sentences or paragraphs.

Participants were randomly assigned into two groups: Experiment Group and Control Group. Both groups completed a 1-hour argumentative essay task on contemporary topics. Essays were evaluated by three raters, an AI scoring engine, a trained researcher, and a certified IELTS instructor, according to IELTS assessment criteria, including fluency, grammatical accuracy, and lexical and grammatical complexity. In addition to writing performance assessments, participants in the experimental group completed a post-task questionnaire measuring motivation and perceptions of LLM utility. The questionnaire responses were analyzed using principal component analysis (PCA) to uncover underlying dimensions of learners' attitudes toward LLMs.

Results

Descriptive statistics suggested only minimal differences between groups, as shown in Table 1. Although the experimental group showed slightly higher scores in grammar and language use, independent samples t-tests revealed no statistically significant differences in any dimension ($p > .05$). Similarly, ANCOVA, controlling for pre-task self-rated writing confidence, confirmed the absence of significant group effects ($p = .35$). PCA results for the post-task questionnaire identified three principal components: non-motivation, motivation, and perceived assistance. Figure 1 presents a summary of the relationships among questionnaire items and components. Participants reported positive perceptions of LLM assistance. Mean ratings were high for motivation and assistance, suggesting learners felt that ChatGPT boosted their confidence and reduced writing anxiety. Nevertheless, no correlation was found between these positive perceptions and actual performance improvements.

Discussion

The findings of this study align with sociocultural theory predictions concerning the Zone of Proximal Development. Since advanced learners already possess strong autonomous writing skills, the scaffolding provided by ChatGPT offered minimal additional advantage. This lack of significant performance improvement contrasts with studies focusing on intermediate-level learners, where LLMs have been shown to yield more pronounced gains (Barrot, 2023;

Table 1. Descriptive statistics for writing outcomes

Rating Aspect	Descriptive Statistics			
	Group	Mean	Std. D	Std. E
Overall	Experiment	92.21	2.28	0.46
	Control	91.47	2.39	0.48
Content	Experiment	22.79	1.30	0.26
	Control	22.87	1.08	0.22
Grammar	Experiment	22.97	1.15	0.23
	Control	22.82	1.32	0.26
Language Use	Experiment	23.26	1.13	0.23
	Control	22.85	1.32	0.26
Coherence and Cohesion	Experiment	23.20	0.97	0.19
	Control	22.94	1.20	0.24

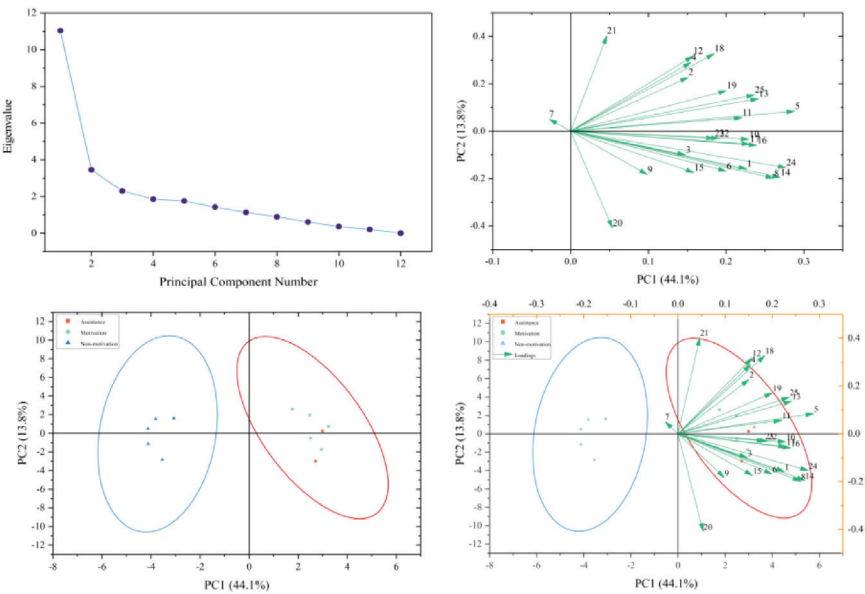


Figure 1. Visualized principal component analysis results.

Song & Song, 2023). Our results suggest that the “one-size-fits-all” application of LLMs may not be effective across different proficiency levels.

While real-time feedback and language suggestions were appreciated, these tools may not have sufficiently challenged learners to catalyze measurable growth. However, the positive motivational effects reported by participants

resonate strongly with earlier studies (Hwang et al., 2023; Song & Song, 2023), indicating that LLMs can enhance learners' emotional engagement even when cognitive gains are limited. This suggests a bifurcated role of AI in language education: improving affective variables without necessarily driving immediate skill acquisition for highly proficient learners.

Moreover, the slight grammatical and lexical advantages observed in the experiment group, although not statistically significant, hint at a possible cumulative benefit over longer durations. Given the short-term nature of this study, it remains plausible that extended interaction with LLMs could yield meaningful improvements for advanced learners. Importantly, the disconnect between learners' perceived benefits and actual performance outcomes raises critical pedagogical questions. While high satisfaction and motivation are valuable, educators must balance these factors with realistic expectations about LLMs' capacity to foster measurable writing improvements at higher proficiency levels. To be truly inclusive, CALL tools must be designed or implemented in a proficiency-sensitive manner.

Several limitations of the current study warrant attention. The homogeneous sample (Chinese advanced EFL learners) limits generalizability across cultural and linguistic backgrounds. Furthermore, the short intervention period may not capture the longitudinal effects of sustained LLM usage.

Conclusion

This study contributes to the emerging literature on AI applications in second language writing by demonstrating that, for advanced EFL learners, ChatGPT-assisted writing primarily enhances motivational factors rather than producing significant immediate improvements in writing performance. These findings highlight the need for proficiency-sensitive design of AI educational tools and call for continued exploration of the long-term impacts of AI integration in language learning. By focusing on a frequently overlooked learner group, this research underscores the importance of a nuanced, proficiency-aware approach to creating truly inclusive CALL environments.

References

- Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, 57, Article 100745. <https://doi.org/10.1016/j.asw.2023.100745>
- Liu, M., Zhang, L. J., & Biebricher, C. (2023). Investigating students' cognitive processes in generative AI-assisted digital multimodal composing

- and traditional writing. *Computers & Education*, 211, Article 104977. <https://doi.org/10.1016/j.compedu.2023.104977>
- Song, C., & 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*, 14, Article 1260843. <https://doi.org/10.3389/fpsyg.2023.1260843>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Artificial Intelligence-Driven Personalized Language Learning: Customizing Content and Feedback to Learners' Needs and Proficiency Levels

Mengkai Wang* and Yifei Gao

Beijing Foreign Studies University, China

Abstract

Artificial Intelligence (AI)-driven personalized language learning holds great promise for customizing content and feedback according to learners' needs and proficiency levels. This study explores the potential of AI, particularly large language models (LLMs), for enhancing personalized language learning experiences. The research examines how LLMs can be leveraged to provide dynamic, tailored content and real-time feedback, adapting to individual learners' progress and learning preferences. A group of 100 language learners participated in a 6-week experiment where they engaged with AI-powered platforms that adjusted the content's difficulty and offered personalized feedback based on their performance and proficiency levels for English grammar and vocabulary acquisition. The study used a randomized controlled trial design, assessing learners' language proficiency, motivation, engagement, and perceived effectiveness of the AI-assisted learning experience before and after the intervention. The results showed that LLMs significantly improved learners' language proficiency, with noticeable increases in their engagement and motivation. Learners reported higher satisfaction with the personalized feedback and content, appreciating the system's ability to adapt to their unique needs. However, challenges such as occasional misinterpretations by the AI and limited content variety were identified. This study contributes to

*Corresponding Author E-mail: tayloricy@bfsu.edu.cn & smxgaoyf@126.com

understanding the effectiveness of AI in creating personalized language learning environments and highlights the potential of integrating such technologies into mainstream language education.

Keywords: AI-driven, feedback, learner needs, proficiency levels, language learning, personalization

Introduction

Personalized learning has been a long-standing goal in education, particularly in language acquisition, where individual differences in proficiency, motivation, and learning strategies are profound (Dong et al., 2023). Traditional classroom-based instruction often struggles to accommodate these individual needs effectively. Recent advancements in artificial intelligence (AI), especially the development of large language models (LLMs) like GPT-4, have opened new possibilities for tailoring learning experiences dynamically to each learner (Prandi et al., 2023). LLMs are capable of generating human-like text, adapting content, and providing real-time feedback based on the users' input, making them particularly suited to personalized language learning environments.

Despite the excitement surrounding these technologies, empirical research into their effectiveness in educational settings, particularly in language learning, remains limited. Zheng et al. (2025) emphasize the need for robust experimental studies to explore how AI-driven platforms can be integrated into mainstream education to enhance learning outcomes. This study addresses this gap by investigating the impact of LLMs on language learners' proficiency, engagement, and motivation over a 6-week intervention.

Methodology

The study employed a randomized controlled trial design to assess the effectiveness of AI-driven personalized learning platforms. A total of 100 participants, aged 18 to 35, were identified as intermediate (CEFR B1 level) learners of English as a foreign language and recruited from a university language program. The CEFR (Common European Framework of Reference for Languages) is an international standard used to describe language proficiency levels, ranging from A1 (beginner) to C2 (mastery). The participants were randomly assigned to either an experimental group (AI-assisted learning) or a control group (traditional online learning without AI adaptation). Participants in the experimental group interacted with an AI-powered platform designed to adjust the content's difficulty and provide

personalized feedback based on real-time performance data. The platform, leveraging a model based on the GPT-4 architecture, was fine-tuned on a large corpus of learner texts with expert annotations to improve its accuracy in detecting grammatical errors and generating feedback. Learners completed adaptive grammar exercises (e.g., gap-filling, sentence transformations) and vocabulary-in-context tasks. In contrast, the control group used standard online language learning materials with static content. They accessed a fixed library of exercises covering the same grammatical and lexical topics as the experimental group. Feedback was limited to pre-programmed, generic responses, such as revealing the correct answer with a brief rule explanation. Pre- and post-intervention assessments were conducted using standardized language proficiency tests measuring grammar and vocabulary knowledge, self-reported motivation and engagement scales, and surveys on learners’ perceptions of the learning experience. Data were analyzed using mixed analysis of variance to assess changes over time between the two groups.

Results

An analysis of the proficiency scores, which measured grammar and vocabulary knowledge, is summarized in Table 1.

A significant group $\times$ time interaction effect was found ($F(1,98) = 23.45, p < 0.001$), indicating that participants in the AI-assisted group improved significantly more than those in the traditional learning group. Additionally, learners in the AI-assisted group reported higher levels of engagement and motivation. Figure 1 illustrates the changes in self-reported motivation scores.

Qualitative feedback revealed that learners appreciated the system’s ability to adapt to their performance, with comments highlighting the relevance of the exercises and the helpfulness of immediate, tailored feedback. However, several participants noted instances where the AI misinterpreted their responses, leading to occasional frustration.

Table 1. Descriptive statistics for pre- and post-test mean between groups

Group	Pre-Test Mean (SD)	Post-Test Mean (SD)	Gain Score (SD)
AI-assisted learning	72.3 (8.4)	82.7 (7.1)	10.4 (3.9)
Traditional learning	71.8 (7.9)	76.5 (7.5)	4.7 (3.5)

SD, standard deviation.

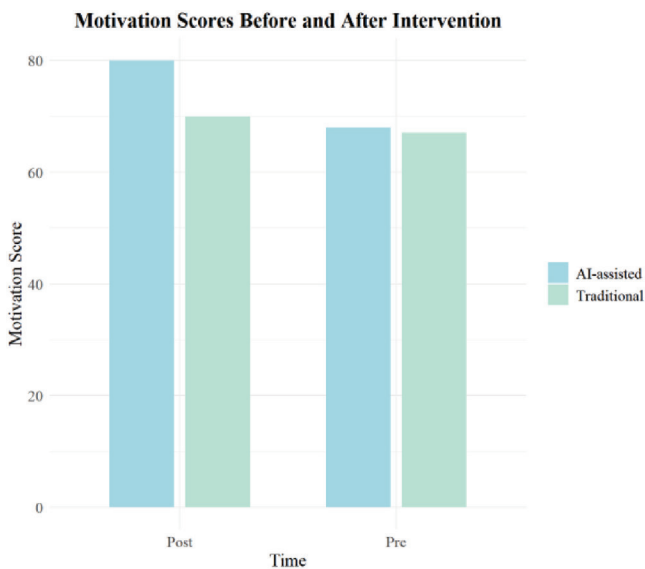


Figure 1. Motivation scores pre- and post-intervention for AI-assisted and traditional groups.

Discussion

The results of this study support the growing belief that AI-driven personalized learning can significantly enhance language acquisition outcomes. The substantial gains in proficiency scores and the heightened engagement and motivation reported by the participants align with prior theoretical expectations about the benefits of personalization (Christodoulou & Angeli, 2022). One major strength of the AI-powered system was its responsiveness to individual progress. Learners felt that the exercises were “just right” in difficulty, a hallmark of effective scaffolding in educational theory (Vygotsky, 1978). The enhanced proficiency can be attributed to this adaptive difficulty combined with personalized feedback; immediate corrections and tailored suggestions helped the learners consolidate their knowledge and address errors more efficiently than static learning materials.

Nonetheless, the study also uncovered challenges. Misinterpretations by the AI, although relatively infrequent, disrupted the learning flow for some participants (e.g., Ji et al., 2023). This highlights a limitation of current LLM-based systems: while powerful, they are not infallible and require careful tuning and perhaps hybrid human-in-the-loop models to mitigate errors. Moreover, the content variety offered by the LLM system was somewhat limited, as noted

by several participants. Learners reported that although the specific sentences and words were novel, the underlying exercise structures (e.g., multiple choice, gap-filling) were sometimes repetitive, which could impact long-term engagement (O'Neill & Satar, 2024). Future systems might incorporate larger and more diverse corpora to enhance content richness.

Conclusion

This study demonstrates the promising potential of LLMs in creating personalized language learning environments that effectively enhance learners' proficiency, engagement, and motivation. Proficiency was enhanced by providing learners with tailored feedback on specific errors and adaptively challenging tasks that aligned with their individual learning pace. While there are challenges to address, particularly regarding misinterpretations and the content's diversity, the findings suggest that the integration of AI in language education could be a transformative development. Future research should explore longer intervention periods, different age groups, and hybrid models combining AI and human feedback to further refine and optimize personalized language learning experiences.

References

- Christodoulou, A. & Angeli, C. (2022). Adaptive learning techniques for a personalized educational software in developing teachers' technological pedagogical content knowledge. *Frontiers in Education*, 7, 789397. <https://doi.org/10.3389/educ.2022.789397>
- Dong, B., Bai, J., Xu, T., & Zhou, Y. (2023). Large language models in education: A systematic review. In *Proceedings of the 6th International Conference on Computer Science and Technologies in Education* (pp. 131–134). <https://doi.org/10.1109/CSTE62025.2024.00031>
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12), 1–38. <https://doi.org/10.1145/3571730>
- O'Neill, E., & Satar, M. (2024). Exploring the impact of task repetition on learner engagement in chatbot-mediated language learning. *Computer Assisted Language Learning*, 37(3), 214–235. <https://doi.org/10.1080/09588221.2023.xxxxxxx>
- Prandi, C., Valva, A., Salomoni, P., Roccetti, M. & Mirri, S. (2023). Incidental language and culture learning through mobile technologies: A multi-case

study. *Multimedia Tools and Application*, 82, 46039–46063. <https://doi.org/10.1007/s11042-023-16095-1>

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Zheng, Y. B., Zhou, Y. X., Chen, X. D., & Ye, X. D. (2025). The influence of large language models as collaborative dialogue partners on EFL English oral proficiency and foreign language anxiety. *Computer Assisted Language Learning*, 1–27. <https://doi.org/10.1080/09588221.2025.2453191>

Developing College Students' Creative Thinking and English Story Writing Through AI-supported Collaborative Learning

Shih-Chiao Wang^{1*} and Yu-Fen Yang¹

¹National Yunlin University of Science and Technology, Yunlin, Taiwan

Abstract

Many English as a foreign language (EFL) students struggle to generate the ideas, express their emotions, and use English to write stories due to the lack of creativity in traditional or teacher-centered writing environment. This study aimed to investigate the effects of AI-supported collaborative learning on EFL college students' creative thinking and English story writing. A total of 38 EFL college students from two classes, with 16 in the experimental group (AI-supported collaborative writing) and 22 in the control group (traditional writing instruction). The results showed the experimental group improved creativity in fluency, flexibility, and elaboration. The control group made progress in fluency and flexibility; however, effect sizes were small. AI-supported collaborative writing includes story chains, word choice, grammar checking, chatbots, peer feedback and evaluation, which promote students' creative thinking and enhance their English story writing. This study suggests that AI-supported collaborative learning helps college students develop their creative thinking and English story writing.

Keywords: EFL college student, creative thinking, English story writing, AI-supported collaborative learning

*Corresponding Author E-mail: M11341015@yuntech.edu.tw, yangy@yuntech.edu.tw

Study Background

Traditional or teacher-centered writing often restricts creative thinking by prioritizing grammatical rules over idea generation (Kawinkoonlasate, 2019). Creativity was defined as fluency, flexibility, originality, and elaboration (Guilford, 1950), which foster students' curiosity, collaboration, risk-taking, and deeper learning, enhancing both academic success and second language acquisition outcomes (Smare, 2022). Story writing serves as a powerful approach for EFL learners to express language, ideas, emotions, and imagination in English. It is the process of taking control of one's creative thinking to produce compelling and dynamic writing (Fitria, 2024). It is not only a language practice but also regarded as an essential skill in education, as it enables college students to express and communicate, which reflects on one's creative thinking (Cheung et al., 2003). For students who struggle with creativity or lack fluency, AI-supported collaborative learning could assist students by generating story elements such as characters, plots, and settings, while also suggesting sentence structures and improving coherence (Biermann et al., 2022; Bai et al., 2021). In this study, AI-supported refers to the integration of digital tools such as story chain, word choice, grammatical checking, chatbot, evaluation and peer feedback, which assist learners in developing ideas, refining language use, and enhancing text coherence. Collaborative writing is defined as a process in which two or more learners jointly plan, draft, and revise a shared text through interaction, negotiation, and mutual feedback (Storch, 2013) to produce good writing. It enhances college students' various learning outcomes (Fang et al., 2023) and improves their creativity, writing skills, engagement, and sense of satisfaction (Chung et al., 2022; Suh & An, 2022; Yuan et al., 2022). Accordingly, this study aimed to investigate the effects of AI-supported collaborative learning on EFL college students' creative thinking and English story writing. The two research questions are as follows.

RQ1. What are the effects of AI-supported collaborative writing on EFL students' creative thinking?

RQ2. What are the effects of AI-supported collaborative writing on EFL students' story writing?

Methodology

38 EFL college students from two classes participated in this study. One class was randomly assigned to the experimental group using an AI-supported collaborative writing method ($N = 16$), while the other class was control group

using traditional collaborative writing method ($N = 22$). Both groups received the same teacher instruction which included creative thinking and story writing exercises, lasting for 18 weeks. The experimental group added AI-supported collaborative learning (story chain, word choice, grammatical checking, chatbot, and evaluation) while the control group using traditional or teacher-centered writing method. Data was collected from the pre- and post-Creative Thinking Scales adapted from Lehmkuhl et al. (2021), English story writing first and final drafts and an open-ended questionnaire.

Result

Figure 1 presents the before and after drafts of English story writing from the experimental group. The experimental group received peer feedback, additionally used AI-supported collaborative learning (story chain, word choice, grammatical checking, chatbot, evaluation and peer feedback) during the process, afterward college students discussed and decided whether to revise their story or proceed to the next step.

Table 1 presents the paired samples t-test for both groups across four dimensions of creativity. In the experimental group, fluency, flexibility, and elaboration improved significantly, with flexibility showing the highest effect size ($d = 0.96$). The control group showed modest gains, with only flexibility ($d = 0.48$) and fluency ($d = 0.54$) reaching significance. These findings suggest that AI-assisted collaborative writing has a positive effect on developing students' creative thinking, particularly in flexibility.

Experimental group's sample 1: before-after version of English story writing			
Process: receiving peer feedback and AI chatbot suggestion		Before and After version	
2025-04-10 by B11341051 Maybe add more conversation between Yi-Jun and her husband to emphasize their incompatible marriage. 2025-04-10 by B11141073 Good plot. 修正建議 對話建議 Can u add some possible dialogues between y-i-jun and her husband? Certainly! Here are some possible dialogue snippets between Yi-Jun and her husband that could help illustrate their relationship: "Some after a particularly rough argument, Yi-Jun confronts her husband about his behavior." "Yi-Jun:" "This isn't how I imagined marriage would be. I feel like I'm walking on eggshells around you. Why can't we talk normally?" "Husband:" "Maybe if you didn't make everything so complicated! You think you're better than everyone else because you go to university!" "Yi-Jun:" "That's not true! I'm just trying to better myself for us, to provide a better future. But I need your support, not your anger!"	peer feedback AI chatbot	Before version As the years passed, Yi-Jun found love—or at least, what she thought was love. She got married, believing she had found a partner who would support her as much as her family had. But things took a dark turn. Her husband had a terrible temper, one that flared at the smallest inconveniences. At first, it was just sharp words, but soon, the anger became something more—constant yelling, cruel remarks, and a presence that suffocated Yi-Jun's once-bright spirit. AI suggestion "This isn't how I imagined marriage would be. I feel like I'm walking on eggshells around you. Why can't we talk normally?" "Maybe if you didn't make everything so complicated! You think you're better than everyone else because you go to university!" Her husband slammed the door of the bedroom. "That's not true! I'm just trying to better myself for us, to provide a better future. But I need your support, not your anger!" Yi-Jun cried desperately. Her family, once so full of laughter, now lived in quiet fear of the man who had become a part of their lives. Her mother tried to talk to her. "Yi-Jun, you don't have to endure this." But Yi-Jun only smiled weakly. "He's just stressed. It'll get better." But deep down, she knew the truth.	After version

Figure 1. An example of English story writing

Table 1. The experiment group and control group's creativity before and after

Dimension	Groups	N	Pre-survey test		Post-survey test		Paired samples t test		
			M	SD	M	SD	t	p	Effect size (d)
Originality	Exp.	16	29.69	5.63	31.56	4.23	-1.297	.214	0.32
	Con.	22	30.05	4.59	32.91	6.30	-2.735	.012	0.58
Fluency	Exp.	16	24.19	4.37	27.81	4.23	-2.897	.011	0.72
	Con.	22	25.55	3.56	28.45	6.11	-2.520	.02	0.54
Flexibility	Exp.	16	17.25	3.02	20.75	2.79	-3.834	.002	0.96
	Con.	22	18.55	2.86	20.23	4.17	-2.256	.035	0.48
Elaboration	Exp.	16	22.19	4.56	24.69	3.38	-2.2	.044	0.55
	Con.	22	23.95	3.66	24.95	5.03	-1.062	.301	0.23

For the open-ended questionnaire, students in the control group reported that paper-based collaborative writing facilitated the sharing of diverse opinions, joint story construction, and collective reflection on areas for improvement. Some expressed that learning was enjoyable because peers' ideas inspired them and fostered growth. However, challenges included the need to merge differing perspectives, sometimes without agreement, and difficulties in fully understanding each other's viewpoints. In contrast, students in the experimental group highlighted that online collaborative writing enabled comparison of writing abilities and styles, allowing them to identify preferred approaches for further development. They emphasized that collaboration fostered creativity by integrating diverse perspectives and imaginations, which expanded potential storylines. Several noted that using AI appropriately inspired new ideas and encouraged experimentation with different story beginnings and character settings, enhancing writing engagement. Moreover, the AI story chain prompted students to transcend habitual thinking patterns by generating unexpected responses and introducing novel scenarios. In one case, the AI enriched the narrative through multiple character dialogues, thereby promoting flexibility and enabling students to refine their stories beyond initial expectations.

Discussion and Conclusion

This study demonstrated that compared to traditional writing instruction, AI-supported collaborative writing effectively promoted EFL students' creative thinking, particularly in fluency, flexibility, and elaboration. The use of AI tools encouraged diverse perspectives and fostered dynamic storytelling.

The students not only learned through peer interaction but also benefited from immediate feedback, idea generation, and language support. The findings suggest that AI-supported collaborative writing can effectively promote students' creative thinking and English story writing.

References

- Bai, B., Wang, J., & Zhou, H. (2022). An intervention study to improve primary school students' self-regulated strategy use in English writing through e-learning in Hong Kong. *Computer Assisted Language Learning*, 35(9), 2265–2290. <https://doi.org/10.1080/09588221.2020.1871030>
- Biermann, O. C., Ma, N. F., & Yoon, D. (2022, June). From tool to companion: Storywriters want AI writers to respect their personal values and writing strategies. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference* (pp. 1209–1227). <https://doi.org/10.1145/3532106.3533506>
- Cheung, W. M., Tse, S. K., & Tsang, H. W. (2003). Teaching creative writing skills to primary school children in Hong Kong: Discordance between the views and practices of language teachers. *The Journal of Creative Behavior*, 37(2), 77–98. <https://doi.org/10.1002/j.2162-6057.2003.tb00827.x>
- Chung, J. J. Y., Kim, W., Yoo, K. M., Lee, H., Adar, E., & Chang, M. (2022, April). TaleBrush: Sketching stories with generative pretrained language models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (pp. 1–19). <https://doi.org/10.1145/3491102.3501819>
- Fang, X., Ng, D. T. K., Leung, J. K. L., & Chu, S. K. W. (2023). A systematic review of artificial intelligence technologies used for story writing. *Education and Information Technologies*, 28(11), 14361–14397. <https://doi.org/10.1007/s10639-023-11741-5>
- Fitria, T. N. (2024). Creative writing skills in English: Developing student's potential and creativity. *EBONY: Journal of English Language Teaching, Linguistics, and Literature*, 4(1), 1–17. <https://doi.org/10.37304/ebony.v4i1.10908>
- Guilford, J. P. (1950). Creativity. *American Psychologist*, 5(9), 444–454. <https://doi.org/10.1037/h0063487>
- Kawinkoonlasate, P. (2019). *A comparative study of e-writing and traditional writing classroom to improve English writing ability and motivate autonomous learning of Thai EFL learners*. Huachiew Chalermprakiet University. <https://doi.org/10.30845/ijll.v6n2p4>
- Kim, H. J., Gray, S., & Lange, C. (2024). Instructional guidance for promoting creativity in English as a foreign language writing classroom: A Korean case study. *RELC Journal*, 55(1), 79–95. <https://doi.org/10.1177/00336882211066624>

- Lehmkuhl, G., von Wangenheim, C. G., Martins-Pacheco, L. H., Borgatto, A. F., & Alves, N. D. C. (2021). SCORE—A model for the self-assessment of creativity skills in the context of computing education in K-12. *Informatics in Education*, 20(2), 231. <https://doi.org/10.15388/infedu.2021.11>
- Mansoor, A. (2010). Teaching creative writing to university level second language learners in Pakistan. *New Writing*, 7(3), 201–218. <https://doi.org/10.1080/14790726.2010.514052>
- Marcos, R. I. S., Fernández, V. L., González, M. T. D., & Phillips-Silver, J. (2020). Promoting children's creative thinking through reading and writing in a cooperative learning classroom. *Thinking Skills and Creativity*, 36, 100663. <https://doi.org/10.1016/j.tsc.2020.100663>
- Smare, Z. (2022). The need to engage students' creative thinking in language learning classes. *TESOL and Technology Studies*, 3(2), 18–33. <https://doi.org/10.48185/tts.v3i2.644>
- Storch, N. (2013). *Collaborative writing in L2 classrooms* (Vol. 31). Multilingual Matters. <https://doi.org/10.2307/jj.27080067>
- Suh, S., & An, P. (2022, March). Leveraging generative conversational AI to develop a creative learning environment for computational thinking. In *Companion Proceedings of the 27th International Conference on Intelligent User Interfaces* (pp. 73–76). <https://doi.org/10.1145/3490100.3516473>
- Wang, A. Y. (2012). Exploring the relationship of creative thinking to reading and writing. *Thinking Skills and Creativity*, 7(1), 38–47. <https://doi.org/10.1016/j.tsc.2011.09.001>
- Yuan, A., Coenen, A., Reif, E., & Ippolito, D. (2022, March). Wordcraft: story writing with large language models. In *Proceedings of the 27th International Conference on Intelligent User Interfaces* (pp. 841–852). <https://doi.org/10.1145/3490099.3511105>

Ethical Principles for Generative AI in Research Paper Writing for EFL Graduate Students in China

Zeng Jianbin  and Zhang Xuhua 

Fudan University, China

Abstract

With DeepSeek and ChatGPT taking the lead in AI technology breakthroughs and applications, AI-assisted writing Apps have demonstrated marvelous efficiency and potential in academic writing for EFL graduate students, sparking concerns over academic misconduct, lack of originality, and privacy risks, and posing severe challenges to academic ethics in universities. Unfortunately, it still remains a universally unsettled debate regarding the accessibility, usability, and inclusivity of AI-assisted writing tools, leading to disparate positions of hesitation, rejection, and inclusion. This paper examines the pros and cons of AI-assisted academic writing and proposes three principles: originality, transparency, and accountability, to guide graduate schools in establishing guidelines for the use of AI in academic paper writing for EFL graduate students in China. It is recommended that the AI-assisted academic writing guidelines focus on AI assistance rather than overindulgence, AI use disclosure and disclaimer, and AI misuse prevention and punishment. The guidelines aim to proactively address the academic ethical challenges brought by AI technological innovation and fill the regulatory gap in AI-assisted academic paper writing for EFL graduate students in China. Our work represents an AI-assisted and CALL-motivated transition from our prior practices and its positive impacts on equity in language education, contributing to the promotion of institutional efforts in providing equitable AI access for EFL graduate students in a more inclusive language learning environment.

Keywords: AI policy, research paper writing, EFL graduate students

E-mail: jbzeng@fudan.edu.cn & zhangxuhua@fudan.edu.cn

AI innovation, such as DeepSeek and ChatGPT, is now transforming the landscape of language education globally, particularly in academic writing. In an experiment involving seven Harvard University general education course papers in the humanities and social sciences, Bodnick (2023) informed the course instructors and teaching assistants beforehand that the submitted course papers (1–9 pages) might be either written by the students themselves or automatically generated by ChatGPT-4, and the professors and TAs provided positive and objective feedback during the grading process on papers that were, in fact, entirely generated by ChatGPT-4. As revealed in Table 1, ChatGPT 4 performed well in the take-home writing assignments, achieving an average grade of A- at Harvard. It is noted that AI writing is challenging traditional university academic writing instruction and assessment, demonstrating potentials for widespread application, especially for EFL graduate students in China, where English is the *de facto* lingua franca of academic paper writing and publication, holding a privileged position in academic evaluation.

A relevant survey (Figure 1) conducted at Fudan University in China revealed that 97% of the 92 participating graduate students—enrolled in an English research paper writing course across humanities, social sciences, and STEM fields—used AI in their academic brainstorming, literature survey, language enhancement, data analysis, editing and reviewing. Their AI tools included DeepSeek, ChatGPT, Grammarly, DeepL, Kimi, Doubao, ERNIE Bot, AI Taiyan, ChemDraw, GitHub Copilot, Connected Papers, Research Rabbit, SemiKong, LaTeX, Writefull, Sider, Claude-Sonnet, etc., while disputes remain concerning the extent and regulations of AI assistance in their academic writing process. An exceptional minority of the participants in the survey (3%) did not use AI due to disciplinary restrictions (e.g., Chinese classics studies) and advisory prohibitions (e.g., modern Chinese history). A significant portion of the participants (35%) explicitly expressed serious concerns about

Table 1. Harvard university GhatGPT-4 course paper transcript (Brodnick, 2023)

No.	Course title	Grade
1	Microeconomics	A-
2	Macroeconomics	A
3	Latin American Politics	B-
4	American Presidency	A
5	Conflict Resolution	A
6	Intermediate Spanish	B
7	Proust Seminar	P

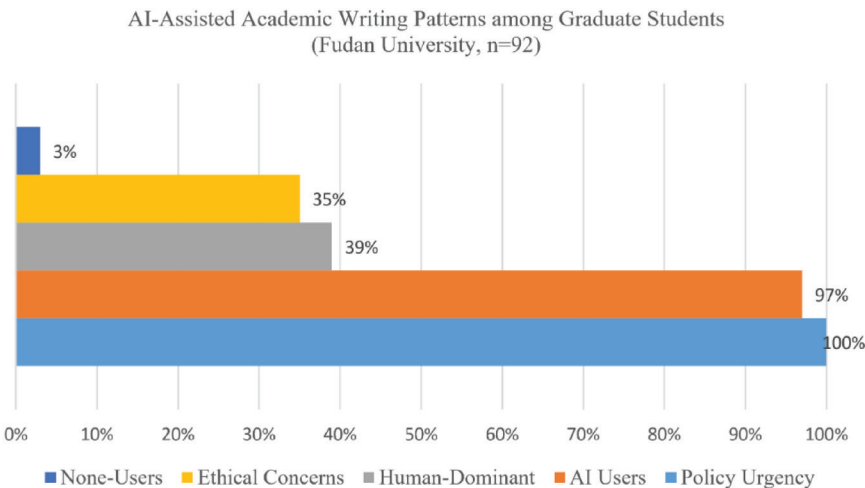


Figure 1. AI-Assisted academic writing patterns among graduate students (Fudan university, n = 92).

AI over-dependence, accessibility disparities and related ethical risks, whereas a larger group (39%) proactively favored a human-dominant approach, critically reviewing their AI-generated content to uphold academic integrity. Unanimously, all participants (100%) emphasized the urgent need for clear AI guidelines in English research paper writing, which are unfortunately lacking in the graduate schools in higher education institutions(HEIs) in China, who are struggling to keep up to date with the rapid pace of AI innovation. (Zeng et al. 2025).

Notably in both Harvard and Fudan cases (Table 1 & Figure 1), AI demonstrates remarkable efficiency in information retrieval, text generation, editing/translation, and data processing. However, AI dependency has sparked concerns over academic misconduct, lack of originality, and privacy risks, compromising scholarly rigor and ethical standards. As best captured in recent peer-reviewed literature (Moorhouse et al. 2025, Rong et al. 2025, Zou et al. 2025), AI has profoundly reshaped the landscape of writing, presenting both opportunities and challenges for L2/EFL writers, especially at advanced academic levels. To make matters worse, it remains a universally unsettled debate regarding the accessibility, usability, and inclusivity of AI as a collaborative or cheating tool, leading to controversial stances of cautious observation, outright ban, and regulated application among graduate students, professors, and educational administrators in graduate schools. Overall, HEIs remain reactive rather than proactive in AI ethical guidance and academic policy development.

To address the controversy over the disruptive impact of AI technologies on academic integrity, when policy development has been outpaced by technological innovation, this study proposes three principles: originality, transparency, and accountability, to facilitate HEIs in establishing AI-assisted writing guidelines. First, originality prioritizes the individual initiative, independent discretion and human authorship of all academic work in topic selection, study design, data analysis, drafting and revision, etc., with AI serving merely as an auxiliary instrument, marking distinct boundaries between AI assistance and replacement. Second, transparency requires the explicit and full disclosure of AI contributions (Appendices 1 & 2, Zeng et al. 2025), specifying the AI tools used, purpose, scope and time of AI assistance, with a highlighted statement acknowledging the author's full responsibility for content accuracy and validity. Third, accountability is ensured by updating AI ethical guidance and academic policy, discipline-specific practices and training programs, including workshops on ethical AI use and detection techniques, writing tutorials, and faculty/peer evaluation requirements, grading penalties for undisclosed AI use, and disciplinary punishments for repeated violations, etc.

Consequently, this study aims to proactively address the disruptive ethical challenges brought by AI innovation and fill the regulatory gap in AI-assisted academic writing for EFL graduate students in China. It is hence urged that AI guidelines and policies will be immediately developed in HEIs, graduate schools, discipline-specific schools, departments and courses, focusing on AI assistance rather than overindulgence, AI use disclosure and disclaimer, and AI misuse prevention and punishment. In other words, HEIs are expected to supervise and regularize AI-assisted writing practices while preserving academic integrity and empowering original research. Graduate schools are recommended to update AI-assisted writing protocols, providing AI disclosure templates and detection resources, and utilizing global networks for AI policy evaluation and improvement. School and department efforts are further encouraged to customize and optimize the AI policies and practices in discipline-specific research programmes. Course-specific requirements are also recommended in AI-assisted writing regulations, mandatory peer-review, writing tutorial and oral defense processes, transparent AI usage statements and originality checks, etc. These ethical principles, guidelines and practices for generative AI in academic writing will empower EFL graduate students in China, as well as students in Harvard and other HEIs to use AI ethically, transforming technological disruption into an opportunity for academic advancement.

In summary, we present an AI/CALL-driven framework for equitable AI integration in academic writing, addressing institutional gaps in ethical principles and academic policy to accommodate AI-assisted English

research paper writing for linguistically underprivileged EFL learners. The ethical guidelines and institutional efforts recommended are expected to accommodate diverse linguistic backgrounds, language proficiency, learning styles, and accessibility needs, empowering all learners equitably by fostering a more inclusive academic ecosystem globally. We also recognize that selection bias, representativeness, or response framing within the Fudan survey remain limitations, and future efforts, such as reporting key demographic distributions within the survey sample, refining sampling frame and method, and sharing data and materials to address these issues, would enhance the transparency and value of the study.

References

- Bodnick, M. (2023). ChatGPT goes to Harvard and does better than you might think! Retrieved May 30, 2025, from <https://www.slowboring.com/p/chatgpt-goes-to-harvard>
- DeepSeek. (2025). Response to author query. Retrieved May 30, 2025, from <https://chat.deepseek.com>
- Moorhouse, B. L., Wan, Y., Wu, C., Wu, M., & Ho, T. Y. (2025). Generative AI tools and empowerment in L2 academic writing. *System*, 133, article 103779. <https://doi.org/10.1016/j.system.2025.103779>
- Rong, M., Yao, Y., Li, Q., & Chen, X. (Winnie). (2025). Exploring student engagement with artificial intelligence-guided chatbot feedback in EFL writing: Interactions and revisions. *Computer Assisted Language Learning*, 1–30. <https://doi.org/10.1080/09588221.2025.2539979>
- Zeng, J., Liao, W., Zeng, M., & Zeng, Y. (2025). Guidelines for English academic writing in the AI era. *Shanghai Graduate Education*, 1, 30–35.
- Zou, M., Kong, D., & Lee, I. (2025). Doctoral student's strategy use in GAI chatbot-assisted L2 writing. *Journal of English for Academic Purposes*, 76, article 101521. <https://doi.org/10.1016/j.jeap.2025.101521>

Appendices

1. No AI tool was used in the writing of this term paper.
2. The author thanks DeepSeek-V3 (2025) for providing language improvement during the final editing phase. The final version of this manuscript represents my own academic insights and research efforts, with the AI's contribution limited to language enhancement.

Using Large Language Models to Enhance Data-driven Learning in Academic Writing

Xuhua Zhang¹  and Jianbin Zeng¹ 

¹Fudan University, China

Introduction

Academic writing instruction in English for Specific Purposes (ESP), particularly for students in science and engineering disciplines, increasingly incorporates genre-based approaches to help learners understand the rhetorical structure and discourse conventions of research writing (Swales, 1990; Hyland, 2004). To support learner autonomy and disciplinary appropriateness, many instructors supplement genre-based instruction with corpus-informed methods such as Data-Driven Learning (DDL) (e.g., Charles, 2007), which encourages learners to discover authentic usage patterns from language corpora (Johns, 1991; Boulton & Cobb, 2017).

While the effectiveness of DDL has been widely acknowledged across educational levels (O’Keeffe, 2021), its classroom implementation often faces practical challenges. These include the cognitive and technical demands of corpus tools, learners’ limited proficiency in interpreting concordance data, and the time constraints inherent in classroom settings (Ädel, 2010). To respond to these challenges, recent advances in artificial intelligence—especially large language models (LLMs) such as DeepSeek, ChatGPT and KIMI—offer new opportunities for supporting writing instruction.

LLMs can provide context-sensitive feedback, diagnose language errors, and assist with paraphrasing, citation checking, and sentence-level revision (Fathi & Rahimi, 2024; Liu et al., 2024; Wang et al. 2025). These capabilities are not merely theoretical; recent classroom-based research demonstrates their practical benefits and highlights the need for AI literacy. A randomized controlled trial involving 259 Chinese undergraduates compared LLM-generated feedback with traditional instructor feedback and found that

E-mail: zhangxuhua@fudan.edu.cn & jbzeng@fudan.edu.cn

students receiving AI feedback showed significantly greater improvements in organization and content development (Zhang, 2025). Another classroom study used structural equation modelling to show that AI literacy and self-regulated learning positively predicted writing performance and digital well-being among university students in generative-AI-supported courses (Shi, Liu & Hu, 2025). These findings, alongside earlier CALL-focused trials (e.g., Fathi & Rahimi, 2024; Liu et al., 2024), underscore both the potential of LLMs to enhance revision and the need for deliberate instruction in AI literacy. AI literacy is now regarded as an extension of digital literacy, encompassing the ability to understand, apply, evaluate and ethically engage with AI technologies (Mansoor et al., 2024). These perspectives informed our integration of LLMs into a genre-based, DDL-supported curriculum. Moreover, there remains a need for empirical classroom-based research to evaluate how LLMs can be meaningfully integrated into genre-based and corpus-assisted academic writing instruction.

This study reports on a two-year teaching practice in an “Academic English for Science and Engineering” course that primarily adopted a genre-based instructional approach supplemented by DDL activities. In a subset of class sections, this combined approach was further enhanced with the integration of LLMs. Focusing on synthesis writing tasks, this study compares student performance between classes using corpus-assisted instruction alone and those additionally supported by LLMs, examining the impact on learners’ ability to identify and revise linguistic and citation-related errors.

Methods

This study was conducted over five consecutive semesters (Spring 2023 to Spring 2025) in an “Academic English for Science Engineering” course offered to undergraduate and graduate students majoring in science and engineering at a leading Chinese university. A total of 673 students from 23 classes participated in the study, all following a genre-based curriculum supplemented by Data-Driven Learning (DDL) methods. Thirteen classes (384 students) used DDL techniques such as analyzing collocations and citation patterns with corpus tools to improve academic writing. The remaining ten classes (289 students), including two at the graduate level, followed the same instructional sequence but also incorporated LLMs, such as DeepSeek, ChatGPT, and KIMI, to support language analysis, error detection, and revision tasks.

In both groups, synthesis writing was introduced in the middle of the semester through model analysis and guided exercises. A key component of the instruction involved corpus consultation using the Corpus of Contemporary

American English (COCA)(Davies, 2009). Students were guided to use COCA's online interface to investigate collocational patterns, phrase structures, and citation-related usage (e.g., reporting verbs, hedging expressions). Specific tasks included searching for high-frequency academic collocations, identifying field-appropriate verbs for integrating sources, and comparing grammatical patterns across registers. These hands-on DDL activities aimed to raise students' awareness of academic language and inform their revisions during writing tasks.

In the LLM-enhanced classes, students completed the same DDL tasks but were also encouraged to use AI tools to diagnose and revise language and citation issues in model synthesis texts. Instructors provided guidance on prompt formulation, AI feedback evaluation, and cross-referencing with corpus findings to validate or challenge AI suggestions.

To evaluate the impact of integrating LLMs into DDL academic writing instruction, we analyzed student performance on a synthesis writing quiz administered one week after classroom practice. The quiz was designed to assess students' ability to identify and correct pre-embedded errors in a synthesis text. It embedded thirteen deliberate errors covering four categories: (1) language mistakes (e.g., grammatical and lexical errors), (2) citation-related issues—including redundant or misformatted reference information, (3) logical inconsistencies in the argument or structure, and (4) mechanical errors such as improper punctuation and spacing. Students received one point for each correctly identified error; identifying ten or more earned full credit.

Data were thus collected to compare performance between the DDL-only and AI-supported groups. Scoring was based on the number of correct identifications. A score of 10 indicated full marks.

Results and Discussion

Across the five semesters, 13 classes (384 students) received genre-based writing instruction with corpus-assisted DDL activities (Referred to as DDL only in Table 1), while 10 classes (289 students) received the same instruction with LLM integration (Referred to as DDL+LLMs in Table 1). The average scores for the corpus-assisted only group ranged from 4.91 to 6.17, with a mean around 5.57. In contrast, the LLM-enhanced group scored between 6.12 and 7.89, with a higher mean of approximately 6.91 (Table 1).

These results suggest that LLM-enhanced instruction led to significantly better performance in error identification tasks, particularly in recognizing subtle citation formatting issues and phraseological inconsistencies that novice

Table 1. Overview of classes, enrollment, and quiz performance by instructional condition

Semester	Instructional Condition	No. of Classes	No. of Students	Mean Quiz Scores
Spring 2023	DDL only	3	83	6.12, 5.16, 5.3
Spring 2023	DDL+LLMs	2	59	6.32, 7.2
Fall 2023	DDL only	3	80	5.12, 5.41, 6.02
Fall 2023	DDL+LLMs	2	60	6.89, 7.43
Spring 2024	DDL only	3	95	6.17, 5.44, 6.0
Spring 2024	DDL+LLMs	2	65	7.14, 7.89
Fall 2024	DDL only	2	60	5.19, 4.91
Fall 2024	DDL+LLMs	2	50	6.54, 6.12
Spring 2025	DDL only	2	66	5.71, 6.01
Spring 2025	DDL+LLMs	2	55	7.21, 7.4
Total	DDL only	13	384	5.57 (avg. across classes)
Total	DDL+LLMs	10	289	6.91 (avg. across classes)

writers often overlook. Students using LLMs could access immediate, context-sensitive feedback, allowing them to test hypotheses, revise iteratively, and internalize accurate academic language patterns more efficiently.

From an instructional perspective, the integration of LLMs addressed two major limitations of traditional DDL: the cognitive load of corpus navigation and the time constraints of in-class feedback. While DDL fosters analytical thinking and pattern discovery, it can be overwhelming for students with limited corpus literacy. LLMs acted as scaffolding agents, enabling broader student engagement with complex tasks such as synthesis writing and source integration.

Importantly, the LLMs did not replace DDL but complemented it. Students were still encouraged to use corpora for verification and to compare with AI output. This hybrid approach maintained the pedagogical rigor of DDL while improving efficiency and inclusivity.

These findings support the use of LLMs as pedagogically meaningful enhancements to corpus-informed writing instruction, particularly for synthesis writing tasks that demand accuracy, integration, and stylistic control.

Conclusion

This study demonstrates that augmenting a genre-based, corpus-assisted academic writing curriculum with LLMs can substantially improve students' ability to identify and correct linguistic and citation-related errors. Over five semesters and 673 participants, classes that used LLM support scored appreciably higher on a synthesis-writing quiz than those relying solely on corpus-based activities. Learners benefitted from immediate, context-sensitive feedback, which enabled iterative revision and deeper engagement with complex academic language patterns. Crucially, LLMs did not replace traditional DDL practices but complemented them, reducing the cognitive load of corpus navigation and providing scaffolding for learners at varying proficiency levels. These findings underscore the promise of hybrid AI-assisted approaches for more inclusive and efficient writing instruction.

However, several limitations must be acknowledged. First, data were collected across five consecutive semesters; although the curriculum was standardized, variations in teaching style, digital proficiency and feedback practices could have influenced student performance. Second, effective use of COCA and LLMs presupposes a level of digital and AI literacy that is unevenly distributed among learners. AI literacy—viewed as an extension of digital literacy—includes understanding, applying and critically evaluating AI technologies. Variations in students' familiarity with prompt engineering or corpus navigation may therefore have contributed to performance differences. Third, group assignment was based on intact class sections rather than randomization, leaving open the possibility of pre-existing differences in language proficiency, writing experience or motivation. Fourth, the synthesis quiz focused on error identification in a single task and may not fully capture sustained improvements in rhetorical control or synthesis skills. Future research should employ randomized class assignments, incorporate longitudinal writing portfolios, and use more comprehensive assessments to evaluate writing development.

References

- Ädel, A. (2010). Using corpora to teach academic writing: Challenges for the direct approach. In O'Keeffe, A., & McCarthy, M. (Eds.), *The Routledge handbook of corpus linguistics* (pp. 371–382). Routledge.
- Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, 67(2), 348–393. <https://doi.org/10.1111/lang.12224>

- Charles, M. (2007). Reconciling top-down and bottom-up approaches to graduate writing: Using a corpus to teach rhetorical functions. *Journal of English for Academic Purposes*, 6(4), 289–302. <https://doi.org/10.1016/j.jeap.2007.09.009>
- Davies, M. (2009). The 385+ million word Corpus of Contemporary American English (1990–2008+): Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics*, 14(2), 159–190. <https://doi.org/10.1075/ijcl.14.2.02dav>
- 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. <https://doi.org/10.1080/09588221.2024.2374772>
- Hyland, K. (2004). *Genre and second language writing*. University of Michigan Press.
- Johns, T. (1991). Should you be persuaded: Two examples of data-driven learning materials. *ELR Journal*, 4, 1–16.
- Liu, Z. M., Hwang, G. J., Chen, C. Q., Chen, X. D., & Ye, X. D. (2024). Integrating large language models into EFL writing instruction: effects on performance, self-regulated learning strategies, and motivation. *Computer Assisted Language Learning*, 1–25. <https://doi.org/10.1080/09588221.2024.2389923>
- O’Keeffe, A. (2021). Data-driven learning—a call for a broader research gaze. *Language Teaching*, 54(2), 259–272. <https://doi.org/10.1017/S0261444820000523>
- Mansoor, H. M., Bawazir, A., Alsabri, M. A., Alharbi, A., & Okela, A. H. (2024). Artificial intelligence literacy among university students—a comparative transnational survey. *Frontiers in Communication*, 9, 1478476.
- Shi, J., Liu, W., & Hu, K. (2025). Exploring how AI literacy and self-regulated learning relate to student writing performance and well-being in generative AI-supported higher education. *Behavioral Sciences*, 15(5), 705.
- Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.
- Wang, Y., Zhang, T., Yao, L., & Seedhouse, P. (2025). A scoping review of empirical studies on generative artificial intelligence in language education. *Innovation in Language Learning and Teaching*, 1–28.
- Zhang, K. (2025). Enhancing critical writing through AI feedback: A randomized control study. *Behavioral Sciences*, 15(5), 600.