

Quarterly journal publishing

Online ISSN: 2790-2110



WORLD SPORTS
UNIVERSITY PRESS
世界體育大學出版社



JOURNAL OF EDUCATIONAL TECHNOLOGY
AND INNOVATION

Scilight

No Challenge
No Change

JETI

JOURNAL OF EDUCATIONAL TECHNOLOGY AND INNOVATION

教育技术与创新

Volume 08 Issue 3, September 2026

Volume 08
Volume 08
Volume 08
Volume 08
Volume 08

Volume 08 Issue 3, 30 September 2026

Table of Contents 目录

Journal of Educational Technology and Innovation JETI

Table of Contents

	Page
1. A Teaching Strategy Based on Observation, Experience and Inquiry Research on Promoting Students' Conceptual Change	1–8
2. Empowerment or Negativeenergy?—Opportunities, Challenges and Reflections of Preschool Teachers under the Application of Artificial Intelligence	9–19
3. Technology-Mediated Individual and Collaborative L2 Academic Writing on Learner Autonomy	20–34
4. Transforming Learning through Productive Failure: A Multi-Dimensional Framework for Fairness in Education	35–51
5. Between Paperwork and Performance: Students' Experiences of Assessment in Industry-Education Integration in Chinese Higher Vocational Education	52–63
6. How Teacher Learning Affects Middle School Students' Learning Motivation and Habits: An Empirical Study Based on PLS-SEM	64–79
7. Do AI Writing Tools Improve L2 Writing Performance? A Meta-Analysis of Experimental Evidence, Moderators, and Methodological Quality	80–103

Article

A Teaching Strategy Based on Observation, Experience and Inquiry Research on Promoting Students' Conceptual Change

Youyou Zhang ¹, Yu Zhao ², Xiaonan Luo ², Danting Deng ², Junjie Yan ², Jing Liu ² and Wenhua Lai ^{2,*}

¹ School of Information Technology in Education, South China Normal University, Guangzhou 510631, China

² School of Education, Wenzhou University, Wenzhou 325035, China

* Correspondence: laiwenhua@wzu.edu.cn

How To Cite: Zhang, Y., Zhao, Y., Luo, X., Deng, D., Yan, J., Liu, J., & Lai, W. (2026). A Teaching Strategy Based on Observation, Experience and Inquiry Research on Promoting Students' Conceptual Change. *Journal of Educational Technology and Innovation*, 8(3), 1–8. <https://doi.org/10.61414/4sge9605>

Received: 19 March 2026

Revised: 28 August 2026

Accepted: 11 September 2026

Published: 30 September 2026

Abstract: The 2022 “Compulsory Education Primary School Science Curriculum Standards” clearly stipulates the specific requirements for students to conduct scientific inquiry in primary school science classes. Based on the experiential learning theory and the concept change theory, the “Observation-Experience-Inquiry” (OEI) teaching strategy was proposed. Taking cell-concept teaching as an example, this strategy guides students to revise their prior misconceptions driven by cognitive conflicts. With tools including structural diagrams and collaborative discussion, students can accomplish scientific conceptual change and develop scientific ideas.

Keywords: primary school science; concept change; experiential learning; observation-experience-inquiry

1. Introduction

The word “inquiry,” when used as a verb, means “to explore, investigate, or study.” When used as a noun, it refers to the exploration and study of natural phenomena and laws through scientific methods and processes. In everyday life, people often conduct investigations or practical activities, such as exploring the conditions under which iron rusts. In scientific research, when scientists discover phenomena that cannot be explained by their existing knowledge, they conduct investigations based on scientific hypotheses. This process is known as scientific inquiry. Scientific inquiry is a systematic and step-by-step process of exploration and practice based on scientific theories and a series of rigorous inquiry procedures directed toward research questions (Kunisch et al., 2023). Since China’s basic education curriculum reform in 2001, scientific inquiry has been included in the primary school science curriculum standards as an important component of science education. Scientific inquiry provides a general framework for raising questions, formulating conjectures and hypotheses, collecting evidence, and drawing conclusions, thereby providing students with a structured direction for scientific exploration (Dewey, 1910; Schwab, 1958).

On this basis, Klahr and Dunbar (1988) proposed the Scientific Discovery as Dual Search model, which conceptualizes scientific inquiry as coordinated searches within a hypothesis space and an experiment space. This model provides a theoretical explanation of the cognitive processes involved in scientific inquiry.

Pedaste et al. (2021) developed a scientific inquiry test for primary school students based on the inquiry-based learning framework proposed by Pedaste et al. (2015). The test encompassed the essential skills aligned with the different phases of inquiry-based learning. These included analytical skills, which are primarily required in the orientation, conceptualization, and investigation phases; planning skills, mainly needed in the investigation phase; and interpretation skills, primarily needed in the conclusion and discussion phases. Chen et al. (2025) examined the effects of ChatGPT- and argument map-supported online argumentation on students’ critical thinking skills



Copyright: © 2026 by the authors. This is an open access article under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Publisher’s Note: Scilight stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

and perceptions, providing insights into the cultivation of scientific reasoning and evaluation skills. These studies offer useful insights into scientific observation, experience, inquiry, and evaluation.

Previous research has established a close relationship between scientific inquiry and conceptual change. Zimmerman (2007) suggested that scientific thinking involves inquiry, experimentation, evidence evaluation, and inference, which contribute to the formation and revision of scientific concepts. However, existing studies provide limited guidance on how these processes can be organized into a coherent classroom sequence to promote primary school students' conceptual change. Therefore, drawing on experiential learning theory and conceptual change theory, this study proposes an Observation–Experience–Inquiry (OEI) teaching model. Specifically, literature analysis was used to clarify the theoretical foundations and construct the OEI model, while teaching case analysis was used to illustrate its classroom application in the teaching of cell concepts.

2. Theoretical Background

2.1. Experiential Learning Theory

The “experience” in pedagogy mainly includes two aspects: one is to regard experience as a way of learning for students, that is, the study of experiential learning; The second is the research on experiential teaching, which regards experience as the mode for teachers to carry out teaching activities.

Experiential Learning theory is a teaching strategy driven by experience and centered on students, which is suitable for students' “learning” (Kolb & Kolb, 2009). It places students' learning in real cognitive situations, helping them complete learning tasks by creating situations that conform to the learning content, thereby experiencing the process of knowledge construction, behavioral evolution, and dynamic emotional development. It is an important means to promote the occurrence of deep learning effectively. Experiential learning theory results from John Dewey's “experience learning” theory, and British educational practices initially adopted it as a way of independent learning and development.

On this basis, American scholar David A. Kolb systematically proposed A complete theory in “Experiential Learning: Making Experience the Source of Learning and Development” and pointed out the characteristics of the experiential learning theory: (1) Process-oriented learning experience. (2) The cognitive development of dialectical opposition. (3) Adaptive multilateral interactivity. (4) Build an integrated learning cycle. These characteristics provide scientific guidance for educational practice (Murrell & Claxton, 1987). Based on these characteristics, the application of experiential learning theory in teaching has gradually expanded, promoting the development of experiential teaching.

Experiential teaching, with the core of stimulating students' learning experiences, is a teaching strategy that guides students to engage in experiential learning (Luo et al., 2023). The research on experiential teaching abroad has a long history, traced back to Socrates' “midwifery”, Aristotle's theory of sensory cognition, and Rousseau's naturalistic education. With the continuous development of experiential teaching, its application scope is constantly expanding. For instance, Spiegel (1975) advocated the “role-playing method”, enabling students to experience and learn through interaction by simulating real situations. This greatly enriches the practical forms of experiential teaching.

2.2. Conceptual Change Theory

Conceptual Change is when learners (students) abandon their original incorrect or incomplete concepts during the learning process and accept and internalize new scientific concepts (Özdemir & Clark, 2007). Its essence is to replace the preconceptions or misconceptions with a scientific concept to achieve a more accurate and deeper understanding of a specific knowledge field or phenomenon. Conceptual Change Model (CCM) was proposed by Posner et al. (1982) from Cornell University in the United States. This model is based on epistemology, Kuhn's theory of scientific revolution (Kuhn, 1970), and Piaget's theory of cognitive development (Piaget, 1976).

(1) Cognitive conflict stage: The concept change begins when students realize a contradiction between their original and scientific concepts, generating cognitive conflicts (Limón, 2001). This conflict prompts students to re-examine and question their preconceived concepts, triggering conceptual change. Research shows that during the stage of cognitive conflict, cooperative communication is regarded as an effective means to trigger cognitive conflict (Zhang & Zhang, 2022). It reveals the irrationality of preconceptions through peer interaction and the collision of viewpoints, causing cognitive conflict.

(2) Concept reconstruction stage: Concept change is not merely a simple acceptance of new concepts, but requires students to understand and internalize them deeply. This process involves the guidance of non-intellectual factors such as emotional engagement and motivational stimulation. Learners (students) must integrate new concepts into their cognitive structure to form a complete knowledge system, achieving a deep understanding of

the concepts. Many researchers believe that through various concept change strategies, such as concept mapping (Amiruddin et al., 2025), analogy strategy (Handee et al., 2025), and dynamic evaluation strategy (Esmailzadeh, 2025), the in-depth understanding of concepts can be effectively promoted. After completing the concept reconstruction, students need to internalize the new concepts further to consolidate and deepen their understanding.

(3) Concept internalization stage: Concept change is continuous and gradual. After accepting new concepts, students must repeatedly apply and verify them in different situations and problems to consolidate and deepen their understanding of the new concepts (Alzheimer et al., 2025). Creating problem situations is an effective teaching strategy to promote the internalization of concepts. For instance, in physics teaching, teachers can design everyday problem scenarios, such as explaining why balloons rise or boats float on the water surface, to help students internalize the concept of buoyancy in solving practical problems (Stănescu et al., 2025).

3. The Design of a Teaching Strategy Based on Observation, Experience and Inquiry

The teaching strategy based on observation, experience, and inquiry is a conceptual teaching that drives students' independent exploration through cognitive conflicts. The teaching strategy constructed in this study is shown in Figure 1, which specifically includes the following three parts.

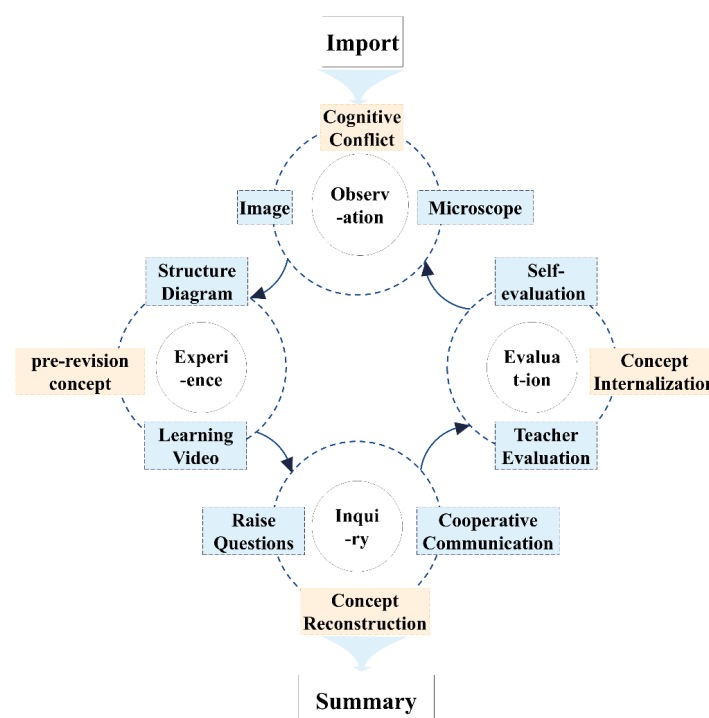


Figure 1. Teaching process of cell concepts.

(1) Import

Teachers first guide students to independently construct based on their initial understanding of knowledge (existing experience) to reveal the initial level of scientific concepts. Then, by creating problem situations that contradict students' experiences, students are made to doubt the new knowledge, form cognitive conflicts, and stimulate their desire to inquire. (Cognitive conflict)

(2) The “Observation-Experience-Inquiry” cycle

In this teaching session, students will carry out inquiry activities driven by cognitive conflicts, constantly revise and improve their existing knowledge structure (new concepts) based on the phenomena and conclusions of the inquiry, and further evaluate and test the construction of new knowledge (new concepts) in the “experience” and “inquiry” sessions. If the construction of new knowledge is consistent with the inquiry's conclusion and the problem situation, it indicates that the new knowledge is internalized with the existing knowledge. Conversely, it indicates that the new knowledge structure (concept) still has some imperfections and requires students to inquire further and correct it. In this cycle, while constantly revising the integration and matching degree of new and old knowledge, students experience the dynamic development process of knowledge construction, enhance their scientific thinking, promote their understanding of concepts, and achieve conceptual change.

(3) Evaluation

Students review the construction process of new knowledge (conceptual structure) and discuss their evaluation and understanding of the final integration/matching degree of new and old knowledge. This section mainly reflects students' understanding of the concepts to consolidate them. (Concept internalization)

4. Model Practice and Case Analysis

Based on the “Compulsory Education Primary School Science Curriculum Standards” and in combination with the core literacy of the subject, this study takes “Observing More Biological Cells” in the first unit of the sixth grade science textbook of the Education Science Edition as an example to further inquiry the application of the “Observation–Experience–Inquiry” teaching strategy in primary school science teaching. The teaching objectives of this course include: (1) Knowledge and skills objectives: Students can accurately describe the morphology, structure, and diversity of cells, including the similarities and differences between animal cells and plant cells, by observing cells under a microscope. (2) Process and Method Objectives: By repeatedly revising the cell structure diagram, students can master the cyclical process of scientific inquiry (observation–experience–inquiry), and enhance their scientific inquiry skills and critical thinking. (3) Emotional attitude and value goals: Recognize scientific knowledge's dynamic and developmental nature, have the courage to question and correct one's preconceived notions, and form correct scientific concepts.

Based on the above teaching objectives, the specific teaching activities designed are as follows:

4.1. Import

Students independently draw and construct cell structure diagrams based on their initial understanding of the knowledge to reveal their understanding of cell science concepts. (Cognitive conflict)

Design intention: Through a series of questions, guide students to construct the initial form of cells and expose their preconceptions. For instance, the shape of a cell is round, and the nucleus is drawn as a small dot. Then, presenting a real case of cell images under a real microscope triggers students' cognitive conflicts, making them question the rationality of the initial cell structure diagram, stimulating their desire to inquire, and laying a foundation for the smooth progress of subsequent teaching links.

Teacher: *“Students, please draw the shape of a cell based on your understanding of it and mark the parts you think are important.”*

Student: *(Draw a diagram of the cell structure. Draw a round cell, and simplify the nucleus to a small dot.)*

Teacher: *The cell structure diagrams you all drew are very interesting, but have you ever thought that the shape of a cell must be round? Is the cell nucleus really just a tiny dot? Next, let us observe the actual shape of the cells through the cell images under a real microscope.*

4.2. Cycle 1: Observation-Experience-Inquiry

- (1) Observation: Students establish connections between things by observing the morphology of various cells and promote conceptual change through practice.

Design intention: Through observation activities, enhance students' scientific inquiry abilities, such as observation and drawing, and provide an intuitive basis for correcting the initial cell structure diagram.

Teacher: *“Now, please observe the slices of animal and plant cells through the microscope, paying attention to their shapes, the positions of the cell nuclei, and other structures.”*

Student: *(Observe and record the cell structure.)*

Teacher: *“What are the differences between the cell structure you observed and the one you drew before?” Does the position and shape of the cell nucleus meet your expectations?*

Student: *“I found that the cell nucleus is not a small dot but has a distinct structure, and the shape of plant cells is quite different from that of animal cells.”*

- (2) Experience: Students gain in-depth experience and reflection on the structure and function of cells by watching videos. (Conceptual Reconstruction)

Design intention: Through students' independent questions after watching the learning video, guide them to think about the structure and function of cells deeply, reflect on the limitations of the initial cell structure diagram,

and complete the construction of a new cell structure diagram, understanding the preconceptions of the specific functions of structures such as the nucleus, cell membrane, and cytoplasm. Then, the students used the new cell structure diagram to explain the question of “how cells maintain life activities” to evaluate the rationality of the new cell structure diagram.

Teacher: *“When watching the video, please think: What is the function of the cell nucleus?” What is the function of the cell membrane?*

Student: *“Teacher, how does the cell nucleus control the activity of cells?”*

Teacher: *“The nucleus is the cell’s control center, and the cell membrane acts as a protective barrier that regulates the passage of substances.” Please correct your cell structure diagram according to the video content.*

Student: *“I added the internal structure of the cell nucleus and also drew the layers of the cell membrane.”*

Teacher: *“Now, please explain how cells maintain life activities with your new cell structure diagram?”*

Student: *“The nucleus controls activities, the cell membrane protects and controls the entry and exit of substances, and the cytoplasm is the site of life activities.”*

- (3) Inquiry: Students redraw and inquire into the cell structure and construct concepts with the teacher’s assistance. (Concept internalization)

Design intention: Create a problem situation with the “structure and function” theme of animal cells to guide students to inquiry in cooperative communication. Subsequently, a new problem situation was created again based on the “research findings of a certain interest group” to help students recognize the limitations of the cell structure diagram: the observation results of only animal cells cannot indicate that cells have similar functions and structures in different organisms, thereby triggering a new round of exploration and further internalization of the concept as well as revision of the cell structure diagram.

Teacher: *“Now, let us inquire into animal cells’ structure and function in groups.” Please discuss the functions of the nucleus, cell membrane, and cytoplasm based on the cell structure diagram you drew, and explain how they work together.*

Students: *(Discuss in groups and correct the cell structure diagram.)*

Teacher: *“A certain interest group discovered that the mitochondria in animal cells are energy factories. So, does a similar structure exist in plant cells as well?” Please further observe the plant cells and correct your cell structure diagrams.*

Student: *(Observe plant cells and correct the cell structure diagram, adding structures such as mitochondria.)*

4.3. Cycle 2: Observation-Experience-Inquiry

Design intention: By guiding students to observe independently, experience, inquiry, and draw diagrams of the structure of plant cells, further enhance students’ exploration abilities. By guiding students to apply the cell structure diagrams they have drawn to explain the structure and function of plant cells, their ability to analyze them is enhanced, enabling them to recognize the social value of biological knowledge.

Teacher: *“Next, we will observe the morphology of plant cells and compare them with animal cells.” Please consider: What are the similarities and differences in structure and function between plant cells and animal cells?*

Student: *(Observe plant cells and record them.)*

Teacher: *“You have found that plant cells have cell walls and chloroplasts, while animal cells do not.” What roles do these structures play in the functions of plant cells? Please explain it using the cell structure diagram you have drawn.*

Student: *“The cell walls of plant cells provide support, and chloroplasts are responsible for photosynthesis. These are all functions that animal cells do not have.”*

4.4. Evaluation: Students Review the Process of Cell Structure Construction, Evaluate and Understand the Final Cell Structure Diagram

Design intention: Review the process of constructing the cell structure diagram. Through students' responses, test their understanding of cell structure and their comprehension of the concept, guide them to consolidate the concept, apply what they have learned, and cultivate their observation, experience, and exploration abilities.

Teacher: *"Please review the cell structure diagrams you have drawn. What changes have occurred in your understanding of cells from the initial simple cell structure diagrams to the current complex ones?"*

Student: *"Initially, I thought the cell nucleus was just a tiny dot. Now I know it has many complex structures, and the cell membrane and cytoplasm functions are also essential."*

Teacher: *"Very good!" Through this learning experience, you not only corrected the previous concepts but also gained a deep understanding of the structure and function of cells. Next, we can use this knowledge to explain more biological phenomena.*

5. Teaching Summary and Strategy Refinement

The cell concept teaching case shows that Observation-Experience-Inquiry (OEI) is not a simple linear sequence of three independent activities, but a cyclical process in which students continuously examine and reconstruct scientific concepts. Observation provides direct empirical evidence and exposes the differences between students' preconceptions and actual cell morphology. Experience helps students interpret the observed phenomena through videos, discussion, and repeated revision of cell structure diagrams, thereby connecting perceptual experience with knowledge of cell structure and function. Inquiry further guides students to raise questions, compare animal and plant cells, explain evidence, and test and revise their existing ideas. The questions generated during inquiry then initiate a new round of observation and experience, promoting the progressive development of students' understanding. Based on this case, the OEI teaching strategy can be summarized as the following instructional path: diagnosing preconceptions, creating cognitive conflict through evidence-based observation, reconstructing concepts through meaningful experience, testing and applying concepts through problem-driven inquiry, and consolidating understanding through reflective evaluation. In this process, students move from passively receiving conclusions to actively observing evidence, explaining phenomena, and revising their ideas. The teacher's role also changes from directly transmitting knowledge to designing situations, providing scaffolding, organizing inquiry, and offering timely feedback. Thus, the value of the OEI strategy lies not only in helping students understand scientific concepts, but also in integrating concept learning with the development of observation, reasoning, inquiry, and reflection.

6. Teaching Reflection and Conclusions

The cell concept teaching case illustrates how the OEI strategy can be applied to primary science concept teaching. It connects the theoretical framework of observation, experience, and inquiry with specific classroom activities. Through microscope observation, students obtain direct evidence of cell structures. Videos, discussions, and diagram comparison then help them connect their observations with scientific concepts. Further questioning and diagram revision encourage students to explain evidence, correct misconceptions, and deepen their understanding. Therefore, the case supports the central idea that scientific concepts can be gradually constructed through the interaction of observation, experience, and inquiry. Several issues should be considered when implementing the strategy. Microscope operation, specimen observation, group discussion, and diagram revision require sufficient classroom time. Differences in students' operational skills and prior knowledge may also affect their participation. Teachers should therefore provide appropriate guidance, assign clear group roles, and use progressive questions to connect the three stages. Evaluation should not focus only on the final cell diagram but should also include students' observation records, diagram revisions, discussions, and explanations as evidence of conceptual development. In conclusion, observation provides empirical evidence, experience supports the interpretation and reconstruction of concepts, and inquiry promotes their application and further development. The interaction of these stages helps students gradually develop a more scientific conception of cells. Future teaching may further adapt the strategy to students' age, prior knowledge, and the characteristics of different scientific topics.

Author Contributions

Y.Z. (Youyou Zhang): Writing—original draft, Methodology, Investigation, Funding acquisition, Case analysis, Conceptualization. Y.Z. (Yu Zhao), X.L. and D.D.: Conceptualization, Writing—original draft,

Writing—review & editing. J.Y., J.L. and W.L.: Conceptualization, Methodology, Investigation. All authors have read and agreed to the published version of the manuscript.

Funding

2025 Zhejiang Provincial Education Science Planning Priority Project: Research on Interdisciplinary Teaching Mode for Promoting Students' Computational Thinking Development under the Perspective of Multiple Analysis (Project No.: 2025SB103).

Informed Consent Statement

Informed consent was obtained from all subjects involved in the study.

Data Availability Statement

The original data supporting the conclusions of this article are not publicly available because of the privacy requirements of the research participants, but may be available from the corresponding author upon reasonable request with the permission of the participants.

Acknowledgments

The authors thank the editors and the anonymous reviewers for their constructive comments.

Conflicts of Interest

The authors declare no conflict of interest.

Use of AI and AI-Assisted Technologies

During the preparation of this work, the authors used ChatGPT to improve language clarity and correct grammatical errors. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

- Alsheimer, S., Schnell, T., Chlebna, C., & Rohe, S. (2025). Competing terms for complementary concepts? Acceptance and legitimacy. *Renewable and Sustainable Energy Reviews*, 207, 114960. <https://doi.org/10.1016/j.rser.2024.114960>.
- Amiruddin, M. Z. B., Samsudin, A., Suhandi, A., Coştu, B., & Prahani, B. K. (2025). Scientific mapping and trend of conceptual change: A bibliometric analysis. *Social Sciences & Humanities Open*, 11, 101208. <https://doi.org/10.1016/j.ssaho.2024.101208>.
- Chen, X., Jia, B., Peng, X., Zhao, H., Yao, J., Wang, Z., & Zhu, S. (2025). Effects of ChatGPT and argument map (AM)-supported online argumentation on college students' critical thinking skills and perceptions. *Education and Information Technologies*, 30(12), 17623–17658. <https://doi.org/10.1007/s10639-025-13471-2>.
- Dewey, J. (1910). Science as subject-matter and as method. *Science*, 31(787), 121–127. <https://doi.org/10.1126/science.31.787.121>.
- Esmailzadeh, P. (2025). The role of ChatGPT in disrupting concepts, changing values, and challenging ethical norms: A qualitative study. *AI and Ethics*, 5, 291–304. <https://doi.org/10.1007/s43681-023-00338-w>.
- Handee, W., Thammaprateep, J., & Suwanjinda, D. (2025). Development of the concept-process model and metacognition via FAR analogy-based learning approach in the topic of metabolism among second-year undergraduates. *Chemistry Teacher International*, 7(2), 359–369. <https://doi.org/10.1515/cti-2024-0104>.
- Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, 12(1), 1–48. [https://doi.org/10.1016/0364-0213\(88\)90007-9](https://doi.org/10.1016/0364-0213(88)90007-9).
- Kolb, A. Y., & Kolb, D. A. (2009). Experiential learning theory: A dynamic, holistic approach to management learning, education and development. In S. J. Armstrong & C. V. Fukami (Eds.), *The SAGE handbook of management learning, education and development* (pp. 42–68). SAGE Publications. <https://doi.org/10.4135/9780857021038.n3>.
- Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd ed., enlarged). University of Chicago Press.
- Kunisch, S., Denyer, D., Bartunek, J. M., Menz, M., & Cardinal, L. B. (2023). Review research as scientific inquiry. *Organizational Research Methods*, 26(1), 3–45. <https://doi.org/10.1177/10944281221127292>.
- Limón, M. (2001). On the cognitive conflict as an instructional strategy for conceptual change: A critical appraisal. *Learning and Instruction*, 11(4–5), 357–380. [https://doi.org/10.1016/S0959-4752\(00\)00037-2](https://doi.org/10.1016/S0959-4752(00)00037-2).

- Luo, X., Zhang, T., Wang, Y., & Liu, C. (2023). A study on the impact of a spherical video-based virtual reality on teaching English writing guided by experiential learning circle theory. *Journal of Educational Technology and Innovation*, 5(2), 69–76. <https://doi.org/10.61414/jeti.v5i2.125>.
- Murrell, P. H., & Claxton, C. S. (1987). Experiential learning theory as a guide for effective teaching. *Counselor Education and Supervision*, 27(1), 4–14. <https://doi.org/10.1002/j.1556-6978.1987.tb00735.x>.
- Özdemir, G., & Clark, D. B. (2007). An overview of conceptual change theories. *Eurasia Journal of Mathematics, Science and Technology Education*, 3(4), 351–361. <https://doi.org/10.12973/ejmste/75414>.
- Pedaste, M., Baucal, A., & Reisenbuk, E. (2021). Towards a science inquiry test in primary education: Development of items and scales. *International Journal of STEM Education*, 8(1), 19. <https://doi.org/10.1186/s40594-021-00278-z>.
- Pedaste, M., Mäeots, M., Siiman, L. A., de Jong, T., van Riesen, S. A. N., Kamp, E. T., Manoli, C. C., Zacharia, Z. C., & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review*, 14, 47–61. <https://doi.org/10.1016/j.edurev.2015.02.003>.
- Piaget, J. (1976). Piaget's theory. In B. Inhelder, H. H. Chipman, & C. Zwingmann (Eds.), *Piaget and his school: A reader in developmental psychology* (pp. 11–23). Springer. https://doi.org/10.1007/978-3-642-46323-5_2.
- Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, 66(2), 211–227. <https://doi.org/10.1002/sce.3730660207>.
- Schwab, J. J. (1958). The teaching of science as inquiry. *Bulletin of the Atomic Scientists*, 14(9), 374–379. <https://doi.org/10.1080/00963402.1958.11453895>.
- Spiegel, C. (1975). *Role-playing sociodramatic simulations in the Montreal area (1971–74)* [Master's thesis, McGill University]. McGill University Repository. <https://mcgill.scholaris.ca/bitstreams/8fe583f5-633e-44ac-b54b-5199b9602a37/download>.
- Stănescu, A.-B., Travadel, S., Rughiniș, R.-V., & Bucea-Manea-Țoniș, R. (2025). Modeling learning outcomes in virtual reality through cognitive factors: A case study on underwater engineering. *Electronics*, 14(17), 3369. <https://doi.org/10.3390/electronics14173369>.
- Zhang, T., & Zhang, Q. (2022). Towards integrated testing approach: An application of cognitive science and deep learning principle to language testing. *Journal of Educational Technology and Innovation*, 3(1), 40–55. <https://doi.org/10.61414/jeti.v3i01.26>.
- Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, 27(2), 172–223. <https://doi.org/10.1016/j.dr.2006.12.001>.

Article

Empowerment or Negativeenergy?—Opportunities, Challenges and Reflections of Preschool Teachers under the Application of Artificial Intelligence

Yanran Wang^{1,2}, Chun Fu^{2,*†}, Ying Liu^{1,†}, Xi Zhang¹, Yun Zheng³, Furong Guo² and Shiwei Hu⁴

¹ School of General Education, Guangdong Finance & Trade Vocational College, Qingyuan Campus, No. 28 Zhongsu Road, Dongcheng Sub-district, Qingcheng District, Qingyuan 511510, China

² Golden Dragon Center, Faculty of Education, City University of Macau, Golden Dragon Campus, 5th Floor, Xianxinghai Avenue, Macau 999078, China

³ Qingxin District No. 2 Kindergarten, Qingquan Road, Qingxin District, Qingyuan, Guangdong 511800, China

⁴ College of Education, Sungshin Women's University, 2, Bomun-ro 34-gil, Seongbuk-gu, Seoul 02844, Republic of Korea

* Correspondence: drfuchun@163.com

† These authors contributed equally to this work.

How To Cite: Wang, Y., Chun, F., Liu, Y., Zhang, X., Zheng, Y., Guo, F., & Hu, S. (2026). Empowerment or Negativeenergy?—Opportunities, Challenges and Reflections of Preschool Teachers under the Application of Artificial Intelligence. *Journal of Educational Technology and Innovation*, 8(3), 9–19. <https://doi.org/10.61414/4a17gm79>

Received: 27 May 2026

Revised: 6 September 2026

Accepted: 20 September 2026

Published: 30 September 2026

Abstract: From the perspective of I-TPACK theory and entanglement and symbiosis human-computer interaction theory, combined with the analysis of narrative interview materials made by 15 preschool teachers in the whole kindergarten, the current practice status of teachers using AIGC or intelligent tools for home communication, lesson plan design, courseware production, and visual observation of children is investigated with the method of narrative inquiry. The study found that AI is an enabling resource to promote preschool teachers to use data to think, design courses, and complete role changes, but at the same time there are three boundary conflicts: human-machine emotion, educational wisdom, and preschool information ethics. In addition, when teachers use relevant tools, they often have practical troubles such as difficulty in writing prompt words, high operating thresholds, technical dependence, and insufficient AI resources for preschool education. The research proposes three professional growth paths: reshaping teachers' professional role cognition, cultivating an AI preschool education dual-core literacy system, and building an artificial intelligence teaching and research community to promote the professionalization of preschool teachers.

Keywords: artificial intelligence; preschool teachers; professional growth; educational boundary

1. Introduction

The deep integration of intelligent technology and education is changing the ecology of preschool education at an unprecedented speed. From intelligent evaluation systems to home interactive apps, from AI picture book generation to children's behavior observation systems, artificial intelligence has inevitably penetrated into the field of preschool education (Nafukho et al., 2004). But at the same time, not all technology embedding is a process of unilaterally improving efficiency. In the practice of using AI in the daily teaching of front-line preschool teachers, this paper finds that there are differences between technical tools and educational practices. For example, teachers want to use AI to compress the working time of copywriting, but due to the cost of learning prompt word instructions and problems such as AI-generated text fact proofreading and logical correction, the total time cost “does not decrease but increases”. For another example, for teaching animations or courseware generated by AI,



Copyright: © 2026 by the authors. This is an open access article under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Publisher's Note: Scilight stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

the feedback is that the content is too rigid and lacks childlike interest, which cannot meet the kindergarten teaching scene; However, relying on AI to build a teaching plan framework and write an activity plan has the problem of weakening the ability of independent teaching design and text creation. This is not a purely technical adaptability issue, but reflects a problem: when AI penetrates into key educational links such as curriculum design, teaching implementation, home communication, and even children's evaluation, the identity of teachers is also undergoing fundamental changes (Putra & Sekti, 2026), technological changes have made preschool teachers stand at a crossroads of identity-not only ushering in the opportunity of technological empowerment, but also facing the deep anxiety of being disciplined or even replaced by technology. (Berson et al., 2025); The core question that emerges from this is: After the deep intervention of AI, will kindergarten teachers be “disabled” or “empowered”? While artificial intelligence brings teaching efficiency dividends, what kind of professional autonomy, educational judgment and creative challenges does it bring to teachers? Although relevant studies have made important contributions to the macro-theoretical construction of artificial intelligence education and the framework of teachers' technical literacy, most of the studies focus on K-12 or higher education, and do not pay special attention to kindergarten teachers. Special groups, and most of the existing research results discuss the application of AI from the perspective of technology functions, and the qualitative investigation of the real practical confusion of front-line teachers and the internal mechanism of professional growth is not yet sufficient. Whether the theory of “human-machine symbiosis” is applicable to the preschool education field also needs to be verified by materials. Therefore, this study is based on the topic of professional growth of preschool teachers from the perspective of artificial intelligence, using the I-TPACK framework and the human-computer interaction theory of entanglement and symbiosis as analysis tools, and on the basis of narrative interviews with 15 front-line preschool teachers, Try to answer how AI technology changes the professional role of preschool teachers; Where is the educational boundary of artificial intelligence; And how teachers can grow professionally rather than degenerate professionally in the process of technological empowerment, so as to provide certain theoretical reference and practical reference for alleviating teachers' professional anxiety, clarifying the path of professional growth, and promoting the healthy advancement of preschool education in the era of artificial intelligence.

2. Literature Review

2.1. Theoretical Context of Artificial Intelligence Education

The research of artificial intelligence education has experienced a process from “human-computer interaction” to “entanglement and symbiosis”. In the early stage, it was mainly about the functional interaction between people and computers, but in recent years, scholars have also begun to pay attention to more the deep-seated interaction between “teacher-student-machine”. “Entanglement Symbiosis” implies the idea of entanglement human-computer interaction theory, and Qiangyu will fully realize a new kind of artificial intelligence (Meniado, 2024). Under this kind of thinking, the machine is no longer a simple teaching method used to help people, but an “actor” who cooperates and coexists with human teachers. At present, there are three main types of thinking about the integration of human intelligence and machine intelligence: the first is a speculative research based on human teachers, focusing on how teachers endogenous artificial intelligence capabilities; The second is the research on the skill change of machine teachers, discussing the shortcomings of AI teachers in terms of deep logic and social emotion; The third is the research of fusion model construction, trying to establish an artificial intelligence teaching that can give full play to the respective advantages of the two. Effective human-machine cooperation involves a series of controls, and ultimately points to “Meaningful Human Control” (Meaningful Human Control), that is to say, humans will always be the real masters of technology.

2.2. Theoretical Framework of Teacher Profession: From TPACK to I-TPACK

The Technology-Pedagogy-Content Knowledge (TPACK) framework has always been a classic tool for analyzing teachers' ability to integrate technology. However, due to the lack of TPACK due to AI's independent agency, epistemological complexity and ethical dimension, Chiu proposed a new intelligent TPACK (I-TPACK), which combines AI technical knowledge, AI content knowledge, The five knowledge areas of AI teaching knowledge, human-machine collaboration knowledge, and ethical knowledge are integrated to make up for the gap in TPACK (Chiu, 2025). And this framework formalizes the knowledge of human-machine collaboration, emphasizes that AI can handle procedural tasks, and allows teachers to have more time to think about higher-level instructional design. At the practical level, the I-TPACK framework is used to guide teacher education, enabling teachers to become designers who can strategically orchestrate human-machine collaborative learning, rather than artificial intelligence users who simply use AI tools. This theory provides a useful analytical tool for understanding and explaining the professional growth of preschool teachers in the AI era (Lakhe Shrestha et al., 2025).

Comparing the interview data, it is found that preschool teachers' cognition and application of AI are still in the operation stage of basic-level AI tools. They only need keyword retrieval and simple copywriting generation, and lack high-level capabilities such as data analysis, artificial intelligence design, and AI ethical judgment. There is still a clear gap from the five-level literacy requirements proposed by the I-TPACK framework.

2.3. Particularity of AI Application in the Field of Preschool Education Research on AI Applications in Preschool Education Is Also Unique

In a systematic literature review on AI-supported teaching human-machine cooperation in early education, international evidence on the mutual contribution of humans and machines is discussed and related issues, and examines the ethical implications of using machine learning, deep learning and learning analysis in early childhood education; It is believed that AI can be meaningfully and ethically integrated into early education, but teacher training and professionalism are the main challenges, and the application of AI must respect children's boundaries (Crescenzi-Lanna, 2023). In the domestic research on the profession of preschool teachers, it is mentioned that the professional growth of preschool teachers is a process that is constantly changing due to the influence of various factors, and needs professional theories as support. Through theoretical learning, teachers can form their own educational beliefs, improve their own professional autonomy and alleviate job burnout (E et al., 2024), but the lack of good learning, cognitive bias, and self-inertia faced by preschool teachers in the kindergarten may be further amplified or alleviated in the AI era.

2.4. Review of Research

On the whole, although the existing research has achieved some results in the macro theory of artificial intelligence education, the framework of teacher AI literacy, and the application of AI in preschool education, there are still some gaps: First, most of the research focuses on K-12 Or in the field of higher education, there are few artificial intelligence studies specifically targeting preschool teachers; Second, most of the existing studies discuss the application of AI from the perspective of technical functions, lacking attention to the deep mechanism of teachers' professional growth, and lacking qualitative empirical support such as the real practical confusion and use of front-line teachers; Third, whether the theory of "synergistic symbiosis" is applicable to the field of early childhood education remains to be tested, and man-machine dialogue learning as a practical path for teachers' profession is still being explored. This study will use the real AI use narrative interviews of 15 preschool teachers in different classes of different teaching ages as empirical materials to make up for the lack of front-line shortcomings of existing research, and conduct exploratory theoretical construction and practical reflection on the above gaps.

3. The "Energy Enhancement" Space for Professional Growth under the Empowerment of Technology

Combined with front-line teacher interview narratives, AI tools form an empowering fulcrum in the whole process of kindergarten home communication, curriculum, teaching implementation, and child assessment, opening up a triple empowerment space for the professional growth of kindergarten teachers.

3.1. From "Style Conservation" to "Data-Driven Decision-Making"

The observation and evaluation of preschool teachers is based on their own and intuition, and the AI-assisted behavior observation system can continuously track and record children's language, social interaction, and action coordination, and transform scattered information into organized data. Let teachers understand the needs of young children more accurately. (Wan & Gu, 2026). The teacher of the second class of middle school shared his practice in the interview: with the help of AI tools to capture the traces of children's activities in the area, generate dynamic visual images with one click and cast them into the classroom, intuitively display the details of children's game behavior and social interaction, and make up for the existence of teachers' single observation blind spot of vision". This is not to replace teachers with algorithms to make judgments, but to provide more evidence support for teachers to make decisions. From the perspective of professional growth, this means cultivating teachers' ability to analyze and interpret data. They have changed from "childcare workers" to "analysts" to achieve role upgrades.

Generative AI can teach strategies by providing personalized learning resources, dynamic classroom activities, and automated feedback to promote student engagement and teacher adaptability (Cimino et al., 2025). For example, the daily work of the sophomore class teachers is to use AI to generate home week summaries and safety reminders in batches, and the work of standardized copywriting has been completed efficiently; Teachers

in Class 4 rely on AI to quickly build teaching plans and activity copywriting frameworks, which greatly increases the time-consuming repetitive writing work (Li et al., 2026).

3.2. From “Course Porter” to “Course Creator”

The emergence of generative AI technology has brought unprecedented assistance to the curriculum creation of preschool teachers (Yang et al., 2025). Teachers can use AIGC to quickly generate personalized first drafts of activity plans, picture book story prototypes, or create inspiration maps, etc. The energy is freed from cumbersome data collection and formatting, and focuses on the design of secondary creativity and the in-depth arrangement of emotional interaction. The teachers of Class 5 and Class 5 mentioned that in their daily work, they will use AI to enter topic keywords to quickly search for corresponding lesson plans, which will be used as reference materials for their course design; The second class of junior high school and the fourth class of senior high school use AI to generate the basic framework of teaching PPT with one click, which greatly increases the threshold for courseware production. The pedagogical significance brought about by this change is that AI undertakes the load of procedural tasks, allowing teachers to return to the most authentic part of education work-creatively responding to the interests and needs of children.

Studies have shown that GenAI supports the continuous growth of the teacher profession, helping teachers optimize their teaching and respond to the needs of diverse learners by providing instant insights (Wu et al., 2025). Preschool teachers are no longer executors of standardized courses, but “curriculum creators” who screen, adapt, and emotionally process materials generated by AI.

3.3. From “Knowledge Authority” to “Co-Learner”

With the intervention of AI tools, the relationship between knowledge and power in kindergartens is also changing. “When teachers and children face the output of AI systems together, teachers are no longer a single knowledge transmitter, but transformed into guides of knowledge exploration and demonstrators of critical thinking. This kind of role transformation advocates a kind of human-machine joint use, and promotes students to have a deeper degree of critical participation in the work. (Al-Abdullatif, 2025). During the interview, the teachers of the third class also took the initiative to share their own growth practices: “In order to conform to the generated content of the kindergarten scene, I took the initiative to learn how to standardize my questioning language, repeatedly debugged the prompt words, and constantly communicated with AI. Exercise your ability to screen information, accurately express and critically judge”. Teachers need to have the ethical reasoning and critical evaluation ability in AI literacy ability to effectively and responsibly integrate AI into students’ educational practice. This “common learning attitude” is also conducive to the establishment of a teacher-child relationship, which is a form of expression of teachers’ lifelong learning ability.

4. “Boundary” Conflict and Ethical Concerns under Technological Intervention

Combined with the interview records of 15 teachers, AI has applied multiple realities in front-line practice, which is essentially a conflict between the triple core boundaries of artificial intelligence, and it is also a realistic obstacle to the professional growth of preschool teachers, as shown in Figure 1.

4.1. The Boundary of Emotional Connection and Machine Algorithms

During the interview, many teachers pointed out that the teaching plans, courseware, and story copywriting generated by AI are generally rigid and cold, and lack the childlike interest, warmth, and empathy perspectives required for early childhood education, making it difficult to meet the emotional needs of young children; Relying too much on AI to evaluate and give feedback on teacher-child interaction may lead to the passivation of teachers’ emotional level-if teachers are used to using data to “indirectly” understand children, then direct observation and empathy will be reduced. degenerate. It is not AI itself that determines whether teachers will be replaced, but the nature of education and the needs of students.

4.2. The Boundary between Educational Wisdom and Fixed Procedures

For preschool teachers, if they rely too much on the activity suggestions and evaluation conclusions provided by the AI system, their own curriculum design ability, educational judgment and creativity will gradually shrink (Eyal, 2025). Under artificial intelligence, the core problems faced by teachers include: the confusion of professional value, how to get along with machines, and how to get along with students. Teachers’ professional growth requires accumulating their own educational wisdom through repeated practical attempts and reflections,

not moving towards “de-specialization” through technical substitution (Kurian, 2025). At the same time, the cost of using the tool and the learning threshold further magnify the contradiction: the free AI courseware template style is relatively single, and most of the high-quality preschool education PPT, animation, etc. need to pay to open a membership, and the use requires high economic costs; The operation process of teachers in the second class of middle school using the AI traces visualization tool is relatively cumbersome. The kindergarten lacks supporting training for this, and there is a lack of independent learning channels. It is difficult for teachers to operate it quickly, and it is difficult to give full play to the teaching value of AI.

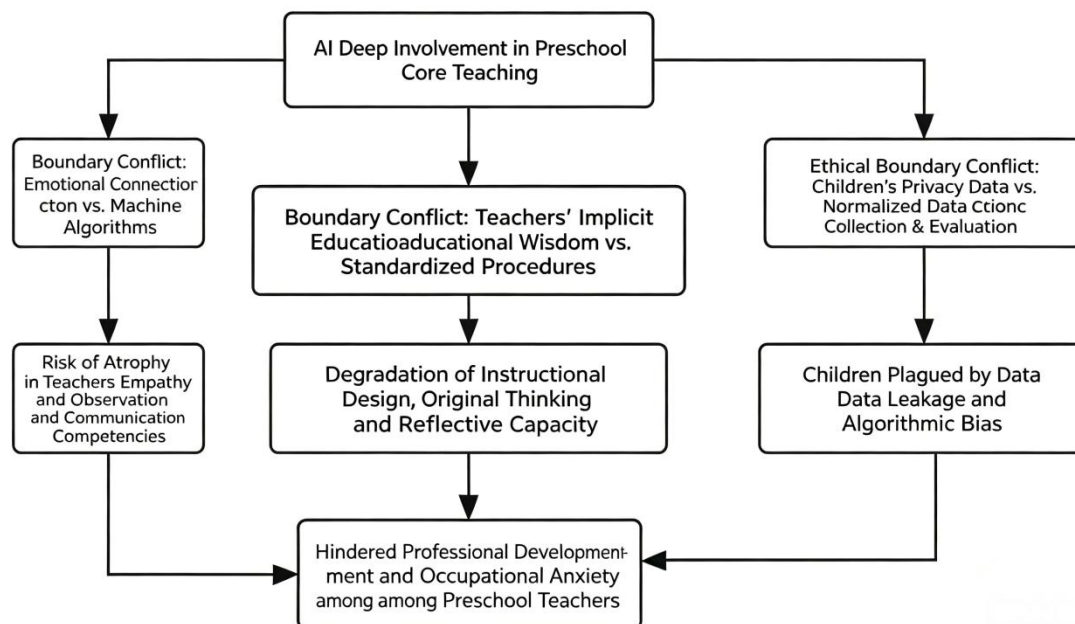


Figure 1. Logic flow chart of artificial intelligence triple conflict generation.

4.3. The Boundary between Ethics and Data

As a special vulnerable group, young children face more urgent issues about data protection than other education stages. AI systems have very profound ethical issues in the collection of children’s behavior data, the preservation of facial recognition information, and the evaluation based on algorithms.

Faced with the current situation of the integration of intelligent technology and education, teacher ethics is facing many practical challenges. The “Artificial Intelligence Education Ethics: Reference Framework” released by the World Digital Education Conference in 2026 established the core concept of “the subject belongs to people, collaborative symbiosis, suitable environment for good, and classified governance”, and clarified “prohibition of access, limited use, Encourage use” three types of behavior boundaries. During the interview, the senior class teachers expanded the application scenarios of AI, using AI to analyze children’s reports and diet plans. This kind of data AI use behavior that involves children’s privacy lacks unified and standardized guidelines, and teachers lack the corresponding ability to judge digital ethics. This requires preschool teachers to have digital ethics literacy and be able to distinguish which data can be collected; Which AI evaluations need to be questioned; Which scenarios should “shut out” technology? The function of AI in education is not to replace the role of human teachers, but as auxiliary beings, and educators play a vital role in shaping ethical and responsible AI integration.

5. Research Design

5.1. Research Side: Narrative Inquiry

The reasons for this study to use Narrative Inquiry as the core orientation for in-depth research are as follows: This method is chosen for the following reasons: First, kindergarten teachers’ experience, feelings and meaning construction of AI are highly personal and situational, and 15 interviewees put forward different use puzzles, practices, etc., which need to be captured with deep-description qualitative research; Secondly, existing studies have shown that narrative inquiry can be used to explore the ethical considerations and professional growth of teachers integrating AI; Finally, the narrative approach can show the reflective process of teachers in artificial intelligence and present a changing timeline.

Recent studies have further confirmed that narrative inquiry is an effective framework for exploring the joint reflection between teachers and AI, and can reveal how teachers negotiate the relationship between reflection and ethics in practice. The design of narrative inquiry is also in line with previous studies in the literature that have used this method to explore GenAI and teachers' professional issues. This study uses 15 teachers' AI use narratives and confusion reflections as core qualitative materials to fully present the real practice picture of artificial intelligence in preschool education.

5.2. Research Objects

This study adopts the method of purposeful sampling to select interviewees, and determines the screening conditions of the sample in combination with various standards. The specific selection criteria are shown in Table 1 below. Regarding the use of AI, the interviewees are required to use intelligent education tools at least once a week for no less than three months to ensure that teachers have real and stable artificial intelligence teaching practices; According to the teaching experience of the interviewees, they are divided into novice teachers (1–3 years), mature teachers (5–10 years) and senior teachers (more than 10 years), covering different occupational stage groups; Considering that there are certain differences in teachers' acceptance of new technologies, two types of teachers, technology-positive and technology-prudent, are included in the sample to capture differentiated teachers' concepts of artificial intelligence; And cover small class, middle class, large class of all ages of class teachers, strive to sample diversity and differences, so as to enrich the level of qualitative data, improve the saturation of research data.

Table 1. Research object selection criteria.

Dimension	Selection Criteria
AI Usage Experience	Have used intelligent educational tools at least once a week for no less than 3 consecutive months
Years of Teaching Experience Distribution	Include novice teachers (1–3 years), experienced teachers (5–10 years), and senior teachers (over 10 years)
Technology Acceptance	Encompass technology-positive teachers and technology-prudent teachers
Kindergarten Type	Cover public kindergartens
Class Coverage	Teachers responsible for classes across all age groups

5.3. Data Collection

5.3.1. In-Depth Interviews

Design a semi-structured interview outline around the following important narrative themes, and interview 15 teachers one-on-one: Please tell the specific scenarios of how you use AI to carry out teaching and home communication in your daily life; What difficult problems and confusions have you encountered in the process of using AI; What changes do you think AI has made to your teaching ability and educational thinking; Which AI technologies do you want to use and related training support for the use of AI in preschool education. Conduct 1–2 interviews for each interviewee, and the time of each interview is controlled within 60–90 min. Detailed records of the details of the teacher's use of AI and real doubts are used as research materials for this paper.

5.3.2. Sexual Recollection Interviews

Invite some teachers to demonstrate their common AI tools on the spot, complete an activity design and implementation combined with AI tools, and video the whole process. Afterwards, key clips were played and asked about the thinking and decision-making process at that time, and accurately captured the recessiveness when teachers used AI, such as prompt word writing, AI content modification, and tool operation difficulties.

5.3.3. Faculty Reflection Journal

Invite the interviewed teachers to record 1–2 weeks of AI use experience logs during the research period, and record in detail the content effect, time cost, and use problems generated by AI as supplementary materials for the interview.

5.4. Information Analysis

Clandinin's "Three-dimensional Narrative Exploration Space" was used for analysis: through cross-analysis of three-dimensional space to sort out the co-existing and personalized stories of 15 teachers, in order to discover the deep mechanism of teachers' professional growth in artificial intelligence (Clandinin, 2006). From the interview materials, four common practical problems can be extracted: non-standard writing of AI prompt words, lack of exclusive AI resources for preschool education, lack of systematic AI training for teachers, and the risk of technology dependence in use. As shown in Figure 2.

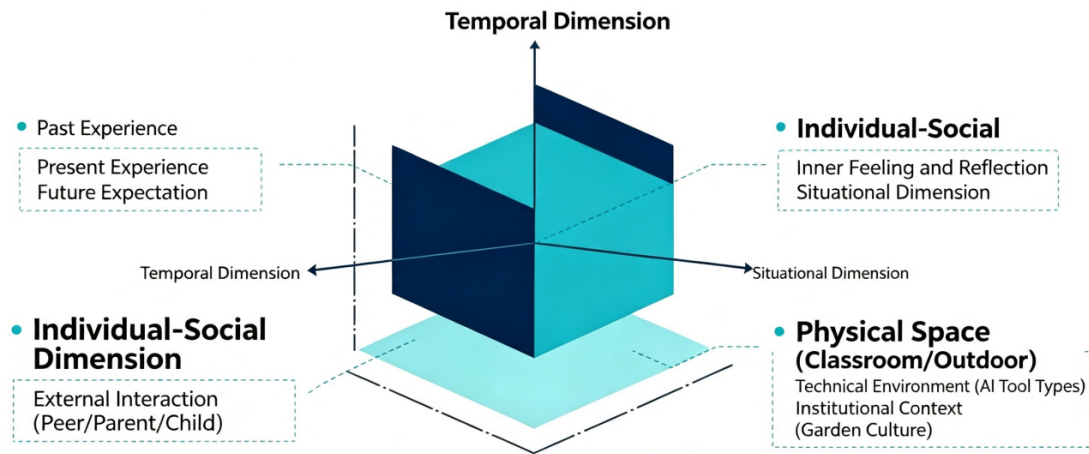


Figure 2. Three-dimensional narrative exploration space.

5.5. Research Ethics and Validity

This study ensures that all teacher participants fully understand the research purpose, interview content, and data use, and voluntarily participate in the interview; At the same time, all kindergartens, teachers' names, and children's related information are anonymized, and interviews involving children's health and behavior data are desensitized; In terms of validity strategy, participant checking (member checking), peer review, immersion and other means were used, and the first draft of the interview was handed over to the interviewed teachers to check the authenticity of the statement.

6. Path Construction: Towards the Professional Growth Strategy of Symbiotic Kindergarten Teachers

Combining the theoretical framework and the reality reflected in the interviews with 15 front-line teachers, based on the core logic of artificial intelligence symbiosis and keeping the boundaries of education, three practical professional growth paths for preschool teachers are constructed, and targeted solutions to teacher prompt word writing, tool learning, technology dependence, Realistic doubts such as lack of resources.

6.1. Reshaping Professional Role Cognition: Establishing the Core of "Irreplaceability"

The premise of professional growth is the cognition of roles. In response to the teacher's question about "whether the use of AI will change my own educational thinking" in the third class of middle school, and "AI will weaken my ability to prepare lessons independently" in the fourth class of primary school. Preschool teachers need to be clear: they are the core undertakers of value education and emotional education, and AI is an auxiliary tool. This kind of cognition does not blindly exclude technology, but determines the core advantages of this paper based on the understanding of the essence of education; Effective use of AI means protecting human professionalism with boundaries, allowing technology to help rather than replace the core role of human teachers. The reshaping of the role of teachers should point to making the teaching process intelligent, digital and personalized, and at the same time make the role of teachers more focused on the hearts and future of students. Preschool teachers should break the myth of "all-round teachers" and accept the efficiency advantages of AI in information retrieval, data processing, and draft copy generation; At the same time, we are soberly aware of AI's lack of empathy, educational tact, and childlike creativity, and we are convinced that we are irreplaceable in terms of emotional companionship, value guidance, individualized teaching, and creative stimulation. In the practical process, establish a fixation of "AI first draft generation-teachers' second original adaptation-teacher-child emotional interaction landing", use teachers' educational wisdom to guide AI output content, and avoid the problem of "de-skilling" caused by technology dependence.

6.2. Improve the Dual-Core Literacy of “AI Education” and Solve the Confusion of Front-Line Practical Operation in a Targeted Manner

The cultivation of AI-related literacy should be included in the teacher’s profession. Based on the “core literacy list required by teachers in the AI era” formed by the Delphi method in the I-TPACK framework, the content of the training is designed, and according to the practical problems concentrated by teachers during the interview process, AI literacy cultivation is built in layers system, as shown in Figure 3.

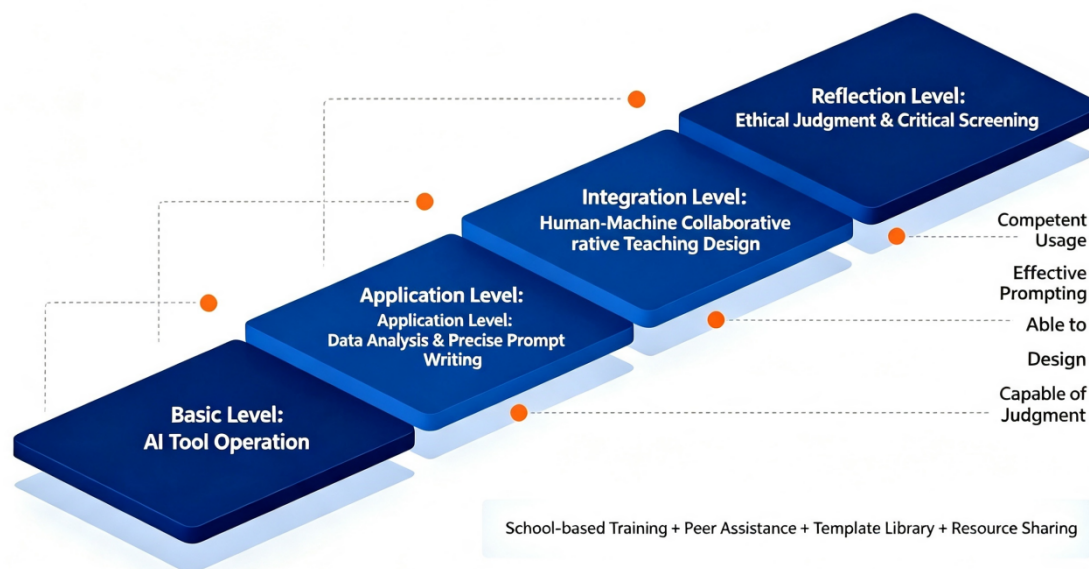


Figure 3. AI education dual-core literacy improvement path.

The specific operation is, at the base layer level; Popularize exclusive AI tool operation training for preschool education. Solve the problems of “few learning channels and complicated operation of AI tools” for teachers in the second class of middle school, and “high payment cost of high-quality preschool education templates” for the second class of primary school. Kindergartens uniformly screen free AIGC, courseware generation, and children’s behavior visualization tools that are suitable for preschool scenarios, and carry out kindergarten-based short-term practical training; The teaching and research team sorts out free preschool education PPT, animation, and picture book material library, and teachers use the economic cost. On the application layer level; Standardized Preschool Education Prompt Word Formula Teaching. Aiming at the high-frequency confusion of teachers in Class 1, Middle 5, and Primary 1 with “vague instructions, AI answers, and general and empty content”, and “AI content lacks childlike interest” in Class 3, a unified promotion of questioning formulas suitable for kindergarten scenarios: Preschool education standard Prompt word template = audience age usage scenario core theme format requirements style limit output length.

On the level of integration; Human-machine cooperation curriculum design ability training. Guide teachers to create courses of “AI framework, teachers fill in the temperature”, solve the problems of “AI teaching plan process does not meet the needs of the class” in the fifth class, and “courseware is dull and childlike” in the third class, and require teachers to combine the materials generated by AI Games, interactions, and emotional guidance links with the characteristics of children in this class. On the level of reflection; Digital Ethics and Critical Screening Competence. For scenarios such as AI analysis of children’s privacy data in the fourth class and algorithm evaluation of children, preschool AI ethics courses are offered to clarify the boundaries of children’s data collection and AI evaluation, cultivate teachers’ ability to criticize and verify AI output content, and respond to the sophomore class teachers “AI proofreading and correction time-consuming” problem, establish an AI text self-examination list, and quickly identify logic, fact, and age adaptation errors. In the pre-service training stage, normal colleges should incorporate AI literacy into the curriculum system, so that prospective teachers can understand the pedagogical meaning of human-machine collaboration from a theoretical level. In the on-the-job training, we should pay attention to sexual professional support, and promote the internalization of literacy through the cycle of “use-reflection-re-practice”. As a new thinking cognition and ethical direction, “teacher wisdom ethics” can help teachers solve problems in artificial intelligence creatively and systematically.

6.3. Build an Artificial Intelligence Teaching and Research Community to Eliminate the Use of Individual Technology

Isolated learning is difficult to support teachers' continuous growth in artificial intelligence. Interviews show that most teachers explore AI tools alone, and lack communication and sharing channels with their peers. Various practical doubts cannot be quickly answered. From the perspective of "symbiosis effect", the establishment of teacher learning community and the development of reflective learning can effectively promote the professional growth of preschool teachers; In the AI era, this community should be further upgraded to an "artificial intelligence teaching and research community". For example, using the AI system to record and generate teaching clips, the teaching and research team jointly discusses the advantages and disadvantages of AI-assisted teaching, shares high-quality prompt words, courseware templates, and AI teaching plan revisions, and focuses on solving common practical problems for teachers; With the help of the AI communication platform, teachers in urban and rural public and private kindergartens can use tools to alleviate the lack of technical training resources in rural and small private kindergartens; Taking teachers' daily AI use confusion as the discussion theme, collectively optimize prompt words, adapt AI-generated teaching plans, distinguish between procedural work that AI can undertake and emotional and creative work that must be completed by teachers, and balance technical efficiency and teachers' professional ability in the community. Artificial intelligence should build a support network to enable teachers to have in-depth reflective dialogues with peers and experts, realize the transition from individual to collective intelligence in the community, and systematically alleviate the anxiety, technical operation problems and problems caused by teachers using AI alone. Ability degradation concerns.

7. Conclusions

7.1. Core Discoveries

This study integrates I-TPACK and entanglement symbiosis theory, combines 11 front-line preschool teachers' AI use narrative interview materials, and conducts analysis through narrative inquiry, and obtains the following three core understandings:

First, technology is a "magnifying glass" rather than a "deformer". AI will not easily change the essence of education, but it will significantly amplify the professional advantages and disadvantages of teachers. Teachers who are good at reflection and artificial intelligence methods will greatly improve teaching efficiency with the help of AI; Teachers who only mechanically rely on AI to generate content and lack critical screening capabilities face the risk of "de-skilled" and solidified educational thinking. Therefore, the key to the professional growth of preschool teachers in the AI era does not lie in the technology itself, but in the teacher's awareness of reflection, system AI literacy, and supporting teaching and research support in kindergartens.

Second, boundaries are formed in dynamic negotiation. There is no fixed answer to artificial intelligence in the specific kindergarten teaching situation, and it is the result of teachers' continuous reflection, trial and error around the core ethical standard of "whether it is beneficial to children's body and mind", combined with the characteristics of children in their own classes. The positioning of AI in education should be a "legal concept and practice", which needs to be systematically integrated into teacher education and professionalism.

Third, the core is the protection of "relationship". The most valuable and irreplaceable professionalism of preschool teachers is precisely reflected in the emotional relationship of deep trust and warm response with children. All AI applications should serve to consolidate rather than weaken this core relationship. The ethical bottom line of artificial intelligence education is "no harm", and the fundamental starting point of teachers' artificial intelligence behavior should be to protect the rights and interests of children.

7.2. Research Limitations and Future Prospects

This study combines the interview narratives of 15 front-line kindergarten teachers to supplement empirical materials, but the sample size is limited and only covers public and private kindergartens in a single region. The differences in AI use by kindergarten teachers in different regions and urban and rural areas need to be further explored. At the same time, this study focuses on qualitative narrative analysis, and has not yet carried out follow-up quantitative research. Future research can be deepened from the following directions: no longer stay in the discussion of generalized technology applications, but anchor several types of representative intelligent tools that have been implemented in the field of preschool education-such as the AI behavior observation system for teacher-child interaction, support for teacher-child The co-created AIGC picture book platform and the preschool education courseware generation tool frequently used by front-line teachers are used as the specific starting point for field investigations, and they enter different types of kindergartens to carry out rooted research.

Early childhood education in the era of artificial intelligence is moving towards a new pattern in which people and technology shape each other and evolve together. In this pattern, AI is a mirror reflecting the core values of teachers. Under the background of human-machine symbiosis, the direction of preschool teachers is to become “reflective practitioners” who are good at controlling technology and have the ability to criticize and reflect. While rationally using AI to improve teaching efficiency, they should keep the warm and vivid humanistic background of preschool education. Realize professional growth empowered by technology.

Author Contributions

Y.W.: conceptualization, methodology, investigation, data curation, formal analysis, validation, writing—original draft preparation, writing—reviewing and editing; Y.L.: conceptualization, methodology, validation, investigation, funding acquisition, supervision, writing—reviewing and editing; C.F.: data curation, formal analysis, validation, visualization, resources, writing—reviewing and editing; X.Z. and Y.Z.: project administration, writing—reviewing and editing; S.H.: supervision, project administration, writing—reviewing and editing; F.G.: resources, validation. All authors have read and agreed to the published version of the manuscript.

Funding

Phased Achievement of the 2023 High-level Talents Scientific Research Startup Project of Guangdong College of Finance and Trade: Research on the Effective Integration and Efficient Support of Life-Education Ideology into the “Five-education Simultaneous Development and Three-all-round Education” in Higher Vocational Colleges (2023GCCXC05).

Institutional Review Board Statement

Ethical review and approval were waived for this study, due to the reason that the study adopted an anonymous questionnaire survey method, all participants participated voluntarily, the research content did not involve sensitive personal information, and no physical or psychological harm would be caused to the participants.

Informed Consent Statement

Informed consent was obtained from all subjects involved in the study.

Data Availability Statement

The raw data of this study cannot be made publicly available due to the need to protect the privacy and confidentiality of the research participants. Aggregated data supporting the conclusions of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare no conflict of interest. The funder had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Use of AI and AI-Assisted Technologies

During the preparation of this work, the authors used Doubao to assist with the Chinese-English translation of the manuscript and English language polishing. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

- Al-Abdullatif, A. M. (2025). Auditing AI Literacy Competency in K–12 Education: The Role of Awareness, Ethics, Evaluation, and Use in Human–Machine Cooperation. *Systems*, 13(6), 490. <https://doi.org/10.3390/systems13060490>.
- Berson, I. R., Berson, M. J., & Luo, W. (2025). Innovating responsibly: Ethical considerations for AI in early childhood education. *AI Brain Child*, 1, 2. <https://doi.org/10.1007/s44436-025-00003-5>.
- Chiu, T. K. (2025). Developing intelligent-TPACK (I-TPACK) framework from unpacking AI literacy and competency: Implementation strategies and future research direction. *Interactive Learning Environments*, 33(7), 4189–4192.
- Cimino, S., Maremmani, A. G. I., & Cerniglia, L. (2025). The Use of Artificial Intelligence (AI) in Early Childhood Education. *Societies*, 15(12), 341. <https://doi.org/10.3390/soc15120341>.
- Clandinin, D. J. (2006). Narrative inquiry: A methodology for studying lived experience. *Research Studies in Music*

- Education*, 27(1), 44–54. <https://doi.org/10.1177/1321103X060270010301>.
- Crescenzi-Lanna, L. (2023). Literature review of the reciprocal value of artificial and human intelligence in early childhood education. *Journal of Research on Technology in Education*, 55(1), 21–33. <https://doi.org/10.1080/15391523.2022.2128480>.
- E, L., Toom, A., Pietarinen, J., Soini, T., Haverinen, K., & Pyhältö, K. (2024). Professional agency in the classroom and burnout among early career teachers in China. *Frontiers in Psychology*, 15, 1412446. <https://doi.org/10.3389/fpsyg.2024.1412446>.
- Eyal, L. (2025). Developing and Validating an AI-TPACK Assessment Framework: Enhancing Teacher Educators' Professional Practice through Authentic Artifacts. *Education Sciences*, 15(11), 1452. <https://doi.org/10.3390/educsci15111452>.
- Kurian, N. (2025). AI's empathy gap: The risks of conversational Artificial Intelligence for young children's well-being and key ethical considerations for early childhood education and care. *Contemporary Issues in Early Childhood*, 26(1), 132–139. <https://doi.org/10.1177/14639491231206004>.
- Lakhe Shrestha, B. L., Dahal, N., Hasan, M. K., Paudel, S., & Kapar, H. (2025). Generative AI in professional development: A narrative inquiry based on the TPACK framework. *Frontiers in Education*, 10, 1550773. <https://doi.org/10.3389/educ.2025.1550773>.
- Li, C., Liu, Y., Wang, Y., Fu, C., Hu, S., & Guo, F. (2026). A Study on the Differences and Influencing Mechanisms of Curriculum Leadership of Urban and Rural Preschool Teachers under Digital Empowerment. *Journal of Educational Technology and Innovation*, 8(2), 35–54. <https://doi.org/10.61414/yhxqnw77>.
- Meniado, J. C. (2024). Human–machine collaboration in language education in the age of artificial intelligence. *RELC Journal*, 55(2), 291–295. <https://doi.org/10.1177/00336882241266629>.
- Nafukho, F. M., Hairston, N., & Brooks, K. (2004). Human capital theory: Implications for human resource development. *Human Resource Development International*, 7(4), 545–551. <https://doi.org/10.1080/1367886042000299843>.
- Putra, D. D. Y., & Sekti, B. A. (2026). Rethinking Education through the Lens of Human-AI Collaboration: Exploring the Shift from Traditional Education Models to AI-Supported Human Collaboration. In *Collaborative Paradigms of Generative AI and Human Intelligence for Equitable Learning Futures* (pp. 423–452). IGI Global. <https://www.igi-global.com/chapter/rethinking-education-through-the-lens-of-human-ai-collaboration/410443>.
- Wan, P., & Gu, X. (2026). Developing teachers' professional abilities: A systematic review of human-machine dialogic learning for teacher education. *Interactive Learning Environments*, 34(2), 615–639. <https://doi.org/10.1080/10494820.2025.2507280>.
- Wu, K., Chen, Y., & Jiang, W. (2025). Exploring the relationship between human teachers and machines in the age of intelligence—Human-machine symbiosis. In *2025 5th International Conference on Artificial Intelligence and Education (ICAIE)* (pp. 289–293). IEEE. <https://doi.org/10.1109/ICAIE64856.2025.11157959>.
- Yang, C., Jiang, J., & Gao, M. (2025). Research on the application and development of artificial intelligence assisted education and teaching under the perspective of human-machine collaboration. In *2025 14th International Conference on Educational and Information Technology (ICEIT)* (pp. 207–212). IEEE. <https://doi.org/10.1109/ICEIT64364.2025.10976098>.

Article

Technology-Mediated Individual and Collaborative L2 Academic Writing on Learner Autonomy

Ziyu Wang¹, Qianhao Chen² and Wilson Cheong Hin Hong^{3,4,*}

¹ Faculty of Education, The University of Hong Kong, Hong Kong SAR, China

² Faculty of Education, The Chinese University of Hong Kong, Hong Kong SAR, China

³ Faculty of Creative Tourism and Intelligent Technologies, Macao University of Tourism, Macao SAR, China

⁴ Institute of Education, University College London, London WC1E 6BT, UK

* Correspondence: wilson.hong.18@ucl.ac.uk

How To Cite: Wang, Z. Y., Chen, Q. H., & Hong, W. C. H. (2026). Technology-Mediated Individual and Collaborative L2 Academic Writing on Learner Autonomy. *Journal of Educational Technology and Innovation*, 8(3), 20–34. <https://doi.org/10.61414/h26xme36>

Received: 16 April 2026

Revised: 31 August 2026

Accepted: 14 September 2026

Published: 30 September 2026

Abstract: This study investigates the impact of individual and collaborative writing approaches on learner autonomy within technology-mediated L2 academic writing contexts. Through thematic analysis of semi-structured interviews conducted with 20 English-major undergraduate students in Macau, we examined how these writing modes—and their intersection with digital tools—influence key dimensions of autonomy: goal setting, self-monitoring, and self-evaluation. Our findings reveal that collaborative writing fostered greater peer support, digital skill complementarity, and reduced writing anxiety. Conversely, individual writing facilitated consistent planning and eliminated interpersonal stress, with students increasingly relying on AI as a “digital peer” for ideation and advanced software for digital workflow management. Notably, while collaborative students demonstrated enhanced peer-mediated self-evaluation, individual writers exhibited autonomy through strategic technological restraint and tech-mediated troubleshooting. A significant finding was the emergence of “complementary autonomy”, where alternating between individual and collaborative approaches—augmented by human-machine interaction—enhanced overall autonomy development more effectively than exclusive reliance on either mode. This study extends sociocultural theory by demonstrating that the quality of interaction—whether with human peers or digital tools—determines its value for autonomy development. The findings challenge dichotomous conceptualizations of learner autonomy, suggesting instead a dynamic interplay between personal agency, social interdependence, and technological mediation. Pedagogical implications include offering students’ flexibility in choosing writing modes while providing appropriate scaffolding for both digital literacy and human collaboration. This research contributes to a more refined understanding of how writing instruction can be designed to foster digital learner autonomy in modern L2 academic contexts.

Keywords: educational technology; collaborative writing; self-regulation; goal setting; self-monitoring

1. Introduction

Over the past decade, the rapid digital transformation of higher education has fundamentally reshaped English language instruction, transitioning from traditional classroom settings to technology-mediated learning environments. This pedagogical evolution, accelerated by the integration of cloud-based collaborative platforms and artificial intelligence (AI), has created unprecedented opportunities for students to exercise agency and make



Copyright: © 2026 by the authors. This is an open access article under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Publisher’s Note: Scilight stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

decisions during their learning process (Hong, 2023). At the core of this transformation is the concept of learner autonomy, which has evolved to encompass digital literacy and the strategic use of educational technologies (Sakai et al., 2008; Wang & Ren, 2024).

Learner autonomy refers to a student's capacity to take charge of their own learning (Holec, 1981; Winne, 2001). In the contemporary domain of English writing, autonomous writers are characterized not only by their ability to set clear goals and independently manage the learning process but also by their proficiency in leveraging digital tools—such as AI writing assistants and collaborative word processors—to support themselves in achieving those goals. Building on this foundation, previous studies have examined various factors affecting learner autonomy in L2 academic writing, including different teaching methods and learning strategies such as individual versus collaborative writing approaches (Sulistiyo & Kamil, 2022).

While much scholarly discussion has focused on conceptualizing learner autonomy and identifying factors that influence it, the advent of generative AI and advanced collaborative software has introduced new variables. There remains a significant gap in our understanding of how different *technology-mediated* learning arrangements—specifically digital individual and collaborative work—impact learner autonomy in L2 academic writing. This gap is particularly noteworthy given that L2 academic writing courses are standard requirements for university EFL learners, and both individual (often AI-assisted) and collaborative (often cloud-mediated) learning approaches are commonly employed in these contexts. The challenges that L2 academic writing poses for students (Paltridge, 2004) further underscore the importance of understanding how different digital learning arrangements might support the development of learner autonomy.

To bridge this gap, this study qualitatively investigates students' perceptions of technology-mediated individual and collaborative work in L2 academic writing, examining how these different approaches, supported by digital tools, affect the development of learner autonomy.

2. Literature Review

2.1. Defining Learner Autonomy

The concept of learner autonomy has evolved significantly since its introduction in language education. Trim's (1978) declaration captured the origins of learner autonomy as a conceptual tool for teachers to plan, structure, and deliver curriculum according to learners' needs, motivations, and characteristics, enabling them to direct and control their learning processes. Holec (1981) further developed this concept in adult education, characterizing it as a shift from "directed teaching" to "self-directed learning". Similarly, Winne (2001) defined learner autonomy as students taking ownership of their learning process, which now heavily involves actively controlling their digital learning environment, managing screen time, and selecting appropriate technological methods (Xu, Hong, et al., 2024; Xu, Li, et al., 2024; Xu, Su, et al., 2024).

More recent conceptualizations emphasize that learner autonomy in digital spaces aims to balance power dynamics by respecting learners' right to participate in decision-making about their learning process, including which digital tools to employ (Balcikanli, 2010; Xu, Hong, et al., 2024). Learner autonomy also encompasses learners' involvement in planning, monitoring, and assessing their learning (Mickwitz, 2020). For the purposes of this study, we define students with high learner autonomy as those who can independently guide and evaluate their academic writing process, confidently share their work with teachers and peers via digital platforms, and effectively apply technological learning strategies (such as AI-assisted or cloud-based co-authoring) to achieve their writing goals.

2.2. Learner Autonomy in L2 Academic Writing

In the specific domain of L2 academic writing, research has consistently demonstrated the significant demand for L2 academic writing skills among university students across disciplines. Academic writing presents unique challenges as it requires learning and following complex rules that vary across disciplines, with each writing context having different audiences and purposes (Paltridge, 2004). Students in the social sciences and humanities frequently need to produce substantial research papers that demonstrate disciplinary knowledge and rhetorical awareness (Hale et al., 1995). These challenges are particularly pronounced for non-native speakers, whose prior writing experiences may differ substantially from the expectations in their current academic environments (Hong, 2021; Zhang et al., 2025).

The development of academic writing skills involves more than just acquiring discipline-specific language; it requires induction into academic discourse communities and their practices (Starfield, 2001). In this challenging context, learner autonomy becomes particularly important for L2 writers who must independently navigate these disciplinary expectations while developing their writing skills (Lin, 2013; Liu & Qi, 2017). Recent studies have

emphasized that successful academic writers must develop self-regulation skills and autonomy to meet the demands of academic writing tasks (Leeming, 2024; Suwannaphima & Vibulphol, 2023). This necessity for autonomy is operationalized through a set of distinct abilities.

Autonomous writers demonstrate distinct characteristics that can enhance L2 learners' writing performance, including determining objectives, defining content and progression, selecting appropriate methods and techniques, monitoring the acquisition process, and evaluating what has been acquired (Holec, 1981, p. 3). More specifically, autonomous writers engage in processes such as goal setting, planning, making decisions about what and how to learn, self-monitoring, and self-evaluation (Mickwitz, 2020). Emotionally, they feel ownership of their writing process and maintain positive attitudes toward taking charge of their education by establishing learning objectives, selecting suitable learning techniques, and assessing their own progress (Head et al., 2014; Mondahl et al., 2009).

While the importance of autonomy in L2 writing is well-established (Cotterall, 1995; Little, 2022), there remains limited research on how different learning arrangements—specifically individual versus collaborative work, especially under technological assistance—impact learner autonomy in L2 academic writing contexts (Mahapatra, 2024; Parnrod & Darasawang, 2018; Wigglesworth & Storch, 2012).

2.3. *The Influences of Individual and Collaborative Work*

2.3.1. Individual and Collaborative Work in L2 Academic Writing

Different teaching methods or learning strategies significantly impact learner autonomy in L2 academic writing (Sulistiyo & Kamil, 2022). Among these approaches, individual and pair work represent two prominent approaches in this domain (Alastuey & de Lizarrondo Larumbe, 2017; Storch, 2002). Research has consistently demonstrated that the impact of these writing arrangements on learner autonomy cannot be overlooked. From a theoretical perspective, Littlewood (1999) emphasized that autonomy should be considered within cultural contexts, noting that East Asian educational traditions may foster different forms of autonomy than Western contexts. His framework distinguishes between proactive autonomy (learners taking charge of their own learning) and reactive autonomy (learners organizing resources autonomously once a direction has been initiated). Similarly, Storch (2002) identified distinct patterns of dyadic interaction in ESL pair work, with collaborative and expert/novice patterns showing the most promise for language development through effective scaffolding and co-construction of knowledge. More recent studies have expanded on these foundations, with Liu et al. (2018) revealing that in digital storytelling, cooperative work yields superior learning outcomes, including higher learner autonomy, compared to individual work. For L2 academic writing specifically, which constitutes a significant component of language learning (Ivanič & Simpson, 2014; Paltridge, 2004), both individual and collaborative approaches play crucial roles in students' development of autonomy (Dobao, 2012; Kaweera et al., 2019; Zabihi & Bayan, 2020). These studies collectively provide important context for understanding how different writing arrangements might influence learner autonomy development in various cultural and educational settings.

Social constructivist theory, based on Vygotsky (1978), provides a foundation for understanding how collaboration enhances autonomy. According to this theory, learning is fundamentally a social process where knowledge is co-constructed through interaction with others. When learners collaborate, they engage in dialogic interactions that help them construct knowledge together, gradually internalizing strategies and skills that they can later apply independently. This process of internalization is central to the development of autonomy, as learners move from other-regulation to self-regulation (Thorne & Lantolf, 2006).

2.3.2. Individual and Pair Work on Learner Autonomy in L2 Academic Writing

Researchers paid more attention to the positive effects that collaborative learning provides on learner autonomy in different portions in language learning. One of the topics Şener and Mede (2023) explored was the role of collaborative learning in promoting learner autonomy. The result indicated that collaborative learning has an enhanced effect on learner autonomy, and learners have positive attitudes towards this learning mode. Through collaboration, students can move from dependence towards independence, thus developing their autonomy (Geary, 1998; Harmer, 2008; Wijaya, 2021). In oral learning, pair work is a significantly useful strategy to promote learners' speaking skills and will enhance learners' independent learning ability, thus increasing learner autonomy (Azam, 2024). Similarly, in the reading field, which is also considered to be essential in language learning (Burt et al., 2003), Timothy et al. (2023) investigated the effect of collaborative strategy on reading ability and reading autonomy of middle school students. This study demonstrated that collaborative work would increase students' independence.

Regarding the field of writing, Tanyeli Zeki and Kuter (2018) analyzed the independence shown by students in different stages of writing, thus concluded that collaborative writing stimulates students to have better manifestations in their autonomy. In addition, collaborative writing in L2 promotes learner autonomy in the context of student

groups, and most students support this activity (Shehadeh, 2011). Wigglesworth and Storch (2012) further demonstrated that collaborative writing can lead to better text quality and language learning opportunities through pooled knowledge and mutual scaffolding. This aligns with Storch's (2002) earlier findings on the benefits of collaborative patterns of interaction, where learners engage in joint problem-solving and knowledge construction.

However, these studies have focused primarily on the impact of cooperation or pair work on learner autonomy. Few studies make a comparative analysis on individual and pair work for a similar task, seldom focusing on the performance of learner autonomy in L2 academic writing.

2.3.3. Technology-Assisted Writing and Digital Autonomy

As higher education increasingly transitions into digital spaces, the mediums through which students engage in both individual and collaborative writing have fundamentally changed. Cloud-based collaborative platforms (e.g., Google Docs, Wikis) have revolutionized pair work by enabling synchronous and asynchronous co-regulation, allowing learners to document iterative revisions and engage in continuous peer-to-peer interactions (Wang & Ren, 2024). In these digital environments, collaborative writing not only enhances language acquisition but also cultivates digital literacy and shared autonomy.

Simultaneously, individual writing has been transformed by the advent of generative AI. AI tools provide immediate, personalized formative feedback, assist in idea generation, and offer linguistic scaffolding (Mahapatra, 2024). Recent research highlights that AI functions as a highly effective “More Knowledgeable Other”, operationalizing the Zone of Proximal Development at scale by providing adaptive scaffolding that allows students to overcome creative blocks and self-edit with unprecedented autonomy (Hong, 2026). Furthermore, for EFL students specifically, AI tools have proven invaluable for structural organization, information management, and language enhancement, effectively reducing the cognitive load associated with writing in a non-native language (Hong & Litwin, 2025). Consequently, individual writers are no longer working in complete isolation; they are exercising a new form of digital autonomy where they must critically evaluate and integrate AI-generated assistance while maintaining their authentic authorial voice.

However, these studies have focused primarily on the impact of the isolated use of specific digital tools. Few studies make a comparative analysis of individual and pair work for a similar task within a modern, technology-mediated environment, seldom focusing on how the integration of digital tools (like AI and cloud platforms) influences the performance of learner autonomy in L2 academic writing. The present study aims to address this gap by investigating both technology-mediated individual and collaborative approaches to L2 academic writing and their respective impacts on learner autonomy.

Specifically, the present study addresses the following research questions:

- (1) What are students' perspectives about individual and pair work in technology-mediated L2 academic writing?
- (2) How do students who have exclusively and recently experienced either individual or pair work perceive the development of their learner autonomy in L2 academic writing?

3. Methods

This study employed a qualitative approach to investigate learner autonomy through the experiences of students who had recently been engaged in L2 academic writing around the period of the investigation. To gain deep insights into students' perceptions and experiences, we adopted a semi-structured interview methodology, which offers the advantage of being “reasonably objective while still permitting a more thorough understanding of the respondent's opinions” (Borg & Gall, 1983, p. 442). This approach enabled us to explore the underlying reasons behind students' responses in ways that would not be possible using quantitative methods such as questionnaires. Our research instruments included semi-structured interviews for data collection and thematic analysis for data analysis.

3.1. Research Context & Participants

The study was conducted at a university in Macau with 20 third-year English majors aged 20–21 years. Prior to data collection, we consulted with the program coordinator and reviewed the curriculum to ensure participants had comparable academic backgrounds. We also verbally confirmed with all participants that this was their first course involving lengthy academic writing projects, ensuring similar prior experience with this type of extended academic writing task. All participants were enrolled in the same L2 academic writing course, receiving identical teaching content, methods, and materials. As a cohort, they were admitted under the same language proficiency requirement,

received the same duration of formal classroom instruction, and also completed a comparable volume and type of academic writing exercises, which allowed for natural control for academic writing experience and proficiency level.

Two groups of participants—10 students who just finished working individually to draft a full-length research proposal and 10 students who worked collaboratively in pairs—were interviewed right after their respective writing activities. With informed consent, the participation was voluntary, and students self-selected their preferred writing approach (individual or collaborative) based on personal preference, which happened to be of exactly equal with both approaches. Throughout the course, all students had access to standard digital writing technologies, including cloud-based word processors (e.g., Google Docs, Microsoft Word Online) for drafting and sharing, as well as generative AI tools (e.g., ChatGPT, Grammarly) which they were permitted to use as supplementary writing aids.

3.2. Measures

Semi-Structured Interview Questions

After the students' writing activity, semi-structured interviews were employed to investigate their individual and collaborative writing processes in depth through self-reflection. This approach allowed for open-ended questions that did not constrain participants' responses (Karatsareas, 2022). The interview questions were initially developed based on an extensive review of literature on learner autonomy and subsequently adapted from Shen et al.'s (2020) questionnaire. The final interview protocol addressed five key dimensions of learner autonomy: (1) learners' expected role of the teacher (5 questions), (2) learners' independence (5 questions), (3) learners' confidence in study ability (3 questions), (4) learners' ability to evaluate the learning process (3 questions), and (5) learning strategies used (4 questions). Crucially, within the "learning strategies used" dimension, specific probes were included to explore how students utilized digital tools, collaborative software, and AI applications to support their writing process. Additional follow-up questions emerged during the interview process. This structured yet flexible approach allowed us to gather comprehensive details and personal narratives while ensuring consistency across responses. All interviews were audio-recorded and later transcribed for analysis. The interviews lasted for 30 to 60 min.

3.3. Data Collection & Analysis

3.3.1. Data Collection

We conducted both online and face-to-face interviews with participants, recording all sessions. To create a comfortable environment and elicit authentic responses, participants were permitted to respond in either Chinese or English according to their preference. Theoretical saturation was found to be reached at the 7–8th participant for each group. Following data collection, we transcribed the audio recordings using iFLYTEK software and carefully reviewed the transcripts to correct any transcription errors.

Initially, we imported the data into NVivo for analysis; however, due to the nuanced nature of our data and the limitations of NVivo's keyword-frequency-based classification (Woolf & Silver, 2017), we determined that manual coding using Excel and Word would provide more detailed and contextually sensitive analysis.

3.3.2. Data Analysis–Thematic Analysis

Thematic analysis is widely employed by qualitative researchers for analyzing rich descriptive data (Naeem et al., 2023). We selected this method to identify common and unique perspectives on learner autonomy in L2 academic writing and to examine themes related to the effects of individual and collaborative approaches. Thematic analysis enabled us to interpret structured, sequential research data, contributing to a comprehensive understanding of participants' experiences.

Our manual coding process followed Naeem et al.'s (2023) six-step framework for conceptual model development in qualitative research. The coding was conducted by three researchers, with any differences in coding resolved through deliberation and consensus-building to ensure consistency (>90%). First, we read the transcripts multiple times to develop a holistic understanding of all responses. Next, we identified recurring words and meanings as keywords. We then transformed these keywords into codes relevant to our research questions. Subsequently, we categorized and organized these codes into groups to identify patterns that provided insights into our research questions, forming preliminary themes. During the conceptualization stage, we refined these themes based on our operational definition of learner autonomy, focusing particularly on goal setting, self-monitoring, and self-evaluation, etc. Open coding was done to identify main themes, sub-themes and code examples. Axial coding was then performed to find relationships among the themes. These collectively address our research questions.

4. Findings

4.1. Themes and Sub-Themes

The open coding of the interview data revealed distinct patterns in how students perceive and experience individual and collaborative approaches to L2 academic writing. Students across both writing conditions recognized that collaborative writing offers increased companionship and idea stimulation during the writing process, while individual writing provides more consistent planning opportunities and reduces interpersonal pressure (see Table 1). Importantly, our findings indicate that collaborative work enhances specific dimensions of learner autonomy in L2 academic writing, particularly in the areas of goal setting, self-monitoring, and self-evaluation.

Table 1. Open coding results.

Theme	Sub-themes	Representative Quotes	Interpretation
Autonomy in Individual Work	Freedom of decision-making	<i>Because individual writing is more conducive to the divergence of my own thinking, that is, I don't need to communicate with others, I can organize the directions I'm interested in myself (HY)</i>	Individual work allows students complete control over their writing direction and interests
	Control over pace and schedule	<i>If I were to do pair collaboration, I might find that the communication between the two of us could delay things in terms of time (XD)</i>	Individual writers can better control their own schedule and pace
	Self-directed learning	<i>I think I am someone who can better control my rhythm for these kinds of tasks. But if you make me do pair writing, I would feel my rhythm being disrupted (XD)</i>	Individual writers feel more in control of their learning process
Autonomy in Collaborative Work	Shared decision-making	<i>I choose pair work because one person's thinking is definitely limited. If it's two people, through discussion, there can be some collisions of ideas (JN)</i>	Collaborative work enables shared decision-making and diverse perspectives
	Mutual support	<i>I think it's also motivation, that is, if two people do it together, the efficiency will be higher (JM)</i>	Partners provide motivation and support to each other
	Complementary skills	<i>I think [my partner] is more technical, I feel he is much more flexible in using such software than me... He very well compensates for my disadvantage of not being familiar with computer operations (LY)</i>	Partners can complement each other's skills and knowledge
Self-regulation	Time management	<i>Formulated, it's not really a plan, writing the outline took about 4 h (PS)</i>	Students develop strategies for managing their time
	Goal setting	<i>My goal is to complete the exploration of a certain issue, that is, I will first search for materials, then in these works or other people's writings, understand the points I'm interested in (HY)</i>	Students set specific goals for their writing
	Self-monitoring	<i>I have changed my strategies, and I find a new method in this time, it is much more suitable for me (PS)</i>	Students reflect on and adjust their strategies
Challenges in Individual Work	Limited perspectives	<i>I think the only benefit is that between us two, there might be more different viewpoints emerging, some of which are parts I hadn't thought of (XD)</i>	Individual writers recognize the limitation of having only one perspective
	Emotional pressure	<i>My disadvantage is that I easily get internally consumed, that is, if I get stuck on this problem, I will panic, I will worry (HY)</i>	Individual writers may experience more emotional pressure when facing challenges
	Technical limitations	<i>First, many professional terms I don't know, some research methods although heard in class, but implementing myself is still difficult (QX)</i>	Individual writers may face technical limitations without partner support
Challenges in Collaborative Work	Coordination difficulties	<i>We might need to revise many versions, to create the final version that both of us agree on (XD)</i>	Collaborative writers need to coordinate and reach consensus
	Interpersonal tensions	<i>Because in group cooperation, you inevitably will have closer interactions with classmates or group members who were originally at a certain distance. And closer interactions actually bring many conflicts that everyone doesn't want (DR)</i>	Collaborative work can lead to interpersonal tensions
	Pace synchronization	<i>Because it's two people, there might be times when the pace is not consistent during writing (DR)</i>	Partners may have different working paces

Table 1. Cont.

Theme	Sub-themes	Representative Quotes	Interpretation
Tech-Assistance	Generating ideas/arguments	<i>That is, now we all have this AI software and such. If you say you want to propose viewpoints, actually AI can also provide them to you (XD)</i>	AI-assisted brainstorming and ideation
	Problem solving	<i>If I really can't find [data/materials], AI can help a bit. (QX)</i>	AI-mediated troubleshooting
	Managing writing	<i>Directly used Mendeley and EndNote to annotate everything... making it easier to match. (PS)</i>	Digital workflow and reference organization
	Avoiding over-reliance	<i>English writing emphasizes logic more... so I will not use a lot of translation or various tools to assist my writing. (YY)</i>	Strategic technological restraint
Teacher's Role	Guidance	<i>Helped me slightly modify my topic (XD)</i>	Teachers provide guidance on topic selection and direction
	Feedback	<i>Then check materials myself, then the teacher can modify my first draft, second draft and such, then gradually perfect it (QX)</i>	Teachers provide feedback on drafts
	Emotional support	<i>Then he will also often care about our emotions, that is, in terms of writing, whether we encountered any problems (HY)</i>	Teachers provide emotional support
Learning Strategies	Seeking help	<i>I will seek help from my mentor, because my mentor in this aspect, many of his previously written papers or published journals are beneficial to my thinking. (HY)</i>	Students seek help from teachers or peers when facing challenges
	Self-help	<i>I would go to YouTube or Bilibili to watch SPSS videos... my partner is not as quick to pick up computer applications as I am. (DR)</i>	Students solve problems on his/her own
	Comparative analysis	<i>That should be the comparative method, um, because my topic is mainly under the comparative perspectives, so I often encounter various opinions. (HY)</i>	Students use specific analytical methods
Peer Influence	Idea generation	<i>Because at the beginning of writing, you actually need many new ideas, regardless of whether they are feasible or not, but you need ideas anyway. So among my partner's many different new ideas, I will also be inspired (DR)</i>	Peers inspire new ideas and perspectives
	Emotional support	<i>Then I think my companion very well compensates for my disadvantage of not being familiar with computer operations (LY)</i>	Peers provide emotional support
	Skill complementarity	<i>I think I might provide more ideas... Then he will be precise to the transition between each sentence, whether it's natural, or whether the logic is smooth like that (DR)</i>	Peers complement each other's skills
Preference Factors	Prior experience	<i>I think now we have ground to a state where both sides feel quite comfortable. So if pairing with a new person, might have to go through such a process again (DR)</i>	Prior experience with partners influences preferences
	Personal traits	<i>I don't hope others urge me, except teachers. Because I am a person with my own rhythm. If you urge me, I will feel annoyed (XD)</i>	Personal traits influence preferences for individual or collaborative work
	Task complexity	<i>At that time, after knowing, we were thinking, because you said this time we also need to collect data, or we feel that one person handling these things would be a large amount, then want to find someone to share (DR)</i>	Task complexity influences preferences

Notes: Individual condition = LV, HY, XD, QX, PS, LJ, PB, YE, YH, YX; Group condition = DR, LY, JN, JM, ZZ, WL, WT, EC, ZE, ZA.

4.2. Relationships between Themes

4.2.1. The Autonomy-Collaboration Spectrum

The analysis reveals a complex relationship between autonomy and collaboration in academic writing. Rather than being opposites, they represent different manifestations of learner autonomy. Individual work offers direct

autonomy through complete control over decisions, pace, and content. In contrast, collaborative work offers a form of “distributed autonomy” where students exercise autonomy through negotiation, mutual decision-making, and complementary contributions.

Students who prefer individual work often value the freedom to follow their own schedule and make independent decisions without compromise. As one student (XD) noted, “因为一个人写的话，他就可以想写你自己选择的主题，而不是需要去迎合别人的这个想法” (Because when writing alone, you can write about your chosen topic without having to accommodate others’ ideas). Interestingly, the integration of technology has expanded this spectrum. For individual writers, AI often acts as a surrogate collaborator for AI-assisted brainstorming and ideation, allowing them to experience the benefits of a sounding board without the interpersonal friction of human collaboration. XD further explained, “因为现在我们都这个 AI 软件什么的。如果你说你要提观点的话，其实 AI 也能给你提供... 就是不需要同伴” (Because now we all have this AI software and such. If you say you want to propose viewpoints, actually AI can also provide them to you... so you don’t need a partner).

Conversely, students who prefer collaborative work recognize the limitations of individual perspectives and value the “intellectual collision” that occurs through collaboration. They see autonomy as enhanced rather than diminished through collaboration, as expressed by JN: “通过讨论可以进行一些思维上的碰撞，你就可以有更好的一些点子去构思你的写作” (Through discussion, you can have intellectual collisions that give you better ideas for your writing).

4.2.2. Self-Regulation as a Core Component of Autonomy

Regardless of writing mode, self-regulation emerges as a core component of learner autonomy. Both individual and collaborative writers demonstrate self-regulation through time management, goal setting, and self-monitoring. However, the manifestation differs, particularly in how they manage their digital workflow:

Individual writers tend to develop more independent self-regulation strategies, often relying on internal motivation, deadlines, and specialized software to manage their writing. For instance, PS highlighted the use of reference management tools to maintain control over the writing process: “直接用 medaling [Mendeley] 和那个 end note 就把所有该批注的地方都批注了，然后就比较匹配这样子” (Directly used Mendeley and EndNote to annotate everything... making it easier to match). Furthermore, autonomous self-regulation also involves strategic technological restraint. Students actively monitor their reliance on tools to preserve their authorial voice. YY noted, “感觉英文写作会更加重逻辑性... 所以说我不会运用大量的。比如说翻译，或者说各种各样的工具去辅助自己写作业” (English writing emphasizes logic more... so I will not use a lot of translation or various tools to assist my writing).

Collaborative writers develop co-regulation strategies, where partners help regulate each other’s learning. LY described how her partner would provide emotional regulation: “然后有时候比如说我很焦虑...他当时就很及时的劝我” (When I was anxious... he promptly advised me). This suggests that both writing modes can foster autonomy, but they develop different aspects of self-regulation, ranging from digital discernment to interpersonal emotional management.

4.2.3. The Role of External Support in Autonomous Learning

The analysis reveals that autonomous learning does not mean learning in isolation. Both individual and collaborative writers seek external support, but in different ways:

Individual writers tend to seek more targeted support from teachers or engage in tech-mediated troubleshooting for specific problems. XD mentioned, “如果遇到问题的话，我就会觉得有点问题...如果那个时候我正好 writing 有问题，并且没有找老师解决，我就会觉得有点问题” (If I encounter problems and haven’t consulted the teacher, I feel troubled). When traditional resources fail, technology serves as a critical fallback. QX noted how AI acts as a safety net for problem-solving: “实在找不了的话可以可以说 AI 可以帮助一下” (If I really can’t find [data/materials], AI can help a bit).

Collaborative writers have built-in support systems through their partners, creating a “first line of support” before seeking teacher assistance. LY explained her problem-solving approach: “我会先自己想...但是如果我实在是想不到了，我就会问他...如果是我们两个实在都没有办法解决，我们才会求助老师” (I’ll think about it myself first... if I really can’t figure it out, I’ll ask him... only if both of us can’t solve it will we seek help from the teacher). This suggests that autonomy is not about complete independence but about making strategic decisions about when and how to seek support, whether human or digital.

4.2.4. The Influence of Personal Traits and Task Complexity

The preference for individual or collaborative writing is strongly influenced by personal traits and task complexity:

Students with strong preferences for controlling their own pace and schedule tend to prefer individual writing. XD stated, “我觉得我是一个比较能掌控我的节奏的这些这种任务。但是如果你让我觉得是双人写作的话, 我就我会我就会觉得我的节奏被打过来” (I feel I can control my own pace for these tasks. If it were pair writing, I would feel my rhythm being disrupted).

Task complexity also influences preferences, particularly when digital literacy is required. For complex tasks involving data collection, software analysis, or formatting, students often opt for collaboration to achieve digital skill complementarity. DR explained the need for shared workload: “因为你有说这次还要收集数据还要收集数据, 或者我们觉得一个人可能就是要处理这些东西的话量很大, 然后想找一个人分担” (Because this time we needed to collect data, we felt it would be too much for one person to handle, so we wanted to find someone to share the workload). Similarly, LY highlighted how her partner scaffolded her technical skills: “一开始是 EndNote... 就很很熟, 然后也是 Rick 一直在帮着... 他完美的弥补了我对电脑操作不熟悉的这个短板” (At first, I was very unfamiliar with EndNote... and Rick [partner] kept helping... he perfectly made up for my shortcoming of being unfamiliar with computer operations). This suggests that autonomy in academic writing is not just about personal preference but also about making strategic choices based on task demands and technological proficiency.

4.2.5. The Development of Autonomy through Experience

The analysis indicates that autonomy develops through experience with both individual and collaborative writing, as well as through the mastery of digital tools:

Students who have had successful collaborative experiences tend to develop confidence in their ability to negotiate, co-construct knowledge, and share technical burdens. DR mentioned wanting to continue with the same partner because “我们已经磨合到了一个两边都觉得还挺舒服的一个状态” (we’ve already adjusted to a state where both sides feel quite comfortable).

Individual writers develop confidence in their independent problem-solving abilities and digital workflows. PS noted, “I’m not a cautious writer, so sometimes I mistakenly take some unnecessary message into molecule, but luckily I’m working individually so I can just delete it or simply replace it.” This suggests that both writing modes can contribute to the development of autonomy, but in different ways and potentially at different stages of learning, ultimately shaping a more resilient and digitally adaptable learner.

5. Discussion

5.1. Students’ Perspectives on Individual and Collaborative Work

Regarding the first research question, our thematic analysis revealed mixed perspectives on both writing approaches, significantly shaped by the integration of modern digital tools. Participants valued individual writing for the complete control it afforded over content, pace, and style. This aligns with Holec’s (1981) conceptualization of autonomy as the ability to take charge of one’s learning. Participants appreciated the freedom to make independent decisions without negotiation or compromise, reflecting Winne’s (2001) emphasis on students taking ownership of their learning process. Interestingly, the traditional challenge of isolation in individual writing was notably mitigated by the use of AI tools, which served as “digital peers” for brainstorming and ideation, allowing students to experience collaborative benefits without interpersonal friction.

However, participants also identified significant challenges in individual writing, including uncertainty, emotional pressure, and technical limitations. These findings resonate with the idea that learner autonomy encompasses planning, monitoring, and assessing learning, which can be difficult to accomplish alone (Mickwitz, 2020). The absence of immediate human feedback often led to anxiety about the quality of work, supporting the notion that confidence is a crucial prerequisite for autonomous learning (Cotterall, 1995).

Collaborative writing elicited equally complex perspectives. On one hand, participants valued the social support, diverse ideas, and immediate feedback that peer collaboration provided. Furthermore, collaboration offered crucial digital skill complementarity, where peers scaffolded each other’s technical shortcomings (e.g., navigating data analysis software or reference managers). This appreciation for collaborative learning supports Vygotsky’s (1978) sociocultural theory and extends Storch’s (2002) findings on the benefits of collaborative patterns of interaction in ESL pair work. The emotional support and reduced anxiety reported by participants suggested that collaborative learning has an enhanced psychological effect on learner autonomy (Şener & Mede, 2023).

On the other hand, participants identified challenges in collaborative writing, including interpersonal conflicts, uneven workload distribution, pace synchronization, and compromised personal voice. These findings echo concerns raised by Wigglesworth and Storch (2012) about the potential complexities of collaborative writing. The frustration expressed by some participants regarding the need to compromise on ideas indicates the importance of learner agency in autonomous learning environments.

Interestingly, our findings revealed that preferences were not static but highly contextualized, influenced by factors such as task complexity, time constraints, and technological proficiency. This contextual nature of preferences supports earlier findings that autonomy should be considered within cultural and environmental contexts (Littlewood, 1999), noting that modern educational environments foster new, digitally-mediated forms of autonomy.

5.2. Impact on Learner Autonomy

Addressing our second research question, the analysis revealed distinct patterns in how individual and collaborative work influenced different dimensions of learner autonomy. In goal setting, individual writing promoted “proactive autonomy”, where learners independently established their objectives (Littlewood, 1999). Collaborative writing, however, facilitated what might be called “reactive autonomy”, where learners organize resources autonomously once a direction has been initiated through social interaction.

For self-monitoring, individual writing necessitated learners’ involvement in monitoring their learning (Mickwitz, 2020). In our study, this was frequently operationalized through digital workflow management (e.g., using Mendeley or EndNote to organize literature) and strategic technological restraint (e.g., deliberately limiting the use of translation tools to preserve logical flow and authorial voice). Collaborative writing introduced peer-mediated regulation, aligning with Vygotsky’s (1978) principles of social learning, where knowledge and technical skills are co-constructed through interaction with others. This supports the concept that learners move from other-regulation to self-regulation through social interaction (Thorne & Lantolf, 2006).

In self-evaluation, individual writing often led to uncertainty, where learners struggled to objectively assess their work, frequently relying on tech-mediated troubleshooting as a fallback before consulting teachers. Collaborative writing provided opportunities for peer assessment, offering multiple perspectives on quality and appropriateness (Tanyeli Zeki & Kuter, 2018). This finding extends research on how collaborative writing stimulates students to have better autonomy manifestations particularly in the evaluation stage of writing.

Our analysis also revealed a phenomenon of “complementary autonomy”, where individual and collaborative approaches mutually reinforced different aspects of learner autonomy. This concept builds upon Littlewood’s (1999) distinction between proactive and reactive autonomy, suggesting that these dimensions are not merely distinct but can be synergistic when appropriately combined. Furthermore, this synergy now extends beyond human interaction to include human-machine collaboration. The benefits of collaborative writing clearly demonstrate the principle that learning is fundamentally a social process (Vygotsky, 1978); however, our study shows that the quality of collaboration—whether with a human peer or an AI tool—significantly impacts autonomy development.

5.3. Theoretical Contributions and Pedagogical Implications

Our research innovates beyond previous studies in several significant ways. While existing literature has primarily focused on text quality and language learning outcomes in collaborative writing (Wigglesworth & Storch, 2012), our study explicitly connects interaction patterns to autonomy development across multiple dimensions. Furthermore, our study expands traditional sociocultural frameworks by highlighting how digital tools and AI act as non-human mediators in the zone of proximal development, fostering a new dimension of “digital autonomy” (Hong, 2026). Previous research has often examined collaborative writing’s general impact on learner autonomy (Shehadeh, 2011), but our findings provide deeper insights into the psychological and technological processes underlying autonomy in both individual and collaborative contexts.

Perhaps most significantly, this research challenges the dichotomous thinking that often characterizes discussions of learner autonomy. Rather than positioning autonomy as either individually or socially constructed, our findings suggest a dynamic interplay between personal agency, social interdependence, and technological mediation (Balcikanli, 2010). This reconceptualization has important implications for L2 writing pedagogy, suggesting that fostering learner autonomy requires a balanced approach that strategically incorporates individual work, collaborative experiences, and digital literacy training tailored to learners’ developmental needs.

By illuminating the complex relationship between writing modes, digital tools, and autonomy development, this study contributes to a more refined understanding of how pedagogical approaches can be optimized to support

learner autonomy in modern L2 academic writing contexts. Previous research has suggested that collaboration can help students move from dependence toward independence (Geary, 1998; Harmer, 2008; Wijaya, 2021). However, our findings indicate that this process is neither linear nor universal. Instead, the development of autonomy in L2 academic writing appears to benefit from a thoughtful integration of both individual and collaborative approaches, augmented by technology.

Hence, at earlier stages of conceptualization, planning, and idea drafting, L2 writers would benefit more from collaborative efforts and AI-assisted ideation. Then, the mid-stages of literature and methodological review, and sectional write-ups would benefit more from individual work supported by digital workflow tools and reference managers. In the later stage, where writers need to put everything together and perform final editing, a collective effort and peer-assisted technical troubleshooting may once again show advantages.

5.4. Limitations

This study has several limitations that should be acknowledged. First, our reliance on semi-structured interviews as the primary data collection method, while theoretically justified for exploring the psychological dimensions of learner autonomy, may limit the comprehensiveness of our findings. Although interviews provided rich insights into students' subjective experiences and perceptions—which are central to understanding autonomy as an internal psychological construct (Holec, 1981)—the absence of triangulation with other data sources represents a methodological limitation. Future research would benefit from a mixed-methods approach that combines interview data with: (a) analysis of actual writing products to examine textual evidence of autonomous behaviors such as goal-setting strategies, self-correction patterns, and evidence of self-monitoring; (b) classroom observation protocols to document real-time collaborative interactions and individual writing behaviors; (c) pre- and post-task questionnaires measuring learner autonomy orientations; and (d) learning analytics data from digital writing platforms to track revision patterns, time allocation, and help-seeking behaviors. Such triangulation would provide a more comprehensive understanding of how different writing arrangements influence both the subjective experience and objective manifestations of learner autonomy.

Second, while the cohort design allowed for natural control over formal academic writing training, former academic writing experience, and language proficiency, we did not formally assess individual differences in communication styles and interpersonal skills that could affect collaborative writing effectiveness, as well as baseline autonomy orientations or learning style preferences that may predispose students toward certain writing arrangements. As stated, future studies could employ pre-task assessments to measure these unmeasured variables and either control for them statistically or examine them as moderating factors in the analysis. Relatedly, the self-selection of writing arrangements (individual or collaborative) introduces potential bias, as students may have chosen approaches that aligned with their pre-existing preferences or autonomy orientations. Future studies might employ random assignment to writing conditions to address this limitation.

Third, while our data analysis involved three researchers with differences resolved through deliberation and consensus-building, we acknowledge that we did not formally calculate inter-coder reliability coefficients or provide comprehensive documentation of our complete coding procedures. This limitation affects the transparency and reproducibility of our analysis process. Future research should implement more systematic reliability verification procedures, including formal inter-coder reliability calculations and detailed documentation of coding manuals and procedures to enhance methodological rigor and reproducibility. Furthermore, our study focused on a specific context—English majors at a university in Macau—which may limit the generalizability of our findings to other educational contexts, disciplines, or cultural settings. Additional research across diverse contexts would enhance our understanding of how individual and collaborative approaches influence learner autonomy in L2 academic writing.

6. Conclusions

This study provides a comparative analysis of how individual and collaborative approaches influence learner autonomy in technology-mediated L2 academic writing, addressing an important gap in the research literature. Our findings reveal that both approaches offer distinct advantages and challenges in shaping students' writing experiences, autonomous behaviors, and digital practices.

Collaborative writing demonstrates significant positive effects on key dimensions of learner autonomy, particularly self-monitoring and self-evaluation. The partnership creates a natural scaffolding environment where students can solve problems, exchange ideas, leverage digital skill complementarity, and critically reflect on their own and each other's performance. Through this process, they mutually explain and internalize writing criteria to achieve their goals. Conversely, individual writing fosters autonomy through consistent personal planning and

freedom from the potential stresses of interpersonal communication and workload negotiation. Crucially, our findings show that the traditional isolation of individual work is increasingly mitigated by the use of AI as a “digital peer” for ideation and advanced software for digital workflow management. This allows solo students to exercise complete control over their time and resources—an important aspect of modern self-direction.

However, rather than advocating for one approach over the other, we suggest that instructors should understand the unique dynamics and benefits of each writing arrangement and provide students with options based on their needs, preferences, and the specific requirements of writing tasks. Additionally, educators should offer appropriate support tailored to each context, such as clear guidelines for human collaboration, training in digital discernment and strategic technological restraint, and strategic feedback for individual writers navigating AI and digital tools.

By recognizing the complementary strengths of individual and collaborative approaches—and the transformative role of digital mediation within them—L2 writing instructors can create learning environments that effectively foster learner autonomy while accommodating diverse student needs and preferences. Future research should continue to explore the complex interplay between writing arrangements, learner characteristics, technological integration, and the development of autonomy in modern L2 academic writing contexts.

Author Contributions

Z.W.: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Writing—original draft, Writing—review & editing; Q.C.: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Writing—original draft; W.C.H.H.: Formal analysis, Validation, Writing—original draft, Writing—review & editing, Supervision. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Institutional Review Board Statement

Ethical approval was obtained from the City University of Macau with ethical approval reference number 2024-RE-31.

Informed Consent Statement

Informed consent was obtained from all subjects involved in the study.

Data Availability Statement

Data will be available upon request.

Conflicts of Interest

The authors declare no conflict of interest.

Use of AI and AI-Assisted Technologies

During the preparation of this work, the authors used ABACUS.AI to identify literature sources, refine the language and compile the reference list. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

- Alastuey, M. C. B., & de Lizarrondo Larumbe, P. M. (2017). Collaborative writing in the EFL secondary education classroom: Comparing triad, pair and individual work. *Huarte de San Juan. Filología y Didáctica de la Lengua*, 17, 254–275. <https://revista-hsj-filologia.unavarra.es/hsj-Filologia/article/view/3017>.
- Azam, A. (2024). The role of pairwork in enhancing speaking skills and learner autonomy among Bangladeshi tertiary level students. *International Journal of English Language Studies*, 6(3), 61–69. <https://doi.org/10.32996/ijels.2024.6.3.9>.
- Balcikanli, C. (2010). Learner autonomy in language learning: Student teachers’ beliefs. *Australian Journal of Teacher Education (Online)*, 35(1), 90–103. <https://search.informit.org/doi/10.3316/ielapa.850642220019550>.
- Borg, W. R., & Gall, M. D. (1983). *Educational research* (4th ed.). Longman.

- Burt, M., Peyton, J. K., & Adams, R. (2003). *Reading and adult English language learners: A review of the research*. National Center for ESL Literacy Education.
- Cotterall, S. (1995). Readiness for autonomy: Investigating learner beliefs. *System*, 23(2), 195–205. [https://doi.org/10.1016/0346-251X\(95\)00008-8](https://doi.org/10.1016/0346-251X(95)00008-8).
- Dobao, A. F. (2012). Collaborative writing tasks in the L2 classroom: Comparing group, pair, and individual work. *Journal of Second Language Writing*, 21(1), 40–58. <https://doi.org/10.1016/j.jslw.2011.12.002>.
- Geary, W. T. (1998, April). *From dependence toward independence via interdependence*. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA, USA.
- Hale, G., Taylor, C., Bridgeman, B., Carson, J., Kroll, B., & Kantor, R. (1995). A study of writing tasks assigned in academic degree programs. *ETS Research Report Series*, 1995(2), i–61. <https://doi.org/10.1002/j.2333-8504.1995.tb01678.x>.
- Harmer, J. (2008). *How to teach English*. Oxford University Press.
- Head, A., Xu, Y., & Wang, J. (2014). ToneWars: Connecting language learners and native speakers through collaborative mobile games. In S. Trausan-Matu, K. E. Boyer, M. Crosby, & K. Panourgia (Eds.), *Intelligent tutoring systems, ITS 2014* (Vol. 8474, pp. 388–393). Springer. https://doi.org/10.1007/978-3-319-07221-0_46.
- Holec, H. (1981). *Autonomy and foreign language learning*. Pergamon Press.
- Hong, W. C. H. (2021). Improving English as a foreign language learners' writing using a minimal grammar approach of teaching dependent clauses: A case study of Macao secondary school students. In B. Reynolds, & M. F. Teng (Eds.), *Innovative approaches in teaching English writing to Chinese speakers* (pp. 67–90). De Gruyter Mouton. <https://doi.org/10.1515/9781501512643-004>.
- Hong, W. C. H. (2023). The impact of ChatGPT on foreign language teaching and learning: Opportunities in education and research. *Journal of Educational Technology and Innovation*, 5(1), 37–45. <https://doi.org/10.61414/jeti.v5i1.103>.
- Hong, W. C. H. (2026). Reflecting on three years of generative AI in higher education: Successes, struggles, and solutions. *Artificial Intelligence Advances in Education*, 1(1), 29–39. <https://aiaie.scs-journals.com/index.php/aiaie/article/view/916>.
- Hong, W. C. H., & Litwin, A. (2025). From human guidance to digital assistance: EFL thesis-writing students' experiences of support before and after the AI revolution. *Journal of Educational Technology and Innovation*, 7(3), 21–34. <https://doi.org/10.61414/a3wy7546>.
- Ivanič, R., & Simpson, J. (2014). Who's who in academic writing? In *Critical language awareness* (pp. 141–173). Routledge. <https://doi.org/10.4324/9781315845661-14>.
- Karatsareas, P. (2022). Semi-structured interviews. In R. Kircher & L. Zipp (Eds.), *Research methods in language attitudes* (pp. 99–113). Cambridge University Press.
- Kaweera, C., Yawiloeng, R., & Tachom, K. (2019). Individual, pair and group writing activity: A case study of undergraduate EFL student writing. *English Language Teaching*, 12(10), 1–13. <https://doi.org/10.5539/elt.v12n10p1>.
- Leeming, P. (2024). The influence of small groups on leader stability and task engagement in the language classroom. *Language Teaching Research*. Advance online publication. <https://doi.org/10.1177/13621688241249288>.
- Lin, Z. (2013). Capitalising on learner agency and group work in learning writing in English as a foreign language. *TESOL Journal*, 4(4), 721–743. <https://doi.org/10.1002/tesj.60>.
- Little, D. (2022). Language learner autonomy: Rethinking language teaching. *Language Teaching*, 55(1), 64–75. <https://doi.org/10.1017/S0261444820000488>.
- Littlewood, W. (1999). Defining and developing autonomy in East Asian contexts. *Applied Linguistics*, 20(1), 71–94. <https://doi.org/10.1093/applin/20.1.71>.
- Liu, H., & Qi, W. (2017). Establishing learner autonomy in China's universities through achievement presentation. *Journal of Language Teaching and Research*, 8(3), 588–593. <https://doi.org/10.17507/jltr.0803.18>.
- Liu, M.-C., Huang, Y.-M., & Xu, Y.-H. (2018). Effects of individual versus group work on learner autonomy and emotion in digital storytelling. *Educational Technology Research and Development*, 66, 1009–1028. <https://doi.org/10.1007/s11423-018-9601-2>.
- Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments*, 11(1), 9. <https://doi.org/10.1186/s40561-024-00295-9>.
- Mickwitz, Å. (2020). Students' writing process and the potential influence of student autonomy. *Educare*, 2020(2), 1–25.
- Mondahl, M., Razmerita, L., & Rasmussen, J. (2009, October). Social software, thinking styles, personalization and case-based foreign language learning: The quest for new pedagogical models in higher education. In *8th European Conference on eLearning 2009* (pp. 314–322). Academic Publishing.
- Naeem, M., Ozuem, W., Howell, K., & Ranfagni, S. (2023). A step-by-step process of thematic analysis to develop a conceptual model in qualitative research. *International Journal of Qualitative Methods*, 22. <https://doi.org/10.1177/16094069231205789>.
- Paltridge, B. (2004). Academic writing. *Language Teaching*, 37(2), 87–105. <https://doi.org/10.1017/S0261444804002216>.

- Parnrod, U., & Darasawang, P. (2018). Group work and learning strategies by EFL students with different cognitive styles: Closing gaps for implementing cooperative learning in language classrooms. *Journal of Institutional Research South East Asia*, 16(1), 24–37.
- Sakai, S., Takagi, A., & Chu, M. P. (2008). Promoting learner autonomy: Student perceptions of responsibilities in a language classroom in East Asia. *Educational Perspectives*, 43(1–2), 12–27. <https://eric.ed.gov/?id=EJ912111>.
- Şener, B., & Mede, E. (2023). Promoting learner autonomy and improving reflective thinking skills through reflective practice and collaborative learning. *Innovation in Language Learning and Teaching*, 17(2), 364–379. <https://doi.org/10.1080/17501229.2022.2047694>.
- Shegadeh, A. (2011). Effects and student perceptions of collaborative writing in L2. *Journal of Second Language Writing*, 20(4), 286–305. <https://doi.org/10.1016/j.jslw.2011.05.010>.
- Shen, B., Bai, B., & Xue, W. (2020). The effects of peer assessment on learner autonomy: An empirical study in a Chinese college English writing class. *Studies in Educational Evaluation*, 64, 100821. <https://doi.org/10.1016/j.stueduc.2019.100821>.
- Starfield, S. (2001). “I’ll go with the group”: Rethinking “discourse community” in EAP. In J. Flowerdew, & M. Peacock (Eds.), *Research perspectives on English for academic purposes* (pp. 132–147). Cambridge University Press.
- Storch, N. (2002). Patterns of interaction in ESL pair work. *Language Learning*, 52(1), 119–158. <https://doi.org/10.1111/1467-9922.00179>.
- Sulistiyo, U., & Kamil, D. (2022). Language learning strategies and learner autonomy: The case of Indonesian tertiary EFL students. *LEARN Journal: Language Education and Acquisition Research Network*, 15(1), 257–281.
- Suwannaphima, S., & Vibulphol, J. (2023). Fostering learner autonomy of Thai lower secondary school students in project-based English instruction. *LEARN Journal: Language Education and Acquisition Research Network*, 16(2), 421–439. <https://so04.tci-thaijo.org/index.php/LEARN/article/view/266946>.
- Tanyeli Zeki, N., & Kuter, S. (2018). Impact of collaborative and reflective writing activities on students’ autonomy in writing. *Quality & Quantity*, 52, 343–360. <https://doi.org/10.1007/s11135-017-0615-3>.
- Thorne, S. L., & Lantolf, J. P. (2006). A linguistics of communicative activity. In S. Makoni, & A. Pennycook (Eds.), *Disinventing and Reconstituting Languages* (pp. 170–195). Multilingual Matters & Channel View Publications. <https://doi.org/10.2307/jj.27939678.11>.
- Timothy, A., Morah, P., Uguma, V., Obiekezie, E., Ironbar, V., & Eba, M. (2023). Using the collaborative learning approach to enhance reading competence, interest, autonomy and reduce reading anxiety in secondary school children. *Ianna Journal of Interdisciplinary Studies*, 5(1), 109–123.
- Trim, J. L. M. (1978). *Some possible lines of development of an overall structure for a European unit/credit scheme for foreign language learning by adults*. Council of Europe.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Wang, L., & Ren, B. (2024). Enhancing academic writing in a linguistics course with generative AI: An empirical study in a higher education institution in Hong Kong. *Education Sciences*, 14(12), 1329. <https://doi.org/10.3390/educsci14121329>.
- Wigglesworth, G., & Storch, N. (2012). What role for collaboration in writing and writing feedback? *Journal of Second Language Writing*, 21(3), 153–173. <https://doi.org/10.1016/j.jslw.2012.09.005>.
- Wijaya, K. F. (2021). Collaborative and problem-based learning in promoting Indonesian EFL learners’ learning autonomy. *Polyglot: Jurnal Ilmiah*, 17(2), 191–211. <https://doi.org/10.19166/pji.v17i2.3590>.
- Winne, P. H. (2001). Self-regulated learning viewed from models of information processing. In B. J. Zimmerman, & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (pp. 153–189). Lawrence Erlbaum Associates.
- Woolf, N. H., & Silver, C. (2017). *Qualitative analysis using NVivo: The five-level QDA® method*. Routledge.
- Xu, X., Hong, W. C. H., Zhang, Y., Jiang, H., & Liu, J. (2024). Learning paths design in personal learning environments: The impact on postgraduates’ cognitive achievements and satisfaction. *Innovations in Education and Teaching International*, 61(4), 748–763. <https://doi.org/10.1080/14703297.2023.2189603>.
- Xu, X., Li, Z., Hong, W. C. H., Xu, X., & Zhang, Y. (2024). Effects and side effects of personal learning environments and personalized learning in formal education. *Education and Information Technologies*, 29(15), 20729–20756. <https://doi.org/10.1007/s10639-024-12685-0>.
- Xu, X., Su, Y., Hong, W. C. H., Zhang, Y., & Zhuang, T. (2024). The impact of a personal learning environment on Chinese postgraduates’ online self-regulated learning skills / Impacto de un Entorno Personal de Aprendizaje en las aptitudes de aprendizaje autorregulado en línea de estudiantes de posgrado en China. *Journal for the Study of Education and Development: Infancia y Aprendizaje*, 47(1), 173–205. <https://doi.org/10.1177/02103702231225382>.

- Zabihi, R., & Bayan, M. (2020). Are two voices better than one? Comparing aspects of text quality and authorial voice in paired and independent L2 writing. *Written Communication*, 37(4), 512–535.
<https://doi.org/10.1177/0741088320939542>.
- Zhang, H., Xu, X., Lei, V. N. L., Hong, W. C. H., & Jie, W. (2025). Academic writing challenges and supports for early-stage Chinese postgraduates: A mixed-methods study on teaching-research integration, supervisor engagement, and self-efficacy. *PLoS ONE*, 20(2), e0317470. <https://doi.org/10.1371/journal.pone.0317470>.

Article

Transforming Learning through Productive Failure: A Multi-Dimensional Framework for Fairness in Education

Jing Liu¹, Junjie Yan¹, Xiaonan Luo¹, Wanni Chen² and Xueliang Li^{3,*}

¹ Department of Educational Technology, College of Education, Wenzhou University, Wenzhou 325035, China

² Wenzhou University Affiliated School, Wenzhou 325035, China

³ Aksu Vocational and Technical College, Wensu Town, Wensu County, Aksu 843100, China

* Correspondence: 1808610551@qq.com; Tel.: +86-15258789292

How To Cite: Liu, J., Yan, J., Luo, X., Chen, W., & Li, X. (2026). Transforming Learning through Productive Failure: A Multi-Dimensional Framework for Fairness in Education. *Journal of Educational Technology and Innovation*, 8(3), 35–51. <https://doi.org/10.61414/bsq6e774>

Received: 8 April 2026

Revised: 12 July 2026

Accepted: 20 August 2026

Published: 30 September 2026

Abstract: High-stakes testing and standardized instruction have long dominated education environments. They often place students with learning difficulties on learning tracks with low-cognitive tasks and low classroom participation. This trend further aggravates educational inequality. Research on fairness in education has gradually shifted its focus from resource allocation to the quality of learning processes. However, most current classroom practices still adopt an “error-prevention instruction” approach oriented toward reducing mistakes and maintaining accuracy rates. This approach limits learners’ deep cognitive development. Against this background, this study takes the theory of productive-failure instruction (Productive Failure, PF) as its core. It re-examines the transition of students with learning difficulties from the perspective of fairness in education. It combines literature analysis, conceptual analysis, and theoretical deduction to systematically explore the internal mechanisms through which productive-failure instruction promotes the cognitive development of students with learning difficulties. This study holds that students with learning difficulties are not a group with fixed ability deficits. They are a group of “developmental differences” that long stay in environments with low-quality cognitive engagement. Compared with traditional error-prevention instruction, productive-failure instruction works through three mechanisms: cognitive conflict, delayed instruction, and knowledge reconstruction. It promotes learners’ deep conceptual understanding and transfer of learning. On this basis, this study further proposes a four-dimensional classroom transformation model of productive failure for fairness in education. The model reshapes the classroom learning ecology. It provides new theoretical explanations and practical pathways for the transition of students with learning difficulties and the practice of fairness in education.

Keywords: productive-failure instruction; transition of students with learning difficulties; fairness in education; classroom reform

1. Introduction

In the process of education modernization, narrowing learning outcome gaps while expanding educational opportunities has become a key topic in research on fairness in education. Education access continues to improve, but ability gaps between learners have not narrowed accordingly. The gaps even widen in education contexts with high-stakes testing systems (Strello et al., 2021) and obvious socioeconomic stratification (Sandsør et al., 2023). Against this background, policy adjustments on education quality and fairness across countries essentially reflect the structural limitations of standardized instruction systems in addressing student diversity.



Copyright: © 2026 by the authors. This is an open access article under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Publisher’s Note: Scilight stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

In specific teaching practices, this problem manifests prominently as a “low-cognitive-load orientation” in classroom logic. Teachers oversimplify knowledge content to prevent students from making mistakes. This approach formally ensures the completion of learning tasks, but keeps students with weak foundations stuck in low-level repetitive training for a long time. They can hardly achieve substantive development of their cognitive structure (Li et al., 2021). This phenomenon is not limited to a single country. It widely exists in education systems that emphasize unified progress and standardized evaluation. A fairness in education report by De Bortoli et al. (2023) shows that the coefficient of academic variation in China’s basic education stage is still higher than the average of Nordic countries. This difference not only reflects uneven regional development but also reveals the general dilemma of unified teaching paradigms in addressing cognitive differences.

As an important theoretical breakthrough in cognitive science, productive-failure instruction follows a three-stage mechanism: “cognitive conflict–independent inquiry–collaborative argumentation” (Kapur, 2015). It provides a new path to break through this dilemma. The theory emphasizes that introducing complex problems at the early stage of learning and allowing learners to experience failure can activate higher-level cognitive processing and thus promote knowledge reconstruction and transfer. However, when this student-centered, trial-and-error-based learning approach is introduced into real classrooms, it will generate a series of structural conflicts with existing teaching logic. On the one hand, traditional instruction treats knowledge as a stable, transferable object. Productive-failure instruction understands knowledge as a dynamic cognitive construction process generated in problem situations (Loibl et al., 2017). This difference changes the definition of “knowledge sources” in the classroom. On the other hand, teacher-centered instruction relies on clear guidance and standard answers to ensure the certainty of learning outcomes. Failure-based learning emphasizes exploration processes and multiple solution paths. Its learning value often emerges gradually in delayed feedback (Schwartz et al., 2011). This requirement clearly mismatches the demand for immediate measurability in high-stakes testing systems. In this context, failure can be both an opportunity for cognitive development and a perceived learning risk. For this reason, people systematically avoid failure in teaching practice. Existing studies have revealed the constraints on the transformation of productive failure theory from the perspectives of emotional culture (Chen et al., 2024), classroom power structure (Lee & Kim, 2019), and examination system constraints (Xu & Lu, 2022). But on the whole, there is still a lack of an analytical framework that integrates the cognitive mechanisms of students with learning difficulties and the goal of fairness in education.

The rapid development of artificial intelligence technology has further changed the ways of knowledge acquisition and learning support structures in classrooms. On one hand, artificial intelligence can provide personalized feedback and dynamic scaffolding based on learners’ behavioral data. It offers more precise cognitive support for students with learning difficulties (Zawacki-Richter et al., 2019). On the other hand, if people only use artificial intelligence for quick error correction and answer generation, it may reduce the space for learners to experience cognitive conflict and generate diverse solutions (Khalil & Er, 2023). This will weaken the deep cognitive processing process on which productive failure depends. Therefore, against the background of deep integration of artificial intelligence into classrooms, balancing the relationship between “support and exploration” and between “efficiency and cognitive generation” has become an important premise for re-examining productive-failure instruction. It is necessary to elevate “productive failure” from a single teaching strategy to a mechanism of fairness in education for investigation.

This study aims to reveal the internal logic of the transition of students with learning difficulties driven by productive-failure instruction, and provide a theoretical basis and a practical framework for classroom reform. Its significance is reflected in three aspects:

- (1) Theoretical level: It breaks through the limitation of traditional research on students with learning difficulties that focuses on compensatory support. It constructs a three-in-one theoretical model of “cognitive reconstruction–social negotiation–fair development”, and expands the practical connotation of fairness in education.
- (2) Strategic level: Based on productive failure theory, formative assessment, and research on artificial intelligence-supported learning, it constructs a “four-dimensional classroom transformation model” covering space, time, technology, and evaluation. It addresses the long-term problem of low cognitive challenge and low classroom participation for students with learning difficulties, and provides operable classroom implementation paths for teachers to carry out deep learning-oriented instruction.
- (3) Practical level: Focusing on teacher education, school evaluation, and classroom implementation mechanisms, it proposes practical pathways such as cultivating implementation capacity for fairness in education, building a value-added fairness assessment system, and optimizing productive-failure instruction. It promotes productive-failure instruction to shift from a partial classroom strategy to a systematic education reform practice.

2. Literature Review

2.1. Causes of Learning Difficulties among Students

The issue of students with learning difficulties is a long-standing topic in education. Its causes show multi-dimensional interactive characteristics. Most existing studies agree that learning difficulties result from the combined effect of individual internal factors and external environmental conditions (Yan et al., 2025). With the continuous penetration of artificial intelligence technology into education, this mechanism is being further reshaped.

From the perspective of internal factors, the lack of learning goals, insufficient self-regulation ability, and weak cognitive structure form the core mechanism of learning difficulties. Vague learning goals weaken the directionality and sustainability of learning behaviors. Goal setting and reflection can significantly improve academic performance (Morisano et al., 2010). Insufficient self-regulation ability manifests as defects in planning, monitoring, and metacognitive regulation. It has a direct impact on the use of learning strategies and learning effectiveness (Panadero et al., 2017). Weak cognitive structure is reflected in insufficient prior knowledge and solidified misconceptions. It limits the integration and transfer of new knowledge (Simonsmeier et al., 2022). In digital learning environments, these problems are often further amplified. For example, online learning requires learners to have stronger self-regulation ability. Otherwise, they easily fall into inefficient participation under information overload and fragmented learning (Broadbent & Poon, 2015). Meanwhile, although artificial intelligence-driven personalized recommendations can make up for learners' knowledge gaps, their adaptation mechanism based on existing performance may solidify low-level learning paths and limit the improvement of students' cognitive level (Pelánek et al., 2024).

From the perspective of external factors, insufficient family support, a single school teaching mode, and uneven distribution of social resources constitute important situational conditions for learning difficulties. Lack of family supervision affects students' learning engagement by reducing learning support and educational expectations (Mao et al., 2020). The teaching mode in schools, characterized by unified progress and mechanical training, keeps some students on low-gain learning tracks for a long time by limiting cognitive challenges and differentiated support (Furtak et al., 2012). Differences in socioeconomic status and educational resources affect learning opportunities at a more macro level and may lead to the continuous widening of academic gaps (Nieuwenhuis & Hooimeijer, 2016). In addition, against the background of technology intervention, the digital divide has gradually shifted from “access differences” to “use quality differences”. Disadvantaged student groups still remain in an unfavorable position in obtaining and effectively using high-quality learning resources (Hadjar & Naumann, 2025).

2.2. Theoretical Progress of Productive-Failure Instruction

Manu Kapur proposed the productive failure (PF) theory in 2006. Its core assumption is that “experiencing failure in problem solving without scaffolding support can promote deep transfer of subsequent conceptual learning”. The development of this theory can be divided into three stages: Mechanism exploration stage (2006–2012): Through mathematics problem-solving experiments, researchers confirmed that the PF group performed significantly better than the direct instruction group in conceptual understanding and transfer ability. Key design principles include “delayed instruction”, “diverse solution generation,” and “cognitive conflict stimulation” (Kapur, 2014). Interdisciplinary verification stage (2013–2020): Research expanded to STEM education, language learning, and maker education (Nachtigall et al., 2020). Refined design stage (2021–present): Focusing on the boundary conditions of PF, scholars put forward the concept of “micro-failure” and a task difficulty calibration model (Sinha & Kapur, 2021).

Although PF shows the potential to promote deep learning in many studies, its applicability still faces important controversies. Cognitive load theory points out that without necessary scaffolding, complex problems may exceed beginners' cognitive capacity and reduce learning efficiency (Kirschner et al., 2006). For example, in learning environments with minimal guidance, beginners perform significantly worse than learners receiving direct instruction in complex problem-solving tasks. The reason is that cognitive resources are overconsumed in ineffective search processes (Ashman et al., 2020). The constructivist perspective argues that failure experience is an important condition for knowledge reconstruction. Moderate cognitive conflict helps activate deep processing (Kapur, 2008). A study by Loibl et al. (2020) shows that moderate cognitive conflict prompts learners to re-examine their original cognitive framework. By identifying structural differences and conceptual gaps between solutions, it enhances learners' understanding of deep features of knowledge and transfer ability. This disagreement essentially reflects the balance between “intensity of guidance” and “space for exploration”.

Further studies show that the effect of PF highly depends on learner characteristics and task design conditions (Nachtigall et al., 2020; Sinha & Kapur, 2021). In recent years, research has gradually shifted from “whether it works” to “under what conditions it works”. For example, learners at different age stages have different perceptions

and coping styles toward failure. Younger students are more likely to attribute failure to insufficient ability and develop avoidance behaviors. Older students are more likely to regard failure as an opportunity for strategy adjustment (Metcalf, 2017). In addition, Loibl et al. (2017) found through experimental research that subsequent instruction can effectively promote conceptual understanding only when learners generate diverse solutions during the failure stage. Otherwise, failure may only bring frustration rather than learning benefits. Today, artificial intelligence technology can provide result-oriented prompts or even directly generate solutions based on learners' immediate performance. Although this can improve task completion efficiency to a certain extent, it may reduce learners' independent exploration time during the failure stage. It weakens their comparison of different solution paths and deep cognitive processing. Conversely, when artificial intelligence is used to record different problem-solving paths, delay key feedback, and support peer comparison of solutions, it helps preserve cognitive conflict in the failure process and thus promotes subsequent conceptual construction (Ouyang & Zhang, 2024).

2.3. Realizing Productive-Failure Classrooms for Fairness in Education

Realizing productive failure in classroom situations depends not only on the teaching strategy itself, but also on the redistribution of learning opportunities and cognitive engagement oriented by fairness in education. Fairness in education is not an external condition for PF. It is the internal mechanism for its realization. Only by expanding learning opportunities at the starting point, improving cognitive engagement in the process, and ensuring development space in evaluation can classrooms truly provide sustainable conditions for “productive failure” to occur.

At the learning starting point level, a fairness orientation requires lowering the threshold for learners to enter complex tasks through the redistribution of resources and support (Tan, 2024). For example, China's “three types of classrooms” (delivery classrooms, famous teacher classrooms, and famous school online classrooms) expand opportunities for students in disadvantaged areas to access high-quality problem situations and diverse solutions through online sharing of high-quality resources (Sun et al., 2024). This provides basic conditions for the “multiple solution generation” on which PF depends. The key to this mechanism does not lie in resource transmission itself. It lies in the opportunity space it provides for students to access complex problems and try multi-path solutions.

At the learning process level, the effective implementation of PF depends on the reconstruction of the learning participation structure. Studies show that with online collaboration platforms and learning management systems, teachers can organize inter-school collaborative learning, share generative tasks, and support multi-solution comparison. This helps enhance the depth of students' cognitive engagement during the failure stage (Meyer et al., 2023). Meanwhile, technology-supported formative assessment can provide immediate feedback for students. It helps them turn failure experience into reflective learning resources, so as to prevent failure from remaining as “ineffective frustration”.

At the evaluation and outcome level, fairness in education requires weakening the constraints of a single summative assessment on learning paths and providing institutional space for exploratory learning (Tai et al., 2018). Relevant studies show that excessive reliance on standardized testing and stratified teaching tends to limit students' opportunities to participate in high-challenge tasks and strengthen the differentiation of learning paths (Terrin & Triventi, 2023). In contrast, an evaluation system combining process evaluation and dynamic feedback is more conducive to supporting students' continuous development in the cycle of attempt–failure–revision.

3. Research Methodology

This study belongs to conceptual/theoretical research. It aims to re-examine the internal logic of the transition of students with learning difficulties driven by productive-failure instruction from the perspective of fairness in education, and construct a practical framework for classroom reform on this basis. Compared with empirical research centered on variable testing, conceptual research emphasizes forming explanatory mechanisms and practical models for complex educational phenomena through systematic literature integration, theoretical deduction, and framework construction (Jaakkola, 2020).

On the basis of literature analysis, the theoretical analysis part mainly adopts a combination of conceptual analysis and theoretical deduction. Relevant studies show that theoretical deduction can help educational research shift from fragmented phenomenon description to mechanistic explanation, thus enhancing the theoretical explanatory power and conceptual innovation of research (Allen-Platt et al., 2021). Therefore, this study does not regard “students with learning difficulties” as a group with fixed ability deficits. It redefines the concept of “students with learning difficulties” by deducing the paradigm of fairness in education, and puts forward the explanatory framework of “developmental differences”. Meanwhile, based on productive failure theory, cognitive conflict theory, formative assessment theory, and fairness in education theory, this study conducts theoretical deduction on the internal logic among “cognitive conflict–knowledge reconstruction–developmental fairness”.

The focus of this part is not to verify variable relationships. It explains why traditional “error-prevention fairness” leads to cognitive impoverishment of students with learning difficulties, and why productive-failure instruction can realize procedural fairness at the classroom level through cognitive activation, delayed instruction, and social negotiation mechanisms.

In the strategic analysis part, this paper mainly adopts the inductive approach. Inductive framework construction in educational research helps transform scattered teaching strategies into systematic, practical models, thus improving the classroom explanatory power and application value of theory (Mu et al., 2022). Therefore, on the basis of the aforementioned theoretical analysis, combined with domestic and foreign research results on classroom cognitive activation, spatial organization, artificial intelligence-supported learning, and formative assessment, this study summarizes and integrates the key implementation elements in productive-failure classrooms. It further proposes the “four-dimensional classroom transformation model”, which systematically reconstructs productive-failure classrooms from four dimensions: space, time, technology, and evaluation. Among them, the spatial dimension emphasizes the reconstruction of classroom power relations. The temporal dimension emphasizes stepped failure teaching design. The technical dimension emphasizes precise support for cognitive exploration by generative artificial intelligence. The evaluation dimension emphasizes the establishment of a value-added fairness assessment system. This model not only responds to the long-standing problem in PF research of “effective in theory but difficult to implement in classrooms”, but also expands the practical path of fairness in education from resource compensation to cognitive empowerment.

In the practical analysis part, this paper adopts a combination of “theoretical deduction–practical derivation” to systematically analyze the implementation path of productive-failure instruction in educational practice. Some studies point out that the important value of educational theory research lies not only in explaining educational phenomena, but also in providing operable institutional paths and practical solutions for education reform (Hanushek, 2003). Therefore, based on the aforementioned theoretical framework and classroom transformation model, this study further carries out practical derivation from the aspects of teacher education curriculum reform, school quality evaluation mechanism reconstruction, and education policy support. It elevates productive-failure instruction from a partial classroom strategy to a practical path of fairness in education with institutional value.

In summary, through the research path combining literature analysis, theoretical deduction, and induction, this study gradually forms a three-in-one research framework of “theoretical mechanism–classroom model–practical path”. It is expected to provide explanatory and practical theoretical support for research on the transition of students with learning difficulties in the context of fairness in education.

4. Theoretical Analysis of the Transition of Students with Learning Difficulties Driven by Productive-Failure Instruction

4.1. Strategic Guidance: The Policy Coordination of the Transition of Students with Learning Difficulties in the Context of Building a Leading Country in Education

4.1.1. Decoding the View of Fairness in Education

The contemporary framework of fairness in education takes “education for all” as its core orientation. People no longer narrowly equate fairness with the allocation of enrollment opportunities and educational resources. Instead, fairness focuses simultaneously on the process quality and outcome quality of teaching. For students with learning difficulties, the view of fairness in education should evolve from coverage-oriented opportunity fairness to “quality justice”. This quality justice centers on the three-in-one process quality of classroom interaction, cognitive activation, and formative assessment. Evidence-based designs such as productive failure serve as actionable approaches to translate fairness into effective learning gains (Sinha & Kapur, 2021).

First, high-quality classroom interaction (including teacher-student emotional support, expectation communication, and constructive interaction) is the basic channel to promote the participation of disadvantaged students, maintain learning motivation, and ensure their access to teaching support. A large body of meta-analytic evidence shows that emotional support and positive teacher-student relationships significantly improve student engagement and academic achievement. The impact is particularly important for low-achieving students (Roorda et al., 2011).

Second, cognitive activation refers to deliberately stimulating students’ cognitive processing through task design (for example, triggering conceptual conflicts through complex problems, promoting generative thinking, and multi-solution exploration). It is the key mechanism to realize deep learning and knowledge transfer (Lazonder & Harmsen, 2016). Productive failure is built on the cognitive activation logic of “first letting students experience cognitive conflict in problem solving in generative tasks, then giving integrated guidance”. Multiple studies show that it has a robust positive effect on students’ conceptual mastery and transfer (Kapur, 2008; Sinha & Kapur, 2021).

Finally, formative assessment and high-quality feedback serve as a mechanistic tool that turns classroom interaction and cognitive activation into sustained learning gains. A large number of reviews and meta-analyses show that explanatory, actionable feedback and continuous formative assessment can significantly improve learning outcomes, and help narrow gaps between students (Black & Wiliam, 1998; Hattie & Timperley, 2007).

4.1.2. Semantic Reconstruction of the Connotation of Students with Learning Difficulties

The terminological shift from “underachievers” to “developmental differences” reflects three paradigm shifts in the concept of fairness in education. The “compensatory fairness” in the 1980s emphasized resource inclination (Borman et al., 2001). Practices in this stage focused on “who receives extra resources” rather than differentiated support in classroom processes. The “procedural fairness” in the early 2000s focused on teaching intervention. It emphasized the importance of classroom interaction quality, teacher-student relationships, formative assessment, and teaching appropriateness in promoting the learning of disadvantaged students (Florian & Black-Hawkins, 2011). This paradigm shifted “compensation” from transfusion-style aid at the institutional level to scaffolded teaching and dynamic monitoring at the classroom level. It significantly reduced the stigmatization effect of “moralized interpretation” on students with learning difficulties. The “developmental fairness” in the new era focuses on activating individual potential. It emphasizes activating the learning potential of every student, rather than only focusing on the equal distribution of resources or processes (Gracia-Calandín & Tamarit-López, 2021). This semantic reconstruction breaks the traditional prejudice that regards academic difficulties as moral defects. It reinterprets the formation mechanism of individual differences among students through the theory of multiple intelligences.

In the process of education modernization, one of the root causes for the systematic neglect of students with learning difficulties is the quantitative and utilitarian tilt of the education evaluation system. It squeezes the essence of education by “measurable indicators” (Polesel et al., 2014). When standardized assessment and test-based accountability systems become the main driving force of teaching, teachers and schools tend to prioritize time and resources into assessed subjects and quantifiable indicators. As a result, learning activities involving practical innovation, art, or cross-domain skills are marginalized. Students with such abilities are judged as “substandard” or “problem students” because they fall outside the evaluation dimensions. This structural dilemma requires education policies to build a differentiated value recognition mechanism, rather than staying at the mechanical allocation of material resources (Sirin, 2005).

4.2. Paradigm Breakthrough: Theoretical Innovation of Productive-Failure Instruction

4.2.1. Localized Integration of Western “Productive Failure” Theory

Kapur (2015) proposed the productive failure theory. This theory shares deep cultural connections with traditional Chinese education. The Book of Learning puts forward a classic teaching principle: guide students instead of pulling them forward, encourage them instead of holding them back. This principle aligns with the “disequilibrium and reconstruction” mechanism in cognitive conflict theory, even though they come from different times and cultural backgrounds. This cultural connection shows itself in three aspects. First, the Confucian idea of “gaining wisdom through hardships and acting with diligence” lays a cultural foundation for the value of frustration education. It gives positive meaning to failure experiences. Second, the dialectical thinking of Taoism, “reversal is the movement of the Dao,” matches the cognitive value-added logic behind failure-focused instructional design. Third, the Chan Buddhism concept of “sudden enlightenment” emphasizes individual differences in the threshold of cognitive breakthroughs. It provides traditional wisdom support for differentiated failure-based teaching design.

To realize the localized transformation of productive-failure instruction theory, educators must address three cultural challenges. The first is how to turn the sense of shame around failure in collectivist traditions into learning motivation (Mesquita & Karasawa, 2004). The second is how to reshape teacher-student interactions in generative learning within the framework of teacher authority (Tweed & Lehman, 2002). The third is how to reconstruct procedural evaluation mechanisms in an exam-oriented culture (Dee & Jacob, 2011).

First, the socialization of failure shame in collectivist cultures makes students avoid risks in public trial-and-error situations. They become less willing to participate or show their mistakes. This weakens the cognitive exploration motivation required for generative tasks (Fang et al., 2023). For productive failure learning to take root effectively in Chinese classrooms, education authorities need to reduce students’ sense of failure shame through emotionally and culturally sensitive interventions.

Second, teacher authority and hierarchical teacher–student relations, rooted in Confucian traditions, create cultural and practical barriers to students’ generative exploration (Sum & Kwon, 2020). In teacher-centered classrooms, teachers find it hard to shift classroom control to student-led exploration spaces. Teachers also hold

professional expectations of maintaining classroom order and avoiding “loss of face”. These expectations limit their willingness to adopt teaching strategies that allow public failure. Therefore, the localization of productive failure is not just about introducing new teaching tasks. It also requires the reshaping of teacher roles and the advancement of teacher professional development.

Finally, the widespread high-stakes testing culture fixes teaching goals and resource allocation on measurable indicators. It squeezes the space for procedural learning and formative assessment. Schools and teachers thus prefer to “teach testable question types” instead of supporting fault-tolerant cognitive exploration (Dee & Jacob, 2011). Research by Polesel et al. (2014) shows that high-pressure testing leads to curriculum narrowing. It tilts teaching toward testable content and changes the priorities of school resources and teaching practices. This creates institutional barriers to fairness in education and teaching innovation. Therefore, we need to reform evaluation and examination mechanisms. This will leave institutional space for formative assessment and bearable failure.

4.2.2. Value Shift of Productive-Failure Instruction

Traditional “error-prevention fairness” provides superficial equality by reducing cognitive difficulty. But it easily leads to “cognitive impoverishment” of students with learning difficulties. Productive-failure instruction can promote three shifts in fairness in education: The focus of teaching shifts from remediating learning outcomes to optimizing cognitive processes (Sinha & Kapur, 2021). Support methods shift from external resource supply to stimulating students’ self-regulation and intrinsic motivation (Zimmerman, 2002). Training objectives shift from guaranteeing the right to acquire knowledge to building transferable learning abilities (Lazonder & Harmsen, 2016). The theoretical basis for this shift lies in the following: moderate frustration experience and delayed instruction can activate students’ reflective learning, promote their knowledge reorganization and transfer, and thus turn “allowing failure” into long-term learning gains (Hattie & Timperley, 2007).

The fairness value of productive-failure instruction design is reflected in its dynamic compensation mechanism: First, by creating cognitive conflict within the zone of proximal development, productive-failure instruction helps students with learning difficulties expose their cognitive misunderstandings in problem solving (Nachtigall et al., 2020). Second, by building a dialogue mechanism of a learning community, productive-failure instruction breaks the discourse hegemony of top students and reconstructs the power structure of knowledge production in the classroom (Mittra & Serriere, 2012). Third, productive-failure instruction guides individuals to achieve cognitive self-regulation through tools such as reflection journals and realizes substantive fairness in differentiated development (Zhang et al., 2022). Productive-failure instruction elevates fairness in education from resource redistribution to the dimension of “thinking rights”. It provides an operable mechanism for realizing procedural justice in classrooms.

5. Strategic Analysis of the Transition of Students with Learning Difficulties Driven by Productive-Failure Instruction

Based on the previous literature review and theoretical analysis, we can find that the formation of learning difficulties is not caused by a single ability defect. It results from the interaction of multiple factors such as cognitive structure, self-regulation, classroom participation, resource acquisition, and evaluation mechanism. Meanwhile, research on productive failure also shows that its learning gains do not come from “failure” itself. They depend on whether learners can complete knowledge reconstruction in moderate cognitive conflict, continuous exploration, and subsequent integrated support (Kapur, 2014). However, the implementation of PF in real classrooms is still restricted by traditional teaching structures: Fixed classroom space easily solidifies students’ participation identities (Vaughn, 2021). Unified teaching rhythm compresses exploration time (Kozanitis & Nenciovici, 2023). Standardized evaluation inhibits trial-and-error behavior (Simon et al., 2024). Improper intervention of technical tools may prematurely terminate the cognitive exploration process (Bastani et al., 2024).

Therefore, regarding PF only as a teaching method can hardly truly respond to the deep-seated needs of the transition of students with learning difficulties and fairness in education. Against this background, based on the aforementioned research on fairness in education, cognitive activation, formative assessment, and artificial intelligence-supported learning, this study further constructs the “four-dimensional classroom transformation model”. It systematically reconstructs productive-failure classrooms from four dimensions: space, time, technology, and evaluation, as shown in Figure 1. The four dimensions are not isolated from each other. They together form the classroom implementation framework for the transition of students with learning difficulties driven by productive-failure instruction. They promote fairness in education to further shift from resource compensation to cognitive empowerment and procedural justice.

5.1. Spatial Dimension: Reconstructing Classroom Space to Eliminate the “Back Row Phenomenon”

Classroom space is a visual projection of the educational power structure. Seating arrangement, sight distribution, and teachers’ spatial attention jointly determine which students can gain the right to speak and cognitive support (Smidekova et al., 2020). A large number of reviews and empirical studies show that teachers tend to place students with distracted attention or low behavioral compliance in the back row of the classroom. This causes this area to gradually evolve into a marginal zone of teaching participation (Dong et al., 2021). This spatial differentiation not only intensifies the identity solidification effect of students but also forms structural obstacles to cognitive development by limiting interaction opportunities (Wannarka & Ruhl, 2008). Building a dynamic learning community requires breaking through traditional spatial barriers and realizing a deep reconstruction of teaching relations. Specific measures can be implemented from the following three aspects:

Spatial deconstruction and reorganization: Adopt U-shaped, round-table, or project-island layouts to eliminate the hierarchical difference between front and back rows. For example, the inter-school learning community practice of Pudong Institute of Educational Development shows that through random grouping and role rotation, students transform from “passive recipients” to “knowledge co-constructors” (Nehyba et al., 2023).

Balanced allocation of cognitive resources: Build virtual learning spaces with platforms such as Moodle, and turn marginal students in physical space into “opinion contribution nodes” in online discussions (Ho et al., 2023). This forms a virtual-real integrated cognitive network. Students can increase their speaking opportunities and peer feedback without being restricted by seating positions.

Reconstruction of power relations: Teachers should transform from “space dominators” into collaborative facilitators. They can break the link between seating position and discourse power by patrolling the classroom and using differentiated questioning strategies. For instance, they can prioritize students in the back row to share their problem-solving ideas.

5.2. Temporal Dimension: Application of the Stepwise Productive-Failure Instruction Design Model

Based on the productive failure theory of Kapur (2008), building a stepped teaching design model requires precise control of the timing of cognitive conflict delivery and support intensity. Teaching design can be carried out in three stages:

First, in the pre-failure stage, teachers should create cognitive conflict situations and deliver cognitive conflicts precisely. By deliberately creating cognitive conflicts and complex problems within learners’ zone of proximal development, productive-failure instruction helps students with learning difficulties expose and become aware of their cognitive misunderstandings. This provides diagnostic information for subsequent guiding instruction and promotes deep knowledge reorganization (Henningsen & Stein, 1997). For example, in mathematics concept teaching, teachers can create students’ cognitive tension through preconception exposure strategies. In the teaching of “fraction size comparison”, teachers can first present an intuitive diagram of “ $1/3 > 1/4$ ”, and guide students to debate the contradiction between empirical cognition and mathematical principles of “the larger the denominator, the smaller the fraction when numerators are the same”. This design forces students with learning difficulties to face knowledge gaps directly and lays the foundation for subsequent conceptual reconstruction.

Second, in the controlled failure stage, teachers should provide externalized thinking and progressive scaffolding while maintaining the complexity of the problem itself (van de Pol et al., 2010). For example, set up stepped task groups in mathematics problem solving. The first-level task retains the complexity of the original problem, but provides 3D modeling tools to assist spatial imagination. The second-level task requires students to analyze wrong solutions against standard answers through data comparison scaffolding.

Finally, in the knowledge reconstruction stage, social negotiation and formative feedback should be completed to turn individual failure into shared cognitive resources (Hattie & Timperley, 2007). For example, hold discussion meetings on failed physics experiment schemes. Students can reconstruct their own cognitive structure by learning from others’ experiences. This process turns individual failure experience into group cognitive resources, and forms a virtuous cycle of “failure–attribution–iteration”.

5.3. Technical Dimension: GAI Empowering Precise Support for Productive Failure

The development of generative artificial intelligence (GAI) has gradually turned productive-failure instruction from a partial teaching strategy that relies on teacher experience into an intelligent learning support system with dynamic adjustment and continuous tracking functions. Traditional digital tools focus on “improving accuracy”. PF-oriented GAI focuses more on preserving learners’ exploration process during the failure stage. It creates personalized failure situations, provides cognitive scaffolding and adaptive assessment, and turns failure into a real resource for cognitive reconstruction.

For the personalized creation of failure situations, GAI can identify learners' preconception misunderstandings based on their historical homework, error records, and classroom behavior data. It then dynamically generates cognitive conflict tasks. For example, in physics learning, the AI system can generate differentiated experimental situations for students who confuse sliding friction and rolling friction (Kieser et al., 2024). In mathematics learning, it can automatically generate counterexample tasks for cognitive biases in fraction size rules (Ross & Andreas, 2024). This dynamic task generation based on learning weaknesses allows students to experience "bearable failure" in problem situations highly related to their own cognitive structure. It increases the possibility of subsequent knowledge reconstruction.

For the cognitive scaffolding link in the failure stage, GAI can maintain students' exploration depth through Socratic questioning, instead of providing answers directly. When students get stuck in complex problems, the AI system does not give standard solutions immediately. It guides them to re-examine their existing ideas through continuous enlightening questions. Lee et al. (2024) point out that a generative AI tutor based on Socratic questioning can guide learners to identify conceptual conflicts and revise reasoning paths independently through continuous reflective questions. This maintains cognitive engagement in complex problem solving. Compared with direct error correction systems, AI support based on heuristic questioning better promotes the development of learning transfer and complex problem-solving abilities (Shabrina et al., 2024; Zhang et al., 2025). This interaction mode of "delayed answers–strengthened reflection" helps preserve the cognitive conflict process on which PF depends. It also helps learners develop metacognitive monitoring ability.

For the adaptive assessment of failure experience, GAI can continuously track the learning process through multimodal learning analytics. It breaks the limitation of traditional evaluation that only focuses on right or wrong answers. For example, in language classrooms, the system can comprehensively analyze students' text expression, voice discussions, and revision trajectories. It identifies weak links in their argumentation structure and language organization (Schneider, 2024). In programming classrooms, the system can judge students' problem-solving strategies and persistence by recording code modification processes, debugging frequencies, and error correction paths (Xu et al., 2023). Some studies have combined facial expression recognition, eye movement trajectories, and behavior log data to conduct a comprehensive analysis of learners' emotional fluctuations, cognitive load, and learning engagement. This helps more accurately distinguish two different learning states: "high engagement but temporary failure" and "low participation and low understanding" (Zawacki-Richter et al., 2019). This feedback mechanism shifts from "error identification" to "growth support". It turns failure experience into a real cognitive resource for the development of learning ability. On this basis, the AI system can generate development-oriented feedback reports instead of simple scores. For example, in writing instruction, when the system finds a lack of logical connection between students' arguments, it not only reminds them of "chaotic structure". It further generates a visual framework of "claim–evidence–connection", and recommends corresponding revision paths and sample texts (Ranalli et al., 2017).

5.4. Evaluation Dimension: Constructing a Value-Added Fairness Assessment System

Value-added assessment takes students' original learning foundation as a reference. It focuses on measuring their growth rate in a specific learning cycle, rather than making horizontal comparisons only based on absolute scores under unified standards. For example, one student improves from 60 to 75 points, and another improves from 90 to 95 points. In traditional evaluation, the latter may still be regarded as "better". But value-added assessment pays more attention to the significant improvement of the former in learning ability and cognitive structure (Morris et al., 2021). For productive-failure instruction, this evaluation method can more truly reflect students' development process in the "attempt–failure–revision" cycle. It prevents students with learning difficulties from being fixed in evaluation due to their long-term low starting point. Therefore, building a "three-in-one" value-added fairness assessment system within the framework of productive-failure instruction requires a fundamental breakthrough from the traditional result-oriented evaluation model. It should turn into an evaluation index system that covers three dimensions of opportunity, process, and outcome, and follows a value-added orientation.

Designing opportunity fairness indicators: Traditional practices often only focus on the average distribution of physical resources, such as equipment, textbooks, and teacher allocation. Value-added assessment breaks through this narrow focus on static resource allocation. It turns to evaluate the actual ability of students with learning difficulties to obtain and use "cognitive opportunities" (OECD, 2012). For example, traditional evaluation may only record whether students participate in classroom discussions. The value-added opportunity fairness indicator evaluates the frequency and quality of students with learning difficulties participating in high-order thinking activities through multi-dimensional dynamic monitoring. Such activities include putting forward hypotheses, participating in reasoning, and conducting self-explanation.

Designing process fairness indicators: Value-added assessment breaks through the traditional limitation of only focusing on unified teaching processes. It turns to focus on the real-time cognitive and emotional state changes of students with learning difficulties during the learning process (Roorda et al., 2011). Previous evaluations tended to focus on “whether formative assessment is carried out”. Now we should consider “to what extent formative assessment promotes the improvement of metacognition and strategy use of students with learning difficulties”. That is, we turn process quality into value-added indicators of students’ learning ability.

Designing outcome fairness indicators: Value-added assessment completely abandons the traditional practice of only taking absolute scores and unified standards as the measurement basis. It adopts the value-added model and focuses on the progress range and growth trajectory of students with learning difficulties relative to their own starting point. Traditional evaluation usually compares students’ achievements through standardized test scores. The value-added outcome fairness indicator tracks the growth of students with learning difficulties in multiple dimensions, such as conceptual understanding, strategy application, and academic resilience. For example, the research of Pounder (2018) emphasizes that outcome evaluation of productive-failure instruction should calculate the ability gain value of students with learning difficulties after experiencing a failure learning cycle. This can objectively reflect their real progress. This evaluation not only reduces prejudice based on horizontal comparison, but also identifies the subtle progress of students with learning difficulties that is not easy to detect. It provides fairer developmental feedback.



Figure 1. Four-dimensional Classroom Transformation Model.

6. Practical Analysis of the Transition of Students with Learning Difficulties Driven by Productive-Failure Instruction

6.1. Value Transmission from the Micro to Macro Level

At the micro level, meaningful failure situations can reconstruct the cognitive schemas of students with learning difficulties and cultivate their resilience and self-efficacy. Social cognitive theory points out that individuals form beliefs about their own abilities through mastery experience and reflective interpretation of failure. They then show higher persistence and coping strategies when facing challenges (Bandura, 1977). Evidence from developmental psychology further shows that resilience is not an innate talent. People can shape it through systematic learning experiences and supportive environments (Masten, 2001). In educational practice, classroom design based on productive failure deliberately exposes students to controllable cognitive conflicts and then provides integrated guidance. Multiple studies have confirmed that this design can improve conceptual understanding and transfer ability. It also enhances learners’ ability to cope with complex tasks (Kapur, 2008; Sinha & Kapur, 2021). Therefore, the core of the transformation path from “frustration tolerance” to “innovative

breakthrough” lies in designing failure as a learnable and reflective experience. This can cultivate cognitive resilience for unknown problems.

At the macro level, international assessments and policy analyses have begun to include “academic resilience/proportion of resilient students” into the indicator system for measuring fairness in education and system compensation effectiveness (Jin et al., 2022). PISA-related studies take the proportion of students who “achieve high achievements despite disadvantaged socioeconomic backgrounds” as an operational indicator of academic resilience. They believe this ratio directly reflects the institutional capacity of a country or region in reshaping the development trajectory of disadvantaged students (Agasisti et al., 2018). Accordingly, China can include classroom-level transition rate of students with learning difficulties, classroom participation balance, and higher-order thinking development index into a micro-meso-macro coupled monitoring framework. It can realize a policy closed loop through the three-level linkage of “school effectiveness–regional balance–national strategy”. Elevating individual gains from classroom interventions (such as PF) to systematic evaluation through value-added and long-term tracking indicators can not only confirm the immediate effect of classroom reform, but also provide operable fairness indicators and evaluation paths for realizing fairness in education (McCaffrey et al., 2004).

6.2. Adding an “Implementation Capacity for Fairness in Education” Module to Teacher Education Curricula

The current teacher training system has an imbalance in ability structure. It overemphasizes teaching skill training, but ignores the practical wisdom of fairness in education. We suggest adding a compulsory module of “implementation capacity for fairness in education” in pre-service teacher training and in-service teacher training. The module focuses on three core abilities: First, differential diagnosis ability. We train teachers to accurately identify the cognitive styles and learning obstacles of students with learning difficulties (van de Pol et al., 2010). Second, culturally responsive ability. We guide teachers to turn the life experiences of students with learning difficulties into teaching resources. For example, they can reconstruct mathematical problem situations through folk culture cases (Villegas & Lucas, 2002). Third, failure of design ability. We train teachers to scientifically plan the threshold of cognitive conflict and design moderate challenges on the premise of protecting the learning motivation of students with learning difficulties (Desimone, 2009). This kind of ability training needs to break through the traditional lecture mode. It adopts a practical path of “clinical diagnosis + situation simulation”. For example, we can develop a case library of urban-rural differences and require teachers to design layered intervention schemes for different types of learning difficulties. This can reshape teachers’ action power for fairness in education.

6.3. Establishing a Linkage Mechanism Between Transition Effectiveness of Students with Learning Difficulties and School Quality Evaluation

The invisible exclusion of students with learning difficulties in the current education evaluation system makes schools more inclined to “cultivate top students” rather than “improve bottom students”. We need to reconstruct school quality evaluation standards and build a “three-dimensional dynamic evaluation model.” The basic dimension examines the academic qualification rate of students with learning difficulties. It focuses on guaranteeing the bottom-line fairness of education (Au, 2007). The development dimension tracks the progress of students with learning difficulties. It uses value-added evaluation to calculate their relative progress value (Koedel & Rockoff, 2015). The innovation dimension introduces resilience assessment. It observes the persistence and strategic flexibility of students with learning difficulties in coping with challenges through situational tasks. Evaluation results should be deeply linked to the allocation of educational resources. For example, schools with a continuously improved transition rate of students with learning difficulties can get targeted support in teacher staffing and teaching research funds. Meanwhile, “pseudo high-quality schools” that simply eliminate students with learning difficulties will receive yellow card warnings (Dee & Jacob, 2011). At the same time, we should also pay attention to the negative effects of this linkage. For example, schools may over-concentrate teaching resources on examination subjects. This will lead to curriculum narrowing and the tendency of exam-oriented teaching (Dee et al., 2013). Some schools may also maintain performance indicators by reducing the participation of low-achieving students or implicit streaming. This will further aggravate educational inequality (Jennings & Lauen, 2016). In addition, continuous performance pressure may also lead schools to form a short-term performance orientation. It will reduce teachers’ teaching autonomy and space for educational innovation (Klein, 2020). Therefore, although the combination of evaluation results and resource inclination can strengthen schools’ motivation for improvement to a certain extent, it may also induce educational alienation and new fairness risks without reasonable supervision and multi-dimensional evaluation mechanisms.

6.4. Optimizing the Design of Productive-Failure Instruction

The promotion of productive-failure instruction needs to solve the double dilemma of “daring not to use failure” and “abusing failure” in practice. We need to optimize the existing standardized teaching design. The optimization can be carried out in three directions: First, in terms of task design, we should follow the layered design logic of “high challenge–accessible”. That is, we maintain the cognitive complexity of tasks and reduce the entry threshold for students with learning difficulties through progressive scaffolding. The research of Roll et al. (2011) points out that progressive scaffolding and layered support for learning tasks can effectively relieve cognitive load in complex tasks. It allows students with learning difficulties to maintain continuous participation in high-challenge tasks. Second, in terms of intervention tools, we provide a standard paradigm of cognitive scaffolding, such as the “four-step guidance method” (exposing misconceptions–comparing conflicts–reconstructing connections–transfer verification). This paradigm helps teachers grasp the intervention timing more accurately. It avoids the interruption of students’ exploration process caused by premature error correction, and improves the quality of cognitive construction in complex problem learning (van de Pol et al., 2019). Third, in terms of evaluation feedback, we should emphasize the continuous recording of “process evidence chain”, rather than only focusing on the final result (Hattie & Timperley, 2007). Teachers can track students’ cognitive changes in the “attempt–failure–revision” process through learning logs, mind maps, scheme iteration records, or classroom dialogue texts. For example, Novak and Treagust (2022) find in science inquiry classrooms that students’ continuous revision of the causes of experimental phenomena in multiple rounds of explanation iteration significantly promotes their scientific reasoning ability and knowledge transfer level. The evaluation goal in PF classrooms is not to reduce errors. It is to identify how students use errors to promote knowledge reconstruction and strategy optimization.

Author Contributions

J.L.: Writing—original draft, Methodology, Investigation, Funding acquisition, Methodology, Formal analysis, Conceptualization; J.Y. and X.L. (Xiaonan Luo): Conceptualization, Writing—original draft, Writing—review & editing; X.L. (Xueliang Li): Conceptualization, Methodology, Investigation, Writing—original draft, Writing—review & editing; W.C.: Methodology, Writing—original draft. All authors have read and agreed to the published version of the manuscript.

Funding

2024 Provincial Teaching Reform Program for Graduate Students in the Second Batch of the 14th Five-Year Plan of Zhejiang Provincial Office of Education: Innovation and Practice of “Six Synergistic” Graduate Teaching Guided by Educator’s Spirit (Project No:JGCG2024406); 2025 Wenzhou University Graduate Student Innovation Fund Project: Research on Improving Junior High School Students’ Practical Knowledge Through Human-Machine Collaborative Learning Decision (3162025003019); 2025 Wenzhou University Graduate Student Innovation Fund Project: Research on Enhancing Junior High School Students’ Deep Knowledge Learning Effect via Data-Driven Learning Decision (3162025004025); 2026 Zhejiang Provincial College Students’ Science and Technology Innovation Activity Plan (Xinmiao Talent Program) College Students’ Scientific and Technological Achievements Promotion Project: Research on Improving Students’ Deep Knowledge Learning Effect Through Data-Driven Learning Decision-Making (Project No.: 2026R444B038).

Institutional Review Board Statement

Ethical approval for this study was granted by the Wenzhou University (WZU-2025-068). All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The research has gained approval from institutional ethics committee.

Informed Consent Statement

Informed consent was obtained from all subjects involved in the study.

Data Availability Statement

The data and materials are available upon request to the corresponding author.

Conflicts of Interest

The authors declare no conflict of interest.

Use of AI and AI-Assisted Technologies

No AI tools were utilized for this paper.

References

- Agasisti, T., Avvisati, F., Borgonovi, F., & Longobardi, S. (2018). *Academic resilience: What schools and countries do to help disadvantaged students succeed in PISA* (OECD Education Working Papers, No. 167). OECD Publishing. <https://doi.org/10.1787/e22490ac-en>.
- Allen-Platt, C., Gerstner, C.-C., Boruch, R., & Ruby, A. (2021). Toward a science of failure analysis: A narrative review. *Review of Research in Education*, 45(1), 223–252. <https://doi.org/10.3102/0091732x20985074>.
- Ashman, G., Kalyuga, S., & Sweller, J. (2020). Problem-solving or explicit instruction: Which should go first when element interactivity is high? *Educational Psychology Review*, 32(1), 229–247. <https://doi.org/10.1007/s10648-019-09500-5>.
- Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, 36(5), 258–267. <https://doi.org/10.3102/0013189x07306523>.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191–215. <https://doi.org/10.1037/0033-295x.84.2.191>.
- Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2024). *Generative AI can harm learning* (SSRN Scholarly Paper No. 4895486). SSRN. <https://doi.org/10.2139/ssrn.4895486>.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>.
- Borman, G. D., Stringfield, S. C., & Slavin, R. E. (2001). *Title I: Compensatory education at the crossroads*. Routledge.
- Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *The Internet and Higher Education*, 27, 1–13. <https://doi.org/10.1016/j.iheduc.2015.04.007>.
- Chen, Y., Chui, H., Huang, Y., & King, R. B. (2024). Adaptation and validation of the Chinese version of fear of failure in learning scale. *Journal of Psychoeducational Assessment*, 42(7), 866–881. <https://doi.org/10.1177/07342829241257867>.
- De Bortoli, L., Underwood, C., & Thomson, S. (2023). *PISA 2022. Reporting Australia's results. Volume I: Student performance and equity in education*. Australian Council for Educational Research. <https://doi.org/10.37517/978-1-74286-725-0>.
- Dee, T. S., & Jacob, B. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management*, 30(3), 418–446. <https://doi.org/10.1002/pam.20586>.
- Dee, T. S., Jacob, B., & Schwartz, N. L. (2013). The effects of NCLB on school resources and practices. *Educational Evaluation and Policy Analysis*, 35(2), 252–279. <https://doi.org/10.3102/0162373712467080>.
- Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, 38(3), 181–199. <https://doi.org/10.3102/0013189x08331140>.
- Dong, Z., Liu, H., & Zheng, X. (2021). The influence of teacher-student proximity, teacher feedback, and near-seated peer groups on classroom engagement: An agent-based modeling approach. *PLoS ONE*, 16(1), e0244935. <https://doi.org/10.1371/journal.pone.0244935>.
- Fang, J., Brown, G. T. L., & Hamilton, R. (2023). Changes in Chinese students' academic emotions after examinations: Pride in success, shame in failure, and self-loathing in comparison. *British Journal of Educational Psychology*, 93(1), 245–261. <https://doi.org/10.1111/bjep.12552>.
- Florian, L., & Black-Hawkins, K. (2011). Exploring inclusive pedagogy. *British Educational Research Journal*, 37(5), 813–828. <https://doi.org/10.1080/01411926.2010.501096>.
- Furtak, E. M., Seidel, T., Iverson, H., & Briggs, D. C. (2012). Experimental and quasi-experimental studies of inquiry-based science teaching. *Review of Educational Research*, 82(3), 300–329. <https://doi.org/10.3102/0034654312457206>.
- Gracia-Calandín, J., & Tamarit-López, I. (2021). Education as a common good from the capability approach. *Journal of Philosophy of Education*, 55(4–5), 817–828. <https://doi.org/10.1111/1467-9752.12575>.
- Hadjar, A., & Naumann, I. K. (2025). The digital divide. *sozialpolitik.ch*, 2025(2). <https://doi.org/10.18753/2297-8224-8576>.
- Hanushek, E. A. (2003). The failure of input-based schooling policies. *The Economic Journal*, 113(485), F64–F98. <https://doi.org/10.1111/1468-0297.00099>.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>.

- Henningsen, M., & Stein, M. K. (1997). Mathematical tasks and student cognition: Classroom-based factors that support and inhibit high-level mathematical thinking and reasoning. *Journal for Research in Mathematics Education*, 28(5), 524–549. <https://doi.org/10.2307/749690>.
- Ho, D. G., Sa'adi, M., He, D., & Hoon, C. Y. (2023). Silence over the wire: Student verbal participation and the virtual classroom in the digital era. *Asia Pacific Education Review*, 24(4), 599–615. <https://doi.org/10.1007/s12564-023-09834-4>.
- Jaakkola, E. (2020). Designing conceptual articles: Four approaches. *AMS Review*, 10(1–2), 18–26. <https://doi.org/10.1007/s13162-020-00161-0>.
- Jennings, J. L., & Lauen, D. L. (2016). Accountability, inequality, and achievement: The effects of the No Child Left Behind Act on multiple measures of student learning. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, 2(5), 220–241. <https://doi.org/10.7758/rsf.2016.2.5.11>.
- Jin, S., Fang, G., Cheung, K. C., & Sit, P. S. (2022). Factors associated with academic resilience in disadvantaged students: An analysis based on the PISA 2015 B-S-J-G (China) sample. *Frontiers in Psychology*, 13, 846466. <https://doi.org/10.3389/fpsyg.2022.846466>.
- Kapur, M. (2008). Productive failure. *Cognition and Instruction*, 26(3), 379–424. <https://doi.org/10.1080/07370000802212669>.
- Kapur, M. (2014). Productive failure in learning math. *Cognitive Science*, 38(5), 1008–1022. <https://doi.org/10.1111/cogs.12107>.
- Kapur, M. (2015). Learning from productive failure. *Learning: Research and Practice*, 1(1), 51–65. <https://doi.org/10.1080/23735082.2015.1002195>.
- Khalil, M., & Er, E. (2023). *Will ChatGPT get you caught? Rethinking of plagiarism detection*. Center for Open Science. <https://doi.org/10.35542/osf.io/fnh48>.
- Kieser, F., Tschisgale, P., Rauh, S., Bai, X., Maus, H., Petersen, S., Stede, M., Neumann, K., & Wulff, P. (2024). David vs. Goliath: Comparing conventional machine learning and a large language model for assessing students' concept use in a physics problem. *Frontiers in Artificial Intelligence*, 7, 1408817. <https://doi.org/10.3389/frai.2024.1408817>.
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1.
- Klein, E. D. (2020). School accountability. In *Oxford research encyclopedia of education*. Oxford University Press. <https://doi.org/10.1093/acrefore/9780190264093.013.677>.
- Koedel, C., & Rockoff, J. E. (2015). Value-added modeling: A review. *Economics of Education Review*, 47, 180–195. <https://doi.org/10.1016/j.econedurev.2015.01.006>.
- Kozanitis, A., & Nenciovici, L. (2023). Effect of active learning versus traditional lecturing on the learning achievement of college students in humanities and social sciences: A meta-analysis. *Higher Education*, 86(6), 1377–1394. <https://doi.org/10.1007/s10734-022-00977-8>.
- Lazonder, A. W., & Harmsen, R. (2016). Meta-analysis of inquiry-based learning. *Review of Educational Research*, 86(3), 681–718. <https://doi.org/10.3102/0034654315627366>.
- Lee, H.-Y., Chen, P.-H., Wang, W.-S., Huang, Y.-M., & Wu, T.-T. (2024). Empowering ChatGPT with guidance mechanism in blended learning: Effect of self-regulated learning, higher-order thinking skills, and knowledge construction. *International Journal of Educational Technology in Higher Education*, 21(1), 27. <https://doi.org/10.1186/s41239-024-00447-4>.
- Lee, J.-A., & Kim, C.-J. (2019). Teaching and learning science in authoritative classrooms: Teachers' power and students' approval in Korean elementary classrooms. *Research in Science Education*, 49(5), 1367–1393. <https://doi.org/10.1007/s11165-017-9659-6>.
- Li, H., Liu, J., Zhang, D., & Liu, H. (2021). Examining the relationships between cognitive activation, self-efficacy, socioeconomic status, and achievement in mathematics: A multi-level analysis. *British Journal of Educational Psychology*, 91(1), 101–126. <https://doi.org/10.1111/bjep.12351>.
- Loibl, K., Roll, I., & Rummel, N. (2017). Towards a theory of when and how problem solving followed by instruction supports learning. *Educational Psychology Review*, 29(4), 693–715. <https://doi.org/10.1007/s10648-016-9379-x>.
- Loibl, K., Tillema, M., Rummel, N., & van Gog, T. (2020). The effect of contrasting cases during problem solving prior to and after instruction. *Instructional Science*, 48(2), 115–136. <https://doi.org/10.1007/s11251-020-09504-7>.
- Mao, M., Zang, L., & Zhang, H. (2020). The effects of parental absence on children development: Evidence from left-behind children in China. *International Journal of Environmental Research and Public Health*, 17(18), 6770. <https://doi.org/10.3390/ijerph17186770>.
- Masten, A. S. (2001). Ordinary magic: Resilience processes in development. *American Psychologist*, 56(3), 227–238. <https://doi.org/10.1037/0003-066x.56.3.227>.

- McCaffrey, D. F., Lockwood, J. R., Koretz, D., Louis, T. A., & Hamilton, L. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics*, 29(1), 67–101. <https://doi.org/10.3102/10769986029001067>.
- Mesquita, B., & Karasawa, M. (2004). Self-conscious emotions as dynamic cultural processes. *Psychological Inquiry*, 15(2), 161–166. <https://www.jstor.org/stable/20447222>.
- Metcalf, J. (2017). Learning from errors. *Annual Review of Psychology*, 68(1), 465–489. <https://doi.org/10.1146/annurev-psych-010416-044022>.
- Meyer, A., Kleinknecht, M., & Richter, D. (2023). What makes online professional development effective? The effect of quality characteristics on teachers' satisfaction and changes in their professional practices. *Computers & Education*, 200, 104805. <https://doi.org/10.1016/j.compedu.2023.104805>.
- Mitra, D. L., & Serriere, S. C. (2012). Student voice in elementary school reform. *American Educational Research Journal*, 49(4), 743–774. <https://doi.org/10.3102/0002831212443079>.
- Morisano, D., Hirsh, J. B., Peterson, J. B., Pihl, R. O., & Shore, B. M. (2010). Setting, elaborating, and reflecting on personal goals improves academic performance. *Journal of Applied Psychology*, 95(2), 255–264. <https://doi.org/10.1037/a0018478>.
- Morris, R., Perry, T., & Wardle, L. (2021). Formative assessment and feedback for learning in higher education: A systematic review. *Review of Education*, 9(3), e3292. <https://doi.org/10.1002/rev3.3292>.
- Mu, J., Bayrak, A., & Ufer, S. (2022). Conceptualizing and measuring instructional quality in mathematics education: A systematic literature review. *Frontiers in Education*, 7, 994739. <https://doi.org/10.3389/feduc.2022.994739>.
- Nachtigall, V., Serova, K., & Rummel, N. (2020). When failure fails to be productive: Probing the effectiveness of productive failure for learning beyond STEM domains. *Instructional Science*, 48(6), 651–697. <https://doi.org/10.1007/s11251-020-09525-2>.
- Nehyba, J., Juhaňák, L., & Cigán, J. (2023). Effects of seating arrangement on students' interaction in group reflective practice. *The Journal of Experimental Education*, 91(2), 249–277. <https://doi.org/10.1080/00220973.2021.1954865>.
- Nieuwenhuis, J., & Hooimeijer, P. (2016). The association between neighbourhoods and educational achievement, a systematic review and meta-analysis. *Journal of Housing and the Built Environment*, 31(2), 321–347. <https://doi.org/10.1007/s10901-015-9460-7>.
- Novak, A. M., & Treagust, D. F. (2022). Supporting the development of scientific understanding when constructing an evolving explanation. *Disciplinary and Interdisciplinary Science Education Research*, 4(1), 2. <https://doi.org/10.1186/s43031-021-00043-w>.
- Organisation for Economic Co-operation and Development. (2012). *Equity and quality in education: Supporting disadvantaged students and schools*. OECD Publishing. <https://doi.org/10.1787/9789264130852-en>.
- Ouyang, F., & Zhang, L. (2024). AI-driven learning analytics applications and tools in computer-supported collaborative learning: A systematic review. *Educational Research Review*, 44, 100616. <https://doi.org/10.1016/j.edurev.2024.100616>.
- Panadero, E., Jonsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review*, 22, 74–98. <https://doi.org/10.1016/j.edurev.2017.08.004>.
- Pelánek, R., Effenberger, T., & Jarušek, P. (2024). Personalized recommendations for learning activities in online environments: A modular rule-based approach. *User Modeling and User-Adapted Interaction*, 34(4), 1399–1430. <https://doi.org/10.1007/s11257-024-09396-z>.
- Polesel, J., Rice, S., & Dulfer, N. (2014). The impact of high-stakes testing on curriculum and pedagogy: A teacher perspective from Australia. *Journal of Education Policy*, 29(5), 640–657. <https://doi.org/10.1080/02680939.2013.865082>.
- Pounder, J. S. (2018). Measuring student success: A value-added approach. In *The future of higher education in the Middle East and Africa* (pp. 205–210). Springer. https://doi.org/10.1007/978-3-319-64656-5_17.
- Ranalli, J., Link, S., & Chukharev-Hudilainen, E. (2017). Automated writing evaluation for formative assessment of second language writing: Investigating the accuracy and usefulness of feedback as part of argument-based validation. *Educational Psychology*, 37(1), 8–25. <https://doi.org/10.1080/01443410.2015.1136407>.
- Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2011). Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system. *Learning and Instruction*, 21(2), 267–280. <https://doi.org/10.1016/j.learninstruc.2010.07.004>.
- Roorda, D. L., Koomen, H. M. Y., Spilt, J. L., & Oort, F. J. (2011). The influence of affective teacher–student relationships on students' school engagement and achievement. *Review of Educational Research*, 81(4), 493–529. <https://doi.org/10.3102/0034654311421793>.

- Ross, A., & Andreas, J. (2024). Toward in-context teaching: Adapting examples to students' misconceptions. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics* (Vol. 1, pp. 13283–13310). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2024.acl-long.718>.
- Sandsør, A. M. J., Zachrisson, H. D., Karoly, L. A., & Dearing, E. (2023). The widening achievement gap between rich and poor in a Nordic country. *Educational Researcher*, 52(4), 195–205. <https://doi.org/10.3102/0013189x221142596>.
- Schneider, B. (2024). Three challenges in implementing multimodal learning analytics in real-world learning environments. *Learning: Research and Practice*, 10(1), 103–112. <https://doi.org/10.1080/23735082.2023.2270611>.
- Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology*, 103(4), 759–775. <https://doi.org/10.1037/a0025140>.
- Shabrina, P., Mostafavi, B., Abdelshiheed, M., Chi, M., & Barnes, T. (2024). Investigating the impact of backward strategy learning in a logic tutor: Aiding subgoal learning towards improved problem solving. *International Journal of Artificial Intelligence in Education*, 34(3), 825–861. <https://doi.org/10.1007/s40593-023-00338-1>.
- Simon, P. D., Fryer, L. K., & Nakao, K. (2024). Assessing class participation in physical and virtual spaces: Current approaches and issues. *Frontiers in Education*, 8, 1306568. <https://doi.org/10.3389/educ.2023.1306568>.
- Simonsmeier, B. A., Flaig, M., Deiglmayr, A., Schalk, L., & Schneider, M. (2022). Domain-specific prior knowledge and learning: A meta-analysis. *Educational Psychologist*, 57(1), 31–54. <https://doi.org/10.1080/00461520.2021.1939700>.
- Sinha, T., & Kapur, M. (2021). When problem solving followed by instruction works: Evidence for productive failure. *Review of Educational Research*, 91(5), 761–798. <https://doi.org/10.3102/00346543211019105>.
- Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, 75(3), 417–453. <https://doi.org/10.3102/00346543075003417>.
- Smidekova, Z., Janik, M., Minarikova, E., & Holmqvist, K. (2020). Teachers' gaze over space and time in a real-world classroom. *Journal of Eye Movement Research*, 13(4), 1. <https://doi.org/10.16910/jemr.13.4.1>.
- Strello, A., Strietholt, R., Steinmann, I., & Siepmann, C. (2021). Early tracking and different types of inequalities in achievement: Difference-in-differences evidence from 20 years of large-scale assessments. *Educational Assessment, Evaluation and Accountability*, 33(1), 139–167. <https://doi.org/10.1007/s11092-020-09346-4>.
- Sum, E. S. W., & Kwon, O. N. (2020). Classroom talk and the legacy of Confucian culture in mathematics classroom. *Teaching and Teacher Education*, 88, 102964. <https://doi.org/10.1016/j.tate.2019.102964>.
- Sun, Y., Zhu, X., Zhou, X., Yao, B., Zhang, K., Wang, D., Chen, J., & He, L. (2024). *Who changed the destiny of rural students, and how? Unpacking ICT-mediated remote education in rural China*. arXiv. <https://arxiv.org/abs/2401.13799>.
- Tai, J., Ajajawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education*, 76(3), 467–481. <https://doi.org/10.1007/s10734-017-0220-3>.
- Tan, C. Y. (2024). Socioeconomic status and student learning: Insights from an umbrella review. *Educational Psychology Review*, 36(4), 99. <https://doi.org/10.1007/s10648-024-09929-3>.
- Terrin, É., & Triventi, M. (2023). The effect of school tracking on student achievement and inequality: A meta-analysis. *Review of Educational Research*, 93(2), 236–274. <https://doi.org/10.3102/00346543221100850>.
- Tweed, R. G., & Lehman, D. R. (2002). Learning considered within a cultural context: Confucian and Socratic approaches. *American Psychologist*, 57(2), 89–99. <https://doi.org/10.1037/0003-066x.57.2.89>.
- van de Pol, J., Mercer, N., & Volman, M. (2019). Scaffolding student understanding in small-group work: Students' uptake of teacher support in subsequent small-group interaction. *Journal of the Learning Sciences*, 28(2), 206–239. <https://doi.org/10.1080/10508406.2018.1522258>.
- 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>.
- Vaughn, M. (2021). *Student agency in the classroom: Honoring student voice in the curriculum*. Teachers College Press.
- Villegas, A. M., & Lucas, T. (2002). Preparing culturally responsive teachers. *Journal of Teacher Education*, 53(1), 20–32. <https://doi.org/10.1177/0022487102053001003>.
- Wannarka, R., & Ruhl, K. (2008). Seating arrangements that promote positive academic and behavioural outcomes: A review of empirical research. *Support for Learning*, 23(2), 89–93. <https://doi.org/10.1111/j.1467-9604.2008.00375.x>.
- Xu, S., & Lu, C. (2022). Hybridization of curriculum integration and high-stakes testing: Chinese teachers' experiences and struggles in handling competing institutional logics. *Teaching and Teacher Education*, 119, 103845. <https://doi.org/10.1016/j.tate.2022.103845>.
- Xu, W., Wu, Y., & Ouyang, F. (2023). Multimodal learning analytics of collaborative patterns during pair programming in higher education. *International Journal of Educational Technology in Higher Education*, 20(1), 3. <https://doi.org/10.1186/s41239-022-00377-z>.

- Yan, J. J., Luo, X. N., Zeng, Q. F., Zhang, H. L., & Wu, J. N. (2025). Investigation and strategy research on the causes of middle school students' learning difficulties in the context of the leading country in education. *Journal of Educational Technology and Innovation*, 7(2), 49. <https://doi.org/10.61414/0shg4n49>.
- Zawacki-Richter, O., Marin, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators? *International Journal of Educational Technology in Higher Education*, 16(1), 39. <https://doi.org/10.1186/s41239-019-0171-0>.
- Zhang, H. L., Yan, J. J., Liu, J., Zou, Z. C., Zeng, Q. F., Jin, Y. L., & Wu, J. N. (2025). An empirical study on the reflective feedback model based on pedagogical agents to improve students' problem-solving ability. *Journal of Educational Technology and Innovation*, 7(1), 66–75. <https://doi.org/10.61414/tykx7317>.
- Zhang, Y., Li, M., Chieh, Y., & Han, S. (2022). On a path to becoming more self-regulated: Reflective journals' impact on Chinese English as a foreign language students' self-regulated writing strategy use. *Frontiers in Psychology*, 13, 1042031. <https://doi.org/10.3389/fpsyg.2022.1042031>.
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, 41(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2.

Article

Between Paperwork and Performance: Students' Experiences of Assessment in Industry-Education Integration in Chinese Higher Vocational Education

Tengteng Zhuang and Yuxin Yang *

Institute of Higher Education, Beijing Normal University, Beijing 100875, China

* Correspondence: yangyuxin@mail.bnu.edu.cn

How To Cite: Zhuang, T., & Yang, Y. (2026). Between Paperwork and Performance: Students' Experiences of Assessment in Industry-Education Integration in Chinese Higher Vocational Education. *Journal of Educational Technology and Innovation*, 8(3), 52–63. <https://doi.org/10.61414/6p41wq42>

Received: 12 May 2026

Revised: 6 September 2026

Accepted: 19 September 2026

Published: 30 September 2026

Abstract: Industry-education integration has become a central strategy for connecting vocational education with occupational practice, yet the assessment of learning across school and workplace settings remains comparatively underexamined. This study investigates how Chinese higher vocational students experience assessment within internships, apprenticeships, enterprise-oriented classes, work-study arrangements, and production-based training. Drawing on semi-structured interviews with 15 students from higher vocational institutions across multiple regions and occupational fields, the study adopts an interpretive qualitative design and reflexive thematic analysis. The findings reveal substantial variation in assessment arrangements. Some students experienced relatively integrated systems combining workplace observation, enterprise ratings, practical tasks, school examinations, reports, attendance, and skills tests. Others encountered assessment dominated by written documentation, weakly consequential supervisor ratings, or standardized forms that students regarded as largely procedural. Across the interviews, students distinguished between school-based indicators of achievement and workplace judgements concerning whether they could actually complete tasks, solve problems, work responsibly, and meet occupational standards. Assessment authenticity therefore depended not simply on whether learning took place in a real enterprise but on whether assessment captured consequential workplace performance and supported learning from that performance. Students also described an asymmetry of evaluative power: schools and enterprises routinely assessed students, whereas students often had limited formal opportunities to evaluate enterprises, supervisors, or placement quality. The study argues that industry-education integration does not necessarily entail assessment integration. Meaningful vocational assessment requires alignment among school requirements, workplace standards, students' learning processes, and reciprocal opportunities for evaluative participation.

Keywords: Chinese higher vocational education; industry-education integration; assessment; work-integrated learning; student voice

1. Introduction

Vocational education is increasingly expected to connect formal education with the knowledge, practices, technologies, and standards of contemporary workplaces. In China, industry-education integration has become an important mechanism through which higher vocational institutions seek to strengthen practical learning and



Copyright: © 2026 by the authors. This is an open access article under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Publisher's Note: Scilight stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

improve students' transitions into employment. These arrangements take diverse forms, including internships, modern apprenticeships, order-oriented classes, enterprise-delivered courses, school-based production facilities, work-study alternation, and joint training projects. Although organizational forms vary, they share a central promise: learning should move closer to occupational practice.

Assessment occupies a particularly important position within this promise. What is assessed communicates what counts as valued knowledge and competence. In conventional educational environments, achievement may be represented through examinations, written assignments, attendance, or course grades. Workplaces operate with different forms of judgement. Workers may be evaluated according to whether they can complete tasks, meet quality specifications, respond to unexpected problems, cooperate with colleagues, follow safety requirements, satisfy clients, or achieve performance indicators. When vocational students move between school and workplace, they therefore encounter not simply two learning environments but potentially two different systems for defining and recognizing competence.

Research on work-integrated learning has increasingly identified assessment as a central but difficult component of educational-workplace integration. Assessment becomes complex when learning occurs beyond the direct control of the educational institution and when college teachers, workplace supervisors, students, and employers may understand desirable performance differently (Ferns & Zegwaard, 2014; Jackson, 2018). In particular, the authenticity of work-integrated assessment depends on alignment among workplace activities, educational assessment, industry participation, and students' emerging professional selves (Ajjawi et al., 2020). Assessment can, in turn, shape whether students interpret workplace participation as meaningful learning and whether they begin to see themselves as developing practitioners (Boud et al., 2023).

Against this background, the diversity of assessment methods used in work-integrated learning becomes important. Workplace-supervisor evaluations, reports, learning logs, self-assessment, oral assessments, and other approaches are commonly combined, frequently with the stated purpose of supporting learning rather than merely certifying performance (Luk & Chan, 2024). Yet the existence of multiple assessment methods does not itself guarantee meaningful assessment. Written reports may become exercises in documentation rather than reflection. Employer ratings may appear authentic because they originate in the workplace but become educationally weak when criteria are unclear or when supervisors routinely award uniformly high scores. Conversely, an apparently conventional practical examination may strongly represent occupational competence when students must independently produce an artefact meeting genuine technical specifications.

For this reason, the problem is not simply whether vocational programs contain practical assessment. More important questions concern what is made visible by assessment, whose judgement is authoritative, how school and enterprise evaluations relate to one another, what consequences assessments carry, and whether students themselves participate in determining the quality of their work and learning environment.

These questions are particularly significant for Chinese higher vocational education because industry-education integration brings together institutions whose purposes overlap but are not identical. At the broader higher-education level, teaching-focused university-industry collaboration in China has increasingly sought to bring industrial resources, technologies, professional expertise, and enterprise participation into educational provision, while persistent tensions remain around enterprises' long-term incentives and the relative primacy of educational and industrial actors (Zhuang et al., 2024). Colleges are responsible for curriculum, academic progression, certification, and student welfare, whereas enterprises are concerned with productivity, work discipline, occupational standards, and the identification of potentially employable workers. Students may simultaneously seek practical competence, occupational experience, qualifications, employment opportunities, and protection against low-quality placements. Thus, school, enterprise, and student enter the same educational arrangement with partially different expectations.

Within such arrangements, assessment is one location in which these different expectations become especially visible. An enterprise may understand a successful student as someone reliable and immediately useful at work; a college may require documentary evidence capable of being recorded within an academic system; a student may judge success according to whether the experience produced genuine competence or a desirable employment opportunity. As a result, the resulting assessment system is not necessarily coherent even when the broader program is institutionally described as integrated.

Even so, existing Chinese vocational-education research has paid considerable attention to school-enterprise cooperation, employability, occupational transitions, and relationships between vocational institutions and employers (Müller, 2024; Wang & Wang, 2023). Comparatively less attention has been directed toward how students themselves experience the assessment arrangements through which their learning and workplace performance are recognized. Student perspectives are especially important because students move directly between

the two evaluative environments. They experience what is formally graded, what supervisors actually notice, what workplace competence demands, and which parts of their learning remain invisible to institutional assessment.

To address this gap, this study examines assessment as experienced by higher vocational students participating in different forms of industry-education integration. Rather than assuming that workplace assessment is necessarily authentic or that dual school-enterprise assessment is necessarily integrated, the analysis investigates how students interpret the relationship among school requirements, enterprise judgements, practical performance, documentary tasks, feedback, and opportunities for student voice. It is guided by three research questions:

- (1) How do Chinese higher vocational students experience the different standards and practices through which their learning is assessed in school and workplace settings?
- (2) How are school-based and enterprise-based forms of assessment combined within different models of industry-education integration?
- (3) What do students perceive existing assessment practices as recognizing, overlooking, or preventing them from evaluating?

By addressing these questions, the study conceptualizes assessment integration as a distinct dimension of industry-education integration. It argues that placing students in workplaces, involving enterprise instructors, or introducing practical activities does not by itself integrate assessment. Assessment becomes meaningfully integrated when educational and occupational standards are connected without reducing workplace learning either to school paperwork or to unexamined employer judgement, and when students are enabled to participate in evaluating the quality of both their own performance and the learning arrangements in which that performance occurs.

2. Theoretical Framework: Authentic Assessment and Evaluative Judgement across School-Work Boundaries

This study understands assessment not only as a mechanism for allocating grades but as a social practice through which competence is defined, made visible, and recognized. This perspective is particularly appropriate for industry-education integration because students participate in overlapping educational and occupational environments in which different actors may apply different standards of quality.

Authentic assessment is useful for examining the school-work boundary because it directs attention to the relationship between assessment tasks and the practices in which competence will eventually be used. Authenticity is not a property of a task in isolation; it depends on the extent to which assessment reproduces meaningful features of professional practice and requires students to exercise judgement in context (Ashford-Rowe et al., 2014). In work-integrated learning, this relationship is especially complex because educational requirements, workplace activity, and students' developing professional identities may not align automatically (Ajjawi et al., 2020). Thus, a workplace location may increase realism without necessarily making the assessment itself educationally authentic (Villarroel et al., 2018).

This distinction is especially important in work-integrated learning. A task completed inside an enterprise is not automatically authentic as an assessment. A workplace-based assignment may be largely administrative, disconnected from the work students actually perform, or scored according to criteria that participants do not understand. Conversely, a structured training task can be highly authentic when it requires students to integrate knowledge and action under standards that closely resemble occupational performance. In this sense, students' perceptions of authenticity depend on alignment among placement activities, assessment activities, industry participation, and their anticipated professional futures (Ajjawi et al., 2020).

A second lens concerns evaluative judgement: students' capacity to make informed decisions about the quality of their own and others' work (Tai et al., 2018). This capacity matters in vocational settings because competence is not exhausted by completing a prescribed task; learners must increasingly recognise standards, interpret feedback, and judge whether performance is adequate for a particular context. Assessment that supports such judgement can therefore contribute to learning beyond the immediate course or placement, an orientation closely related to sustainable assessment (Boud, 2000).

This is particularly relevant in vocational contexts. The ability to identify whether a manufactured component meets tolerance requirements, whether a vehicle fault has been properly diagnosed, whether an operational plan responds to market data, or whether a construction procedure satisfies safety and quality requirements involves evaluative capability as well as technical skill. Workplace mentors may contribute to this development because their feedback is grounded in occupational experience. Employer participation in assessment can provide students with a real-world perspective, although its educational value depends on clear expectations, useful feedback, and coordination between educational and industry actors.

The third element of the framework concerns the school-work boundary. Industry-education integration places assessment across organizational settings with different purposes. Educational institutions require

comparability, documentation, progression decisions, and evidence of curricular achievement. Workplaces operate through situated judgements responsive to productivity, quality, customer needs, deadlines, safety, and organizational expectations. Assessing workplace learning therefore generates unavoidable tensions between standardization and situated performance (Jackson, 2018).

Rather than treating one side as inherently superior, this study examines how their relationship is organized. A school examination can provide useful evidence of conceptual or technical knowledge that a temporary placement cannot expose. Workplace observation can reveal responsiveness, reliability, judgement, communication, and problem solving that are difficult to capture through written examination. Reports can encourage reflection, yet they can also become detached from the complexity of actual performance. Employer ratings can provide occupational legitimacy, yet they can become ceremonial if virtually every student receives the same result.

From this perspective, meaningful assessment integration has at least four dimensions. First, alignment concerns whether assessment corresponds to what students actually learn and do. Second, plurality of judgement concerns how school, enterprise, and student perspectives contribute to decisions about quality. Third, consequence concerns whether assessment influences learning, progression, responsibility, or occupational opportunity rather than existing merely as documentation. Fourth, reciprocity concerns whether students are only objects of assessment or are also enabled to judge their own performance and the quality of the environments in which they learn.

These concepts informed later stages of interpretation without operating as predetermined coding categories. The analysis first attended to how students themselves described assessment practices and then used authenticity, evaluative judgement, and school-work alignment to interpret patterned differences across cases.

3. Methodology: Research Design and Participants

This study employed a qualitative interview design to investigate how Chinese higher vocational students understood assessment within industry-education integration. Qualitative interviews were appropriate because assessment is experienced not merely through formal instruments but through students' interpretations of expectations, feedback, workplace relationships, perceived fairness, meaningfulness, and consequences.

An interpretive approach was adopted. Participants' accounts were treated as situated understandings of how assessment operated in their particular programs rather than as objective audits of institutional assessment policy. This distinction is important. A student may not know the precise mathematical weighting assigned to an enterprise evaluation, yet their understanding of whether the evaluation matters, whether supervisors distinguish performance, and whether feedback affects learning remains analytically important.

The dataset comprised 15 semi-structured interviews with students from higher vocational institutions in different Chinese regions. Participants represented nursing, accounting, numerical-control technology, e-commerce, water engineering, landscape/agricultural studies, energy and automotive technology, finance, electric power, mechatronics, statistics and big-data analysis, tourism, rail transportation, and computing. Most were in the final year of three-year higher vocational programs, while several were second-year students and one was completing a five-year integrated vocational program.

Their forms of industry-education integration also varied. Some undertook conventional or school-mediated workplace placements. Others participated in modern apprenticeship, order-oriented education, work-study alternation, enterprise-oriented project classes, or production-based training located within the college. This variation enabled comparison between assessment arrangements with different levels of school-enterprise coupling.

The interviews were conducted in Mandarin. Students were asked about the form of industry-education integration they experienced, their activities and responsibilities, relationships with school and enterprise teachers, connections between school knowledge and workplace practice, difficulties and support, perceived outcomes, assessment practices, feedback mechanisms, and overall evaluation of their programs.

Assessment questions addressed school assessment, enterprise assessment, practical testing, examination, attendance, reports and logs, performance criteria, student evaluation of enterprises, and feedback mechanisms. Because the interviews examined the wider experience of industry-education integration rather than assessment alone, assessment accounts were interpreted in relation to students' descriptions of actual workplace tasks, mentoring, curricular alignment, and employment expectations.

The analysis adopted thematic analysis to explore how students understood assessment within industry-education integration. This approach provided a flexible yet systematic way to identify and interpret patterns of meaning across the interview data (Braun & Clarke, 2006). It took a reflexive orientation, treating these

understandings as situated and acknowledging the researcher's interpretive role rather than prioritising coding reliability (Braun & Clarke, 2021).

Interviews were transcribed in Chinese. Analysis was undertaken using the Chinese transcripts, and quotations were translated into English for reporting. Translations prioritize substantive meaning and readability rather than literal word-for-word equivalence.

All participants were informed about the research and provided consent. Names, specific workplaces, and identifying institutional details should be anonymized in publication. The present participant codes are used solely to distinguish accounts analytically.

Table 1. Anonymized Characteristics of Interview Participants.

Participant ID	Field of Study	Academic Status	Main Form of Industry-Education Integration
P01	Nursing	Third year	Workplace placement
P02	Accounting	Third year	School-based practical training and workplace experience
P03	Numerical-Control Technology	Second year	School-based enterprise production and apprenticeship
P04	E-commerce	Third year	Workplace placement with enterprise mentoring
P05	Water Engineering	Third year	Workplace placement in construction
P06	Landscape/Agricultural Studies	Third year	Research-institute placement
P07	Energy/Automotive Technology	Third year	Work-study alternation and enterprise placement
P08	Finance	Third year	Modern apprenticeship/workplace placement
P09	Electric Power Technology	Third year	Workplace placement
P10	Mechatronics	Third year	Order-oriented class and workplace placement
P11	Mechatronics	Second year	Work-study alternation/apprenticeship
P12	Statistics and Big-Data Analysis	Second year	Enterprise-oriented project class
P13	Tourism Management	Fifth year	Order-oriented/industry-linked class
P14	Rail Transportation	Second year	Order-oriented training and workplace practice
P15	Computing	Third year	Enterprise training project/workplace placement

4. Findings

Students did not experience a single model of assessment associated with industry-education integration. Instead, their accounts ranged from technically detailed performance assessment to loosely monitored internships and largely procedural documentation. Across this diversity, five themes demonstrated how assessment operated at the school-work boundary.

4.1. Two Standards of Being “Good”: From Academic Achievement to Work Performance

A recurring distinction concerned what counted as successful performance in school and what counted in the workplace. Students did not necessarily reject examinations or theoretical knowledge, but they understood that workplace participation introduced another basis for judgement.

P09, an electric-power student, expressed the difference particularly clearly. Reflecting on what made someone a good student compared with a good employee, she questioned whether school judgement was ultimately based on grades. In the enterprise, she argued, the project itself became the criterion: if the project was completed well, the worker had demonstrated competence.

This distinction appeared in more concrete forms in technical fields. P03, studying numerical-control technology, described a semester assessment in which students independently programmed and produced a component. The resulting object was judged according to characteristics such as precision and surface roughness. The task was created for assessment rather than commercial sale, but the standards applied to it resembled those relevant to actual production.

For P14 in rail transportation, assessment similarly centred on whether students could actually do what they claimed to know. Enterprise mentors and school staff considered performance during practice, the completion of assigned tasks, and a final practical task in which students had to demonstrate that they could perform the work. He contrasted this with merely describing what one knew.

These accounts suggest that occupational assessment changed the object of evaluation. Rather than primarily demonstrating remembered knowledge, students were required to transform knowledge into an acceptable product, action, decision, or completed task.

P05 experienced an especially consequential version of this shift in a construction placement. Although still an intern, he described occupying the position of a trainee construction technician and being evaluated against monthly indicators similar to those applied to formal workers.

“After around two months, our technical summaries and the company’s monthly KPI indicators were basically the same as those for formal construction staff.” (P05)

4.2. When School and Enterprise Both Assess: Integration through Multiple Judgements

Several programs attempted to connect these evaluative environments through multiple assessors and forms of evidence. The most elaborate arrangements combined enterprise observation, school records, practical testing, reports, and attendance.

P04, an e-commerce student, described one such system. The enterprise completed an evaluation covering daily attendance, work attitude, and task completion. Students also submitted internship logs. Enterprise and school information was then combined within the student’s academic record. Practical competence was assessed twice: the enterprise considered workplace operations such as backend procedures and routine e-commerce processes, while the college organized a professional skills test after the student returned to school.

For this student, assessment therefore operated across time and location. The enterprise evaluated performance during participation in actual work; the school translated aspects of this experience back into an academic judgement.

P07 described a comparable arrangement in an automotive-related apprenticeship. School staff considered attendance, enterprise feedback, and students’ weekly or daily records, while workplace masters evaluated practical ability. The student understood the final result as a composite rather than as the judgement of a single actor.

P11, a second-year mechatronics student, similarly distinguished between school and enterprise responsibilities. School assessment included course examinations, training reports, and practical-training scores, while the enterprise conducted everyday skills assessment. Rather than replacing one system with the other, the program retained both.

These cases demonstrate that assessment integration can involve combining heterogeneous evidence rather than searching for one universal measure of competence. Written knowledge, sustained workplace behaviour, task completion, and practical performance reveal different aspects of learning. However, the presence of multiple assessments does not itself establish integration. The educational value of a dual system depends on whether the different judgements inform one another.

4.3. “It Is Still a Form”: When Workplace Assessment Becomes Procedural

Some of the strongest critiques concerned assessment that appeared formally complete but substantively weak.

P09 described an enterprise evaluation form covering performance, work attitude, and theoretical knowledge. In principle, such an instrument represented direct employer participation. In practice, she regarded it as largely ceremonial.

She reported no skills test, no internship report at that stage, and little assessment beyond the final score. The presence of an enterprise assessor therefore did not make the assessment meaningfully occupational. Because scores did not discriminate among levels of performance, the rating provided little information about competence or development.

P15 in computing described a different form of weak coupling. The college required an overall training report summarizing what had been done during the company project, but the experience did not enter the regular semester grade. There was no formal skills test and the school did not monitor workplace attendance. The enterprise observed individual and group performance but did not issue an explicit numerical score.

These cases reveal two forms of proceduralization. In the first, assessment exists visibly as a score but has little discriminating meaning. In the second, documentation confirms that participation occurred without strongly representing the quality of occupational learning.

The problem was not paperwork itself. Logs and reports could support reflection and provide evidence across a placement. The issue arose when documentary completion substituted for judgement of what students had actually learned or could do.

P06’s research-institute placement illustrated this issue. Her school evaluated students primarily through internship reports, weekly reports, and experimental records. She characterized this approach as more written. The documentation provided an educational record of activity but only indirectly represented the developing experimental competence she described elsewhere in the interview.

The contrast across cases suggests that authentic workplace learning can coexist with relatively inauthentic assessment. Students may participate in genuine experiments, commercial operations, construction, care, or production while the educational institution continues to recognize that learning largely through reports and completion records.

“They are all positive evaluations. If ten is the full score, everyone gets ten. It is still a formality.” (P09)

4.4. What the Final Product Cannot See: The Learning Process as an Assessment Blind Spot

Another important distinction concerned the difference between assessing an output and understanding how learning occurred.

P12, studying statistics and big-data analysis in an enterprise-oriented workshop class, completed professional courses through projects rather than conventional final examinations. Enterprise teachers designed a concluding project and students’ completion of that project functioned as the end-of-course assessment.

At first sight, this represented strong alignment between project-based teaching and assessment. Yet the student identified an important limitation. Students completed work orders or project records showing what they understood after a training activity, but she believed these documents mainly highlighted the final product. Smaller pieces of knowledge acquired during the process, difficulties encountered, and adaptation to the enterprise teacher’s instructional approach were much less visible.

This account complicates a simple celebration of performance assessment. A final occupational product can be more authentic than a conventional written examination while still obscuring important learning. Students may arrive at similar outputs through different levels of independence, different amounts of mentoring, trial and error, or conceptual understanding. Assessment focused only on completion may therefore underrepresent development.

P04 expressed a related concern from the perspective of program evaluation rather than personal grading. Her internship questionnaire asked about overall experience, satisfaction with school-enterprise services, and suggestions for improvement. She regarded the questions as broadly relevant but too general. They could capture whether the internship was satisfactory overall but could not adequately represent specific problems encountered during practical work or differences between classroom training and actual enterprise operations.

Students thus drew attention to a broader assessment problem: standardized instruments privilege what institutions have already decided is important. Incidental learning, changing confidence, difficulties, mentor relationships, contradictions between curriculum and current workplace technology, and other emergent experiences may remain outside formal categories.

From the students’ perspective, meaningful assessment needed to capture not only whether a task was completed, but how competence was developing through engagement with real and changing work.

4.5. Assessed by Everyone, but Who Assesses the Program? The Asymmetry of Student Voice

A final theme concerned the direction of assessment. Schools and enterprises possessed multiple mechanisms for evaluating students, but students frequently described far fewer mechanisms for evaluating the enterprises, supervisors, or industry-education arrangements themselves.

P15 reported that the school did not ask students formally to evaluate the company. Although students could approach a teacher when practical problems arose and the teacher could communicate with the company, this was an informal problem-resolution channel rather than a systematic evaluation mechanism.

P10 likewise described school visits, workplace mentor evaluation, practical scores, reports, attendance, and skills assessment, but did not recall a formal questionnaire through which students evaluated the company.

P14 said that students could discuss the internship orally and communicate problems to teachers or masters, but there was no project evaluation form. The existence of communication therefore did not necessarily give students a formal role in judging placement quality.

The strongest example came from P13, a tourism student in an industry-linked program. Enterprise teachers could teach courses and assess students through examinations, but students were not formally asked by the school to evaluate the enterprise arrangement. The student was highly critical of the program, particularly discrepancies between earlier employment promises and later outcomes, and argued that an effective feedback channel was necessary to prevent unresolved problems from escalating.

Not all programs lacked reciprocal evaluation. P03 reported a more explicitly two-way arrangement in his school-based production facility: enterprises evaluated students, students evaluated enterprises, and open-ended items invited suggestions about teaching time and content. P04 also reported a student questionnaire, although she criticized its limited capacity to capture detailed experience.

The comparison demonstrates that student voice was possible but unevenly institutionalized. This matters theoretically because students were not merely recipients of training. They occupied the intersection of school and enterprise and were often uniquely positioned to identify misalignment between classroom knowledge, workplace tasks, mentoring, assessment criteria, and promised outcomes.

The interviews therefore reveal an asymmetry of evaluative power. Students were expected to demonstrate competence to schools and enterprises, while enterprises did not always have to demonstrate educational quality to students through an equally structured process.

5. Discussion

This study examined how Chinese higher vocational students experienced assessment within different forms of industry-education integration. The findings demonstrate that assessment cannot be understood simply as a technical component added after practical learning. It is one of the principal mechanisms through which the relationship between education and work is organized. Students encountered heterogeneous arrangements ranging from technically detailed performance assessment to loosely monitored reporting and apparently ceremonial enterprise ratings. Across these differences, they distinguished educational judgements concerning examinations, documentation, and course completion from occupational judgements concerning whether they could perform tasks, solve problems, meet quality requirements, work responsibly, and function within actual organizational routines.

The central argument is therefore that industry-education integration and assessment integration are related but distinct. A curriculum can be highly integrated with a workplace while its assessment remains largely school-centred. An enterprise can participate in assessment while providing little meaningful judgement. Conversely, a carefully designed practical task can connect educational and occupational standards even when conducted in a controlled training context.

5.1. Industry-Education Integration Does Not Automatically Produce Assessment Integration

The findings extend research emphasizing the complexity of work-integrated assessment. Perceived authenticity depends on alignment among workplace experience, assessment activity, and stakeholder participation (Ajjawi et al., 2020). At the same time, Chinese teaching-focused university-industry collaboration depends on sustained enterprise engagement and on the productive coordination of educational and industrial resources, interests, and responsibilities (Zhuang et al., 2024). The present interviews show what misalignment can look like from the student perspective when organizational collaboration is visible but assessment remains only weakly connected across settings.

In particular, several participants undertook authentic and consequential workplace activities while school assessment relied primarily on reports, logs, or completion requirements. The physical location of learning in an enterprise therefore did not ensure that educational assessment represented occupational learning.

This distinction is important for vocational policy. Industry-education integration is often described organizationally: colleges partner with enterprises, students enter workplaces, enterprise personnel teach, or training facilities replicate production. However, assessment asks a different question: whether the standards, evidence, judgements, and consequences used to recognize student learning are themselves connected across school and work.

Where an institution primarily requires attendance and a report, its formal message may remain that participation and documentation are sufficient evidence of learning. Where enterprise ratings automatically produce full marks, employer involvement is visible institutionally but weak educationally. Where students are required to complete realistic technical tasks and receive feedback tied to occupational standards, assessment provides stronger evidence that school and workplace expectations have actually been connected.

Assessment integration should therefore be treated as an independent indicator of the depth of industry-education integration rather than assumed from partnership structures alone.

5.2. Authenticity as Consequential Judgement Rather than Workplace Location

The data also refine the concept of authentic assessment. Authenticity was strongest when students understood why a standard mattered in occupational practice. For the numerical-control student, precision and surface quality were not arbitrary grading categories; they represented whether a component had been produced acceptably. For the construction student, KPI evaluation carried meaning because similar indicators were used for regular workers. For the e-commerce student, workplace assessment concerned actual backend operation, routine processes, task completion, and work attitude.

By contrast, an enterprise form completed with uniformly high marks was workplace-based without being strongly authentic. Its organizational origin was industrial, but its educational information value was minimal.

This pattern supports the view that authentic assessment should not be reduced to surface resemblance to professional practice (Ashford-Rowe et al., 2014; Villarroel et al., 2018). In particular, the present findings suggest that, in vocational industry-education integration, authenticity has a consequential dimension. Students experienced assessment as authentic when criteria corresponded to occupationally meaningful consequences: a component either met tolerance, a task was completed to workplace standard, a construction responsibility affected real work, or performance influenced employment opportunity.

An assessment that cannot meaningfully distinguish satisfactory from weak work tells students little about occupational quality even when completed by an employer.

5.3. From Dual Assessment to Translated Judgement

A second contribution concerns school-enterprise dual assessment. Several cases included both educational and employer judgement. However, calling an arrangement dual does not explain how the two sides relate.

The strongest examples involved some translation between environments. Enterprise assessments of work behaviour and practical operations entered school records; school-based skills tests attempted to use materials linked to workplace experience; teachers communicated with enterprise mentors and considered workplace feedback alongside student documentation.

Other arrangements were more additive than integrated. School and enterprise each evaluated students, but students were uncertain about how the judgements were combined or what consequences they carried.

This distinction corresponds to longstanding concerns about the role of multiple stakeholders in work-integrated assessment (Ferns & Zegwaard, 2014; Winchester-Seeto et al., 2016). Employer participation can enhance authenticity because workplace supervisors observe situated performance that educational staff cannot easily see. Yet enterprise involvement alone is insufficient; the broader collaboration also depends on how educational and industrial actors coordinate their different resources and priorities (Zhuang et al., 2024).

The present study suggests that effective dual assessment requires translation across evaluative systems. Schools need to understand what workplace ratings represent; enterprises need sufficiently clear criteria to distinguish training judgement from routine managerial impressions; and students need to understand the relationship between workplace expectations and academic outcomes. Without such translation, multiple assessors create administrative plurality rather than genuine integration.

5.4. From Assessment of Learning to Assessment for Occupational Judgement

The students' concerns about learning processes point toward another distinction: whether assessment merely certifies completed work or develops the capacity to judge quality.

Assessment in work-integrated learning frequently serves assessment-for-learning purposes rather than certification alone (Luk & Chan, 2024). More broadly, higher education should develop evaluative judgement so that graduates can determine the quality of their own and others' work beyond the period of formal instruction (Tai et al., 2018).

The interviews illustrate both the presence and absence of such development. Students developed evaluative judgement informally when they learned what a mentor considered acceptable, compared classroom procedures with workplace requirements, diagnosed why a task failed, or began to understand the standards governing an actual product. Yet these forms of learning were not always explicitly represented within formal assessment.

P12's observation that the project form emphasized final results while overlooking process is particularly important. The messy learning that occurs through uncertainty, correction, negotiation, and adaptation may be precisely what prepares students for future occupational judgement. A final grade may record competence without helping the student understand how to reproduce, monitor, and improve it independently.

Vocational assessment should therefore create opportunities for students to articulate why work meets a standard, diagnose deficiencies, compare alternative approaches, and reflect on differences between educational and workplace expectations. Reports and logs could contribute to this purpose, but only when designed as tools for interpretation rather than routine evidence of attendance.

5.5. Student Voice and Evaluative Reciprocity

Perhaps the most significant finding concerns reciprocity. Much assessment literature asks how teachers and employers can validly judge students. The interviews suggest that industry-education integration also requires attention to how students judge educational and workplace provision.

This is not merely a satisfaction issue. Students observe whether a mentor actually teaches, whether tasks are occupationally relevant, whether equipment is outdated, whether workload is educationally justified, whether

employment promises are credible, and whether school learning prepares them for workplace demands. Their evaluations therefore contain information about the quality of integration itself.

Yet several participants described arrangements in which enterprises possessed formal authority to judge students while students lacked an equivalent structured opportunity to judge enterprise participation. Complaints could sometimes be communicated to teachers, but problem resolution is different from systematic evaluation.

Developing evaluative judgement should therefore involve more than self-assessment of individual performance. Students should also learn to judge educational and occupational environments against explicit standards: whether tasks are developmentally appropriate, whether feedback is useful, whether supervision is adequate, whether promised learning opportunities are delivered, and whether workplace practices meet agreed expectations.

Reciprocal evaluation would reposition students from passive recipients of institutional judgement toward participants in the governance of industry-education integration.

5.6. Toward Meaningful Assessment Integration

At an aggregate level, the findings suggest that meaningful assessment integration contains four interrelated principles. In the first place, performance alignment requires assessment to represent capabilities actually needed in occupational practice while retaining relevant conceptual knowledge. Practical assessment should not simply imitate work superficially but should require students to make, decide, diagnose, communicate, or perform according to recognizable occupational standards. Following the above, multi-source judgement requires meaningful contributions from educational and occupational actors. Different assessors can observe different dimensions of competence, but criteria and consequences must be intelligible across settings. Then, developmental visibility requires assessment to recognize learning processes as well as final outputs. Students should receive feedback and opportunities to interpret mistakes, progression, mentoring, and changing independence. In addition, evaluative reciprocity requires students to participate as judges as well as objects of judgement. They need opportunities to evaluate their own work, understand quality standards, and provide consequential feedback on enterprises and programs.

These dimensions allow assessment arrangements to be understood along a continuum. At one end lies administrative assessment, dominated by attendance, required documentation, generic reports, or non-discriminating forms. Such assessment confirms that participation occurred but provides limited evidence about learning. A second form is parallel assessment, in which school and enterprise both evaluate the student but operate largely independently. A third is integrated performance assessment, in which school and workplace evidence are deliberately connected and occupational performance contributes substantively to educational judgement. The most educationally developed form could be described as reciprocal developmental assessment. Here assessment not only integrates school and occupational standards but also develops students' capacity to judge quality and provides structured opportunities for students to evaluate the learning environment itself.

The interviews contained examples of elements of all four forms, but no single case consistently embodied the final form. This suggests that assessment may remain one of the less integrated components even within relatively mature industry-education partnerships.

5.7. Implications for Vocational Education Practice

The findings have several implications for the design of industry-education integration. First, colleges should distinguish between evidence of participation and evidence of competence. Attendance, internship logs, and reports may be necessary for administrative accountability, but they should not substitute for assessment of what students can actually do.

Seen in this light, enterprise assessment should avoid generic rating forms that encourage uniformly positive scores. Criteria should relate to observable dimensions of work, and assessors should be able to distinguish stages of performance rather than simply classify students as satisfactory.

In addition, practical assessment should be linked to feedback. A score for technical performance is more educationally useful when students understand why a product or action met or failed to meet the standard and how their next performance can improve.

At the level of school-enterprise coordination, staff on both sides should jointly clarify which aspects of competence each is best positioned to assess. Educational staff may be better placed to evaluate conceptual integration and reflection, while workplace supervisors may be better positioned to evaluate situated performance, reliability, communication, and occupational judgement. Such coordination is consistent with the broader need to align the distinctive resources and responsibilities contributed by educational institutions and enterprises in teaching-focused collaboration (Zhuang et al., 2024).

At the same time, student evaluation of the placement should be treated as part of quality assurance rather than an optional satisfaction survey, since students' engagement in such evaluation plays an important role in quality assurance in higher education (Zhang et al., 2024). Instruments should include opportunities to report specific issues concerning task relevance, mentoring, learning opportunities, technology, safety, workload, workplace treatment, and alignment with curriculum to strengthen students' right to voice their concerns (Zhang et al., 2025).

Finally, these arrangements should give students explicit opportunities to develop evaluative judgement. Students should learn not only to perform occupational tasks but also to explain what good work looks like, interpret criteria, judge examples, evaluate their own performance, and identify when workplace arrangements support or obstruct learning.

5.8. Limitations and Future Research

Several limitations should be recognized. First, the study is based on student interviews and therefore examines experienced rather than formally documented assessment systems. Institutional assessment policies, employer rubrics, marking schemes, and course documentation were not systematically analysed. Future research could compare student accounts with official assessment instruments and interviews with teachers and enterprise supervisors.

Second, we would note that assessment was one component of broader semi-structured interviews rather than the sole topic. Future studies could examine whether students increasingly develop evaluative judgement, how they reinterpret earlier grades after entering employment, and whether particular assessment designs improve occupational learning and transition.

6. Conclusions

This study shows that students experience assessment in industry-education integration as movement between different ways of defining competence. School assessment may prioritize examinations, reports, attendance, and formally recordable evidence, whereas workplace assessment makes visible practical ability, task completion, reliability, responsibility, and response to real problems.

Across the cases, the two evaluative environments could complement one another, but they were not automatically integrated. Some students experienced productive combinations of enterprise observation, skills testing, school assessment, and documentary reflection. Others encountered reports disconnected from workplace competence, employer ratings with little discriminating value, or limited opportunities to evaluate the quality of enterprise participation.

The central implication is that authentic learning does not necessarily produce authentic assessment. Industry-education integration becomes educationally deeper when assessment itself crosses the school-work boundary: occupational standards become meaningfully connected to academic judgement; multiple assessors contribute intelligible and consequential evidence; learning processes as well as outputs become visible; and students participate in evaluation rather than remaining only its objects.

For this reason, assessment integration should be understood not as a technical afterthought but as a central criterion for judging the quality of industry-education integration.

Author Contributions

T.Z.: conceptualization, methodology, data curation, writing—original draft preparation; Y.Y.: investigation, data curation, visualization, writing—reviewing and editing. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Institutional Review Board Statement

The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethics Committee of Faculty of Education, Beijing Normal University (BNU202109100025, 28 September 2021)

Informed Consent Statement

Informed consent was obtained from all subjects involved in the study.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare no conflict of interest.

Use of AI and AI-Assisted Technologies

No AI tools were utilized for this paper.

References

- Ajjawi, R., Tai, J., Huu Nghia, T. L., Boud, D., Johnson, L., & Patrick, C.-J. (2020). Aligning assessment with the needs of work-integrated learning: The challenges of authentic assessment in a complex context. *Assessment & Evaluation in Higher Education*, 45(2), 304–316. <https://doi.org/10.1080/02602938.2019.1639613>.
- Ashford-Rowe, K., Herrington, J., & Brown, C. (2014). Establishing the critical elements that determine authentic assessment. *Assessment & Evaluation in Higher Education*, 39(2), 205–222. <https://doi.org/10.1080/02602938.2013.819566>.
- Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education*, 22(2), 151–167. <https://doi.org/10.1080/713695728>.
- Boud, D., Costley, C., Cranfield, S., Desai, J., Nikolou-Walker, E., Nottingham, P., & Wilson, D. (2023). The pivotal role of student assessment in work-integrated learning. *Higher Education Research & Development*, 42(6), 1323–1337. <https://doi.org/10.1080/07294360.2022.2152981>.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>.
- Braun, V., & Clarke, V. (2021). One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology*, 18(3), 328–352. <https://doi.org/10.1080/14780887.2020.1769238>.
- Ferns, S., & Zegwaard, K. E. (2014). Critical assessment issues in work-integrated learning. *Asia-Pacific Journal of Cooperative Education*, 15(3), 179–188.
- Jackson, D. (2018). Challenges and strategies for assessing student workplace performance during work-integrated learning. *Assessment & Evaluation in Higher Education*, 43(4), 555–570. <https://doi.org/10.1080/02602938.2017.1378618>.
- Luk, L. Y. Y., & Chan, C. K. Y. (2024). Assessing student learning in work-integrated learning (WIL): A systematic review on assessment methods and purposes. *Studies in Educational Evaluation*, 81, 101345. <https://doi.org/10.1016/j.stueduc.2024.101345>.
- Müller, A. (2024). Cooperation between colleges and companies: Vocational education, skill mismatches and China's turnover problem. *The China Quarterly*, 260, 986–1004. <https://doi.org/10.1017/S0305741023001868>.
- Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education*, 76, 467–481. <https://doi.org/10.1007/s10734-017-0220-3>.
- Villarroel, V., Bloxham, S., Bruna, D., Bruna, C., & Herrera-Seda, C. (2018). Authentic assessment: Creating a blueprint for course design. *Assessment & Evaluation in Higher Education*, 43(5), 840–854.
- Wang, G., & Wang, Z. (2023). Vocational education: A poor second choice? A comparison of the labour market outcomes of academic and vocational graduates in China. *Oxford Review of Education*, 49(3), 408–427. <https://doi.org/10.1080/03054985.2022.2096583>.
- Winchester-Seeto, T., Rowe, A., & Mackaway, J. (2016). Sharing the load: Understanding the roles of academics and host supervisors in work-integrated learning. *Asia-Pacific Journal of Cooperative Education*, 17(2), 101–118.
- Zhang, L., Xiang, Y., & Zheng, L. (2025). Institutional restructuring and practical pathways for quality assurance in Macau higher education: A marketization transition perspective. *Journal of Educational Technology and Innovation*, 7(4), 14–25. <https://doi.org/10.61414/nsf2s091>.
- Zhang, L., Zheng, L., & Zhou, Y. (2024). Assessing and improving the quality of undergraduates in preschool education programs in China: The course experience. *Journal of Educational Technology and Innovation*, 6(3), 31–46. <https://doi.org/10.61414/jeti.v6i3.206>.
- Zhuang, T., Zhou, H., & Sun, Q. (2024). Ushering in industrial forces for teaching-focused university-industry collaboration in China: A resource-dependence perspective. *Studies in Higher Education*, 49(12), 2357–2375. <https://doi.org/10.1080/03075079.2024.2306343>.

Article

How Teacher Learning Affects Middle School Students' Learning Motivation and Habits: An Empirical Study Based on PLS-SEM

Junjie Yan ¹, Jing Liu ¹, Xiaonan Luo ¹, Qunfang Zeng ², Huilin Zhang ¹ and Jiafeng Xu ^{3,*}

¹ Laboratory of Ecological Civilization and Environmental Management, University Laboratory of Humanities (Zhejiang), Wenzhou University, Wenzhou 325035, China

² Experimental Teaching Center for Teacher Education, Wenzhou University, Wenzhou 325035, China

³ School of Immigration Management, China People's Police University (Guangzhou), 399 Aoti South Road, Tianhe District, Guangzhou 510000, China

* Correspondence: xujiafeng@cppu.edu.cn; Tel.: +86-13631390490

How To Cite: Yan, J., Liu, J., Luo, X., Zeng, Q., Zhang, H., & Xu, J. (2026). How Teacher Learning Affects Middle School Students' Learning Motivation and Habits: An Empirical Study Based on PLS-SEM. *Journal of Educational Technology and Innovation*, 8(3), 64–79. <https://doi.org/10.61414/3tafbv16>

Received: 19 March 2026

Revised: 2 September 2026

Accepted: 16 September 2026

Published: 30 September 2026

Abstract: There is currently a lack of sufficient empirical support regarding how teachers' learning affects students' learning motivation and habits. This research is based on a survey of teachers and students in 23 middle schools in Wenzhou City, China. Nine hundred questionnaires were distributed in this research, and 895 were returned. We used partial least squares structural equation modeling (PLS-SEM) to reveal the relationship between teacher learning and student learning motivation, learning habits. The research found that "teacher learning" has a direct and significant effect on teacher professional development, teachers' working status, and learning task design. At the same time, teacher learning indirectly affects students' learning motivation and learning habits through learning task design, teachers' working status, and teacher professional development. This research reminds educational administration agencies that they should focus on and support teacher learning, which will enhance teacher professional development, improve teachers' working status, and student learning.

Keywords: teacher learning; teacher professional development; teachers' working status; learning motivation; learning habits

1. Introduction

In the current process of teacher professional development, teachers need to not only continuously update their subject knowledge but also improve their teaching skills to respond to the diverse learning needs of students (You et al., 2025; Smit & Humpert, 2012). Teacher learning has become an integral part of modern education systems and is considered a key factor in improving teaching quality and student learning effectiveness (Darling-Hammond et al., 2017; Ruthven et al., 2004). With the development of the education system and global trends placing higher demands on teachers' professional abilities, the importance of teacher learning has become increasingly apparent. Although teacher learning is widely recognized in theory as an important means of improving educational effectiveness, there are still many unresolved issues regarding how to improve student learning effectiveness. Especially in the context of limited educational resources, how to achieve a win-win situation for teachers and students through systematic teacher learning projects still needs to be explored in depth (Borko, 2004).

First, teacher professional development demonstrates the individual growth effects of teacher learning. Avalos (2011) pointed out that teacher learning is the core of teacher professional development, and through a



combination of deep reflection and practice, it can promote the improvement of teachers' teaching abilities. Second, teachers' working status reveals the emotional motivation mechanism of teacher learning. Madigan & Kim (2021b) pointed out that teachers can enhance their sense of commitment to work and reduce the risk of professional burnout by participating in teacher learning activities. Finally, the design of student learning tasks demonstrates the application value of teacher learning. When teachers participate in training and learning about task design, they are better able to understand student needs and design learning tasks that are more suited to student abilities (Smets & Struyven, 2020).

With the increasing attention paid to the effects of teacher learning, especially the relationship between the quality of teacher learning and student learning efficiency. For example, Desimone (2009) emphasized that teacher learning can indirectly enhance students' learning motivation and engagement by affecting teaching confidence and the quality of classroom interaction. Many researchers also point out that positive learning motivation can improve students' learning engagement and achievement (Martínez et al., 2019; Wu et al., 2022; Yan et al., 2025). At the same time, positive learning habits, such as effective time management and proactive learning strategies, are also closely related to academic success (Pérez-González et al., 2022).

Previous researchers have shown that teacher professional development, teachers' working status, and learning task design are all likely to be important factors affecting students' learning motivation and learning habits (Holzberger & Prestele, 2021; Desimone, 2009; Reeve & Cheon, 2021; Guskey, 2002). However, although there has been research focusing on the effect of teacher learning on student learning, there are still some shortcomings in the research (Opfer & Pedder, 2011; Day et al., 2016). First, most research focuses on the direct effects of teacher learning and neglects its indirect effects. Second, existing research is mainly based on European and American education systems, with little research on teacher learning in other cultural contexts. For example, in the Confucian cultural context, teacher learning may have different mechanisms and effects (Huang & Asghar, 2018). Third, empirical research on teacher learning also has methodological shortcomings, such as sample size limitations and complex relationships between variables that have not been fully revealed (Zhang et al., 2024).

This research explores the mechanism of teacher learning on professional development and student learning. It uses partial least squares structural equation modeling (PLS-SEM) to reveal the direct effect of teacher learning on professional development, and the indirect effect on students through mediating variables such as teacher professional development, teachers' working status, and learning task design. Finally, it examines the moderating effects of moderating variables such as class manager and professional position. Furthermore, in the Chinese education field, class managers and professional positions will become important indicators that affect student learning and reflect teacher professional levels (Shi & Leuwerke, 2010). Therefore, the class manager and professional position were introduced as moderating variables to reveal the mechanism of teacher learning more comprehensively. Based on the above research results, this research proposes the following three important issues:

- (1) How does teacher learning directly affect professional development?
- (2) How does teacher learning indirectly affect student learning?
- (3) Do professional position and class manager have a moderating effect?

2. Literature Review and Model Development

2.1. How Does Teacher Learning Directly Affect Professional Development?

Teacher learning typically takes the form of ongoing participation in professional development programs and reflective practice, through which teachers refine their knowledge base and instructional approaches (Vescio et al., 2008). In our research, teacher learning refers to teachers participating in school-organized training, seminars, school-based teaching and research, and other activities to continuously improve their professional skills, teaching knowledge, management abilities, and learning task design abilities, with the aim of promoting professional development and student learning.

In educational practice, teacher professional development is commonly seen in teachers' intentional efforts to refine their teaching methods, boost their instructional impact, and meet higher standards of educational excellence (Guskey, 2002). Teacher learning plays a key role in teacher professional development. Research has shown that there is a direct and significant relationship between teacher learning and teacher professional development. For example, Desimone (2009) pointed out that teachers' participation in deep learning activities such as peer collaboration, reflective teaching, and professional learning communities can directly improve their teaching practices and promote their professional development. In addition, teacher learning is closely related to the effectiveness of teacher professional development. Through continuous knowledge renewal and teaching skill enhancement, teachers can not only improve the spread of teaching content, but also enhance classroom management and the quality of teaching interaction (Desimone, 2009). According to research by Zhou et al. (2023),

continuous teacher learning, especially professional development activities that can be integrated with actual teaching issues, significantly improves teachers' teaching confidence and effectiveness. In our research, teacher professional development refers to the process by which teachers improve their knowledge and skills in subject knowledge, teaching skills, student evaluation, information technology application, classroom management, personalized teaching methods, task design, etc., according to their teaching requirements. Based on this analysis, our research proposes the following hypothesis:

H1a: *Teacher learning has a direct and significant effect on teacher professional development.*

Teachers' working status is typically manifested in their high levels of energy, dedication, and optimism when carrying out teaching and other school-related responsibilities (Schaufeli et al., 2002). Teachers' working status includes high levels of work engagement and enthusiasm for teaching activities, and this status is closely related to teachers' work performance, work satisfaction, and level of professional burnout. This not only reflects teachers' psychological commitment to their work but also involves their emotional identification with their professional role and their sustained efforts. Recent research has emphasized the substantial effect of teacher learning on teachers' working status. For example, Goldschmidt and Phelps (2010) found that through reflective learning and sustained professional training, teachers were able to not only improve their teaching efficiency but also stimulate their creativity and positivity in the teaching process. This positive working status helped reduce stress in teaching and increased teachers' professional satisfaction. Similarly, Madigan & Kim (2021b) pointed out that teachers can enhance their sense of work engagement and reduce the risk of professional burnout by participating in various forms of learning activities such as training courses, peer collaboration, reflective teaching, etc. Especially when teachers' learning is relevant to teaching work, they will significantly improve work engagement and satisfaction. In our research, teachers' working status refers to the psychological and behavioral performance of teachers who are energetic, effectively respond to issues at work, and are positively engaged in classroom teaching. Based on this analysis, our research proposes the following hypothesis:

H2a: *Teacher learning has a direct and significant effect on teachers' working status.*

Learning task design usually refers to teaching designing tasks and activities that can effectively promote student learning based on learning objectives, learner characteristics, and contextual requirements (Roure & Pasco, 2018). Extensive research has identified that teacher learning plays a crucial role in teaching design of learning tasks for students. On the one hand, Smets and Struyven (2020) argued that teachers continuously adjust and improve students' learning task design methods through practical reflection and participation in teaching activities. On the other hand, according to the research of Wake et al. (2016), teachers' decision-making in student learning task design is usually affected by their professional learning. When teachers enhance their pedagogical knowledge through participation in professional development activities, collaboration with colleagues, and self-reflection, they are able to design more challenging and meaningful learning tasks. Some researchers even claim that through professional learning and development, teachers can better understand students' learning requirements and translate these requirements into specific student learning task design solutions (Wake et al., 2016). In our research, learning task design refers to the design of homework by teachers targeting the knowledge, concepts, and skills that students learn. Based on this analysis, our research proposes the following hypothesis:

H3a: *Teacher learning has a direct and significant effect on the design of student learning tasks.*

2.2. How Does Teacher Learning Indirectly Affect Student Learning?

In educational settings, learning motivation is commonly observed when students show persistent effort, purposeful engagement, and self-directed learning, driven by internal psychological processes (Urhahne & Wijnia, 2023). In our research, learning motivation refers to the internal driving forces that enable students to set clear learning goals, take an interest in classroom activities, actively participate in classroom activities, and stimulate their initiative and positive attitude toward learning.

Learning habits are typically manifested in students' regular use of structured approaches to learning, such as time management, note-taking, and self-monitoring, which help sustain academic performance (Urhahne & Wijnia, 2023). In our research, learning habits refer to the normalization of good learning behaviors such as actively preparing learning materials before class, independent learning in class, and promptly correcting mistakes after class.

Research shows that there is a positive correlation between teacher professional development and student learning motivation. For example, Datnow & Hubbard (2015) found that the degree of teacher participation in

professional development was significantly related to student engagement and learning motivation in the classroom. On the one hand, teachers who have received professional development training are more likely to adopt student-centered teaching methods, encourage independent learning, and enhance students' intrinsic motivation. On the other hand, teacher professional development can also improve their sensitivity to student requirements and promote the improvement of classroom management skills, thereby creating a learning environment conducive to student active participation. This improvement in the environment further enhances students' learning motivation and academic achievement. Based on this analysis, our research proposes the following hypotheses:

H1b: *Teacher professional development has a direct and significant effect on student learning motivation.*

H1: *Teacher learning indirectly has a positive effect on student learning motivation by enhancing teacher professional development.*

Research shows that teachers' working status, including their professional burnout, work satisfaction, and emotional engagement, directly affects students' learning motivation. Research found that there is a significant correlation between teachers' emotions and students' emotions and motivation (Frenzel et al., 2009). When teachers display positive emotions, students are more likely to show higher learning motivation. On the other hand, teacher professional burnout has also been shown to have a negative effect on student learning motivation (Madigan & Kim, 2021a). When teachers feel tired, indifferent, or unfulfilled, students may sense these negative emotions, which can reduce their interest and engagement in learning. Based on this analysis, our research proposes the following hypotheses:

H2b: *Teachers' working status has a direct and significant effect on students' learning motivation.*

H2: *Teacher learning indirectly has a positive effect on students' learning motivation by improving teachers' working status.*

Research has shown that teachers can stimulate students' interest, enhance students' intrinsic motivation, and strengthen students' engagement and participation through effective learning task design (Guskey, 2002). Guskey (2002) pointed out that professional learning enables teachers to understand better how to design tasks that are suitable for student learning, thereby enhancing student learning motivation. When teachers are able to design challenging and engaging learning tasks based on student requirements and interests, students are more likely to develop positive learning attitudes and motivation. Based on this analysis, our research proposes the following hypotheses:

H3b: *Effective learning task design has a direct and significant effect on students' learning motivation.*

H3: *Teacher learning indirectly has a positive effect on students' learning motivation by affecting learning task design.*

Research shows that teachers play a key role in cultivating students' learning habits (Darling-Hammond et al., 2017; Smets & Struyven, 2020). On the one hand, Darling-Hammond et al. (2017) believed that teachers learning how to design effective learning tasks helped students form positive learning habits. On the other hand, when designing learning tasks, teachers could enhance students' learning motivation and further consolidate good learning habits by combining students' interests and real-life situations. Smets and Struyven's (2020) research showed that contextualized learning task design could enhance student engagement and learning effectiveness. Based on this analysis, our research proposes the following hypotheses:

H3c: *Effective learning task design has a direct and significant effect on students' learning habits.*

H4: *Teacher learning indirectly has a positive effect on students' learning habits by affecting learning task design.*

In this research, we attempted to use class manager and professional position as moderating variables because relevant research has pointed out that the multiplicity of teachers' identities significantly expands their scope of responsibilities. For example, in the Chinese education system, class managers usually play multiple roles, including teaching, student management, and home-school communication, which contrasts sharply with the single role of subject teachers. Clifton and Jordan (2019) pointed out that teachers with mixed identities, such as class managers, are more likely to adopt innovative strategies to satisfy diverse requirements. Therefore, the class manager's identity will have an effect on teacher professional development and learning task design. Wettstein et

al. (2021) found that class managers need to devote more energy to switching roles when balancing discipline management, emotional support, and subject teaching. Therefore, the class manager's identity has an effect on teachers' working status and a moderating effect. In addition, Zheng et al. (2023) found that teacher quality, as measured by academic qualifications and professional titles, significantly affects students' cognitive achievement. It shows that professional experience accumulated during the professional position promotion process directly affects teaching innovation capability and the depth of teaching practice. Therefore, our research proposes to use class manager and professional position as moderating variables. These two moderating variables may affect the model predictors and the direct effect of teacher learning on teacher professional development, teachers' working status, and learning task design, thereby further affecting the indirect effect of teacher learning on students' learning motivation and learning habits.

Based on the above hypothesis, this research model includes one predictive variable (teacher learning), three mediating variables (teacher professional development, teachers' working status, and learning task design), two dependent variables (learning motivation and learning habits), and two moderating variables (professional position and class manager). A model of teacher learning affecting students will be formed. As shown in Figure 1.

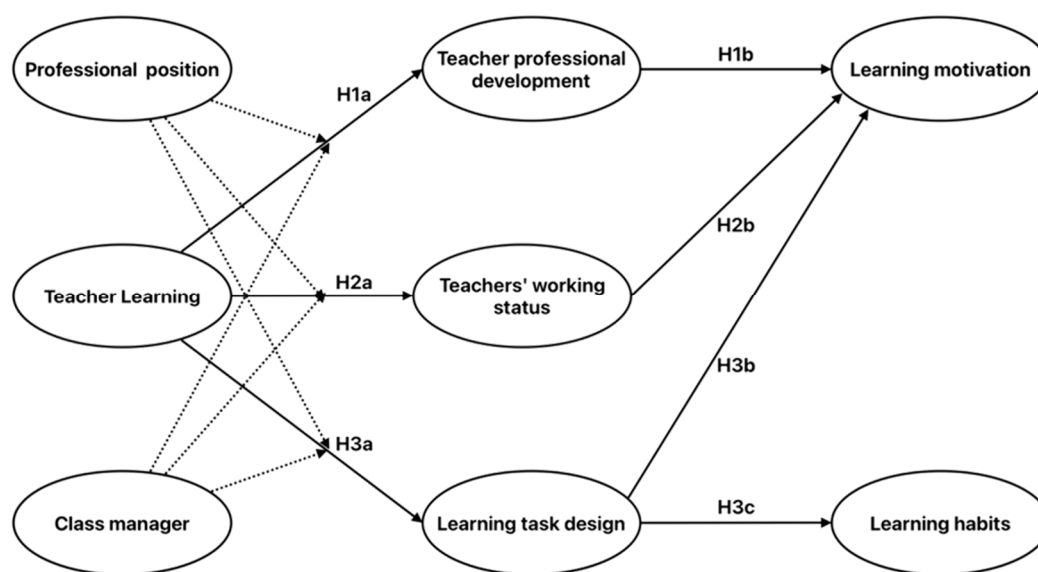


Figure 1. Model of teacher learning affecting students.

3. Method

3.1. Sample Selection and Data Collection

In terms of sample selection size, some researchers suggest that the ratio of sample size to indicators should be at least 5:1 or 10:1 or that the sample size should be at least 100–200 observations (Kock, 2018). For this research, since 27 indicators were used, a critical sample size of at least 270 observations was required. Therefore, this research selected a sample size of 895.

The survey questionnaire was distributed in 2024 via Wenjuanxing (<https://www.wjx.cn>, a widely used online survey platform). A total of 23 middle schools in Wenzhou City, China, were surveyed. The survey participants consisted of teachers and the students they teach. The survey lasted for approximately one month. A total of 895 valid responses were collected, with a response rate of 94.91%. The teacher sample included 353 class managers (39.4%) and 542 non-class managers (60.5%). In terms of professional positions, 60 teachers had junior professional positions (6.7%), 221 teachers had second-level professional positions (24.6%), 370 teachers had first-level professional positions (41.3%), and 244 teachers had senior professional positions (27.2%). The student sample included 456 female students (50.9%) and 439 male students (49.1%), with a roughly equal gender ratio overall. Participants obtained consent prior to participation and were informed of the purpose of the research, their rights, and data protection measures.

3.2. Measurement Scale

The measurement tool uses a five-point Likert scale. To standardize the model estimates for each option and make them consistent with the five-point Likert scale, we created a numerical scale from 1 to 5 for each option. Detailed results of measurement scales and descriptive statistical information are shown in Table 1.

Table 1. Measurement scale and factor loadings.

Construct	Item	Items	Loading	Mean	St.dev.	Adapted from
Teacher Learning	TL1	How often do schools organize training and lecture activities on teachers' professional skills?	0.836	0.182	0.006	Desimone (2009)
	TL2	How often do the teaching and research groups and lesson planning groups conduct seminars on teaching and learning topics?	0.876	0.188	0.005	
	TL3	Teaching and research teams organize training and learning for teachers on the design of learning tasks for students.	0.850	0.174	0.005	
	TL4	Teaching and research teams regularly check students' implementation of learning tasks.	0.833	0.163	0.005	
	TL5	Teachers will comply with the school. Management system for teaching and learning.	0.756	0.155	0.007	
	TL6	Teaching and research groups showcase teachers' teaching performance.	0.767	0.155	0.006	
	TL7	My school conducts regular school-based workshops on teaching and learning.	0.866	0.189	0.005	
Teacher professional development	TPD1	Please select the extent to which you currently need professional development in terms of knowledge and teaching skills in the subject area you are teaching.	0.868	0.237	0.013	Guskey (2002)
	TPD2	Please select the extent to which you currently need to evaluate student approach expertise.	0.884	0.207	0.010	
	TPD3	Please select the extent to which you currently need professional development in multimedia information technology and big data technology.	0.765	0.147	0.015	
	TPD4	Please select the extent to which you currently need professional development in student behavior and classroom management.	0.844	0.197	0.014	
	TPD5	Please select the extent to which you currently need professional development in teaching methods for personalized learning.	0.880	0.205	0.010	
	TPD6	Please select the extent to which you currently require professional development in assignment research and design.	0.833	0.182	0.012	
Teachers' working status	TWS1	I feel energized on workdays.	0.735	0.223	0.014	Toropova et al. (2021)
	TWS2	I can effectively deal with issues brought to me by students and coworkers during the workday.	0.828	0.261	0.012	
	TWS3	I engage my students in the classroom with a positive state of mind.	0.789	0.254	0.013	
	TWS4	I believe my work positively impacts others	0.823	0.269	0.012	
	TWS5	I like my current workplace and workmates.	0.758	0.261	0.014	
Learning task design	LTD1	I assign tasks that require students to think critically	0.777	0.391	0.017	Roure & Pasco (2018)
	LTD2	I give students the necessary scaffolding for assignments based on the difficulty of the assignment.	0.796	0.409	0.017	
	LTD3	I transfer the theoretical knowledge I have learned about assignment design to actual assignment design practice.	0.814	0.455	0.018	
Learning motivation	LM1	I know exactly what the learning objectives are	0.747	0.351	0.019	Pintrich et al. (1993)
	LM2	I am interested and willing to participate in classroom learning activities.	0.913	0.388	0.012	
	LM3	My learning initiative and motivation will be stimulated during the learning process.	0.930	0.412	0.011	
Learning habit	LH1	Before each lesson, I will prepare my own textbook and school supplies.	0.775	0.414	0.035	Zimmerman (2002)
	LH2	In class, even if the teacher is not there, I will read or do my homework carefully.	0.819	0.400	0.032	
	LH3	I will check for errors and correct them as soon as I get my homework back.	0.821	0.427	0.030	

Among them, teacher learning (TL) contains seven items adapted from Desimone's (2009) research. The response options and their corresponding definitions are as follows: "Always" corresponds to 5, "Often" corresponds to 4, "Sometimes" corresponds to 3, "Rarely" corresponds to 2, and "Never" corresponds to 1.

Teacher professional development (TPD) contains six items adapted from Guskey's (2002) research. The response options and their corresponding definitions are as follows: "Not needed at all" corresponds to 1, "Not needed at present" corresponds to 2, "Low level of need" corresponds to 3, "Moderate level of need" corresponds to 4, and "High level of need" corresponds to 5.

Teachers' working status (TWS) contains five items adapted from the research of Toropova et al. (2021). The response options and their corresponding definitions are as follows: "Always" corresponds to 5, "Often" corresponds to 4, "Sometimes" corresponds to 3, "Rarely" corresponds to 2, and "Never" corresponds to 1.

Learning task design (LTD) contains three items adapted from Roure & Pasco's (2018) research. The response options and their corresponding definitions are as follows: "Always" corresponds to 5, "Often" corresponds to 4, "Sometimes" corresponds to 3, "Rarely" corresponds to 2, and "Never" corresponds to 1.

Learning motivation (LM) contains three items, adapted from the research of Pintrich et al. (1993); Learning habit (LH) contains three items, adapted from the research of Zimmerman (2002). The response options and their corresponding definitions are as follows: "Very much agree" is 5, "Agree" is 4, "Somewhat agree" is 3, "Disagree" is 2, and "Strongly disagree" is 1. Detailed results of measurement scales and descriptive statistical information are shown in Table 1.

3.3. Data Analysis

The PLS-SEM method was used for data analysis in the research. To estimate the model, we used the PLS-SEM algorithm with the weighting path scheme in SmartPLS 4 software (Version 4.1.0.9) with a maximum of 3000 iterations and default initial weights, and employed bootstrapping, a nonparametric procedure, running 5000 samples, to determine the statistical significance of the PLS-SEM results, as recommended by Ringle et al. (2022). Reflectively specified constructs were analyzed using the indicator loadings, and an indicator loading above 0.7 suggested that the construct accounted for more than 50% of the variance in the indicator, indicating an acceptable level of item reliability. Therefore, the loadings of this research meet the criteria. See Table 1 for details.

4. Results

4.1. Questionnaire Validity and Reliability Analysis

Composite reliability is a standard for evaluating reliability, with results ranging from 0.70 to 0.95, indicating acceptable to good levels of reliability (Hair et al., 2022). Internal consistency reliability was measured using Cronbach's alpha, which uses similar thresholds as composite reliability (ρ_c). Another reliability coefficient (ρ_A), based on Dijkstra (2010), was also used to provide an exact and consistent alternative. The average variance extracted (AVE) from all items linked to a specific reflective variable was used to assess the convergent validity of the measurement models, and an AVE threshold of 0.50 or higher was deemed acceptable (Sarstedt et al., 2022). The Cronbach's alpha, Reliability coefficient, Composite reliability, and AVE values of this research model meet the standards. See Table 2 for details.

The preferred method for analyzing discriminant validity in PLS-SEM is the heterotrait–monotrait ratio of correlations (HTMT) introduced by Henseler et al. (2015). A HTMT threshold of 0.90 is recommended to ensure discriminant validity, particularly when constructs are conceptually similar, whereas a threshold of 0.85 is more appropriate for more distinct constructs. In Table 3, all values fall below the 0.85 threshold, indicating good discriminant validity.

Table 2. Construct reliability and validity.

	Cronbach's Alpha	Reliability Coefficient (ρ_A)	Composite Reliability (ρ_c)	AVE
Learning habit	0.728	0.729	0.847	0.648
Learning motivation	0.830	0.841	0.900	0.753
Learning task design	0.711	0.715	0.838	0.633
Teacher learning	0.923	0.927	0.938	0.684
Teacher professional development	0.921	0.932	0.938	0.717
Teachers' working status	0.846	0.849	0.891	0.620

Table 3. HTMT values.

	LH	LM	LTD	TL	TPD	TWS
LH						
LM	0.256					
LTD	0.458	0.527				
TL	0.496	0.553	0.735			
TPD	0.291	0.320	0.248	0.209		
TWS	0.414	0.423	0.624	0.528	0.107	

Note: LH: Learning habit; LM: Learning motivation; LTD: Learning task design; TL: Teacher Learning; TPD: Teacher professional development; TWS: Teachers' working status.

4.2. The Significance Tests of the Structural Model Path Coefficients and Hypothesis Verification

The coefficient of determination (R^2) is then assessed to determine the explanatory power of each construct and the overall model. R^2 ranges from 0 to 1, and higher values indicate larger explanatory power. As a general rule of thumb, R^2 values of 0.25, 0.50, and 0.75 are considered weak, moderate, and substantial, respectively (Hair et al., 2011). To determine the effect size of a variable, f^2 values of 0.35, 0.15, and 0.02 correspond to large, medium, and small effects, respectively, and values below 0.02 suggest an absence of impact (Sarstedt et al., 2022).

Figure 2 shows the results of the PLS-SEM analysis. Through the analysis of standardized regression coefficients, the results show that the standardized regression coefficient of teacher learning on learning task design is the highest, with a value of 0.620, followed by teachers' working status (0.465) and teacher professional development (0.195). The strongest predictor of learning motivation is learning task design, with a standardized regression coefficient of 0.253, followed by teachers' working status (0.219) and teacher professional development (0.217). Therefore, the analysis results show that teacher learning has the most significant indirect effect on students' learning motivation through its effect on learning task design. Similarly, the standardized regression coefficient of learning task design on learning habits is 0.332, indicating that teacher learning also has a positive indirect effect on students' learning habits through its effect on learning task design.

The R^2 value of learning motivation in Figure 2 is 0.240, which indicates that teacher learning has a weak positive explanatory power on learning motivation through "teacher professional development, teachers' working status, and learning task design. The R^2 value of learning habits in Figure 2 is 0.110, which indicates that teacher learning has a weak positive explanatory power on learning habits through teacher professional development, teachers' working status, and learning task design.

The commonly used significance levels for p -values are 0.05, 0.01, and 0.001. These levels correspond to the standards of * $p < 0.05$ (statistically significant), ** $p < 0.01$ (highly significant), and *** $p < 0.001$ (extremely significant) (Biau et al., 2010; Wasserstein & Lazar, 2016). Table 4 shows that, except for H1, whose p -value is 0.001, the p -values of the remaining hypotheses are all less than 0.001, which means that each hypothesis is extremely significant.

The analysis results indicate that all hypotheses are valid and significant. Table 4 lists the significance tests of the structural model path coefficients and hypothesis verification.

We have incorporated moderating relationships of class manager and professional position into the model, which were hypothesized a priori and specifically tested. The results of the moderating effect of class manager and professional position are presented in Table 5. The results show that neither of the two moderating variables had a significant effect on the relationships between the predictors and the criterion variables across the hypothesized paths.

Table 4. Path coefficients and the results of the significance tests.

Hypothesis	Path	Coefficient	p Values	f^2	Confirmed
H1a	TL→TPD	0.195	0.000	0.025	Yes
H2a	TL→TWS	0.465	0.000	0.169	Yes
H3a	TL→LTD	0.620	0.000	0.359	Yes
H1b	TPD→LM	0.217	0.000	0.059	Yes
H2b	TWS→LM	0.219	0.000	0.048	Yes
H3b	LTD→LM	0.253	0.000	0.062	Yes
H3c	LTD→LH	0.332	0.000	0.124	Yes
H1	TL→TPD→LM	0.042	0.001	/	Yes
H2	TL→TWS→LM	0.102	0.000	/	Yes
H3	TL→LTD→LM	0.157	0.000	/	Yes
H4	TL→LTD→LH	0.206	0.000	/	Yes

Note: LH: Learning habit; LM: Learning motivation; LTD: Learning task design; TL: Teacher Learning; TPD: Teacher professional development; TWS: Teachers' working status.

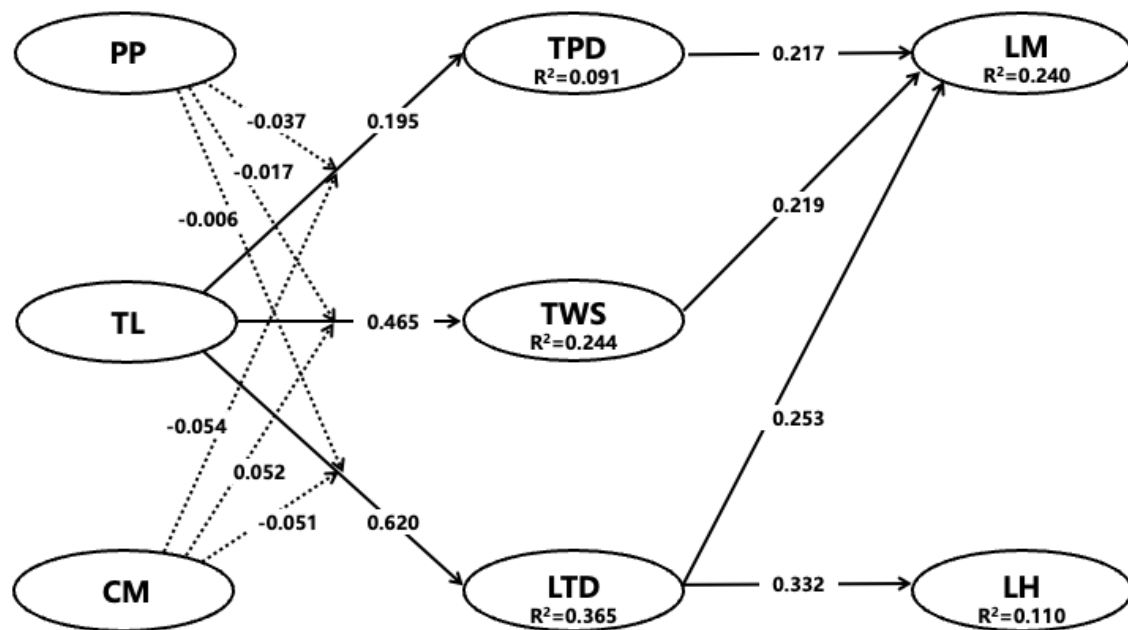


Figure 2. The results for the Teacher Learning Effect Model. Note: LH: Learning habit; LM: Learning motivation; LTD: Learning task design; TL: Teacher Learning; TPD: Teacher professional development; TWS: Teachers' working status; CM: Class manager; PP: Professional position.

Table 5. Moderating effects.

Path	Coefficient	<i>p</i> Values	<i>f</i> ²	Moderating Effect
CM × TL → TPD	−0.054	0.442	0.001	No
CM × TL → TWS	0.052	0.416	0.001	No
CM × TL → LTD	−0.051	0.375	0.001	No
PP × TL → TPD	−0.037	0.283	0.001	No
PP × TL → TWS	−0.017	0.532	0.000	No
PP × TL → LTD	−0.006	0.827	0.000	No
CM × TL → TPD → LM	−0.012	0.453	/	No
CM × TL → TWS → LM	0.011	0.426	/	No
CM × TL → LTD → LM	−0.013	0.367	/	No
CM × TL → LTD → LH	−0.017	0.365	/	No
PP × TL → TPD → LM	−0.008	0.288	/	No
PP × TL → TWS → LM	−0.004	0.538	/	No
PP × TL → LTD → LM	−0.001	0.829	/	No
PP × TL → LTD → LH	−0.002	0.829	/	No

Note: LH: Learning habit; LM: Learning motivation; LTD: Learning task design; TL: Teacher Learning; TPD: Teacher professional development; TWS: Teachers' working status; CM: Class manager; PP: Professional position.

5. Discussion

5.1. How Does Teacher Learning Directly Affect Professional Development

The results of our research indicate that teacher learning has a significant effect on teacher professional development ($p = 0.000 < 0.001$), which means that hypothesis H1a is valid. This research also confirms the view of Xie et al. (2021) that the teacher learning process, including formal professional development activities and informal practice learning, provides teachers with the necessary knowledge and skills to respond more effectively to challenges in teaching practice and promote student learning outcomes. Research has shown that when teachers engage in continuous learning and reflection in different practices, they are better able to exercise leadership, which is of great significance in promoting educational reform and improving teaching quality (Oppi & Eisenschmidt, 2022). In the context of Confucian culture in China, teachers' continuous learning and professional recognition are considered important factors in promoting successful education reform (Huang & Asghar, 2018). This result indicates that teacher learning plays an indispensable role in both individual professional development and systemic educational reform.

The results of our research indicate that teacher learning has a significant effect on teachers' working status ($p = 0.000 < 0.001$, $f^2 = 0.169 > 0.15$), which means that hypothesis H2a is valid. When teachers participate in meaningful learning experiences, they are more likely to develop innovative teaching methods, experience higher work satisfaction, and demonstrate stronger professional commitment (Opfer & Pedder, 2011). At the same time, professional communities for teachers, as a collaborative learning environment, can significantly enhance teachers' motivation and sense of belonging (Vangrieken et al., 2017). Obviously, participation in professional communities for teachers not only enriches teachers' teaching abilities but also alleviates feelings of isolation and professional burnout, thereby improving their overall working status. These studies demonstrate the key role of professional development through teacher learning in enhancing teacher performance, satisfaction, and overall professional effectiveness.

The results of our research indicate that teacher learning has a significant effect on learning task design ($p = 0.000 < 0.001$, $f^2 = 0.359 > 0.35$), which means that hypothesis H3a is valid. First, teacher learning enables them to design more targeted and effective learning tasks by promoting the improvement of professional knowledge and skills. Darling-Hammond et al. (2017) pointed out that effective teacher professional development can provide teachers with in-depth subject knowledge and teaching strategies, thereby enhancing their ability to design learning tasks. Second, teacher learning not only enhances teachers' teaching confidence but also changes their educational philosophy, thereby further promoting innovation in learning task design. As described by Borko (2004), professional development activities provide teachers with a platform for reflection and collaboration, enabling them to critically examine their teaching practices and learning tasks that are more suited to students' requirements. Obviously, this reflective practice helps teachers to more effectively combine teaching objectives and students' actual circumstances, thereby optimizing the content and form of learning tasks.

5.2. How Does Teacher Learning Indirectly Affect Student Learning

The results of our research indicate that teacher learning has a significant effect on learning motivation through teacher professional development ($p = 0.001 < 0.01$), which means that hypothesis H1 is valid.

Although our research did not directly demonstrate the effect of teacher learning on learning motivation, it provides indirect evidence of a positive effect via improvements in teachers' professional abilities and teaching levels. Avalos (2011) pointed out that this enhancement of professional capacity not only improved teachers' instructional practice but also affected how they motivated students through their teaching behavior. At the same time, Desimone (2009) pointed out that the effectiveness of teacher professional development is often closely related to the depth and continuity of teacher learning. Teacher learning helps them update teaching content and methods, thereby enhancing the quality of classroom interaction. This positive teaching atmosphere can stimulate students' interest and enhance their learning motivation. Hattie and Timperley (2007) further pointed out that the quality of teacher feedback is related to teacher learning. Through learning, teachers can better understand students' requirements and provide more accurate and constructive feedback. This high-quality feedback enhances students' self-efficacy and promotes their engagement and interest in learning (Zhang et al., 2025). Therefore, teacher learning indirectly enhances students' learning motivation by improving the quality of teacher feedback and classroom interaction.

The results of our research indicate that teacher learning has a significant effect on learning motivation through teachers' working status ($p = 0.000 < 0.001$), which means that hypothesis H2 is valid.

This is consistent with previous research findings that teachers' working status directly affects their teaching effectiveness and indirectly affects students' learning motivation. For example, Banerjee et al. (2017) found that there was a significant positive relationship between teachers' work satisfaction and students' learning motivation. Improving teachers' working status can help teachers maintain high enthusiasm for teaching and a positive classroom atmosphere, which is crucial for stimulating students' learning motivation. The process of teacher learning is often accompanied by an increase in self-efficacy, which in turn enhances teachers' working status and enables them to respond to challenges and pressures in teaching with a more positive attitude (Tschannen-Moran & Hoy, 2001). On the other hand, Schaufeli and Bakker (2004) noted that teachers' professional burnout is negatively associated with the quality of instruction and students' learning motivation. When teachers experience poor working conditions, they are more likely to suffer from excessive stress and burnout, which in turn undermines students' learning experiences. In this sense, teacher learning can play a crucial role in alleviating professional burnout, enabling teachers to regain enthusiasm for their work, thereby enhancing instructional quality and fostering students' learning motivation. Teacher learning, therefore, is not only an improvement of individual professional skills but also an indirect pathway to promoting students' motivation to learn (Schaufeli & Bakker, 2004). Such improvement not only advances teachers' instructional effectiveness but also fosters both

pedagogical and emotional support for students, thereby serving as a crucial mechanism for enhancing their learning motivation.

The results of our research indicate that teacher learning has a significant effect on learning motivation through learning task design ($p = 0.000 < 0.001$), which means that hypothesis H3 is valid.

Teacher learning not only has a direct effect on learning task design but also has an indirect effect on learning motivation. Therefore, we need to focus on how teachers can enhance their task design abilities through learning methods, thereby improving students' learning motivation. On the one hand, teacher learning provides both theoretical grounding and practical support for the design of student learning tasks. Through such learning, teachers can adjust task difficulty in line with students' learning characteristics and cognitive levels, thereby improving the effectiveness of these tasks. You et al. (2025) argue that by engaging in professional learning activities, teachers not only strengthen their instructional abilities but also build confidence in designing learning tasks. Such confidence allows them to develop tasks that align more closely with students' cognitive development, which in turn stimulates students' interest and fosters their self-directed motivation. On the other hand, Wijnia et al. (2024) further demonstrate that when teachers design learning tasks that are both challenging and attainable, students' learning motivation is significantly enhanced.

The results of our research indicate that teacher learning has a significant effect on learning habits through learning task design ($p = 0.000 < 0.001$), which means that hypothesis H4 is valid.

Teacher learning not only has a direct effect on learning task design but also has an indirect effect on learning habits. Specifically, this indirect effect is reflected in teachers' improvement of teaching strategies and task design skills through learning. Such improvement enables them to provide students with more structured and challenging learning tasks. This task design not only enhances students' engagement but also develops students' effective learning habits. For example, Borko (2004) argues that teacher learning enables teachers to design tasks that go beyond knowledge acquisition. These tasks also promote students' autonomous learning and self-management, which contribute to the development of effective learning habits. When students engage in tasks that are both challenging and aligned with their cognitive development, they are more likely to adopt positive learning behaviors, thereby enhancing the consistency and regularity of learning. Obviously, teacher learning can strengthen the cultivation of students' learning habits in task design. Hattie (2008) pointed out that learning task design is not only related to students' academic achievement, but also helps them develop effective learning habits by promoting reflection and self-regulation. Teacher learning encourages teachers to pay more attention to the long-term effects of task design, thereby helping students form long-term effective learning habits and self-management skills (Michalsky, 2024). These factors play a crucial role in the development of students' learning habits.

5.3. Do Professional Position and Identity of Class Manager Have a Moderating Effect

In PLS-SEM, moderating effects are tested by introducing interaction terms between independent variables and moderator variables. Moderating effects are crucial for understanding the boundary conditions of relationships between variables (Hair et al., 2014). The advantage of this method is that it handles non-normal data and complex models containing composite constructs (Shmueli et al., 2019).

The results of our research show that neither identity of class manager nor professional position has a moderating effect. Institutional roles and formal positions may not directly affect teachers' psychological or behavioral mechanisms in educational environments. For example, Mansfield et al. (2016) found that in research on values-based education, teacher attitudes and school culture (rather than administrative roles) were the main drivers of teaching outcomes. The results indicate that identity of class manager and professional position may lack direct psychological significance. Therefore, they cannot moderate the relationship chain of "teacher learning → teacher professional development."

Beyond internal psychological mechanisms, the absence of moderating effects may also reflect structural features of the Chinese middle-school system (Zhang et al., 2024). Because all schools operate under unified national curriculum standards (e.g., the 2022 Compulsory Education Curriculum Standards) and are evaluated through standardized, high-stakes metrics, the pedagogical environment is relatively homogeneous across teachers regardless of whether they serve as class managers or hold different professional positions. Such standardization can flatten between-teacher variation in instructional practice, leaving less room for institutional roles to condition how teacher learning translates into professional development, working status, or task design. This contextual reading complements the psychological explanation above and is consistent with the cross-cultural stability of the model noted in Section 5.4.

In addition, empirical research in the field of teacher learning has found many other factors that do not have a moderating effect, such as years of teaching experience (Hair et al., 2017), school type (public vs. private)

(Rigdon et al., 2017), gender (Venkatesh et al., 2003), and so on. These factors further support the rationality of the insignificant moderating effect in this research, emphasizing that institutional and demographic variables may not be able to reshape the core mechanisms of teacher learning.

However, some have pointed out that teacher learning is related to teacher professional development, teachers' working status, and learning task design, as well as the following factors: peer guidance and interdisciplinary cooperation (Shand & Goddard, 2025), expectations and motivation (Rubie-Davies et al., 2006), emotional resilience, and institutional trust (Mansfield et al., 2016).

5.4. Cross-Cultural Implications: Confucian Context versus Western Findings

A central question is whether the mechanisms identified here are distinctive to the Confucian cultural context or consistent with research conducted in Western systems. Importantly, the core pathway—teacher learning enhancing student outcomes via teachers' professional development, working status, and learning task design—aligns closely with studies from Europe and North America (e.g., Desimone, 2009; Guskey, 2002; Reeve & Cheon, 2021; Darling-Hammond et al., 2017), suggesting that this indirect mechanism is largely universal rather than culturally specific. At the same time, the Confucian heritage of China (Huang & Asghar, 2018) may accentuate certain links: a strong societal valuation of education and collectivist, school-based learning traditions (e.g., lesson-study and teaching-research groups) can amplify the translation of teacher learning into classroom practice. We therefore conclude that the fundamental mechanism generalizes across cultural contexts, while its strength and expression are shaped by local institutional and cultural conditions—a point underscored by the limited moderating role of formal teacher roles discussed below.

6. Conclusions

This research focuses on the mechanism of teacher learning's effect on teachers' self-development and student learning. First, regarding the research question, "How does teacher learning directly affect their own development?". The research found that teacher learning has a significant direct effect on teacher professional development, working status, and learning task design. Second, with respect to the research question, "How does teacher learning indirectly affect student learning?". The path analysis indicates that teacher learning indirectly promotes students' learning motivation and learning habits through three mediating paths. Among these, the mediating effect of learning task design is the strongest, highlighting its central role in transforming teacher learning outcomes into observable student behaviors. Finally, concerning the research question, "Do professional positions and identity of class manager have a moderating effect," neither class manager identity nor professional positions showed a significant moderating effect on the examined paths. This suggests that the core mechanism of teacher learning may depend more on teachers' internal psychological motivations, such as self-efficacy and work engagement, rather than differences in institutional roles.

This research reveals the multilevel effect pathways of teacher learning on student development from an empirical perspective. It provides important insights for education policymakers: First, priority should be given to supporting teacher professional development projects, ensuring the continuity of project implementation through institutional safeguards, and providing special training to strengthen learning task design. Second, attention should be paid to teacher learning and the establishment of cross-school learning communities to promote the sharing of teaching experience and professional growth. Finally, school-based teaching and research systems should be strengthened. Online teaching research and virtual teaching research should be promoted. Additionally, technology-supported teaching and research models should be developed to facilitate the integration of resources.

To make this recommendation concrete, two exemplars are worth highlighting. First, a school-based lesson-study cycle in which teachers collaboratively design, teach, observe, and revise a single learning task—explicitly targeting critical-thinking prompts and differentiated scaffolding (Roure & Pasco, 2018; Smets & Struyven, 2020)—directly strengthens the learning task design pathway identified as the strongest mediator in our model. Second, a short workshop module can train ordinary middle-school teachers to translate task-design principles into daily assignments by (a) formulating tasks that require students to think critically, (b) providing scaffolding matched to task difficulty, and (c) transferring design theory into practice—the exact behaviors captured by our learning task design items. Such practical, low-cost training turns the abstract recommendation into an actionable routine.

7. Limitations

First, the research sample is relatively concentrated and regional, and the results may not be consistent with those of other regions. It is recommended that future research be expanded to other cultural contexts and educational systems to explore the cross-cultural applicability of teacher learning. Second, longitudinal research designs can be

further integrated to reveal the long-term effects of teacher learning. Finally, the study emphasizes quantitative statistical analysis and cannot provide further insight into the inner thoughts and feelings of teachers during the learning process. Therefore, in future research, qualitative data can be obtained through semi-structured interviews with special samples randomly selected from the total sample for a more comprehensive and in-depth analysis.

Author Contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by J.Y., J.L., H.Z., X.L. and J.X. Project administration were performed by J.Y. Methodology and supervision were performed J.Y. and J.L. The first and final draft of the manuscript was written by J.Y. and J.L. All authors commented on previous versions of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding

2024 Provincial Teaching Reform Program for Graduate Students in the Second Batch of the 14th Five-Year Plan of Zhejiang Provincial Office of Education: Innovation and Practice of “Six Synergistic” Graduate Teaching Guided by Educator’s Spirit (Project No.: JGCG2024406); 2025 Wenzhou University Graduate Student Innovation Fund Project: Research on Enhancing Junior High School Students’ Deep Knowledge Learning Effect via Data-Driven Learning Decision (Project No.: 3162025004025); 2025 Wenzhou University Graduate Student Innovation Fund Project: Research on Improving Junior High School Students’ Practical Knowledge Through Human-Machine Collaborative Learning Decision (Project No.: 3162025003019); 2026 Zhejiang Provincial College Students’ Science and Technology Innovation Activity Plan (Xinmiao Talent Program) College Students’ Scientific and Technological Achievements Promotion Project Research on Improving Students’ Deep Knowledge Learning Effect Through Data-Driven Learning Decision-Making (Project No.: 2026R444B038).

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

The participants all agreed to take part in this study.

Data Availability Statement

The data and materials are available upon request to the corresponding author.

Conflicts of Interest

The authors declare no conflict of interest.

Use of AI and AI-Assisted Technologies

No AI tools were utilized for this paper.

References

- Avalos, B. (2011). Teacher professional development in teaching and teacher education over ten years. *Teaching and Teacher Education*, 27(1), 10–20. <https://doi.org/10.1016/j.tate.2010.08.007>.
- Banerjee, N., Stearns, E., Moller, S., & Mickelson, R. A. (2017). Teacher job satisfaction and student achievement: The roles of teacher professional community and teacher collaboration in schools. *American Journal of Education*, 123(2), 203–241. <https://doi.org/10.1086/689932>.
- Biau, D. J., Jolles, B. M., & Porcher, R. (2010). P value and the theory of hypothesis testing: An explanation for new researchers. *Clinical Orthopaedics and Related Research*, 468(3), 885–892. <https://doi.org/10.1007/s11999-009-1164-4>.
- Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher*, 33(8), 3–15. <https://doi.org/10.3102/0013189x033008003>.
- Clifton, J., & Jordan, K. (2019). Who is the hybrid teacher educator? Understanding professional identity in school–university partnership. In *Professionalism and teacher education: Voices from policy and practice* (pp. 71–90). Springer. https://doi.org/10.1007/978-981-13-7002-1_4.

- Darling-Hammond, L., Hyster, M. E., & Gardner, M. (2017). *Effective teacher professional development*. Learning Policy Institute. <https://doi.org/10.54300/122.311>.
- Datnow, A., & Hubbard, L. (2015). Teachers' use of assessment data to inform instruction: Lessons from the past and prospects for the future. *Teachers College Record: The Voice of Scholarship in Education*, 117(4), 1–26. <https://doi.org/10.1177/016146811511700408>.
- Day, C., Gu, Q., & Sammons, P. (2016). The impact of leadership on student outcomes: How successful school leaders use transformational and instructional strategies to make a difference. *Educational Administration Quarterly*, 52(2), 221–258. <https://doi.org/10.1177/0013161x15616863>.
- Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, 38(3), 181–199. <https://doi.org/10.3102/0013189x08331140>.
- Dijkstra, T. K. (2010). Latent variables and indices: Herman Wold's basic design and partial least squares. In *Handbook of partial least squares* (pp. 23–46). Springer. https://doi.org/10.1007/978-3-540-32827-8_2.
- Frenzel, A. C., Goetz, T., Lüdtke, O., Pekrun, R., & Sutton, R. E. (2009). Emotional transmission in the classroom: Exploring the relationship between teacher and student enjoyment. *Journal of Educational Psychology*, 101(3), 705–716. <https://doi.org/10.1037/a0014695>.
- Goldschmidt, P., & Phelps, G. (2010). Does teacher professional development affect content and pedagogical knowledge: How much and for how long? *Economics of Education Review*, 29(3), 432–439. <https://doi.org/10.1016/j.econedurev.2009.10.002>.
- Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching*, 8(3), 381–391. <https://doi.org/10.1080/135406002100000512>.
- Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2022). *A primer on partial least squares structural equation modeling (PLS-SEM)*. SAGE.
- Hair, J., Hollingsworth, C. L., Randolph, A. B., & Chong, A. Y. L. (2017). An updated and expanded assessment of PLS-SEM in information systems research. *Industrial Management & Data Systems*, 117(3), 442–458. <https://doi.org/10.1108/imds-04-2016-0130>.
- Hair, J. F., Jr., Sarstedt, M., Hopkins, L., & Kuppelwieser, V. G. (2014). Partial least squares structural equation modeling (PLS-SEM). *European Business Review*, 26(2), 106–121. <https://doi.org/10.1108/ebrev-10-2013-0128>.
- Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing Theory and Practice*, 19(2), 139–152. <https://doi.org/10.2753/mtpl069-6679190202>.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>.
- Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge. <https://doi.org/10.4324/9780203887332>.
- Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43(1), 115–135. <https://doi.org/10.1007/s11747-014-0403-8>.
- Holzberger, D., & Prestele, E. (2021). Teacher self-efficacy and self-reported cognitive activation and classroom management: A multilevel perspective on the role of school characteristics. *Learning and Instruction*, 76, 101513. <https://doi.org/10.1016/j.learninstruc.2021.101513>.
- Huang, Y. S., & Asghar, A. (2018). Science education reform in Confucian learning cultures: Teachers' perspectives on policy and practice in Taiwan. *Cultural Studies of Science Education*, 13(1), 101–131. <https://doi.org/10.1007/s11422-016-9762-4>.
- Kock, N. (2018). Minimum sample size estimation in PLS-SEM: An application in tourism and hospitality research. In *Applying partial least squares in tourism and hospitality research* (pp. 1–16). Emerald Publishing Limited. <https://doi.org/10.1108/978-1-78756-699-620181001>.
- Madigan, D. J., & Kim, L. E. (2021a). Does teacher burnout affect students? A systematic review of its association with academic achievement and student-reported outcomes. *International Journal of Educational Research*, 105, 101714. <https://doi.org/10.1016/j.ijer.2020.101714>.
- Madigan, D. J., & Kim, L. E. (2021b). Towards an understanding of teacher attrition: A meta-analysis of burnout, job satisfaction, and teachers' intentions to quit. *Teaching and Teacher Education*, 105, 103425. <https://doi.org/10.1016/j.tate.2021.103425>.
- Mansfield, C. F., Beltman, S., Broadley, T., & Weatherby-Fell, N. (2016). Building resilience in teacher education: An evidenced informed framework. *Teaching and Teacher Education*, 54, 77–87. <https://doi.org/10.1016/j.tate.2015.11.016>.

- Martínez, I. M., Youssef-Morgan, C. M., Chambel, M. J., & Marques-Pinto, A. (2019). Antecedents of academic performance of university students: Academic engagement and psychological capital resources. *Educational Psychology, 39*(8), 1047–1067. <https://doi.org/10.1080/01443410.2019.1623382>.
- Michalsky, T. (2024). Metacognitive scaffolding for preservice teachers' self-regulated design of higher order thinking tasks. *Heliyon, 10*(2), e24280. <https://doi.org/10.1016/j.heliyon.2024.e24280>.
- Opfer, V. D., & Pedder, D. (2011). Conceptualizing teacher professional learning. *Review of Educational Research, 81*(3), 376–407. <https://doi.org/10.3102/0034654311413609>.
- Oppi, P., & Eischmidt, E. (2022). Developing a professional learning community through teacher leadership: A case in one Estonian school. *Teaching and Teacher Education: Leadership and Professional Development, 1*, 100011. <https://doi.org/10.1016/j.tatelp.2022.100011>.
- Pérez-González, J. C., Filella, G., Soldevila, A., Faiad, Y., & Sanchez-Ruiz, M. J. (2022). Integrating self-regulated learning and individual differences in the prediction of university academic achievement across a three-year-long degree. *Metacognition and Learning, 17*(3), 1141–1165. <https://doi.org/10.1007/s11409-022-09315-w>.
- Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the motivated strategies for learning questionnaire (MSLQ). *Educational and Psychological Measurement, 53*(3), 801–813. <https://doi.org/10.1177/0013164493053003024>.
- Reeve, J., & Cheon, S. H. (2021). Autonomy-supportive teaching: Its malleability, benefits, and potential to improve educational practice. *Educational Psychologist, 56*(1), 54–77. <https://doi.org/10.1080/00461520.2020.1862657>.
- Rigdon, E. E., Sarstedt, M., & Ringle, C. M. (2017). On comparing results from CB-SEM and PLS-SEM: Five perspectives and five recommendations. *Marketing ZFP, 39*(3), 4–16. <https://doi.org/10.15358/0344-1369-2017-3-4>.
- Ringle, C. M., Wende, S., & Becker, J. M. (2022). *SmartPLS 4*. SmartPLS GmbH.
- Roure, C., & Pasco, D. (2018). The impact of learning task design on students' situational interest in physical education. *Journal of Teaching in Physical Education, 37*(1), 24–34. <https://doi.org/10.1123/jtpe.2017-0046>.
- Rubie-Davies, C., Hattie, J., & Hamilton, R. (2006). Expecting the best for students: Teacher expectations and academic outcomes. *British Journal of Educational Psychology, 76*(3), 429–444. <https://doi.org/10.1348/000709905x53589>.
- Ruthven, K., Hennessy, S., & Brindley, S. (2004). Teacher representations of the successful use of computer-based tools and resources in secondary-school English, mathematics and science. *Teaching and Teacher Education, 20*(3), 259–275. <https://doi.org/10.1016/j.tate.2004.02.002>.
- Sarstedt, M., Ringle, C. M., & Hair, J. F. (2022). Partial least squares structural equation modeling. In *Handbook of market research* (pp. 587–632). Springer. https://doi.org/10.1007/978-3-319-57413-4_15.
- Schaufeli, W. B., & Bakker, A. B. (2004). Job demands, job resources, and their relationship with burnout and engagement: A multi-sample study. *Journal of Organizational Behavior, 25*(3), 293–315. <https://doi.org/10.1002/job.248>.
- Schaufeli, W. B., Salanova, M., González-romá, V., & Bakker, A. B. (2002). The measurement of engagement and burnout: A two sample confirmatory factor analytic approach. *Journal of Happiness Studies, 3*(1), 71–92. <https://doi.org/10.1023/a:1015630930326>.
- Shand, R., & Goddard, R. D. (2025). The relationship between teacher collaboration and instructional practices, instructional climate, and social relations. *Educational Policy, 39*(4), 912–937. <https://doi.org/10.1177/08959048241278290>.
- Shi, Q., & Leuwerke, W. C. (2010). Examination of Chinese homeroom teachers' performance of professional school counselors' activities. *Asia Pacific Education Review, 11*(4), 515–523. <https://doi.org/10.1007/s12564-010-9099-8>.
- Shmueli, G., Sarstedt, M., Hair, J. F., Cheah, J. H., Ting, H., Vaithilingam, S., & Ringle, C. M. (2019). Predictive model assessment in PLS-SEM: Guidelines for using PLSpredict. *European Journal of Marketing, 53*(11), 2322–2347. <https://doi.org/10.1108/ejm-02-2019-0189>.
- Smets, W., & Struyven, K. (2020). A teachers' professional development programme to implement differentiated instruction in secondary education: How far do teachers reach? *Cogent Education, 7*(1), 1742273. <https://doi.org/10.1080/2331186x.2020.1742273>.
- Smit, R., & Humpert, W. (2012). Differentiated instruction in small schools. *Teaching and Teacher Education, 28*(8), 1152–1162. <https://doi.org/10.1016/j.tate.2012.07.003>.
- Toropova, A., Myrberg, E., & Johansson, S. (2021). Teacher job satisfaction: The importance of school working conditions and teacher characteristics. *Educational Review, 73*(1), 71–97. <https://doi.org/10.1080/00131911.2019.1705247>.
- Tschannen-Moran, M., & Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education, 17*(7), 783–805. [https://doi.org/10.1016/s0742-051x\(01\)00036-1](https://doi.org/10.1016/s0742-051x(01)00036-1).
- Urhahne, D., & Wijnia, L. (2023). Theories of motivation in education: An integrative framework. *Educational Psychology Review, 35*(2), 45. <https://doi.org/10.1007/s10648-023-09767-9>.
- Vangrieken, K., Meredith, C., Packer, T., & Kyndt, E. (2017). Teacher communities as a context for professional development: A systematic review. *Teaching and Teacher Education, 61*, 47–59. <https://doi.org/10.1016/j.tate.2016.10.001>.

- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified View. *MIS Quarterly*, 27(3), 425–478. <https://doi.org/10.2307/30036540>.
- Vescio, V., Ross, D., & Adams, A. (2008). A review of research on the impact of professional learning communities on teaching practice and student learning. *Teaching and Teacher Education*, 24(1), 80–91. <https://doi.org/10.1016/j.tate.2007.01.004>.
- Wake, G., Swan, M., & Foster, C. (2016). Professional learning through the collaborative design of problem-solving lessons. *Journal of Mathematics Teacher Education*, 19, 243–260. <https://doi.org/10.1007/s10857-015-9332-9>.
- Wasserstein, R. L., & Lazar, N. A. (2016). The ASA statement on *p*-values: Context, process, and purpose. *The American Statistician*, 70(2), 129–133. <https://doi.org/10.1080/00031305.2016.1154108>.
- Wettstein, A., Ramseier, E., & Scherzinger, M. (2021). Class- and subject teachers' self-efficacy and emotional stability and students' perceptions of the teacher–student relationship, classroom management, and classroom disruptions. *BMC Psychology*, 9(1), 103. <https://doi.org/10.1186/s40359-021-00606-6>.
- Wijnia, L., Noordzij, G., Arends, L. R., Rikers, R. M. J. P., & Loyens, S. M. M. (2024). The effects of problem-based, project-based, and case-based learning on students' motivation: A meta-analysis. *Educational Psychology Review*, 36, 29. <https://doi.org/10.1007/s10648-024-09864-3>.
- Wu, J., Zhong, Y., & Qi, S. (2022). *Intrinsic motivation, need for cognition, grit, growth mindset and academic achievement in high school students: Latent profiles and its predictive effects*. PsyArXiv. <https://doi.org/10.31234/osf.io/mheg7>.
- Xie, C., Song, P., & Hu, H. (2021). Measuring teacher leadership in different domains of practice: Development and validation of the Teacher Leadership Scale. *The Asia-Pacific Education Researcher*, 30(5), 409–419. <https://doi.org/10.1007/s40299-020-00527-9>.
- Yan, J. J., Luo, X. N., Zeng, Q. F., Zhang, H. L., & Wu, J. N. (2025). Investigation and strategy research on the causes of middle school students' learning difficulties in the context of the leading country in education. *Journal of Educational Technology and Innovation*, 7(2), 1–10. <https://doi.org/10.61414/0shg4n49>.
- You, H., Park, S., Hong, M., & Warren, A. (2025). Unveiling effectiveness: A meta-analysis of professional development programs in science education. *Journal of Research in Science Teaching*, 62(4), 971–1005. <https://doi.org/10.1002/tea.21985>.
- Zhang, H. L., Yan, J. J., Liu, J., Zou, Z. C., Zeng, Q. F., Jin, Y. L., & Wu, J. N. (2025). An empirical study on the reflective feedback model based on pedagogical agents to improve students' problem-solving ability. *Journal of Educational Technology and Innovation*, 7(1), 66–75. <https://doi.org/10.61414/tykx7317>.
- Zhang, H., Lyu, X., & Qiu, Y. (2024). Effective teacher professional development and its influencing factors: A cross-national comparison of the United States, China, Finland, and Singapore. *Journal of Teacher Education*, 75(5), 502–515. <https://doi.org/10.1177/00224871241254779>.
- Zheng, L., Qi, X., & Zhang, C. (2023). Can improvements in teacher quality reduce the cognitive gap between urban and rural students in China? *International Journal of Educational Development*, 100, 102781. <https://doi.org/10.1016/j.ijedudev.2023.102781>.
- Zhou, X., Shu, L., Xu, Z., & Padrón, Y. (2023). The effect of professional development on in-service STEM teachers' self-efficacy: A meta-analysis of experimental studies. *International Journal of STEM Education*, 10, 37. <https://doi.org/10.1186/s40594-023-00422-x>.
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, 41(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2.

Article

Do AI Writing Tools Improve L2 Writing Performance? A Meta-Analysis of Experimental Evidence, Moderators, and Methodological Quality

Jinming Du

Languages and Cultures Department, University of Otago, Dunedin 9016, New Zealand; duji6714@student.otago.ac.nz

How To Cite: Du, J. (2026). Do AI Writing Tools Improve L2 Writing Performance? A Meta-Analysis of Experimental Evidence, Moderators, and Methodological Quality. *Journal of Educational Technology and Innovation*, 8(3), 80–103. <https://doi.org/10.61414/awr9xk46>

Received: 21 May 2026

Revised: 24 August 2026

Accepted: 15 September 2026

Published: 30 September 2026

Abstract: Artificial intelligence (AI) writing tools—including large language model (LLM)-based systems, automated writing evaluation (AWE) platforms, and AI-powered writing assistants—have become an increasingly prominent feature of second language (L2) writing pedagogy. Yet whether these tools reliably improve L2 writing performance, and under what instructional conditions, remains an open question. This study reports a systematic meta-analysis of experimental and quasi-experimental evidence on the effects of AI writing tools on L2 writing performance. A multi-database search of Scopus, Web of Science, ERIC, and ProQuest Dissertations & Theses, supplemented by forward citation tracing and targeted hand-searching, identified 34 eligible focal effect sizes from studies published between 2021 and 2026. Effect sizes were expressed as Hedges' g . A verified subset of 22 effects with directly extractable statistics was estimated using restricted maximum-likelihood (REML) with Knapp–Hartung adjustment; an expanded model ($k = 34$) incorporating conversion-based effects served as a robustness analysis. The verified model produced a pooled effect of $g = 0.714$, 95% CI [0.592, 0.836], with low observed heterogeneity ($I^2 = 24.3\%$). Exploratory moderator analyses indicated that multidimensional feedback designs, sustained interventions, particularly those lasting eight weeks or longer, were associated with larger effects, while single-session and lower-order-only feedback showed smaller gains. Methodological quality was generally adequate but uneven: many studies lacked delayed transfer measures and transparent effect-size reporting. These findings position AI writing tools as pedagogically contingent rather than inherently effective—their educational value depends critically on how feedback literacy, teacher mediation, and instructional design shape opportunities for noticing, revision, and transferable writing development.

Keywords: AI writing tools; second language writing; ChatGPT; generative AI; meta-analysis; automated writing evaluation.

1. Introduction

Artificial intelligence (AI) writing tools are now central to computer-assisted language learning (CALL), yet their effects on L2 writing development remain theoretically and empirically contested. Rather than treating them simply as automated correctors or text generators, this article conceptualises AI writing tools as CALL feedback systems that may reshape learner engagement, revision practices, teacher mediation, and human–technology agency. It examines whether controlled intervention evidence supports claims that these tools improve L2 writing performance. Their appeal is especially strong in L2 writing, where learners must coordinate linguistic, rhetorical, cognitive, and affective demands, including grammar, vocabulary, organisation, coherence, content development,



Copyright: © 2026 by the authors. This is an open access article under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Publisher's Note: Scilight stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

audience awareness, and disciplinary voice (Hyland, 2019; Ferris, 2011). Because teachers often have limited capacity to provide timely and individualised feedback, AI writing tools are increasingly positioned as scalable feedback resources that can supplement teacher response, expand revision opportunities, and offer immediate language support (Godwin-Jones, 2022; Mahapatra, 2024; Wiboolyasarin et al., 2024).

1.1. AI Writing Tools and L2 Writing Development

AI writing tools may support L2 development through several routes: rapid corrective feedback on grammar and vocabulary helps learners address surface concerns (Alsaweed & Aljebreen, 2024; Ding et al., 2026; Du & Chen, 2026); generative AI extends this to higher-order concerns such as idea development and rhetorical refinement (Su et al., 2023; Wang & Ren, 2024; Wiboolyasarin et al., 2024); and iterative dialogue allows learners to ask follow-up questions and refine drafts progressively (Koltovskaia et al., 2024; Yan & Zhang, 2024). When feedback is used reflectively, AI-supported environments also show gains in motivation and self-regulated learning (Du et al., 2026; Jin et al., 2025; Liu et al., 2026).

At the same time, generative AI can produce fluent but inaccurate, generic, or rhetorically inappropriate text, and learners may accept suggestions without sufficient evaluation (Barrot, 2023; Cingillioglu, 2023; Lingard, 2023). L2 writers do not always possess the metacognitive or critical AI literacy needed to regulate AI feedback effectively (Koltovskaia et al., 2024; Ou et al., 2024; Wang & Wang, 2025; Yan & Zhang, 2024), so effectiveness likely depends on pedagogical design, feedback focus, learner level, task type, and teacher or peer mediation. At the broader educational-ecology level, recent evidence on acceptance of large language models in education also shows that perceived usefulness coexists with concerns about data privacy, over-reliance, bias, and unequal AI knowledge (Stroud & Du, 2025; Wan et al., 2026), reinforcing the need to treat AI writing support as a sociotechnical rather than purely technical intervention.

1.2. Fragmented Empirical Evidence and the Need for Quantitative Synthesis

A growing body of intervention studies has examined AI writing tools' effects on L2 writing performance: chatbot-based practice improved elementary EFL learners' writing (Kwon et al., 2023); LLM-supported strategy instruction improved EFL writing, self-regulated learning, and motivation (Liu et al., 2026); ChatGPT-based formative feedback was associated with improved ESL/EFL academic writing (Ghafouri et al., 2024; Mahapatra, 2024; Tsai et al., 2024); and AI-supported collaborative writing environments showed positive effects on specific writing aspects (Hu et al., 2025; Wiboolyasarin et al., 2024).

Despite this accumulation of evidence, the field remains fragmented. Studies use different labels—ChatGPT, LLM-supported instruction, AI chatbots, AWE systems, automated feedback systems—making cross-study comparisons difficult (Godwin-Jones, 2022; Shi et al., 2026), and measured outcomes vary from holistic writing scores to accuracy, complexity, grammar, and coherence across designs ranging from RCTs to small-scale classroom studies. These variations make it hard to determine from individual studies whether AI writing tools produce reliable improvements and under what conditions.

Existing reviews have mapped AI-assisted writing opportunities but are predominantly narrative, scoping, or perception-focused. The present study synthesises eligible experimental evidence quantitatively, reports study-disposition decisions transparently, and distinguishes the primary model from records retained for background only.

1.3. The Present Study

The present study conducts a meta-analysis of experimental, quasi-experimental, and intervention-based studies examining the effects of AI writing tools on L2 writing performance. The study is positioned as a CALL-focused quantitative synthesis rather than a broad review of all AI-in-writing scholarship. Its contribution is threefold: it estimates the overall effect of AI writing tools on L2 writing performance, examines moderator patterns that matter for CALL design, and evaluates the methodological maturity of this rapidly developing intervention literature. Three research questions guided the study:

RQ1. What is the overall effect of AI writing tools on L2 writing performance in controlled intervention studies?

RQ2. Do effects vary by AI architecture, feedback focus, outcome type, learner level, or intervention duration?

RQ3. What is the methodological quality of existing experimental evidence on AI writing tools and L2 writing performance, and what does the observed evidence map suggest for future CALL research design?

2. Literature Review

2.1. L2 Writing as a Cognitively and Pedagogically Demanding Activity

Second language writing is a complex cognitive, linguistic, and socio-rhetorical activity (Ferris, 2011; Hyland, 2019), and L2 learners face additional constraints from limited lexical resources, incomplete grammatical control, and restricted feedback access. Teachers frequently cannot provide detailed formative feedback to large classes, creating a strong rationale for technology-supported writing assistance. Research on L2 writing feedback has shifted toward viewing it as a process requiring learner engagement, interpretation, and revision (Hyland & Hyland, 2006): if AI feedback merely produces a polished text without learner understanding, it may improve the immediate product while contributing little to transferable competence, but when learners compare alternatives, diagnose weaknesses, and revise deliberately, AI tools may function as scaffolds for language development and metacognitive regulation (Koltovskaia et al., 2024; Yan & Zhang, 2024).

2.2. From Automated Writing Evaluation to Generative AI Writing Support

AI-supported writing assistance did not begin with ChatGPT. Earlier technologies such as AWE, automated written corrective feedback, and grammar checkers supported L2 writing through automated scoring, error detection, and revision prompts (Godwin-Jones, 2022; Nazari et al., 2021; Xu & Zhang, 2022). These constrained systems were strongest in lower-order concerns such as grammar and sentence-level accuracy, but were criticised for limited contextual sensitivity, formulaic feedback, and insufficient attention to higher-order concerns such as argumentation and disciplinary voice.

The expanded evidence map reinforces this pre-ChatGPT lineage: AWE and AI-writing-assistant studies (Fan, 2023; Nazari et al., 2021; Wei et al., 2023; Xu & Zhang, 2022) show that research has long addressed the same design questions now raised by LLMs: feedback timing, learner uptake, teacher mediation, and whether automated feedback supports durable learning.

Generative AI and LLMs have expanded AI-assisted writing's functional scope: such tools produce explanations, propose outlines, reformulate sentences, and engage in iterative dialogue, so empirical studies have examined ChatGPT and related systems as feedback providers, brainstorming partners, and self-regulated learning supports (Chen & Gong, 2025; Ghafouri et al., 2024; Liu et al., 2026; Mahapatra, 2024; Shi et al., 2026). This broader functionality also complicates evaluation because AI tools may simultaneously influence text quality, writing process, learner agency, and academic integrity.

The newer generative-AI literature broadens this trajectory from automated correction to interactive feedback: recent studies compare ChatGPT/GenAI feedback against teacher feedback, AWE feedback, restricted-access conditions, hybrid teacher-AI feedback, and AI-supported peer feedback (Chan et al., 2024; Teng et al., 2026; Escalante et al., 2023; Guo et al., 2024; Hong & Shin, 2025; Luo et al., 2025; Wang & Ren, 2024; Zou et al., 2026), shifting the field from asking whether automated feedback can identify errors to asking how learners orchestrate tool suggestions, teacher guidance, and their own rhetorical intentions.

2.3. Empirical Evidence on AI Tools and L2 Writing Performance

Experimental and quasi-experimental studies increasingly report positive effects: chatbot-based practice improved young learners' L2 writing (Kwon et al., 2023); LLM-supported strategy instruction enhanced EFL writing, self-regulated learning, and motivation (Liu et al., 2026); and ChatGPT-supported interventions improved overall writing quality, grammar, organisation, and coherence (Chen & Gong, 2025; Ghafouri et al., 2024; Hongxia & Razali, 2025; Mahapatra, 2024; Shi et al., 2026; Wiboolyasarin et al., 2024).

Across the expanded corpus, several partially overlapping research traditions emerge: Grammarly/AWE studies focus on accuracy, grammar, or complexity; ChatGPT/LLM studies examine holistic quality, revision quality, or self-regulated learning; hybrid studies ask whether AI feedback should replace or supplement teacher feedback. A positive grammar-accuracy effect in a Grammarly study is not equivalent to a positive argument-organisation effect in a ChatGPT study: effects are mediated by what kind of writing problem the tool is designed to address.

The evidence is also methodologically heterogeneous: studies differ in tool type, learner population, outcome measurement, and comparison condition, making it difficult to infer overall effectiveness from individual studies alone.

2.4. Potential Moderators in AI-Assisted L2 Writing Effects

The effects of AI writing tools are unlikely to be uniform because technologies differ in feedback architecture, interactional capacity, and learner control (Zimmerman, 2000). Earlier AWE and automated corrective feedback

systems typically provide constrained diagnostic or error-level feedback, whereas generative AI and LLM-based systems offer dialogic explanations, revision alternatives, and metalinguistic justifications—changing not only the feedback provided but how learners question, negotiate, and regulate it.

Several moderators therefore map directly onto CALL design principles: tool architecture distinguishes constrained AWE from dialogic LLM systems (Godwin-Jones, 2022; Nazari et al., 2021; Xu & Zhang, 2022); feedback source distinguishes AI-only, teacher-only, peer-supported, and hybrid teacher–AI designs; outcome focus distinguishes linguistic accuracy, syntactic complexity, coherence, rhetorical quality, revision quality, self-efficacy, and engagement; and duration distinguishes one-shot revision tasks from sustained writing instruction. Together, these dimensions capture learner–tool interaction, teacher mediation, feedback-cycle depth, and opportunities for revision and transfer. Feedback focus may also shape outcomes. Lower-order feedback on grammar and mechanics, higher-order feedback on content and organisation, and multidimensional feedback are likely to support different aspects of writing development (Ferris, 2011; Hyland, 2019). Outcome type is equally important: immediate holistic scores may be more responsive to AI-assisted revision than delayed measures of independent accuracy, complexity, or rhetorical organisation. Learner level and duration add further variation, as novice writers may benefit from scaffolding but remain more vulnerable to uncritical uptake, while longer interventions may allow feedback to be internalised rather than merely applied (Schmidt, 1990). These moderators link tool affordances, feedback-mediated noticing, self-regulated revision, and writing performance.

2.5. Risks, Methodological Gaps, and Rationale for the Present Meta-Analysis

Scholars caution that AI writing tools introduce pedagogical and ethical risks: generative AI may produce inaccurate content, learners may over-rely on suggestions, and extensive AI-mediated revision may obscure authorship (Barrot, 2023; Cingillioglu, 2023; Du & Kohnke, 2026; Lingard, 2023; Tsai et al., 2024). L2 writers' engagement varies substantially by proficiency and critical AI literacy (Koltovskaia et al., 2024; Ou et al., 2024; Wang & Wang, 2025; Yan, 2025; Yan & Zhang, 2024), so pedagogical value depends on how learners interpret, regulate, and learn from AI assistance.

The expanded corpus further shows that risks span both AWE and generative AI contexts. In AWE/Grammarly studies, risks include over-attention to surface form and uncritical acceptance of suggestions; in ChatGPT/GenAI studies, risks extend to rhetorical homogenisation, loss of authorial voice, and untraceable revision histories. The underlying question—whether AI improves the submitted text or the writer—cannot be answered without designs separating assisted performance from independent competence.

Existing reviews have mapped opportunities and challenges of AI tools in L2 writing, but many remain conceptual, narrative, or scoping. A quantitative synthesis is therefore useful for estimating the direction and magnitude of effects while exposing methodological limitations; the present study synthesises eligible experimental evidence and reports study-disposition decisions transparently, distinguishing the primary quantitative model from records retained only for background interpretation.

3. Methods

3.1. Review Design

This study adopted a systematic search and meta-analytic design to synthesise empirical evidence on the effectiveness of AI writing tools for L2 writing, estimating a pooled effect size across experimental and quasi-experimental studies. Given the limited number of eligible studies, the analysis was conducted as a random-effects meta-analysis with sensitivity and descriptive moderator analyses rather than a full multilevel model.

The focal construct was AI-supported L2 writing performance: digital systems using AI to support planning, drafting, feedback, revision, evaluation, or instruction, including generative AI systems such as ChatGPT and large language models (LLM)-supported systems, AI chatbots, AWE systems, automated written corrective feedback systems, and AI-powered digital writing assistants, reflecting the literature's overlapping traditions of AWE research, chatbot-assisted writing, generative AI feedback, and LLM-supported strategy instruction (Barrot, 2023; Godwin-Jones, 2022; Su et al., 2023; Warschauer et al., 2023).

The synthesis prioritised studies estimating changes in actual writing performance rather than perceptions alone, focusing on experimental, quasi-experimental, intervention, counter-balanced, and closely comparable pre-post designs. Studies reporting qualitative engagement, learner perceptions, AI literacy, feedback-seeking behaviour, or AI-detection performance were retained for contextual interpretation but excluded from the primary pooled effect unless they reported an extractable writing-performance effect.

3.2. Study Identification and Screening

A systematic database search (see Figure 1) was conducted in Scopus, Web of Science, ERIC, and ProQuest Dissertations & Theses, with the final search completed in December 2025. These databases were selected for their broad coverage of applied linguistics, CALL, educational technology, and education research. The initial search returned 1595 records (Scopus $n = 801$; Web of Science $n = 590$; ERIC $n = 120$; ProQuest $n = 76$; Call for Unpublished Work $n = 8$). The search string combined terms for AI writing tools, generative AI, automated writing evaluation, ChatGPT, L2 writing, ESL/EFL, intervention, experimental and quasi-experimental designs, and writing performance. The core Boolean structure was: (“AI writing tool*” OR “artificial intelligence” OR “generative AI” OR “ChatGPT” OR “large language model*” OR “automated writing evaluation” OR AWE OR “AI feedback” OR “AI writing assistant*”) AND (“second language writing” OR “L2 writing” OR ESL OR EFL OR “foreign language writing”) AND (intervention OR experiment* OR “quasi-experiment*” OR “control group” OR “writing performance” OR “writing quality” OR “writing score*”). After removing 982 duplicates, 613 unique records were screened by title and abstract. Of the 613 screened records, 541 were excluded at title-and-abstract screening because they were not focused on L2 writing, did not examine AI writing tools, were conceptual/review articles, or did not report an intervention or writing-performance outcome. The remaining 72 reports were sought for retrieval. A further 21 reports were excluded for not focusing on L2 writing or AI writing tools, 10 for lacking an intervention or comparison design, and 7 for not reporting extractable writing-performance data, leaving 34 reports from the database search. Reference searching of included publications identified 5 additional reports, and supplementary searching identified 29 further reports; all 34 of these passed full-text screening with no exclusions. Combining both pathways, 68 reports met full eligibility criteria; after removing 34 duplicate or overlapping datasets reported across multiple publications, 34 independent focal effects were retained for the quantitative synthesis.

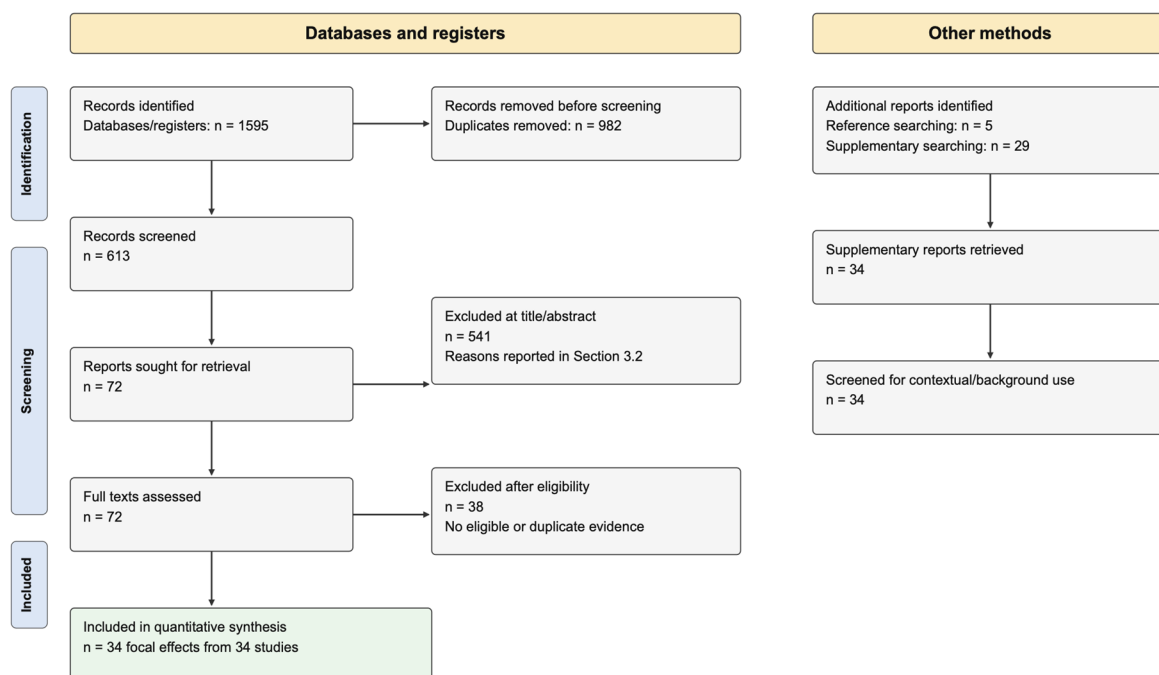


Figure 1. PRISMA flow diagram for study identification, screening, and inclusion.

3.3. Inclusion and Exclusion Criteria

Studies were included in the primary meta-analysis if they met four criteria. The study had to investigate an AI writing tool used in a second or foreign language writing context. This criterion allowed the inclusion of tools such as ChatGPT, AI KAKU, LLM-supported CALLA instruction, generative AI conversational agents, chatbot-based writing practice systems, and AWE or AI-powered feedback systems.

The study also had to report at least one writing-performance outcome (overall writing score, writing quality, accuracy, grammar, vocabulary, complexity, fluency, coherence, organisation, or other rubric-based dimension). Studies focusing solely on motivation, engagement, self-efficacy, or learner attitudes were excluded from the primary model unless they also reported a quantifiable writing-performance measure.

A third requirement was an experimental, quasi-experimental, intervention, counter-balanced, or comparable pre-post design. Studies with AI-supported and non-AI comparison groups were prioritised; studies with paired before-and-after comparisons or group-level outcomes were coded but flagged for sensitivity analysis because their effect sizes are not fully equivalent to independent-group standardised mean differences (Borenstein et al., 2009; Morris, 2008).

Finally, the study had to report sufficient statistical information to calculate or approximate an effect size. Eligible statistics included sample sizes, means, SDs, $t/F/p$ values, partial eta-squared, Mann-Whitney statistics, or author-reported Cohen's d ; studies were excluded when no defensible conversion was possible.

Studies were excluded if they were conceptual articles, reviews, purely qualitative studies, perception-only surveys, AI-detection studies, or studies in which all groups used AI with no AI-versus-comparison contrast (e.g., Koltovskaia et al., 2024; Ou et al., 2024; Wang & Wang, 2025; Yan & Zhang, 2024; Yang & Lin, 2025; Zou & Huang, 2024); these remained important for interpreting the field but did not provide primary writing-performance effects.

3.4. Coding Procedure

A structured coding sheet extracted bibliographic, methodological, intervention, outcome, and statistical information from each study, with each study receiving a unique study ID and each extractable outcome or comparison an effect ID. Extracted fields included design, participant population, educational level, AI tool, intervention description, comparison condition, outcome construct, duration, sample size, means, standard deviations, test statistics, computed effect size, standard error, quality score, and sensitivity flag.

When studies reported multiple outcomes, time points, or comparison groups, all potentially relevant effects were initially coded, then one focal effect was selected per study following an a priori hierarchy: (a) the most direct AI-versus-non-AI comparison; (b) immediate post-test over delayed outcomes; (c) holistic over narrowly defined subcomponents; (d) independent-group over paired effects; and (e) for multi-arm studies, the contrast most aligned with the research question. Thus for Shi et al. (2026), the ChatGPT-versus-control contrast was selected as focal because this review centres on generative AI, with the AWE contrast retained as supplementary evidence.

This rule avoided cherry-picking by magnitude; selected effects are rule-based focal estimates, not the largest or most favourable available.

3.5. Moderator Coding

Six moderators were coded to capture theoretically meaningful variation in how AI writing tools operate as affordances. This coding scheme (Table 1) is consistent with the tool-architecture and feedback-focus distinctions established in Sections 2.2 and 2.4 (Ferris, 2011; Godwin-Jones, 2022; Hyland, 2019; Nazari et al., 2021; Xu & Zhang, 2022). *AI_tool_type* (the user-facing implementation, e.g., chatbot, AWE system) and *AI_generation_type* (the underlying architecture, e.g., generative/LLM-based versus constrained automated systems) were defined at different analytic levels; a system could be a conversational agent at the tool-type level and generative AI at the architecture level. Feedback focus was recoded into multidimensional versus single-dimensional linguistic suggestion, as the small dataset did not support fine-grained comparisons across many one-study categories.

Given the relatively small number of focal effects, moderator analyses were treated as exploratory and descriptive rather than confirmatory, since small subgroup sizes can produce unstable estimates and low statistical power (Borenstein et al., 2009; Harrer et al., 2021).

Key coding decisions followed original-study descriptions: Hu et al. (2025) were coded as a GenAI conversational agent (tool type) and generative AI (generation type); Liu et al. (2026) as an LLM-supported writing system and generative AI; Kwon et al. (2023) as a chatbot/conversational agent but traditional AI generation type because the Dialogflow implementation functioned as a designed conversational practice system rather than an open-ended LLM generator. These decisions are operational rather than ontological and should be revisited as newer platforms integrate LLM capabilities.

3.6. Methodological Quality Assessment

Claims about AI-supported writing gains also depend on whether studies provide enough methodological detail for differences to be interpreted with confidence. Quality was therefore assessed using a nine-point checklist informed by the MMAT (Hong et al., 2018), the JBI critical appraisal approach (Munn et al., 2020), risk-of-bias principles (Higgins et al., 2019; Sterne et al., 2019), and applied linguistics research synthesis recommendations (Norris & Ortega, 2006; Plonsky, 2014).

The checklist assessed nine criteria: (1) participant description, (2) comparison condition, (3) pre-test/baseline equivalence, (4) intervention replicability, (5) writing-performance measures specified, (6) assessment reliability, (7) statistics sufficient for effect-size extraction, (8) attrition addressed, and (9) delayed or transfer evidence. Each item was coded as present, partially present, or absent.

Each criterion was scored 0 or 1 (maximum 9). Quality scores were used descriptively rather than as automatic exclusion criteria; studies with lower comparability or less direct effect-size extraction were flagged for sensitivity analysis, recognising that this emerging field is methodologically innovative but not yet uniformly reported.

Table 1. Moderator coding scheme.

Moderator	Categories Coded	Purpose and Operational Meaning
AI tool type	AWE; AI grammar/corrective feedback tool; AI chatbot/conversational agent; generative AI/LLM tool; intelligent writing assistant/AI tutor; mixed/unclear	Indexes the user-facing pedagogical implementation or functional format of the writing support, not necessarily the underlying AI architecture.
AI generation type	Generative AI; traditional AI/automated feedback system; mixed; unclear	Indexes the underlying generation architecture as reported by the authors; systems were coded as generative only when LLM/open-ended language generation was part of the implemented intervention.
Feedback focus	Grammar/mechanics; vocabulary/lexical; content/ideas; organisation/coherence; holistic; multiple; unclear	Identifies whether the tool primarily targets lower-order concerns, higher-order concerns, or multiple writing dimensions.
Writing outcome type	Overall writing score; accuracy; complexity; fluency; lexical quality; coherence/organisation; content quality; other	Classifies the writing construct used to compute the focal effect size.
Learner level	Primary; secondary; undergraduate; postgraduate; adult/other; mixed/unclear	Captures developmental and educational-level variation across samples.
Intervention duration	One-shot/single session; 2–4 weeks; 5–8 weeks; >8 weeks; unclear	Represents the intensity and temporal depth of the intervention.

3.7. Effect-Size Calculation

For Chan, Lo, and Wong (2024), the effect size was calculated using the sample size implied by the focal t-test degrees of freedom ($df = 192$, implying $n \approx 194$), rather than the total reported sample of 918, because the focal contrast appeared to involve a subset of participants. The study was retained with a sensitivity flag; omitting it did not alter the direction of the pooled result ($g = 0.751$, 95% CI [0.653, 0.849]).

The primary effect-size metric was Hedges' g , which estimates the standardised mean difference while correcting for small-sample bias (Hedges & Olkin, 1985). For independent-group studies reporting post-test means, standard deviations, and sample sizes, Hedges' g was calculated from the difference between AI-supported and comparison conditions. When baseline differences were addressed using ANCOVA or adjusted means, the adjusted statistic was preferred where available.

When studies reported t values, F values, partial eta-squared, Mann-Whitney statistics, paired t tests, or author-reported Cohen's d , effect sizes were converted into approximate Hedges' g using standard transformations. Converted effects were retained only when the source statistic reflected a writing-performance contrast relevant to the focal research question. Because such conversions can introduce additional uncertainty, all conversion-based effects were flagged for sensitivity analysis.

The largest effect in the dataset, Ghafouri et al. (2024), was derived from an ANCOVA group effect rather than raw post-test standard deviations. The study reported $F(1, 45) = 44.055$ for the post-test writing outcome after controlling for pre-test performance. The conversion used $t = \sqrt{F} = 6.637$ and $d = t * \sqrt{(1/24 + 1/24)} = 1.916$, followed by the small-sample Hedges correction, yielding $g = 1.885$. This effect was therefore treated as statistically extractable but potentially influential, and it was examined explicitly in the leave-one-out analysis.

3.8. Statistical Analysis

The model included all 34 focal effects. Substantive consistency between the two models was interpreted as evidence of robustness, with 95% prediction intervals reported for both. Sensitivity analyses omitted the largest original converted effect, the largest newly added effect, and studies below the highest quality ratings. Moderator analyses were exploratory subgroup tests using Q -between statistics. Publication bias and small-study effects were assessed through funnel-plot inspection, exploratory Egger regression (Egger et al., 1997), and influence diagnostics, interpreted cautiously given the inclusion of converted effects.

4. Results

4.1. Study Selection and Analytic Dataset

Systematic screening identified 34 eligible focal effect sizes from independent studies. Twenty-two were directly extractable from complete descriptive statistics (means, SDs, sample sizes), constituting the verified model; the remaining 12 were derived through standard conversion from t , F , or partial η^2 and are included in the expanded model ($k = 34$), flagged in Table 2 with a sensitivity indicator. Studies spanned ChatGPT-supported writing feedback, LLM-supported writing instruction, AWE and grammar-checking systems, AI-based writing assistants, chatbot-based writing practice, and AI-enhanced formative feedback across EFL, ESL, and CSL contexts (Aladini et al., 2025; Chen & Gong, 2025; Gayed et al., 2022; Ghafouri et al., 2024; Hongxia & Razali, 2025; Hu et al., 2025; Kwon et al., 2023; Liu et al., 2026; Mahapatra, 2024; Shi et al., 2026; Wiboolyasarin et al., 2024).

A further group of studies was excluded because they did not estimate AI-versus-comparison writing-performance effects, including qualitative engagement studies (Koltovskaia et al., 2024; Yan, 2025; Yan & Zhang, 2024), large-scale modelling without intervention contrasts (Jin et al., 2025), perception-focused studies (Al-Raimi et al., 2024; Bibi & Atta, 2024; Meniado et al., 2024), AI literacy studies (Ou et al., 2024; Wang & Wang, 2025), and AI-detection studies (Cingillioglu, 2023). These remain important for theorising pedagogical and ethical conditions.

4.2. Overall Effect of AI Writing Tools on L2 Writing Performance

The primary verified model ($k = 22$; REML/Knapp–Hartung; Knapp & Hartung, 2003) indicated a positive overall effect: $g = 0.714$, 95% CI [0.592, 0.836], $I^2 = 24.3\%$, $\tau^2 = 0.019$. The expanded model ($k = 34$; DerSimonian–Laird) produced a closely consistent estimate: $g = 0.731$, 95% CI [0.631, 0.830], $I^2 = 29.9\%$, $\tau^2 = 0.023$. Please see Figure 2. Using Plonsky and Oswald’s (2014) L2 benchmarks, both estimates fall in the large range. The prediction interval [0.414, 1.047] suggests moderate-to-large effects across future implementations; convergence between models supports robustness.

The forest plot in Figure 2 shows that most individual effects were positive. Although effect magnitudes varied, the direction of evidence was consistently favourable to AI-supported writing across diverse tool types, learner populations, and instructional designs (Barrot, 2023; Gayed et al., 2022; Godwin-Jones, 2022; Mahapatra, 2024; Su et al., 2023).

4.3. Heterogeneity

The primary verified model showed low observed heterogeneity ($I^2 = 24.3\%$, $\tau^2 = 0.019$). The expanded model showed similarly modest heterogeneity ($I^2 = 29.9\%$, $\tau^2 = 0.023$). Some between-study variance is expected given variation in tool type, duration, learner population, and outcome measure, reflecting genuine diversity in how AI writing tools operate. The prediction interval [0.414, 1.047] communicates this variability.

The relatively low rather than extreme heterogeneity suggests that quantitative synthesis was appropriate, but it also indicates that AI writing tools do not constitute a uniform instructional category. Their effects likely depend on how tools are pedagogically integrated, what kind of feedback they provide, and whether learners are guided to process AI-generated suggestions critically rather than merely accept them.

4.4. Influence Diagnostics and Sensitivity Analysis

Influence diagnostics were conducted to assess whether the pooled effect was disproportionately shaped by the largest effects. Ghafouri et al. (2024) remained the largest original effect ($g = 1.88$), while Khany and Barzan (2026) contributed the largest newly added effect ($g = 1.61$). When Ghafouri et al. were omitted, the pooled estimate remained positive, $g = 0.716$, 95% CI [0.629, 0.803]. When Khany and Barzan were omitted, the pooled

estimate also remained positive, $g = 0.713$, 95% CI [0.623, 0.804]. These checks indicate that the direction of the finding was not dependent on either large effect.

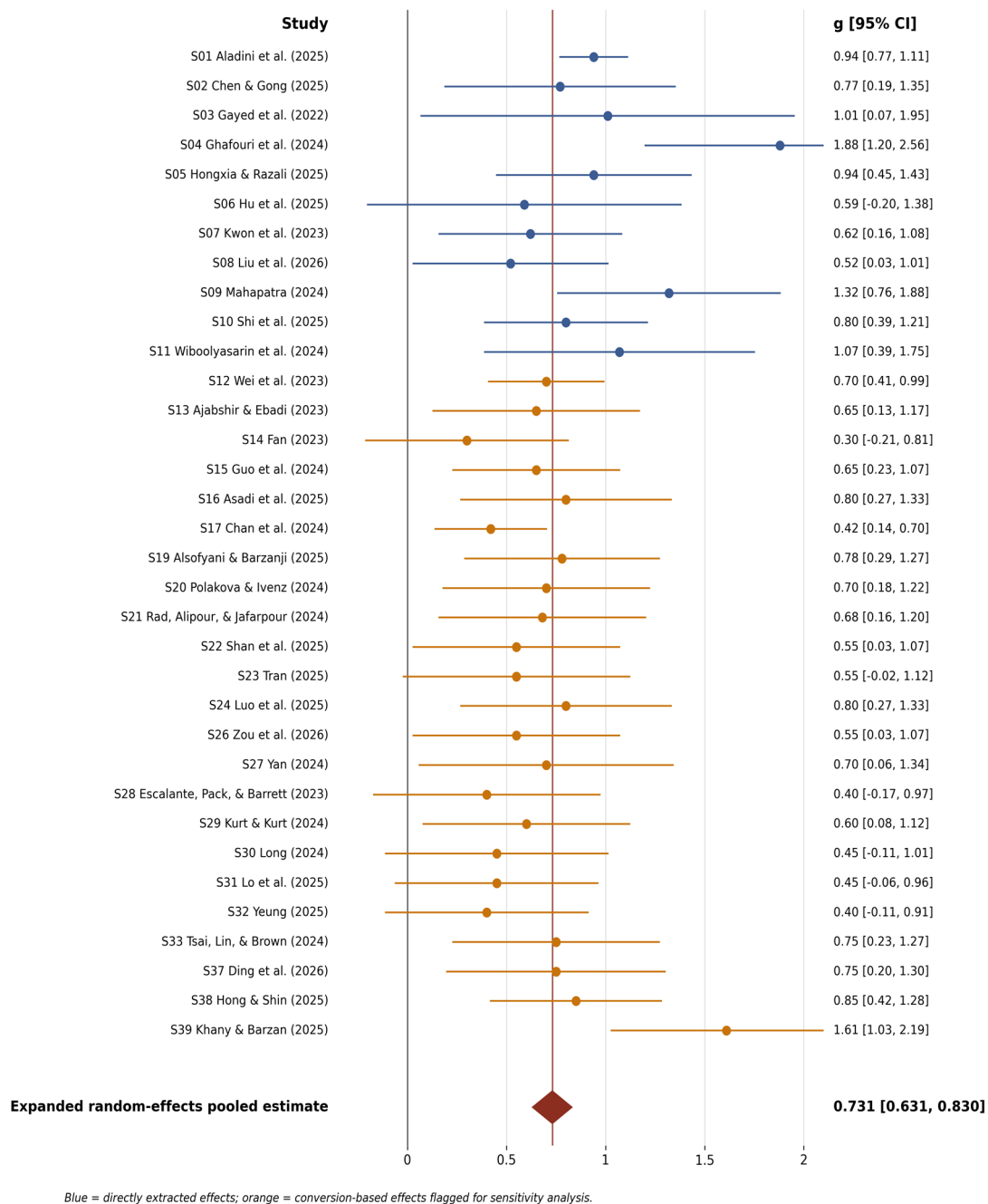


Figure 2. Forest plot of the effects of AI writing tools on L2 writing performance. Note: Effect sizes are reported as Hedges' g . The pooled estimate was obtained from the expanded $k = 34$ random-effects model.

The two large effects were retained because they addressed theoretically relevant AI-supported writing interventions and had extractable focal effects. They were also examined through leave-one-out sensitivity checks. The pooled effect remained positive after omitting Ghafouri et al. (2024) and after omitting Khany and Barzan (2026), indicating that the overall finding was not dependent on either large-effect study.

A higher-quality sensitivity model restricted the dataset to studies scoring 8 or 9 on the methodological quality checklist. This model included 24 focal effects and produced a positive pooled estimate, $g = 0.742$, 95% CI [0.631, 0.854], indicating that the expanded finding remained positive when lower-scored studies were removed.

Table 2. Characteristics of studies included in the primary meta-analysis.

Study	Participants	AI Condition/Comparison	Design/Duration	Outcome	<i>g</i>	Quality/9
Aladini et al. (2025)	Adult/institute L2 learners; EG = 276, CG = 285	AI self-directed writing tasks	Quasi-experimental; Unclear	Overall writing score	0.940	8
Chen & Gong (2025)	CSL higher-education learners; EG = 25, CG = 25	ChatGPT-assisted learning	Mixed-methods experimental; Unclear	Academic writing score	0.770	9
Gayed et al. (2022)	Adult English language learners; n = 10	AI KAKU writing assistant	Counter-balanced; One session	Fluency/ syntactic complexity	1.010	7
Ghafouri et al. (2024)	EFL learners; EG = 24, CG = 24	ChatGPT writing protocol	Quasi-experimental ANCOVA; >8 weeks	Overall writing skills	1.880	9
Hongxia & Razali (2025)	Chinese EFL undergraduates; EG = 34, CG = 36	ChatGPT feedback	Quasi-experimental; Unclear	Academic writing ability	0.940	9
Hu et al. (2025)	Undergraduate collaborative-writing groups; 13 groups per condition	GAI conversational agent	Quasi-experimental group-level; Short	Group essay score	0.590	9
Kwon et al. (2023)	Elementary EFL learners; EG = 37, CG = 38	Dialogflow chatbot	Randomized intervention; >8 weeks	Adjusted post-test writing score	0.620	9
Liu et al. (2026)	Elementary EFL learners; EG = 32, CG = 33	LLM-supported CALLA	Randomized controlled trial; 5 weeks	Composition post-test	0.520	9
Mahapatra (2024)	Undergraduate ESL learners; two classes	ChatGPT formative feedback	Mixed-methods intervention; Short-medium	Academic writing post-test	1.320	9
Shi et al. (2026)	Undergraduate EFL learners; about 50 per condition	ChatGPT vs AWE vs control	Three-arm intervention; >8 weeks	Writing performance	0.800	9
Wiboolyasarini et al. (2024)	Thai-major exchange students; EG = 18, CG = 21	Wiki + ChatGPT feedback	Quasi-experimental; >8 weeks	Collaborative writing score	1.070	8

Table 2. Cont.

Study	Participants	AI Condition/Comparison	Design/Duration	Outcome	g	Quality/9
Wei et al. (2023)	Chinese EFL learners; EG = 95, CG = 95	Grammarly/AWE	Randomized controlled trial; 12 weeks	L2 writing skills	0.700	8
Ajabshir & Ebadi (2023)	EFL learners; n: see Table A1 [†]	Automatic writing evaluation	AWE vs teacher feedback; Genre-based treatment	CALF and global writing quality	0.650	8
Fan (2023)	EFL university students; n: see Table A1 [†]	Grammarly AWCF	Quasi-experimental; Short	Writing quality/CAF	0.300	7
Guo et al. (2024)	University EFL students; n: see Table A1 [†]	AI-supported peer feedback	Experimental/quasi-experimental; Course intervention	Writing ability	0.650	8
Asadi et al. (2025)	EFL learners; n: see Table A1 [†]	ChatGPT + teacher feedback	Quasi-experimental; Intervention	Argumentative writing skills	0.800	8
Chan et al. (2024)	Hong Kong university L2 English students; focal t-test df = 192 implies analytic n $\approx$ 194; full article also reports FG = 342, CG = 576	Generative AI automated feedback	Randomized feedback/no-feedback intervention; One session	Essay revision gain score	0.420	8
Alsofyani & Barzanji (2025)	Saudi EFL tertiary; n: see Table A1 [†]	ChatGPT-generated feedback	Mixed-methods intervention; Tertiary-level treatment	Writing skills	0.780	8
Polakova & Ivenz (2024)	EFL students; n: see Table A1 [†]	ChatGPT feedback	Intervention study; Unclear	Writing skills	0.700	7
Rad et al. (2024)	EFL learners; n: see Table A1 [†]	Wordtune	Intervention study; Unclear	Writing outcome	0.680	8
Shan et al. (2025)	Chinese L2 learners; n: see Table A1 [†]	GPT-4 and native-speaker interaction	Comparative intervention; Unclear	Chinese L2 writing outcomes	0.550	8
Tran (2025)	EFL learners; n: see Table A1 [†]	AI- and teacher-generated feedback	Comparative intervention; Unclear	Writing revision practices	0.550	7

Table 2. Cont.

Study	Participants	AI Condition/Comparison	Design/Duration	Outcome	g	Quality/9
Luo et al. (2025)	EFL learners; EG = 30, CG = 30	AI and teacher feedback	Quasi-experimental; 12 weeks	Argumentative writing	0.800	8
Zou et al. (2026)	EFL writers; n: see Table A1 †	Teacher- and ChatGPT-generated feedback	Comparison study; Unclear	Feedback uptake and writing	0.550	8
Yan (2024)	L2 writers; n: see Table A1 †	Collaborative ChatGPT feedback processing	Intervention/mixed-methods; Short	Task improvement and learning	0.700	7
Escalante et al. (2023)	ENL/EFL students; n = 48; converted effect flagged	AI-generated feedback	Quasi-experimental/comparative; Semester	Writing improvement	0.400	8
Kurt & Kurt (2024)	L2 writers; n: see Table A1 †	ChatGPT automated feedback	Intervention study; Unclear	Writing skills	0.600	7
Long (2024)	EFL learners; n: see Table A1 †	ChatGPT written corrective feedback	Classroom study; Short	Revision/writing outcomes	0.450	7
Lo et al. (2025)	GenAI users; n: see Table A1 †	Generative AI essay revision support	Intervention/revision study; Unclear	Essay revision and engagement	0.450	7
Yeung (2025)	University students; n: see Table A1 †	Generative-AI-supported AWE feedback	Engagement study; Unclear	AWE feedback engagement and writing	0.400	7
Tsai et al. (2024)	EFL English majors; n: see Table A1 †	ChatGPT-assisted writing	Feasibility/intervention study; Unclear	Writing performance	0.750	7
Teng et al. (2026)	Chinese university EFL learners; focal contrast ERNIE Bot GenAI feedback (EG = 29) vs no-feedback control (CG = 26); AWE arm n = 29 not pooled	AWE vs ERNIE Bot GenAI feedback	Quasi-experimental three-condition; 3 months	Writing performance and SRL	0.750	8

Table 2. Cont.

Study	Participants	AI Condition/Comparison	Design/Duration	Outcome	g	Quality/9
Hong & Shin (2025)	University L2 academic writing students; focal contrast combines guided/free ChatGPT use (EG = 71) vs restricted use (CG = 32)	Restricted/guided/free ChatGPT use	Three-group classroom comparison; Course intervention	Essay scores	0.850	8
Khany & Barzan (2026)	L2 English learners; EG = 30, CG = 30	Grammarly/AI-powered writing assistant	Mixed-methods quasi-experimental; 12 weeks	Writing proficiency, accuracy, coherence, complexity	1.610	7

Note: EG = experimental group; CG = comparison/control group. g = Hedges' g for the focal AI-versus-comparison contrast. [†] Effect size derived from converted statistics (t, F, or partial η^2); included in the expanded model (k = 34) only. These studies are included in sensitivity analyses but excluded from the verified primary model. For Chan, Lo, and Wong (2024), g was calculated using the sample size implied by df = 192 (analytic $n \approx 194$); the full article also reports EG = 342, CG = 576. For Teng et al. (2026), the focal effect is the ERNIE Bot GenAI vs. no-feedback contrast (EG = 29, CG = 26); the AWE arm is not pooled.

4.5. Publication Bias and Small-Study Effects

A funnel plot (Please see Figure 3) was generated to inspect possible small-study effects. Visual inspection did not reveal an obvious single-sided pattern, though this should be interpreted cautiously given converted effects and the publication pressures of a rapidly emerging field. Exploratory Egger regression produced an intercept of -0.192 ($SE = 0.584$), $z = -0.328$, $p \approx 0.743$, indicating no strong small-study-effect signal but not ruling out publication bias.

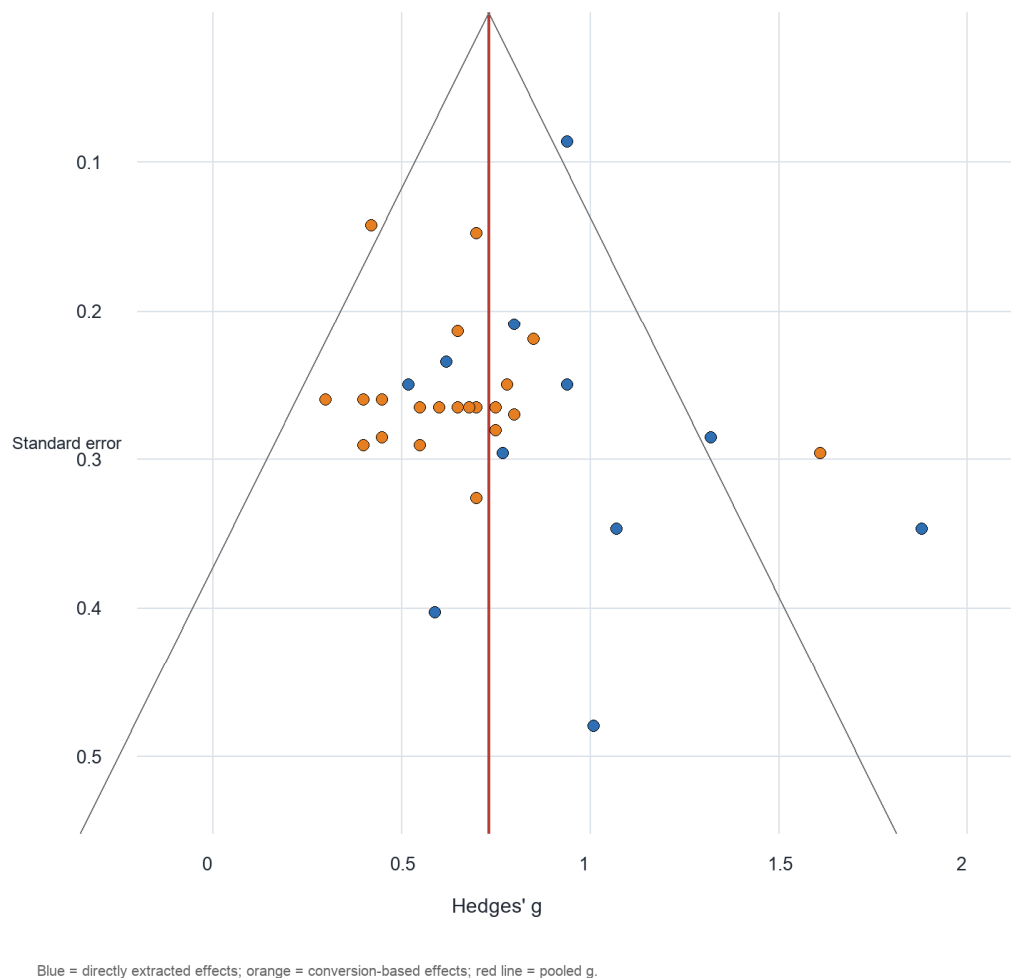


Figure 3. Funnel plot for assessing small-study effects in the expanded $k = 34$ model. Note: Effect sizes are plotted against their standard errors; the vertical line marks the pooled effect estimate.

4.6. Moderator Patterns

Because the expanded dataset contains 34 focal effects, moderator patterns are reported with exploratory Q between tests, but should still be interpreted cautiously. Table 3 focuses on contrasts most relevant to CALL theory and AI-assisted writing pedagogy: generative/LLM-based versus earlier non-LLM AI systems, multidimensional feedback/scaffolding, learner level, and intervention duration. Sparse cells, converted effects, and mixed classroom designs prevent strong causal claims about moderation.

The technology categories should be interpreted as an evidence map rather than a ranking of effectiveness. Generative and LLM-based tools accounted for most available effects, whereas earlier non-LLM systems were represented by only two heterogeneous studies. The dataset cannot determine whether generative AI is more effective than traditional AWE or earlier chatbot systems; it only shows that current intervention evidence is concentrated around generative and LLM-supported designs.

This distinction remains theoretically important because these systems differ in feedback architecture, learner agency, dialogic capacity, and cognitive demands. However, meaningful comparison requires larger samples that can model tool architecture while controlling for intervention duration, learner level, outcome type, and feedback focus. Feedback focus also warrants caution. Most interventions combined language feedback with planning,

revision, strategy instruction, or holistic guidance, suggesting that AI support may be most pedagogically plausible when it creates repeated opportunities for noticing and regulation rather than offering isolated corrections. Yet too few single-focus studies were available to test this contrast reliably.

Learner level and intervention duration were examined descriptively. Evidence was concentrated in higher education, though positive effects also appeared in school contexts. Interventions ranged from one-shot revision tasks to multi-week designs. Longer, instructionally embedded studies may provide stronger evidence of writing development because they are more likely to capture learning transfer rather than immediate text improvement.

Table 3. Main descriptive moderator patterns with at least two focal effects.

Moderator	Category	k	Pooled g [95% CI]	Exploratory Moderator Test/Interpretation
AI generation architecture	Constrained/AWE	6	0.771 [0.451, 1.092]	Q_between(1) = 0.021, $p \approx 0.884$; subgroup I2 = 58.8%
AI generation architecture	GenAI/LLM	28	0.728 [0.625, 0.831]	Q_between(1) = 0.021, $p \approx 0.884$; subgroup I2 = 22.7%
Feedback focus	Lower-order	3	0.457 [0.107, 0.807]	Q_between(1) = 2.744, $p \approx 0.098$; subgroup I2 = 0.0%
Feedback focus	Multidimensional	31	0.748 [0.646, 0.849]	Q_between(1) = 2.744, $p \approx 0.098$; subgroup I2 = 29.6%
Writing outcome type	CAF/accuracy	4	0.517 [0.227, 0.808]	Q_between(1) = 2.558, $p \approx 0.110$; subgroup I2 = 0.0%
Writing outcome type	Overall	30	0.751 [0.646, 0.855]	Q_between(1) = 2.558, $p \approx 0.110$; subgroup I2 = 31.7%
Learner level	Primary/young	2	0.573 [0.235, 0.912]	Q_between(1) = 1.053, $p \approx 0.305$; subgroup I2 = 0.0%
Learner level	University/adult	32	0.741 [0.637, 0.845]	Q_between(1) = 1.053, $p \approx 0.305$; subgroup I2 = 32.5%
Intervention duration	5–8 weeks	7	0.777 [0.593, 0.960]	Q_between(3) = 8.955, $p \approx 0.030$; subgroup I2 = 0.0%
Intervention duration	>8 weeks	9	0.912 [0.646, 1.178]	Q_between(3) = 8.955, $p \approx 0.030$; subgroup I2 = 60.3%
Intervention duration	Short/unclear	6	0.471 [0.270, 0.673]	Q_between(3) = 8.955, $p \approx 0.030$; subgroup I2 = 0.0%
Intervention duration	Unclear	12	0.770 [0.653, 0.887]	Q_between(3) = 8.955, $p \approx 0.030$; subgroup I2 = 0.0%

Note: Subgroup estimates are based on the expanded $k = 34$ random-effects dataset. Q_between values are exploratory moderator tests and should not be interpreted as definitive causal comparisons because several cells still combine heterogeneous designs and converted effects.

4.7. Methodological Quality of Included Studies

The methodological quality assessment indicated that the evidence base is promising but still developing. Stronger studies included comparison groups, pre-test measures, clearly described interventions, and explicit writing-performance assessments. The quality scores in **Table 2** evaluate reporting and design quality, whereas sensitivity flags evaluate effect-size comparability; the two dimensions are distinct, as a well-designed study may still require analytic caution if its effect size was converted from ANCOVA or group-level units.

Several limitations were common across the literature. Sample sizes were often small and intervention periods frequently short, making it difficult to determine whether gains represent durable development or immediate task-level improvement. Many studies reported overall writing scores without disaggregating accuracy, complexity, fluency, or coherence; the role of teacher mediation was also underspecified, even though how students are guided to interpret and revise AI feedback shapes whether it supports learning.

The field is moving beyond perception-focused research toward intervention-oriented evidence, but more rigorous studies are needed: larger samples, delayed post-tests, pre-registration, and clearer descriptions of AI tool use and instructional mediation.

4.8. Summary of Findings

Overall, the expanded $k = 34$ results provide stronger quantitative evidence. The random-effects model produced a positive pooled effect, and the larger evidence base permits exploratory moderator estimates and

Q_between tests that were not meaningful in the original model. At the same time, the expanded evidence base should be interpreted with caution because several newly added effects required converted extraction from available statistics and therefore carry greater uncertainty than effects extracted directly from complete means, standard deviations, and group sizes. The evidence supports the conclusion that AI writing tools can enhance L2 writing performance, particularly when integrated into structured instructional designs. However, it does not support a simplistic claim that AI tools automatically improve independent writing competence.

5. Discussion

The expanded evidence map shows a field expanding faster than its reporting conventions: many recent AI-writing studies report promising classroom outcomes but do not consistently provide the exact group sizes, means, standard deviations, or contrast definitions needed for reproducible meta-analysis. This mismatch is itself a substantive finding: as AI-writing interventions proliferate, the field needs not only more experiments but better reporting standards that make effect-size extraction and moderator coding possible.

5.1. RQ1: Overall Effect of AI Writing Tools on L2 Writing Performance

The size of the pooled effect suggests AI writing tools can provide more than superficial language polishing when embedded in pedagogically structured activities. These tools may operate as scaffolded affordances: making alternative formulations available, prompting learners to notice gaps, and creating repeated opportunities for planning, monitoring, and revision (Schmidt, 1990; Vygotsky, 1978), consistent with studies linking AI support to drafting, feedback uptake, and strategy use rather than one-off proofreading.

The range of tools in the expanded corpus reinforces this reading: the positive pattern is not limited to one platform, spanning Dialogflow chatbots, LLM-supported CALLA instruction, ChatGPT formative feedback, wiki-based collaborative writing, Grammarly/AWE systems, teacher-AI hybrid feedback, and GenAI revision support. What effective implementations share is not a brand name but a pedagogical function—diagnosing problems, prompting noticing, supporting revision, or providing repeated feedback-uptake opportunities.

The finding also aligns with feedback-as-noticing and process-oriented accounts of L2 writing (Schmidt, 1990). According to the noticing hypothesis, input or feedback contributes to acquisition only when learners consciously attend to the gap between their current output and the target form, rather than merely receiving a corrected version. Timely and usable feedback can promote learning when learners engage with it cognitively and revise purposefully, but it may have limited developmental value; that is, it may improve the immediate text without contributing to internalised, transferable writing competence: when learners simply accept suggestions without understanding their linguistic or rhetorical basis (Ferris, 2011; Hyland, 2019). This distinction helps explain why positive effects on writing products should not be equated automatically with internalised writing competence.

The large pooled effect nonetheless calls for caution. Many included studies assessed short-term or task-specific gains rather than long-term transfer; improvements in revised drafts or post-test scores may indicate enhanced task performance without necessarily reflecting durable independent competence. Learners may produce better texts with tool support without internalising the underlying linguistic, rhetorical, or strategic knowledge that produced them (Barrot, 2023; Warschauer et al., 2023).

5.2. RQ2: Descriptive Moderator Patterns and Pedagogical Conditions

The moderator patterns point to pedagogical conditionality rather than uniform effectiveness. Tools providing multiple forms of support—grammar, vocabulary, organisation, content, holistic feedback—appeared especially promising, consistent with L2 writing’s dual demands of lower-order accuracy and higher-order rhetorical control (Hyland, 2019). Tools focused narrowly on surface correctness may improve grammatical accuracy while leaving argumentation, coherence, and authorial voice underdeveloped; tools supporting planning, interpretation, and reflection have broader developmental potential.

The distinction between generative AI and traditional AI-supported writing tools is not merely a technology label but a CALL design issue. Earlier AWE and corrective feedback tools typically operated within constrained feedback architectures and were often strongest in lower-order concerns such as grammar, mechanics, and sentence-level accuracy (Godwin-Jones, 2022; Nazari et al., 2021; Xu & Zhang, 2022), whereas LLM-based tools can generate explanations, alternatives, examples, and dialogic support. This shift changes the learner’s role from receiving automated correction to managing an interactive writing ecology in which prompting, evaluating, rejecting, revising, and justifying AI feedback become part of the writing task. This synthesis frames AI writing

tools as configurable affordance systems (Gibson, 1979; Hutchby, 2001) whose value depends on how interaction design, task structure, and learner feedback literacy shape opportunities for noticing and transferable development.

The expanded corpus also complicates any simple GenAI-versus-AWE contrast. Traditional AWE systems can be pedagogically powerful when they are embedded in sustained revision cycles and teacher explanation; for instance, Ajabshir and Ebadi (2023) found that automatic writing evaluation combined with teacher-focused feedback improved CALF measures and overall writing quality across genres more effectively than either source alone. However, LLMs offer richer dialogic affordances, but only useful when learners know how to question and evaluate suggestions. The discussion should therefore move beyond tool typology toward implementation typology: AI-only, teacher-mediated, peer-AI, hybrid AWE-human, and strategy-focused designs likely produce different learning even with superficially similar tools.

Learner level and intervention duration also merit attention. Most studies were conducted in higher education, with fewer involving younger or lower-proficiency learners, though included school-level studies suggest AI-supported writing practice can benefit younger EFL learners when structured appropriately (Kwon et al., 2023; Liu et al., 2026). Sustained interventions may provide more meaningful evidence of learning than one-shot studies: short-term use demonstrates feasibility, but longer interventions are needed to determine whether learners develop transferable writing strategies and self-regulation.

5.3. RQ3: Methodological Quality and Reporting Implications

The methodological quality assessment indicates that AI-assisted L2 writing research is moving toward stronger intervention designs, but the evidence base remains uneven. Several studies used randomised or quasi-experimental designs, control groups, pre-test measures, and rubric-based assessments (Chen & Gong, 2025; Kwon et al., 2023; Liu et al., 2026; Shi et al., 2026), representing an important maturation beyond perception-only studies, though many still suffer from small samples, short intervention periods, and inconsistent effect-size reporting. The expanded study list reveals a tension: many new studies investigate motivation, engagement, or emotional response alongside writing performance, but multidimensional designs complicate synthesis when performance outcomes are not reported separately. Future studies should report writing outcomes and learner-experience outcomes distinctly.

Future studies should improve transparency by reporting group means, standard deviations, sample sizes, attrition, scoring rubrics, and complete post-test and delayed post-test statistics—essential for cumulative meta-analytic evidence building (Borenstein et al., 2009; Harrer et al., 2021; Viechtbauer, 2010)—and should distinguish writing performance with AI assistance from independent performance afterwards, since without this distinction it remains unclear whether observed gains reflect competence, AI-mediated text quality, or both.

Another issue concerns outcome comparability: some studies measured holistic writing quality, others accuracy, fluency, or engagement-related outcomes. This diversity reflects L2 writing's complexity but complicates synthesis; future research would benefit from clearer outcome taxonomies, more consistent assessment frameworks, and examination of whether gains transfer to new tasks, genres, and delayed post-test conditions.

5.4. Pedagogical Implications

The findings carry several implications for L2 writing pedagogy. AI writing tools work best as feedback and scaffolding resources rather than autonomous substitutes for instruction, and positive effects were consistently associated with structured pedagogical designs—writing protocols, strategy instruction, collaborative tasks, teacher-mediated feedback—rather than unrestricted tool access (Ghafouri et al., 2024; Liu et al., 2026; Shi et al., 2026; Wiboolyasarin et al., 2024).

Classroom integration should be organised around feedback routines rather than tool access alone: designing tasks in which learners draft independently, receive AI feedback, annotate which suggestions they accept or reject, explain their rationale, and transfer feedback to a new task. Such routines make AI use visible and assessable while supporting feedback literacy and metacognitive regulation.

Cultivating AI feedback literacy deserves explicit attention. Learners need to understand how to prompt AI systems, evaluate feedback accuracy, reject inappropriate suggestions, and maintain authorship; students vary considerably in their capacity to regulate AI use critically (Koltovskaia et al., 2024; Ou et al., 2024; Wang & Wang, 2025; Yan, 2025; Yan & Zhang, 2024), so AI-supported writing instruction should include metacognitive prompts, reflective logs, and discussions of academic integrity.

Assessment also needs to adapt. When learners use AI tools to revise extensively, traditional take-home writing products may no longer provide valid evidence of independent ability. Rather than banning AI tools

universally, instructors may need task-level policies specifying whether AI use is permitted for brainstorming, language polishing, feedback interpretation, or final drafting, and assessment designs should align accordingly.

A related implication is that assessment should distinguish at least three products: unaided writing, AI-supported revision, and post-intervention independent writing. Many studies demonstrate improvement in AI-supported texts, but fewer show transfer to new genres or no-AI conditions; future classroom designs should include delayed or transfer tasks and revision logs documenting how learners interpret AI feedback.

6. Limitations and Future Research

Although the search covered Scopus, Web of Science, ERIC, and ProQuest with supplementary forward citation tracing, it did not include trial registries or Chinese-language databases such as CNKI, which may hold relevant EFL intervention studies. Future updates should pre-register the protocol and report each PRISMA stage with complete item-level rationale.

The model contains 34 focal effects, but some remain converted estimates. The larger dataset improves moderator coverage, yet subgroup findings should stay exploratory until future full-text extraction can confirm effect sizes and cell sizes.

Where studies reported ANCOVA results, *t* values, partial eta-squared, or group-level outcomes rather than complete descriptive statistics, standard conversion procedures were used, but each conversion step introduces additional uncertainty. Future primary studies should report group means, standard deviations, sample sizes, adjusted means, inter-rater reliability, and attrition so that effect-size extraction is direct rather than approximate.

Variation in design, tool type, writing task, learner population, and outcome measurement reflects the emerging nature of the field but limits causal inference. Small-study and publication-bias checks should be read as exploratory rather than definitive reassurance.

The most important gap is the absence of delayed or unaided transfer measures. Informative designs would pair AI-supported intervention with a delayed post-test completed without AI assistance, examining whether learners can independently apply revised strategies, lexical choices, or grammatical knowledge. Think-aloud protocols, stimulated recall, and revision-trace analysis could reveal whether learners understand AI feedback or merely accept it. Pre-registered longitudinal studies spanning a semester or more are needed to test whether scaffolded performance becomes internalised competence.

7. Conclusions

This meta-analysis provides quantitative evidence that AI writing tools can improve L2 writing performance. Across 34 retained focal effects, the pooled effect was positive and robust to sensitivity analyses, suggesting that AI-supported writing instruction, feedback, and revision tools can make a meaningful contribution to L2 writing development when embedded in structured pedagogical designs.

At the same time, the findings should not be read as evidence that AI tools automatically produce writing development. Educational value depends on how learners use them, how teachers mediate them, and whether learners critically engage with AI-generated suggestions. AI writing tools can enhance feedback access and revision opportunities, but can also encourage overreliance without pedagogical regulation. Future research must separate assisted text improvement from independent competence development through delayed, unaided transfer measures and transparent effect-size reporting.

Funding

This research received no external funding.

Institutional Review Board Statement

Not applicable. This study is a meta-analysis of previously published research and collected no primary human-participant data.

Informed Consent Statement

Not applicable.

Data Availability Statement

The data supporting this meta-analysis are available in Appendix A.

Conflicts of Interest

The authors declare no conflict of interest.

Use of AI and AI-Assisted Technologies

Large language model-based AI tools (ChatGPT 5.0) were used for proofreading and checking internal consistency of statistical reporting. They were not used to generate research ideas, conduct searches, code effects, perform analyses, interpret findings, or produce substantive intellectual content. The author verified all factual claims, citations, statistics, and arguments and accepts full responsibility for the submitted manuscript.

Appendix A

Table A1. Sample-size and effect-size extraction details for studies abbreviated.

Study	AI Condition/ Comparison	EG/AI n	CG/ Comparison n	Design/Duration	Outcome	g
Ajabshir & Ebadi (2023)	Automatic writing evaluation	30	30	AWE vs teacher feedback; Genre-based treatment	CALF and global writing quality	0.650
Fan (2023)	Grammarly AWCf	30	30	Quasi-experimental; Short	Writing quality/CAF	0.300
Guo, et al. (2024)	AI-supported peer feedback	45	45	Experimental/ quasi-experimental; Course intervention	Writing ability	0.650
Asadi et al., (2025)	ChatGPT + teacher feedback	30	30	Quasi-experimental; Intervention	Argumentative writing skills	0.800
Alsofyani & Barzanji (2025)	ChatGPT-generated feedback	35	35	Mixed-methods intervention; Tertiary-level treatment	Writing skills	0.780
Polakova & Ivenz (2024)	ChatGPT feedback	30	30	Intervention study; Unclear	Writing skills	0.700
Rad et al. (2024)	Wordtune	30	30	Intervention study; Unclear	Writing outcome	0.680
Shan, et al. (2025)	GPT-4 and native-speaker interaction	30	30	Comparative intervention; Unclear	Chinese L2 writing outcomes	0.550
Tran (2025)	AI- and teacher-generated feedback	25	25	Comparative intervention; Unclear	Writing revision practices	0.550
Zou et al. (2026)	Teacher- and ChatGPT-generated feedback	30	30	Comparison study; Unclear	Feedback uptake and writing	0.550
Yan (2024)	Collaborative ChatGPT feedback processing	20	20	Intervention/mixed-methods; Short	Task improvement and learning	0.700
Kurt & Kurt (2024)	ChatGPT automated feedback	30	30	Intervention study; Unclear	Writing skills	0.600
Long (2024)	ChatGPT written corrective feedback	25	25	Classroom study; Short	Revision/writing outcomes	0.450
Lo et al., (2025)	Generative AI essay revision support	30	30	Intervention/revision study; Unclear	Essay revision and engagement	0.450
Yeung (2025)	Generative-AI-supported AWE feedback	30	30	Engagement study; Unclear	AWE feedback engagement and writing	0.400
Tsai, Lin, & Brown (2024)	ChatGPT-assisted writing	30	30	Feasibility/intervention study; Unclear	Writing performance	0.750

Note: EG/AI n and CG/comparison n are the analytic sample-size values used for Hedges' g variance estimation. Coding summary reports generation type, feedback focus, outcome type, and duration category; Abbreviations: EG = experimental or AI-supported group; CG = comparison group; AWE = automated writing evaluation; CAF = complexity, accuracy, and fluency; GenAI = generative artificial intelligence.

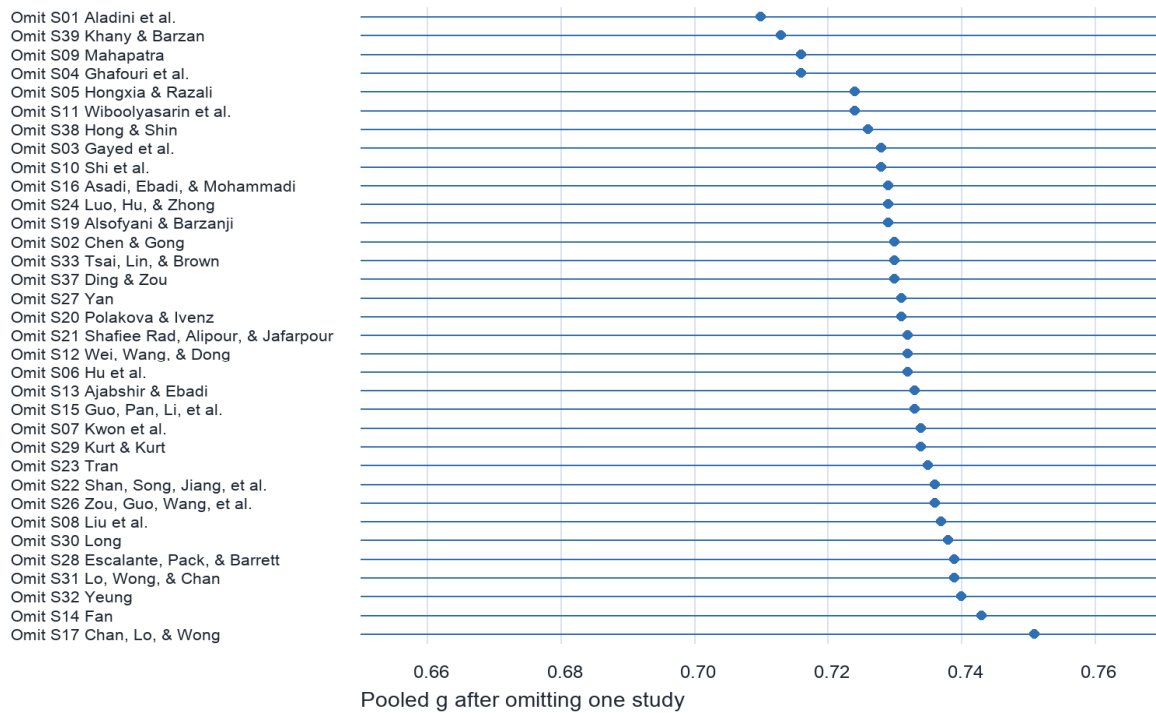


Figure A1. Leave-one-out sensitivity analysis for the expanded $k = 34$ model.

References

- Ajabshir, Z. F., & Ebadi, S. (2023). The effects of automatic writing evaluation and teacher-focused feedback on CALF measures and overall quality of L2 writing across different genres. *Asian-Pacific Journal of Second and Foreign Language Education*, 8(1), 26. <https://doi.org/10.1186/s40862-023-00201-9>.
- Al-Raimi, M., Mudhsh, B. A., Al-Yafaei, Y., & Al-Maashani, S. (2024). Utilizing artificial intelligence tools for improving writing skills: Exploring Omani EFL learners' perspectives. *Forum for Linguistic Studies*, 6(2), 1177. <https://doi.org/10.59400/fls.v6i2.1177>.
- Aladini, A., Ismail, S. M., Khasawneh, M. A. S., & Shakibaei, G. (2025). Self-directed writing development across computer/AI-based tasks: Unraveling the traces on L2 writing outcomes, growth mindfulness, and grammatical knowledge. *Computers in Human Behavior Reports*, 17, 100566. <https://doi.org/10.1016/j.chbr.2024.100566>.
- Alsaweed, W., & Aljebreen, S. (2024). Investigating the accuracy of ChatGPT as a writing error correction tool. *International Journal of Computer-Assisted Language Learning and Teaching*, 14(1), 1–18. <https://doi.org/10.4018/ijcallt.364847>.
- Alsofyani, A. H., & Barzanji, A. M. (2025). The effects of ChatGPT-generated feedback on Saudi EFL learners' writing skills and perception at the tertiary level: A mixed-methods study. *Journal of Educational Computing Research*, 63(2), 431–463. <https://doi.org/10.1177/07356331241307297>.
- Asadi, M., Ebadi, S., & Mohammadi, L. (2025). The impact of integrating ChatGPT with teachers' feedback on EFL writing skills. *Thinking Skills and Creativity*, 56, 101766. <https://doi.org/10.1016/j.tsc.2025.101766>.
- Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, 57, 100745. <https://doi.org/10.1016/j.asw.2023.100745>.
- Bibi, Z., & Atta, A. (2024). The Role of ChatGPT as AI English Writing Assistant: A Study of Student's Perceptions, Experiences, and Satisfaction. *Annals of Human and Social Sciences*, 5(I), 433–443. [https://doi.org/10.35484/ahss.2024\(5-i\)39](https://doi.org/10.35484/ahss.2024(5-i)39).
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Wiley. <https://doi.org/10.1002/9780470743386>.
- Chan, S. T. S., Lo, N. P. K., & Wong, A. M. H. (2024). Enhancing university level English proficiency with generative AI: Empirical insights into automated feedback and learning outcomes. *Contemporary Educational Technology*, 16(4), ep541. <https://doi.org/10.30935/cedtech/15607>.
- Chen, C., & Gong, Y. F. (2025). The role of AI-assisted learning in academic writing: A mixed-methods study on Chinese as a second language students. *Education Sciences*, 15(2), 141. <https://doi.org/10.3390/educsci15020141>.
- Cingillioglu, I. (2023). Detecting AI-generated essays: The ChatGPT challenge. *The International Journal of Information and Learning Technology*, 40(3), 259–268. <https://doi.org/10.1108/ijilt-03-2023-0043>.

- Ding, L., Zou, D., & Kohnke, L. (2026). Enhancing L2 writing with AI: Effects of ChatGPT feedback and the role of student engagement. *Artificial Intelligence in Language Education*, 2, 103356. <https://doi.org/10.29140/aile.2026.103356>.
- Du, J. L., & Chen, Y. (2026). Exploring English for academic purposes learners' engagement with ChatGPT-4 in academic writing revision: A case study from a private language institute in China. *Frontiers in Artificial Intelligence*, 9, 1802007. <https://doi.org/10.3389/frai.2026.1802007>.
- Du, J. L., Liu, Z., & Liao, N. (2026). Predicting higher education students' continued intention and self-reported use of generative AI tools in academic work: An extended UTAUT mixed-methods study. *Acta Psychologica*, 269, 107471. <https://doi.org/10.1016/j.actpsy.2026.107471>.
- Du, J. L., & Kohnke, L. (2026). Repositioning teachers in the age of AI: A cross-disciplinary study of perceptions and practices in China. *Teaching Education*, 37(3), 281–305. <https://doi.org/10.1080/10476210.2025.2590694>.
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, 315(7109), 629–634. <https://doi.org/10.1136/bmj.315.7109.629>.
- Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, 20, 57. <https://doi.org/10.1186/s41239-023-00425-2>.
- Fan, N. (2023). Exploring the effects of automated written corrective feedback on EFL students' writing quality: A mixed-methods study. *SAGE Open*, 13(2), 1–17. <https://doi.org/10.1177/21582440231181296>.
- Ferris, D. (2011). *Treatment of error in second language student writing* (2nd ed.). University of Michigan Press. <https://doi.org/10.3998/mpub.2173290>.
- Gayed, J. M., Carlon, M. K. J., Oriola, A. M., & Cross, J. S. (2022). Exploring an AI-based writing Assistant's impact on English language learners. *Computers and Education: Artificial Intelligence*, 3, 100055. <https://doi.org/10.1016/j.caeai.2022.100055>.
- Ghafouri, M., Hassaskhah, J., & Mahdavi-Zafarghandi, A. (2024). From virtual assistant to writing mentor: Exploring the impact of a ChatGPT-based writing instruction protocol on EFL teachers' self-efficacy and learners' writing skill. *Language Teaching Research*. <https://doi.org/10.1177/13621688241239764>.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton Mifflin.
- Godwin-Jones, R. (2022). Partnering with AI: Intelligent writing assistance and instructed language learning. *Language Learning & Technology*, 26(2), 5–24. <https://doi.org/10.64152/10125/73474>.
- Guo, K., Pan, M., Li, Y., & Lai, C. (2024). Effects of an AI-supported approach to peer feedback on university EFL students' feedback quality and writing ability. *The Internet and Higher Education*, 63, 100962. <https://doi.org/10.1016/j.iheduc.2024.100962>.
- Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2021). *Doing meta-analysis with R: A hands-on guide*. Chapman and Hall/CRC. <https://doi.org/10.1201/9781003107347>.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press.
- Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2019). *Cochrane handbook for systematic reviews of interventions* (2nd ed.). Wiley. <https://doi.org/10.1002/9781119536604>.
- Hong, S., & Shin, Y. K. (2025). Effects of three levels of AI integration on second language academic writing: Evaluating restricted, guided, and free use of ChatGPT. *System*, 134, 103820. <https://doi.org/10.1016/j.system.2025.103820>.
- Hong, Q. N., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M. P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M. C., Vedel, I., & Pluye, P. (2018). The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. *Education for Information*, 34(4), 285–291. <https://doi.org/10.3233/efi-180221>.
- Hongxia, H., & Razali, A. B. (2025). Impact of ChatGPT on English academic writing ability and engagement of Chinese EFL undergraduates. *International Journal of Instruction*, 18(2), 323–346. <https://doi.org/10.29333/iji.2025.18218a>.
- Hu, W., Tian, J., & Li, Y. (2025). Enhancing student engagement in online collaborative writing through a generative AI-based conversational agent. *The Internet and Higher Education*, 65, 100979. <https://doi.org/10.1016/j.iheduc.2024.100979>.
- Hutchby, I. (2001). Technologies, texts and affordances. *Sociology*, 35(2), 441–456. <https://doi.org/10.1017/S0038038501000219>.
- Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching*, 39(2), 83–101. <https://doi.org/10.1017/s0261444806003399>.
- Hyland, K. (2019). *Second language writing*. Cambridge University Press. <https://research-portal.uea.ac.uk/en/publications/second-language-writing/>.
- Jin, F., Lin, C. H., & Lai, C. (2025). Modeling AI-assisted writing: How self-regulated learning influences writing outcomes. *Computers in Human Behavior*, 165, 108538. <https://doi.org/10.1016/j.chb.2024.108538>.

- Khany, R., & Barzan, P. (2026). Leveraging AI-powered writing assistants to enhance L2 writing proficiency: A mixed-methods study. *AI and Tech in Behavioral and Social Sciences*, 3(4), 1–14. <https://doi.org/10.61838/kman.aitech.4287>.
- Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine*, 22(17), 2693–2710. <https://doi.org/10.1002/sim.1482>.
- 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>.
- Kurt, G., & Kurt, Y. (2024). Enhancing L2 writing skills: ChatGPT as an automated feedback tool. *Journal of Information Technology Education: Research*, 23, 024. <https://doi.org/10.28945/5370>.
- Kwon, S. K., Shin, D., & Lee, Y. (2023). The application of chatbot as an L2 writing practice tool. *Language Learning & Technology*, 27(1), 1–19. <https://doi.org/10.64152/10125/73541>.
- Lingard, L. (2023). Writing with ChatGPT: An Illustration of its Capacity, Limitations & Implications for Academic Writers. *Perspectives on Medical Education*, 12(1), 261–270. <https://doi.org/10.5334/pme.1072>.
- Liu, Z. M., Hwang, G. J., Chen, C. Q., Chen, X. D., & Ye, X. D. (2026). Integrating large language models into EFL writing instruction: Effects on performance, self-regulated learning strategies, and motivation. *Computer Assisted Language Learning*, 39(3), 466–490. <https://doi.org/10.1080/09588221.2024.2389923>.
- Lo, N., Chan, S., & Wong, A. (2025). Evaluating teacher, AI, and hybrid feedback in English language learning: Impact on student motivation, quality, and performance in Hong Kong. *SAGE Open*, 15(3). <https://doi.org/10.1177/21582440251352907>.
- Long, H. S. (2024). Exploring the use of ChatGPT as a tool for written corrective feedback in an EFL classroom. *The Journal of AsiaTEFL*, 21(2), 397–412. <https://doi.org/10.18823/asiatefl.2024.21.2.8397>.
- Luo, M., Hu, X., & Zhong, C. (2025). The collaboration of AI and teacher in feedback provision and its impact on EFL learner's argumentative writing. *Education and Information Technologies*, 30(12), 17695–17715. <https://doi.org/10.1007/s10639-025-13488-7>.
- Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments*, 11, 9. <https://doi.org/10.1186/s40561-024-00295-9>.
- Meniado, J. C., Huyen, D. T. T., Panyadilokpong, N., & Lertkomolwit, P. (2024). Using ChatGPT for second language writing: Experiences and perceptions of EFL learners in Thailand and Vietnam. *Computers and Education: Artificial Intelligence*, 7, 100313. <https://doi.org/10.1016/j.caeai.2024.100313>.
- Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, 11(2), 364–386. <https://doi.org/10.1177/1094428106291059>.
- Munn, Z., Barker, T. H., Moola, S., Tufanaru, C., Stern, C., McArthur, A., Stephenson, M., & Aromataris, E. (2020). Methodological quality of case series studies: An introduction to the JBI critical appraisal tool. *JBI Evidence Synthesis*, 18(10), 2127–2133. <https://doi.org/10.11124/jbisrir-d-19-00099>.
- Nazari, N., Shabbir, M. S., & Setiawan, R. (2021). Application of Artificial Intelligence powered digital writing assistant in higher education: Randomized controlled trial. *Heliyon*, 7(5), e07014. <https://doi.org/10.1016/j.heliyon.2021.e07014>.
- Norris, J. M., & Ortega, L. (Eds.). (2006). *Synthesizing research on language learning and teaching*. John Benjamins. <https://doi.org/10.1075/llt.13>.
- Ou, A. W., Khuder, B., Franzetti, S., & Negretti, R. (2024). Conceptualising and cultivating critical GAI literacy in doctoral academic writing. *Journal of Second Language Writing*, 66, 101156. <https://doi.org/10.1016/j.jslw.2024.101156>.
- 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>.
- Plonsky, L. (2014). Study quality in quantitative L2 research (1990–2010): A methodological synthesis and call for reform. *The Modern Language Journal*, 98(1), 450–470. <https://doi.org/10.1111/j.1540-4781.2014.12058.x>.
- Polakova, P., & Ivenz, P. (2024). The impact of ChatGPT feedback on the development of EFL students' writing skills. *Cogent Education*, 11(1), 2410101. <https://doi.org/10.1080/2331186x.2024.2410101>.
- Rad, H. S., Alipour, R., & Jafarpour, A. (2024). Using artificial intelligence to foster students' writing feedback literacy, engagement, and outcome: A case of Wordtune application. *Interactive Learning Environments*, 32(9), 5020–5040. <https://doi.org/10.1080/10494820.2023.2208170>.
- Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–158. <https://doi.org/10.1093/applin/11.2.129>.
- Shan, Z., Song, Z., Jiang, X., Chen, W., & Chen, L. (2025). Complementing but not replacing: Comparing the impacts of GPT-4 and native-speaker interaction on Chinese L2 writing outcomes. *Behavioral Sciences*, 15(4), 540. <https://doi.org/10.3390/bs15040540>.

- Shi, H., Chai, C. S., Zhou, S., & Aubrey, S. (2026). Comparing the effects of ChatGPT and automated writing evaluation on students' writing and ideal L2 writing self. *Computer Assisted Language Learning*, 39(4), 1148–1175. <https://doi.org/10.1080/09588221.2025.2454541>.
- Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H. Y., Corbett, M. S., Eldridge, S. M., Emberson, J. R., Hernán, M. A., Hopewell, S., Hróbjartsson, A., Junqueira, D. R., Jüni, P., Kirkham, J. J., Lasserson, T., Li, T., ... Higgins, J. P. T. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. *Bmj*, 366, l4898. <https://doi.org/10.1136/bmj.l4898>.
- Stroud, R., & Du, J. (2025). Student acceptance of ChatGPT in higher education: A mixed-method study of attitudes, intentions, and concerns. *Journal of Applied Learning & Teaching*, 8(2), 35–48. <https://doi.org/10.37074/jalt.2025.8.2.7>.
- Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing*, 57, 100752. <https://doi.org/10.1016/j.asw.2023.100752>.
- Teng, L. S., Deng, X., & Yang, J. (2026). Can GenAI-empowered feedback promote L2 learners' self-regulation strategic behavior and writing performance? *System*, 138, 103970. <https://doi.org/10.1016/j.system.2026.103970>.
- Tran, T. T. T. (2025). Enhancing EFL writing revision practices: The impact of AI- and teacher-generated feedback and their sequences. *Education Sciences*, 15(2), 232. <https://doi.org/10.3390/educsci15020232>.
- Tsai, C. Y., Lin, Y. T., & Brown, I. K. (2024). Impacts of ChatGPT-assisted writing for EFL English majors: Feasibility and challenges. *Education and Information Technologies*, 29(17), 22427–22445. <https://doi.org/10.1007/s10639-024-12722-y>.
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1–48. <https://doi.org/10.18637/jss.v036.i03>.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Wan, X., Liu, Q., & Adebayo, S. A. (2026). Explore Parents' Views and Acceptance of Large Language Models Applied to Education. *Journal of Educational Technology and Innovation*, 8(2), 85–99. <https://doi.org/10.61414/wbgfyr12>.
- Wang, L., & Ren, B. (2024). Enhancing academic writing in a linguistics course with generative AI: An empirical study in a higher education institution in Hong Kong. *Education Sciences*, 14(12), 1329. <https://doi.org/10.3390/educsci14121329>.
- Wang, C., & Wang, Z. (2025). Investigating L2 writers' critical AI literacy in AI-assisted writing: An APSE model. *Journal of Second Language Writing*, 67, 101187. <https://doi.org/10.1016/j.jslw.2025.101187>.
- Warschauer, M., Tseng, W., Yim, S., Webster, T., Jacob, S., Du, Q., & Tate, T. (2023). The affordances and contradictions of AI-generated text for writers of English as a second or foreign language. *Journal of Second Language Writing*, 62, 101071. <https://doi.org/10.1016/j.jslw.2023.101071>.
- Wei, P., Wang, X., & Dong, H. (2023). The impact of automated writing evaluation on second language writing skills of Chinese EFL learners: A randomized controlled trial. *Frontiers in Psychology*, 14, 1249991. <https://doi.org/10.3389/fpsyg.2023.1249991>.
- Wiboolyasarin, W., Wiboolyasarin, K., Suwanwihok, K., Jinowat, N., & Muenjanchoey, R. (2024). Synergizing collaborative writing and AI feedback: An investigation into enhancing L2 writing proficiency in wiki-based environments. *Computers and Education: Artificial Intelligence*, 6, 100228. <https://doi.org/10.1016/j.caeai.2024.100228>.
- Xu, J., & Zhang, S. (2022). Understanding AWE feedback and English writing of learners with different proficiency levels in an EFL classroom: A sociocultural perspective. *The Asia-Pacific Education Researcher*, 31(4), 357–367. <https://doi.org/10.1007/s40299-021-00577-7>.
- Yan, D. (2024). Comparing individual vs. collaborative processing of ChatGPT-generated feedback: Effects on L2 writing task improvement and learning. *Language Learning & Technology*, 28(1), 1–19. <https://doi.org/10.64152/10125/73597>.
- Yan, D. (2025). Feedback seeking abilities of L2 writers using ChatGPT: A mixed method multiple case study. *Kybernetes*, 54(7), 3757–3781. <https://doi.org/10.1108/k-09-2023-1933>.
- Yan, D., & Zhang, S. (2024). L2 writer engagement with automated written corrective feedback provided by ChatGPT: A mixed-method multiple case study. *Humanities and Social Sciences Communications*, 11, 1086. <https://doi.org/10.1057/s41599-024-03543-y>.
- Yang, W., & Lin, C. (2025). Translanguaging with generative AI in EFL writing: Students' practices and perceptions. *Journal of Second Language Writing*, 67, 101181. <https://doi.org/10.1016/j.jslw.2025.101181>.
- Yeung, S. (2025). University students' engagement with generative AI-supported automated writing evaluation (AWE) feedback. *Journal of Second Language Writing*, 68, 101203. <https://doi.org/10.1016/j.jslw.2025.101203>.
- Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). Academic Press. <https://doi.org/10.1016/b978-012109890-2/50031-7>.

- Zou, M., & Huang, L. (2024). The impact of ChatGPT on L2 writing and expected responses: Voice from doctoral students. *Education and Information Technologies*, 29, 13201–13219. <https://doi.org/10.1007/s10639-023-12397-x>.
- Zou, S., Guo, K., Wang, J., & Liu, Y. (2026). Investigating students' uptake of teacher- and ChatGPT-generated feedback in EFL writing: A comparison study. *Computer Assisted Language Learning*, 39(4), 1062–1091. <https://doi.org/10.1080/09588221.2024.2447279>.