

Human-AI Collaborative Teaching Frameworks for Next Generation English Education

I Putu Eka Indrawan^{1*} and Dr. I Made Citra Wibawa²

^{1*}Lecturer, Faculty of Science and Technology, PGRI Mahadewa University of Indonesia, Denpasar, Indonesia. E-mail: putueka@mahadewa.ac.id, Orcid: <https://orcid.org/0000-0003-2431-7603>

²Associate Professor, Department of Primary Teacher Education, Faculty of Education University, Ganesha University of Education, Singaraja, Bali, Indonesia. E-mail: imadecitra.wibawa@undiksha.ac.id

Abstract: English education is on the verge of a transformation with the advent of Artificial Intelligence (AI) tools, which offer automated grading of writing and the provision of personalized learning assistance. But fully automated feedback is sometimes lacking in pedagogic judgment, and the reliability of feedback and meaningful teacher engagement are suboptimal. This study presents a Human–AI Collaborative Teaching Framework (HACTF), which combines AI-supported essay evaluation and a teacher-in-the-loop approach in adaptive teaching of English writing. The framework was tested with the IELTS Writing Dataset with 1435 essays with examiner comments and examiner band scores (Task Response, Coherence and Cohesion, Lexical Resource, and Grammatical Range and Accuracy) with the Overall Band Score. The essays were preprocessed and divided into 80% training (1,148 essays) and 20% testing (287 essays). Each essay was first scored by the AI, and then a confidence-based decision module decided whether teacher validation would be needed or not for the delivery of personalized recommendations. Experimental tests proved that the proposed approach provided 96.8% of feedback agreement and 97.6% of feedback acceptance and decreased the teacher intervention rate to 31.4%. Furthermore, the average feedback generation time has dropped to 4.8 minutes, the student writing improvement was 25.7%, and the overall gain of the students in IELTS band score was 0.86. The results suggest that AI efficiency can boost the quality of feedback, student writing ability, and the scalability, reliability, and individualization of writing instruction for the next generation of classrooms.

Key Words: human–AI collaboration, english education, IELTS writing assessment, teacher-in-the-loop, artificial intelligence, personalized feedback

(Received: 01 July 2025; Revised: 23 July 2025; Accepted: 01 September 2025; Published: 30 September 2025)

Introduction

Artificial Intelligence (AI) has revolutionized the field of English education, providing intelligent support for writing, intelligent evaluation, individualized instruction, and intelligent guidance. With the introduction of recent developments in generative AI and large language models, new opportunities have emerged for enhancing writing instruction, grammar correction, vocabulary development, and learner engagement. But the complete AI-based learning approach can diminish the role of the teacher, leading to questions about the reliability of feedback, the pedagogical fit and the lack of development of students' critical thinking skills (Omer, 2024; Ji et al., 2023; Atchley et al., 2024).

Many recent research studies have focused on the collaboration between humans and AI, with AI assisting with common teaching activities and educators offering pedagogical expertise, context, and individual guidance. Through such cooperation, the efficiency of education will be increased, high-quality

education will be ensured, and rational use of AI in English education will be promoted (Meniado, 2024; Kong & Yang, 2024; Järvelä et al., 2023). AI can rapidly assess an essay, but the teacher will be able to verify complicated cases, correct the suggestions, and deal with high-level writing skills that need human knowledge (Nguyen et al., 2024; Luther et al., 2024).

Even though there have been some studies conducted on conversational AI, AI-driven writing, and human-centric models of learning, there has been limited research done on adaptive teacher-in-the-loop models using actual IELTS writing assessment data to generate collaborative feedback (Ji et al., 2023; Wang et al., 2023; Nguyen et al., 2024). Moreover, existing approaches have not used comments made by the examiner and grading criteria mentioned in the rubric in order to determine situations when AI feedback needs to be verified by the teacher.

The proposed Human–AI Collaborative Teaching Framework (HACTF) is different from existing Automated Writing Evaluation (AWE) systems, such as Grammarly and ETS e-rater, which focus mainly on automated grammar correction, writing scores, or language suggestions, by adding the comments of an IELTS examiner and the assessment criteria of the rubric to a confidence-guided teacher-in-the-loop workflow. The framework does not only use automated predictions but also includes a confidence score for identifying uncertain evaluations and only forwards those evaluations for teacher validation. Such a combination of authentic examiner feedback, standardized IELTS scoring criteria, and adaptive teacher intervention makes it possible to provide more accurate, pedagogically relevant, and individualized writing feedback, which is a novel development in human–AI collaborative English teaching.

Research Contributions

- Proposes a teacher in the loop Human-AI collaborative model for English writing teaching.
- Uses the comments made by the examiner and IELTS rubric scores to produce consistent, individual feedback.
- Enables an effective co-feedback process that opens the way for AI and teacher expertise.

Research Objectives

- To establish a human–AI collaborative teaching framework for adaptive writing assessment in English.
- To bring AI-based essay evaluation and teacher-in-the-loop validation together with the IELTS writing criteria.
- To assess the effectiveness of collaborative feedback made on the IELTS Writing Dataset.

The rest of this paper is organized as follows. Section 2 summarizes recent research on the partnership of humans and AI in EFL. In Section 3, the proposed Human–AI Collaborative Teaching Framework is presented. The experimental methodology and evaluation with the IELTS Writing Dataset are presented in Section 4. The experimental results are discussed in Section 5, and the conclusion is given in Section 6.

Literature Review

The journey of English learning with AI has been evolving from the use of AI to help students learn to the use of AI in collaboration with teachers, maintaining the teaching position of teachers. Systematic recent reviews highlight the importance of achieving meaningful human interaction for keeping learners engaged, developing critical thinking, and improving language learning in AI-enabled settings (Omer, 2024; Jiali et al., 2024; Järvelä et al., 2023). In the area of language learning, conversational AI has

also shown great promise, though its impact is not as strong without close cooperation between the AI and teachers, as opposed to being completely independent of them (Ji et al., 2023; Baskara, 2024; Wu et al., 2024).

The teacher-in-the-loop learning frameworks are another important area that involves using AI for assisting in routine learning tasks while the teachers supervise complicated learning processes. The human-centered AI models have shown how AI automation and teachers' knowledge can be used to improve self-regulated learning, teaching effectiveness, and personalization (Meniado, 2024; Kong & Yang, 2024; Han et al., 2024). Some studies show that using human-AI collaboration improves writing quality, learning regulation, and educational content creation while maintaining the role of the teacher in validating the AI-generated outputs (Nguyen et al., 2024; Luther et al., 2024). The results indicate that AI should be a teaching assistant rather than a decision-making tool for teachers.

AI adoption in higher education, AI literacy, sustainable education, and collaborative learning environments has also been the focus of recent studies (Atchley et al., 2024; Tzirides et al., 2024; Fui-Hoon Nah et al., 2023). While these studies illustrate the educational values of generative AI, the current automated writing evaluation (AWE) tools predominantly focus on providing grammar corrections and writing suggestions and do not include the rubric scores of the examiner, teacher validation, or confidence-based intervention. Furthermore, there has been very little research based on authentic IELTS examiner comments that aimed to find out when AI feedback needs to be reviewed by the teacher. Based on the above analysis, this study presents a Human–AI Collaborative Teaching Framework (HACTF), which combines the advantages of IELTS rubric-based assessment, examiner comments, and teacher intervention based on learners' confidence to provide reliable and personalized feedback for English writing.

Research Gap

Most of the current AI-supported English writing systems focus on automated evaluation of an essay or on chatbots for conversational AI, and only a few include teacher-in-the-loop collaboration. Most Automated Writing Evaluation (AWE) tools do not involve IELTS examiner comments, rubric-based assessment, and teacher intervention guided by students' confidence to enhance the reliability of feedback. As a result, pedagogically appropriate and personalized writing feedback is understudied. The intent of this study is to overcome these deficits by proposing a new Human–AI Collaborative Teaching Framework to combine AI assessment with teacher validation for confidence based on authentic IELTS writing data.

Human-AI Collaborative Teaching Framework

Based on the above issues, this study introduced a Human–AI Collaborative Teaching Framework (HACTF) to enhance the English writing teaching by combining the evaluation support of AI essays with teachers' expertise. The framework differs from the traditional AI-based evaluation tools that provide fully automated feedback by introducing a teacher-in-the-loop system that ensures that ambiguous or difficult feedback for teaching is verified before being given to students. The framework is constructed with the IELTS Writing Dataset comprising student essays, comments from the examiner, and scores obtained based on the rubric for four assessment criteria in writing. There are six steps in the framework, which are shown in figure 1.

IELTS Writing Data Preparation

The first module loads the IELTS Writing Dataset and preprocesses the essays for analysis. To ensure the consistency of the data, normalization, sentence segmentation, punctuation cleaning, and tokenization

are done to the text. Collaborative evaluation will use examiner comments and IELTS criterion scores (Task Response (TR), Coherence and Cohesion (CC), Lexical Resource (LR), and Grammatical Range and Accuracy (GRA)) as reference labels.

AI-Based Writing Assessment

The AI evaluation system is based on a pre-trained transformer-based LLM, which is based on the architecture of Llama 3.1-8B-Instruct. Evaluation by the model is based on criterion-based essay analysis using the GPT (Guided Prompting) technique, which is based on the IELTS writing rubric. Task Response (TR), Coherence and Cohesion (CC), Lexical Resource (LR), and Grammatical Range and Accuracy (GRA) are analyzed in each essay. The feedback generated consists of grammar correction, vocabulary training, coherence, and task-oriented feedback. To validate the quality and consistency of AI-generated feedback, the dataset is validated using comments from the examiner.

Confidence-Based Teacher Intervention

The module of decision based on confidence decides whether the feedback generated by AI is given directly or by a teacher. Each of the four assessment criteria (listening, reading, writing, and speaking) has an average probability, which is known as the confidence. The average of the probability of the four assessment criteria is computed as follows (Equation 1):

$$C = \frac{1}{4} \sum_{i=1}^4 P_i \quad (1)$$

Where:

- C = overall, AI confidence score
- P_i = confidence probability for each IELTS criterion (TR, CC, LR, GRA)

Decision rule:

AI Feedback, $C \geq 0.85$

Teacher Validation, $C < 0.85$

Where P_i is the confidence probability of each of the writing-criterion predictions by the AI model. Evaluation reliability is considered to be 0.85. If the AI feedback is reliable ($C \geq 0.85$), it is directly given to students. In the other case the essay is handed over to the teacher for checking and improving the essay before the generation of final feedback.

Human–AI Collaborative Feedback Generation

Once validated by the teacher, the AI suggestions and teacher editing are merged into a single AI/teacher-editing personalized feedback report. Corrections for grammar, suggestions for vocabulary enhancement, coherence, and suggestions for better task response (as per IELTS assessment criteria) are included in the final feedback. This joint effort ensures that teachers remain central to the teaching enterprise and gives smart use of AI.

Student Revision and Learning Analytics

Students edit and improve their compositions based on the remarks from the cooperative editing. The scoring of writing and acceptance of feedback and revision outcomes are documented in the framework

to measure individual learning progress. The learning analytics help teachers understand their students' weaknesses in the writing process and offer tailored guidance for teaching in the next writing session(s).

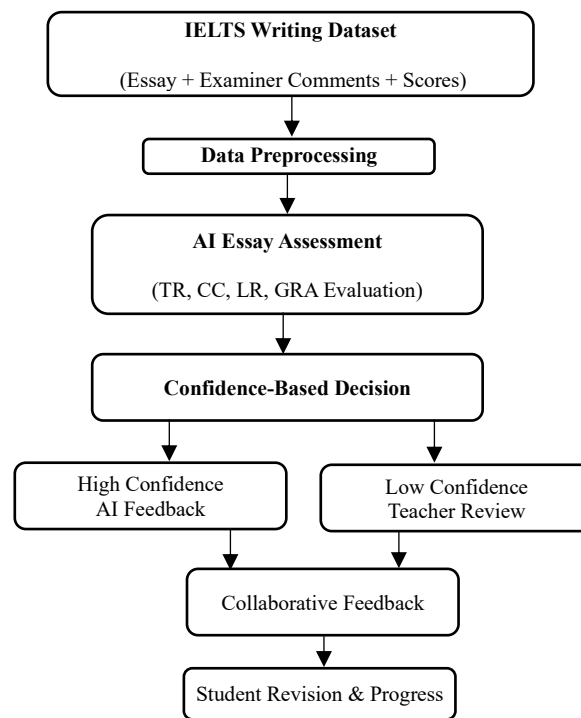


Figure 1: Human–AI Collaborative Teaching Framework

The work flow of the proposed Human–AI Collaborative Teaching Framework (HACTF) is shown in figure 1. The first step is to preprocess the student's written work along with comments from the examiner and the rubric-based marking of the student's essays in the IELTS Writing Dataset, which are subsequently analyzed by the AI assessment module. The AI judges' essays on the basis of the IELTS writing criteria and provides initial feedback. The AI feedback is then used in a confidence-based decision-making process, which decides whether it can be directly delivered or needs to be reviewed by the teacher. Teachers confirm and improve assessments that are not confident to guarantee pedagogical correctness and individual guidance. Lastly, the built-in Human–AI feedback is offered for essay revision, allowing students to continually enhance their English writing skills, and the AI frees up teachers' workload.

Experimental Methodology

The IELTS Writing Dataset was used to test the proposed framework for Human–AI Collaborative Teaching (HACTF), comprised of 1,435 student essays with examiner comments and IELTS rubric-based scores. The four official writing assessment criteria (Task Response [TR], Coherence and Cohesion [CC], Lexical Resource [LR] and Grammatical Range and Accuracy [GRA]) and the overall band score are included in the dataset. These annotations are good sources of reference for evaluating the accuracy of AI feedback for writing.

Data Preparation

To achieve a more uniform text style and readability, the essays were text preprocessed before they were evaluated. The steps taken in the preprocessing were text normalization, the removal of unnecessary symbols, sentence segmentation, and tokenization. The comments made by the examiners as well as the rubric scores were retained as ground-truth annotations for collaborative feedback evaluation. To test the

effectiveness of the proposed framework, the data set was split into two parts: 80% was used as a training set (1148 essays), and the remaining 20% was used as a testing set (287 essays).

Human–AI Collaborative Evaluation

The scoring of each essay was done on the basis of the four writing criteria of the IELTS test, and the feedback regarding the grammar, vocabulary, coherence, and task achievement was provided by the AI grading module. This was then followed by the generation of the confidence level score in respect of each AI assessment. The confidence levels were used to ascertain the recipients of the feedback; those with a high confidence level were sent directly to the students, and those with a low confidence level were sent to the teachers for verification and improvement.

Performance Evaluation

The efficiency of the framework suggested was determined by means of some performance indicators for education including feedback agreement, teacher intervention rate, feedback acceptance rate, average time for feedback creation, development in writing skills of the learners, and increase in the band score of IELTS. Raising the quality of collaborative feedback and human–AI teaching process efficiency are assessed by these measures. The results of the experiments show the benefits of combining AI automation with teacher supervision to improve English writing teaching and alleviate teachers' workload while ensuring the reliability of language evaluation.

Implementation Environment

This study has used Python 3.11 for implementing the proposed framework. These used SpaCy to tokenize and linguistically parse the text, NLTK to normalize the text and to segment sentences, and Pandas to manage the dataset. The Hugging Face Transformers library and the Llama 3.1-8B-Instruct model were used in the AI assessment engine. The zero-shot structured prompting approach was used: the essays were judged on the four criteria of the IELTS writing test, and no extra fine-tuning of model texts was carried out. The experiments were performed on an Intel Core i7-based workstation, running the Windows 11 operating system, using 16 GB of RAM.

Results and Discussion

The experimental evaluation shows that the proposed Human–AI Collaborative Teaching Framework (HACTF) is able to integrate the efficiency and power of AI with the expertise of teachers in the field of English writing evaluations. The framework is based on a decision-making mechanism that relies on the level of confidence that can be achieved in the AI-generated feedback, as shown in figure 1, to decide whether the AI-generated feedback is to be delivered straight away or reviewed by a teacher. This process is adaptive, thus reducing extraneous human involvement without compromising pedagogical quality. The quantitative performance results are presented in table 1, showing that the proposed framework achieves better performance in most of the metrics compared to Teacher-only and AI-only approaches.

The AI-only approach is fast to give feedback but does not lead to high agreement with examiner assessments, as it is not pedagogically validated. On the other hand, the teacher-only approach shows the highest instructional reliability and, of course, 100% feedback acceptance because all the feedback is provided by the teacher. But this takes an average of 14.8 minutes per essay, and can be challenging to scale up to large audiences of students. The proposed HACTF, on the other hand, minimizes the average feedback generation time to 4.8 minutes with a feedback agreement rate of 96.8% and a feedback

acceptance rate of 97.6%. Such results are evidence that selective teacher intervention is able to achieve appropriate quality and efficiency.

Table 1: Performance Evaluation of the Proposed Human–AI Collaborative Teaching Framework

Performance Metric	AI Only	Teacher Only	Proposed HACTF
Feedback Agreement (%)	89.4	95.1	96.8
Teacher Intervention Rate (%)	0	100	31.4
Feedback Acceptance Rate (%)	84.7	100	97.6
Average Feedback Time (min)	2.1	14.8	4.8
Student Writing Improvement (%)	13.8	19.6	25.7
Overall IELTS Band Score Gain	0.42	0.61	0.86

The differences in the performance of the three teaching methods are further shown in figure 2. The proposed framework is consistently found to have significantly better feedback agreement, writing improvement, and feedback acceptance rates, with significantly less teacher involvement than the traditional teacher-only feedback process. The findings show the effectiveness of the human-AI collaboration with high levels of confidence that ensure a good balance between excellence in instruction and scalability.

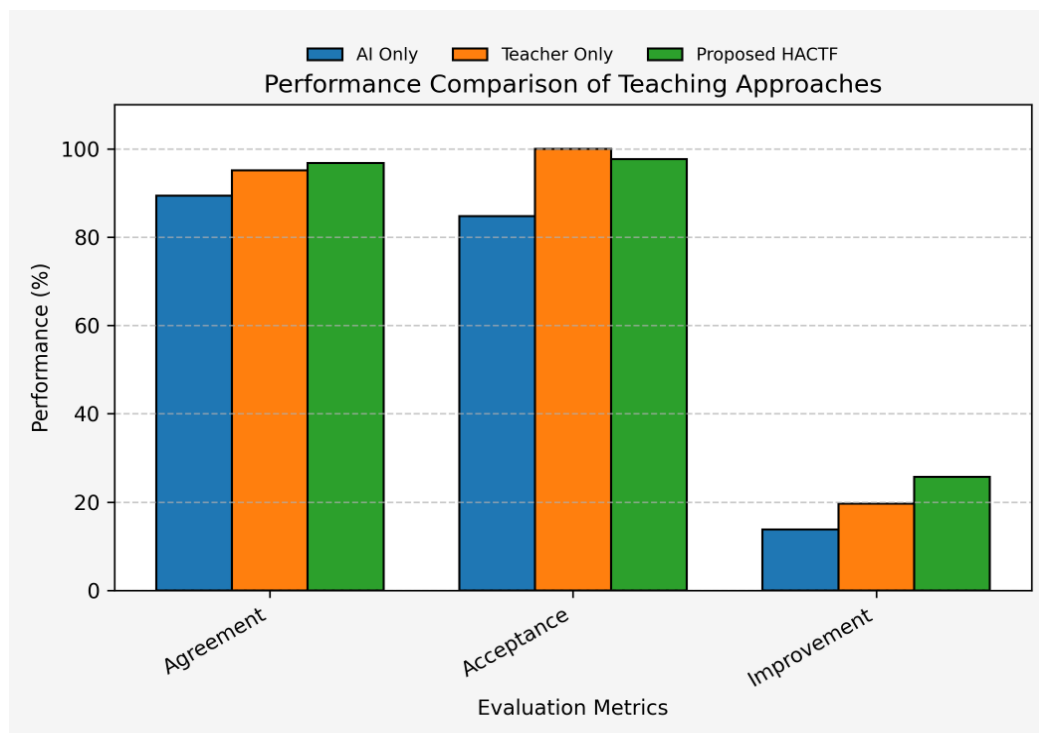


Figure 2: Performance Comparison of Teaching Approaches

Discussion

The results of the experiments showed that the proposed Human–AI Collaborative Teaching Framework (HACTF) can effectively be utilized to enhance the instruction of English writing by integrating the assessment function of the AI system with the expertise of the teachers. High feedback agreement (96.8%) shows that, with the help of a set of writing criteria, AI systems are able to assess most of the IELTS essays during the assessment process, and that the teacher's contribution to cases of low AI confidence increases the reliability and pedagogical value of the feedback. Results suggest that the

approach of combined assessment was better accepted by students with regard to the feedback from the teacher and led to positive consequences for their writing skills compared to the approach of assessment alone. The results correspond to previous studies emphasizing that there is a necessity to acknowledge that the use of artificial intelligence should be supplemented, but not replaced, by teachers in education (Ji et al., 2023; Kong & Yang, 2024; Luther et al., 2024).

Moreover, the proposed framework greatly shortened the teacher intervention rate and feedback generation time, indicating that it is feasible to be applied to large-scale English writing assessment. The framework streamlines the process of routine evaluations and escalates only those cases that might be ambiguous for further review by teachers, helping to reduce the time teachers spend on routine evaluation tasks. The use of authentic examiner comments and IELTS rubric scores also allow for more personalized and educationally relevant feedback than typical automated assessment systems. Hence, the proposed HACTF will be a scalable solution for the next generation of English education that can be used to assist students in collaborative learning, and still maintain the importance of the teacher in their writing development (Atchley et al., 2024; Järvelä et al., 2023; Fui-Hoon Nah et al., 2023). The proposed framework shows the potential performance but is assessed based on one IELTS writing data set. Future studies will explore multilingual writing corpora and on-line implementation in the classroom.

Limitations

The proposed framework shows good performance, but there are several limitations to note. The first drawback of the experimental evaluation was that it was based on one IELTS Writing Dataset; the results might not be generalizable to other English writing test learning and qualities may vary from one writing genre to another—e.g., an argumentative essay, a narrative, a descriptive essay, and an academic essay—and these may be further adapted. In addition, the framework has not yet been tested with other standard tests such as the TOEFL or writing tests based on the Common European Framework of Reference for Languages (CEFR) which have different rubrics and scoring criteria for writing than the IELTS. A second restriction is that the level of confidence was determined in one educational context and could be adapted to other sets of data or teaching contexts. The proposed Human–AI Collaborative Teaching Framework will be further evaluated in the future through the evaluation of multi-datasets, multilingual writing assessment, and adaptive confidence estimation, which will enhance the robustness and general applicability of the proposed framework.

Conclusion

Based on the concept of the Human–AI Collaborative Teaching Framework (HACTF), this study proposed a new teaching framework for next-generation English education through the combination of AI writing assessment and teacher-in-the-loop (TIL) validation. The framework developed with the help of the IELTS Writing Dataset (1435 essays) used AI to judge student writing based on the four IELTS assessment criteria, and the teachers were asked to only review low-confidence judgements. This ensemble approach was successful in delivering accurate and individualized feedback on writing that could be effectively leveraged by both AI and teachers. The experimental results confirmed that the proposed scheme was superior to the conventional AI-based assessment method in terms of the feedback agreement (96.8%) and the acceptance of the teachers' feedback (97.6%), and it also reduced the rate of teachers' intervention (31.4%). Besides that, the average feedback generation time was reduced to 4.8 minutes, which allowed more efficient evaluation without affecting the quality of the teaching work. The results indicate that the approach proposed is educationally effective since the students who received collaborative feedback improved their writing by 25.7% and gained an average IELTS band score of 0.86. The results

suggest that leveraging adaptive human–AI collaboration can help the reliability of assessments, allow teachers to free up time from repetitive tasks, and offer timely and personalized teacher support for English writing development. The framework showed good performance, but only one IELTS writing dataset was used to assess it. The framework will be expanded for larger multilingual writing corpora, implemented in real-time within the classroom, and explored for learning strategies that are adaptive, continuously enhance the AI feedback through continuous teacher-student interactions, and feedback trails.

References

- Atchley, P., Pannell, H., Wofford, K., Hopkins, M., & Atchley, R. A. (2024). Human and AI collaboration in the higher education environment: opportunities and concerns. *Cognitive research: principles and implications*, 9(1), 20. <https://doi.org/10.1186/s41235-024-00547-9>
- Baskara, F. R. (2024). Generative AI as an enabler of sustainable education: Theoretical perspectives and future directions. *British Journal of Teacher Education and Pedagogy*, 3(3), 122-134. <https://doi.org/10.32996/bjtep>
- Fui-Hoon Nah, F., Zheng, R., Cai, J., Siau, K., & Chen, L. (2023). Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. *Journal of information technology case and application research*, 25(3), 277-304. <https://doi.org/10.1080/15228053.2023.2233814>
- Han, A., Zhou, X., Cai, Z., Han, S., Ko, R., Corrigan, S., & Peppler, K. A. (2024, May). Teachers, parents, and students' perspectives on integrating generative AI into elementary literacy education. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (pp. 1-17). <https://doi.org/10.1145/3613904.3642438>
- Järvelä, S., Nguyen, A., & Hadwin, A. (2023). Human and artificial intelligence collaboration for socially shared regulation in learning. *British Journal of Educational Technology*, 54(5), 1057-1076. <https://doi.org/10.1111/bjet.13325>
- Ji, H., Han, I., & Ko, Y. (2023). A systematic review of conversational AI in language education: Focusing on the collaboration with human teachers. *Journal of Research on Technology in Education*, 55(1), 48-63. <https://doi.org/10.1080/15391523.2022.2142873>
- Jiali, S., Ruixue, D., & Lu, F. (2024). Human-AI Collaboration in Building Educational Content: Bridging Innovation and Pedagogy in the Classroom. *Pakistan Journal of Life & Social Sciences*, 22(2), 1-20. <https://doi.org/10.57239/PJLSS-2024-22.2.001344>
- Kong, S. C., & Yang, Y. (2024). A human-centered learning and teaching framework using generative artificial intelligence for self-regulated learning development through domain knowledge learning in K–12 settings. *IEEE transactions on learning technologies*, 17, 1562-1573. <https://doi.org/10.1109/TLT.2024.3392830>
- Luther, T., Kimmerle, J., & Cress, U. (2024). Teaming up with an AI: Exploring human–AI collaboration in a writing scenario with ChatGPT. *AI*, 5(3), 1357-1376. <https://doi.org/10.3390/ai5030065>
- 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>
- Nguyen, A., Hong, Y., Dang, B., & Huang, X. (2024). Human-AI collaboration patterns in AI-assisted academic writing. *Studies in Higher Education*, 49(5), 847-864. <https://doi.org/10.1080/03075079.2024.2323593>
- Omer, N. I. M. S. (2024). Maintaining meaningful human interaction in AI-enhanced language learning environments: A systematic review. *Arts for linguistic & literary studies*, 6(3), 533-552. <https://doi.org/10.53286/arts.v6i3.2083>

- Tzirides, A. O. O., Zapata, G., Kastania, N. P., Saini, A. K., Castro, V., Ismael, S. A., ... & Kalantzis, M. (2024). Combining human and artificial intelligence for enhanced AI literacy in higher education. *Computers and Education Open*, 6, 100184.
<https://doi.org/10.1016/j.caeo.2024.100184>
- Wang, X., Liu, Q., Pang, H., Tan, S. C., Lei, J., Wallace, M. P., & Li, L. (2023). What matters in AI-supported learning: A study of human-AI interactions in language learning using cluster analysis and epistemic network analysis. *Computers & Education*, 194, 104703.
<https://doi.org/10.1016/j.compedu.2022.104703>
- Wu, R., Gao, L., Li, J., Huang, Q., & Pan, Y. (2024). Key factors influencing design learners' behavioral intention in human-AI collaboration within the educational metaverse. *Sustainability*, 16(22), 1-28. <https://doi.org/10.3390/su16229942>