AI-Powered Language Learning in ESP: A Quasi-Experimental Study on TalkPal versus Conventional Instruction among Tourism Students

Putu Tika Virginiya[1]*, I Putu Yoga Laksana[2], Ni Nyoman Sri Astuti[1], and Sang Ayu Made Krisna Dewi Natalia[1]

[1] Tourism Department, Politeknik Negeri Bali, Bali, Indonesia
[2] Business Administration Department, Politeknik Negeri Bali, Bali, Indonesia
tikavirginiya@pnb.ac.id

Abstract. This study investigates the effectiveness of TalkPal, a Generative AI-powered conversational application, in improving English for Specific Purposes (ESP) learning outcomes among Tourism Department students at Politeknik Negeri Bali, compared with conventional instruction. Using a quasi-experimental two-group pre-test/post-test design, students were assigned to an experimental group receiving TalkPal-integrated ESP instruction (n = 30) and a control group receiving conventional instruction (n = 30). English proficiency was measured using a Complexity-Accuracy-Fluency (CAF) speaking task, and students' perceptions of the use of TalkPal were explored through semi-structured interviews analyzed thematically. As the data were not normally distributed (Shapiro Wilk, $p < .05$), non-parametric statistics were used. The experimental group showed a significant increase in ESP proficiency ($p < .001$), with a mean N-Gain of 39.83% (moderate category), significantly outperforming the control's group N-Gain of 17.32% (low category). The between-group difference at post-test was also significant (Mann Whitney $U = 4.000$, $p < .001$), corresponding to a rank-biserial effect size of $r = .99$, indicating a very large effect. Thematic analysis corroborated these quantitative gains, highlighting real-time feedback, reduced speaking anxiety, and increased motivation as contributing mechanisms, alongside reported drawbacks associated with the application free-tier limitations. The study concludes that TalkPal is effective interactive medium for developing ESP speaking, grammar, and pronunciation skills when embedded with a structured task-based instructional sequence, though findings should be interpreted in light of the study's short duration, single-institution sample, and reliance on the free version of the application.

Keywords: English for Specific Purposes, Generative AI, TalkPal

1 Introduction

Artificial Intelligence has restructured the educational landscape, opening new possibilities for language learning across academic contexts. Holmes & Tuomi (2022) position AI in Education (AIED) as bridging administration, teacher support, and emerging pedagogical approaches. Within English for Specific Purposes (ESP),

© The Author(s) 2026
W. G. S. Parwita et al. (eds.), *Proceedings of the International Conference on Sustainable Green Tourism Applied Science - Social Applied Science 2026 (ICoSTAS-SAS 2026)*, Advances in Social Science, Education and Humanities Research 1041,
https://doi.org/10.2991/978-94-6239-809-2_41

however, its integration conveys particular urgency, because ESP learners require not only general communicative competence, but also domain-specific linguistic proficiency tied to occupational performance (Aguaguiña Pilla et al., 2025; Costa & Mastellotto, 2022). For tourism students, that proficiency is tangible and visible, from greeting and handling guests at the front office, describing an itinerary to visitors, responding to complaints, and mediating across cultures with clients of varied linguistic backgrounds. These are spoken, interactional, and largely impromptu demands.

Yet, in Indonesian polytechnic institutions, a persistent gap remains between the oral competence graduates are expected to hold and what they demonstrate in practice (Sintia et al., 2025; Suherman, 2023; Suherman & Kertawijaya, 2023). Learners' speaking time is abruptly reduced due to large, fixed, multilevel cohorts. Meanwhile, teacher-centered delivery and outdated materials focusing on grammatical form over the technical registers of the workplace cause students to pass examinations yet remain deficient in the very interactions their profession demands.

Conversational Generative AI is well positioned to address this limitation: AI-mediated tools can simulate occupational scenarios (Taheri et al., 2026), supply immediate corrective feedback (Wang, 2024), adapt to individual proficiency (Chaudhry & Kazim, 2022), and sustain dialog interaction (Tang & Putra, 2025) which affordances are difficult to deliver at scale in a crowded classroom. TalkPal is one such platform, offering interactive spoken practice across customizable scenarios with in-conversation error flagging and contextually appropriate lexical suggestions (Aras & Taufan, 2026; Hidayatullah, 2024; Nguyen, 2025; Purbasari, 2025).

However, no prior study has evaluated it within a vocational tourism ESP setting: existing work examines TalkPal in general EFL rather than ESP (Dikaprio & Diem, 2024), and comprises literature reviews (Nguyen, 2025) or perception studies (Purbasari, 2025) rather than controlled comparisons against conventional instruction. The present study addresses this gap by combining a two-group quasi-experimental design, scored against the Complexity-Accuracy-Fluency (CAF) framework on tourism-specific tasks, with thematic analysis of learner perceptions, to establish not only whether TalkPal improves ESP outcomes, but why it does so and under what constraints.

2 Methodology

The study employed a mixed-methods approach combining a quasi-experimental design with qualitative interviews. A two-group pre-test and post-test design was used to compare the effectiveness of TalkPal-integrated ESP instruction (n = 30) and conventional classroom instruction among Tourism Department students at Politeknik Negeri Bali (n = 30), while interviews explored students' perceptions of using TalkPal. A quasi-experimental design was considered appropriate given the practical constraints inherent in institutional educational research because students were already enrolled in intact class sections prior to the commencement of the study; random assignment at the individual level was not feasible (Creswell, 2009).

The intervention lasted four weeks, comprising four instructional sessions (3 x 50 minutes per session) with equivalent instructional time for both groups. Following Bui & Skehan (2018), instruction was organized into pre-task, during-task, and post-task

phases. In the experimental group, TalkPal was integrated into guided classroom activities and independent practice, with students completing at least 10 minutes of daily interaction using the application's free version. The experimental group used TalkPal's free version for at least 10 minutes daily, integrated into guided classroom activities and independent practice, receiving immediate grammatical and lexical feedback during conversations; the control group followed the same instructional sequence without AI-assisted speaking practice. Following the intervention, semi-structured interviews were conducted with purposively selected students from the experimental group (30 students) and analyzed thematically following Braun & Clarke (2013) six-phase process. As coding was conducted by one researcher, formal inter-coder reliability statistics were not applicable. QDA Miner Lite was used to facilitate coding and organize qualitative data.

English proficiency was assessed via an audio-recorded guided narration of a local tourist attraction, administered as both pre-test and post-test and scored using the Complexity, Accuracy, and Fluency (CAF) rubric (Bui & Skehan, 2018), a framework aligned with the course's ESP objectives and covering grammatical/lexical range (complexity), correctness (accuracy), and smoothness of delivery (fluency). Pre-test data were collected before the intervention, and post-test data were collected at the end of the four-week treatment period. Quantitative data were analyzed using IBM SPSS Statistics (Version 27). Data normality was assessed using the Shapiro-Wilk test; homogeneity of variance was examined using Levene's test. As the data were not normally distributed (Sig. < 0.05), non-parametric analyses were employed, including: (1) Wilcoxon Signed-Rank test, to examine within-group improvement; (2) Mann-Whitney U test, to compare post-test scores between groups; and (3) Normalized Gain (N-Gain) analysis, to calculate learning effectiveness using the formula below:

$$\text{N-Gain}=(\text{Posttest Score}-\text{Pretest Score})/(\text{Maximum Score}-\text{Pretest Score}) \quad (1)$$

N-Gain values were classified as low (<30%), moderate (30% - 70%), or high (>70%).

3 Result and Discussion

3.1 Result

Normality Test. In this study, given that the sample size in this study was small (N=30 per group), the Shapiro-Wilk test was used to assess normality (see Table 1).

Table 1. Normality Test Result

Group	Sig.	Interpretation
Pre-Test Control Group	<0.001	Not normal
Post-Test Control Group	<0.001	Not normal
Pre-Test Experimental Group	<0.001	Not normal
Post-Test Experimental Group	<0.001	Not normal

Source: Processed data, 2026

The consequences of this failure to meet the assumption of normality underpin the researchers' decision to use non-parametric statistical analysis in the subsequent hypothesis testing.

Homogeneity Test. To assess the equality of variances between the control group and the experimental group in the pre-test and post-test data, a homogeneity test was conducted (see Table 2).

Table 2. Homogeneity Test Result

Score	Sig.	Interpretation
Based on Mean	0.322	Homogenous
Based on Median	0.222	Homogenous

Source: Processed data, 2026

Based on the results of Levene's Test, a significance value of 0.322 was obtained. As this significance value is greater than 0.05 (0.322 > 0.05), it can be concluded that the variances of the post-test data between the control group and the experimental group are homogeneous.

Hypothesis Testing. As the data were not normal, a Wilcoxon signed-rank test was conducted (see Table 3).

Table 3. Wilcoxon Signed-Rank Testing

Group	Negative Ranks	Positive Ranks	Ties	Sum of Ranks	Z/Sig.	r
Experimental Group	0	30	0	465.00	Z=-4.885 p <.001	.89
Control Group	0	29	1	435.00	Z=-5.231 p<.001	.96

Source: Processed data, 2026

The analysis shows an Asymp. Sig. (2-tailed) value of <.001 for both groups. In the experimental group, all 30 students (100%) showed an increase in scores from pre-test to post-test (positive ranks = 30, ties = 0, negative ranks = 0), with a sum of ranks of 465.00 (Z = -4.885, p < .001). In the control group, 29 of 30 students improved, with one student showing no change between pre-test and post-test (ties = 1), implied by the reported rank totals (sum of ranks = 435.00, Z = -5.231, p < .001). The effect size was r = .89 for the experimental group and r = .96 for the control group; both very large conventional benchmarks. That produced a statistically reliable improvement, and the more informative comparison for evaluating TalkPal's added value is the between-group difference at post-test (Mann-Whitney test, below) together with the N-Gain analysis, which corrects for each group's starting point.

Comparison of final scores between groups (Mann-Whitney test)

Table 4. Mann-Whitney tests

	Score
Mann-Whitney U	4.000
Wilcoxon W	469.000
Z	-6.571
Asymp. Sig.	<.001
Rank-biserial r (from U)	.99 (very large)
r (Z/√N, N = 60)	.87 (very large)

Source: Processed data, 2026

The test results yielded an Asymp. Sig. (2-tailed) Value of 0.001 and a Z of -6.571 (see Table 4). This indicates that there is a significant difference in final performance between the group of students who studied using the TalkPal app and the group of students who used conventional teaching methods. Two effect sizes were computed for this comparison: the rank biserial correlation, derived directly from the reported U and group sizes, and r, which gave r = .87. Both indicate a very large effect by conventional benchmarks and lead to the same substantive conclusion, though they are not numerically identical, since the two formulas standardize the test statistic differently. Reporting both make clear that near-complete separation between the two groups' post-test rank distributions reflects a substantial difference in outcomes.

N-Gain Score. To strengthen the analysis of effectiveness and eliminate the bias arising from students' prior ability, the researchers calculated the Normalised Gain (N-Gain) scores for both groups. The results of the descriptive analysis of the N-Gain data are presented in the Table 5.

Table 5. N-Gain Score

Group	Pre-Test Mean	Post-Test Mean	N-Gain	Interpretation
Control Group	6.033	7.067	17.32%	Low
Experimental Group	7.900	9.533	39.83%	Moderate

Source: Processed data, 2026

Based on the table above, the average N-Gain Percentage in the control group was only 17.32%, which, according to classification, falls into the low category. Meanwhile, the average N-Gain Percentage for the experimental group reached 39.83%, which falls into the moderate category. Descriptively, the effectiveness of the improvement in ability in the experimental class using TalkPal was found to be more than twice as high as that of the control class.

Thematic Analysis. Using the QDA Miner Lite application, thematic analysis was done following Braun & Clarke (2013). The Table 6 shown below are the top codes/sub-themes among the six themes developed.

Table 6. Top Codes of Thematic Analysis Result

Themes	Sub-themes	Count	Cases
Affective experience	Motivation to keep learning	16	13
	Heightened skill awareness	25	13
	Increased confidence	17	12
	Enjoyment & fun	14	12
Perceived learning gains	Instant / real-time feedback	34	17
	Grammar improvement	21	12
	Speaking & fluency gains	15	12
	Error awareness & self-correction	16	11
Usability	Easy / user-friendly	21	20
	Simple, clear interface	10	10
	Multimple modes & features	16	9
Access, autonomy & task support	Task & assignment support	14	8
	Saves time & effort	9	6
Access constraints	Paywall / premium-locked features	14	13
	Limited free feature set	14	12
Limits of the AI	Repetitive / generic responses	4	4
	Internet dependency	3	3

Source: Processed data, 2026

3.2 Discussion

Based on the research results, this study suggests that TalkPal-integrated instruction yields better ESP learning outcomes than conventional instruction among Tourism Department students. However, the value of these results lies not merely in the statistical significance achieved but in what they reveal about how Generative AI conversational practice produces higher, more measurable language gains. A more qualified picture of the use of TalkPal perceptions that they jointly reveal about mechanisms aligns with existing theoretical accounts of second language acquisition and vocational skills development through Generative AI conversational practice. These shed light on the results within the broader context of AI research in language education and consider the theoretical and practical implications that follow.

The Effectiveness of Talkpal-Integrated Instruction. The clearest finding of this study is that integrating TalkPal into ESP instruction led to stronger gains than the conventional teaching did. Students in the experimental group improved their post-test scores by an N-Gain of 39.83% (a moderate gain), more than double the 17.32% recorded by the control group (a low gain), and the difference between the two groups at post-test was statistically significant (Mann-Whitney U = 4.000, p.<.001). What gives this result weight beyond the numbers is how consistent it was that every one of the thirty students in the experimental group improved, with no ties and no student declining. In a fixed, multilevel cohort of the kind that characterizes Indonesian polytechnics ESP, gains are more usually concentrated among students who already

speak with some confidence, while the hesitant remain where they started. Improvement of this kind, spread evenly across a whole class rather than concentrated among already-confident learners, is what one would hope to see from a tool meant to reach learners a crowded classroom tends to leave behind. That this did not happen here suggests the intervention reached across the proficiency range rather than amplifying an existing advantage, which is precisely the problem, identified at the outset of this paper, that large-class ESP instruction has struggled to solve.

The result also echoes earlier work by Hidayatullah (2024) and Aras & Taufan (2026), who, for instance, likewise found TalkPal effective for developing university students' speaking skills, and the present study extends that finding from general EFL into a vocational ESP setting. What differs is the setting and the measure, as prior studies examine general EFL proficiency, whereas the gains reported here were scored on tourism-specific communicative tasks against the Complexity, Accuracy, and Fluency framework (Bui & Skehan, 2018). The improvement therefore concerns occupational language use rather than general conversational competence. This matters for ESP as a field because the discipline's central claim is that domain-specific proficiency is not simply general proficiency applied to a new topic. Demonstrating that a general-purpose conversational AI produces measurable gains on domain-specific tasks is a modest but genuine extension of what was previously known.

Why it Works: Pedagogical Mechanisms. The quantitative findings demonstrate that TalkPal contributed to significant improvements in students' speaking performance; however, the qualitative data provide important insights into the mechanisms underlying these gains. Rather than merely identifying features that students appreciated, the interview findings reveal an interconnected learning process through which the application supported speaking development.

A major theme was enhanced metalinguistic awareness. Students frequently reported becoming more aware of their language use and recognizing errors they had previously overlooked. This supports Schmidt's (2012) Noticing Hypothesis, which emphasizes conscious attention to linguistic forms as a prerequisite for acquisition, and Swain's (2005) Output Hypothesis, which argues that language production helps learners identify gaps between intended meaning and actual performance. Unlike conventional classrooms, where feedback is often delayed, TalkPal enabled learners to produce language, receive corrective feedback, and revise errors within the same interaction.

The timing of feedback emerged as another critical factor influencing learning. "Instant real-time feedback" was the most frequently cited benefit in the interviews, appearing 34 times across 17 participants, while "grammar improvement" was reported 21 times by 12 participants. These perceptions correspond closely with quantitative improvements observed in the accuracy component of the CAF framework. Previous studies have suggested that the effectiveness of AI-generated corrective feedback is partly attributable to its immediacy. By minimizing the delay between language production and correction, learners can revise inaccurate forms before they become reinforced and can immediately apply the corrected forms in subsequent interactions (Wang, 2024). Such rapid feedback is likely to strengthen procedural knowledge of

linguistic forms that learners have partially required, thereby contributing more substantially to improvements in grammatical accuracy and fluency than to linguistic complexity. This interpretation is consistent with the pattern of gains observed in the present study.

The interviews further revealed that TalkPal reduced speaking anxiety by providing a private, non-judgmental environment for practice. This finding aligns with Huang & Zou (2024) who suggests that Willingness to Communicate with AI significantly and positively influenced learners' perceived usefulness and satisfaction, which in turn directly drove their continuance intention; and Krashen's Affective Filter Hypothesis (Gidman, 2025), which suggests that lower anxiety encourages greater language production. Increased motivation to continue practicing indicates that students were more willing to engage in repeated speaking activities, particularly those who might otherwise remain silent in classroom settings.

Overall, the findings suggest that TalkPal's primary contribution lies not in introducing new pedagogical approaches but in substantially increasing opportunities for meaningful speaking practice through repeated cycles of production, noticing, and immediate corrective feedback. However, these results support the effectiveness of TalkPal only when integrated into a structured task-based instructional (Bui & Skehan, 2018), rather than as a standalone learning tool. This conclusion is consistent with recent evidence indicating that AI-assisted speaking applications are most effective when embedded within classroom pedagogy (Bhar, 2026).

Implications and limitations. The findings suggest that the effectiveness of TalkPal depends not only on its features but also on its pedagogical integration. In this study, TalkPal functioned as a rehearsal tool within a task-based instructional sequence: explicit instruction introduced the target language, AI-supported practice provided intensive repetition, and classroom activities offered authentic communicative performance. Consequently, AI practice should be aligned with occupation-specific ESP tasks, such as hotel check-in, complaint handling, site narration, and itinerary explanation, rather than generic conversation. Furthermore, a minimum amount of guided practice should be specified and monitored to ensure sufficient speaking opportunities.

These findings have important implications for vocational ESP instruction, particularly in contexts characterized by limited instructional time, large classes, and mixed language proficiency. Conversational AI can expand individual speaking opportunities without increasing teaching workload, suggesting that AI-supported practice should be incorporated into the curriculum as structured learning activities rather than optional enrichment. In this context, instructors remain essential in designing authentic tasks, monitoring learner progress, and facilitating classroom interaction. The results also highlight the need to reconsider speaking assessment. Reliance on a small number of high-stakes oral tests may overlook the gradual improvements fostered through frequent AI-supported practice, particularly in fluency, self-correction, and learner engagement. More frequent formative speaking assessments may therefore better capture students’ language development.

Several limitations should be acknowledged. First, students used only the free version of TalkPal, with limited features and restricted interaction time, which may have constrained learning opportunities. Second, participants reported repetitive conversations as well as dependence on stable internet access and personal devices, challenges similarly identified by Purbasari (2025). These technological and financial constraints may affect the scalability of AI integration in Indonesian vocational education.

Finally, the findings should be interpreted within the context of a four-week intervention conducted at a single institution. Future studies should examine whether the observed improvements are sustained over longer periods and determine whether the benefits reflect durable language development rather than the novelty of using AI-assisted learning.

4 Conclusion

This study demonstrates that TalkPal significantly enhanced tourism students' ESP speaking proficiency compared with conventional instruction, with the experimental group achieving higher CAF-based post-test scores and more than double the N-Gain of the control group. Extending previous TalkPal research in general EFL, the findings suggest that its effectiveness stems from increasing opportunities for spoken output, providing immediate corrective feedback, and reducing speaking anxiety when integrated into a structured task-based instructional cycle. These results indicate that conversational AI can help address common challenges in vocational ESP, including limited contact hours, large classes, and diverse learner proficiency. Nevertheless, the findings should be interpreted with caution due to the quasi-experimental design, small sample from a single institution, short intervention period, and the use of the free version of TalkPal, all of which limit generalizability. Furthermore, the study assessed speaking performance only in test settings, leaving the transferability of learning to authentic workplace communication unexamined. Future research should employ longitudinal, multi-institutional designs with larger samples, compare free and premium AI implementations, and investigate learners' actual interactions with AI to better understand the mechanisms and long-term effectiveness of AI-assisted ESP learning.

References

Aguaguiña Pilla, M. P., Gordón Fiallos, B. A., Casa Molina, P. E., & Nuñez Melo, N. N. (2025). Development of communicative competence in specific contexts: Teaching English for Specific Purposes (ESP) in fields such as Medicine, Business, and Technology. *Ciencia Latina Revista Científica Multidisciplinar*, *9*(1), 10262–10276. https://doi.org/10.37811/cl_rcm.v9i1.16636

Aras, M. S., & Taufan, M. Y. (2026). Using TalkPal AI to Enhance Speaking Skills: an Experimental Study among EFL Learners. *Al-Irsyad: Journal of Education Science*, *5*(1), 348–359.

Bhar, S. K. (2026). Artificial Intelligence in EFL Speaking Instruction: A Systematic Review of Pedagogical Design, Affective Conditions and Instructional Input. In *Encyclopedia* (Vol. 6, Number 4, pp. 1–22). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/encyclopedia6040074

Braun, V., & Clarke, V. (2013). *Successful Qualitative Research: A Practical Guide for Beginners* (M. Carmichael, Ed.). SAGE Publications Ltd.

Bui, G., & Skehan, P. (2018). Complexity, Accuracy, and Fluency. In *The TESOL Encyclopedia of English Language Teaching* (pp. 1–7). Wiley. https://doi.org/10.1002/9781118784235.eelt0046

Chaudhry, M. A., & Kazim, E. (2022). Artificial Intelligence in Education (AIEd): a high-level academic and industry note 2021. *AI and Ethics*, *2*(1), 157–165. https://doi.org/10.1007/s43681-021-00074-z

Costa, F., & Mastellotto, L. (2022). The role of English for specific purposes (ESP) in supporting the linguistic dimension in English-medium instruction (EMI). *CLIL Journal of Innovation and Research in Plurilingual and Pluricultural Education*, *5*(2), 37–52. https://doi.org/10.5565/rev/clil.91

Creswell, J. W. (2009). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (Third Edition). Sage Publications, Inc.

Dikaprio, V., & Diem, C. D. (2024). How Effective is Talkpal.ai in Enhancing English Proficiency? Insights from an Experimental Study. *Language, Technology, and Social Media*, *2*(1), 48–59. https://doi.org/10.70211/ltsm.v2i1.48

Gidman, H. H. (2025). *The role of krashen's affective filter hypothesis on language learning effectiveness, teacher training and professional development*.

Hidayatullah, E. (2024). The Impact of Talkpal.AI on English Speaking Proficiency: An Academic Inquiry. *Journal of Insan Mulia Education*, *2*(1), 19–25. https://doi.org/10.59923/joinme.v2i1.98

Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. *European Journal of Education*, *57*(4), 542–570. https://doi.org/10.1111/ejed.12533

Huang, F., & Zou, B. (2024). English speaking with artificial intelligence (AI): The roles of enjoyment, willingness to communicate with AI, and innovativeness. *Computers in Human Behavior*, *159*, 108355. https://doi.org/10.1016/j.chb.2024.108355

Nguyen, C. Van. (2025). Employing TALKPAL.AI to Enhance Speaking Proficiency for Vietnamese Adult Learners: A Literature Review. *International Journal of AI in Language Education*, *2*(2), 40–54. https://doi.org/10.54855/ijaile.25223

Purbasari, D. A. (2025). Students' Perspective and Challenge towards Utilization of Talkpal.AI in ESP Classroom. *Proceeding International Conference on Economy, Education, Technology, and Environment*, *3*(1), 18–30. https://prosiding.utp.ac.id/index.php/ICEETE

Schmidt, R. (2012). Attention, awareness, and individual differences in language learning. *Perspectives on Individual Characteristics and Foreign Language Education*, *6*(27), 27–49.

Sintia, D., Perdana, I., & Karani, E. (2025). *ESP in Practice: A Literature Review of Challenges Faced by Teachers and Learners* (Vol. 9, Number 2).

Suherman, A. (2023). Students' Perceptions of the Challenges of ESP Courses: A Study from an Indonesian University. *Indonesian Journal of EFL and Linguistics*, 193–205. https://doi.org/10.21462/ijefl.v8i2.645

Suherman, A., & Kertawijaya, L. (2023). Challenges Perceived by Indonesian ESP Teachers in Post-Pandemic Era: A Survey from a Higher Education Institution. *Scope : Journal of English Language Teaching*, *7*(2), 129. https://doi.org/10.30998/scope.v7i2.15526

Swain, M. (2005). The output hypothesis: Theory and research. In *Handbook of research in second language teaching and learning* (pp. 471–483). Routledge.

Taheri, M., Tan, K., Flaten, O. E., & Cyma-Wejchenig, M. (2026). Enhancing Education Training with A Virtual Reality-Based Simulator Integrated Generative AI. In L. T. De Paolis, P. Arpaia, & M. Sacco (Eds.), *Extended Reality* (pp. 394–406). Springer Nature Switzerland.

Tang, K. S., & Putra, G. B. S. (2025). Generative AI as a Dialogic Partner: Enhancing Multiple Perspectives, Reasoning, and Argumentation in Science Education with Customized Chatbots. *Journal of Science Education and Technology*. https://doi.org/10.1007/s10956-025-10240-1

Wang, D. (2024). Teacher-Versus AI-Generated (Poe Application) Corrective Feedback and Language Learners' Writing Anxiety, Complexity, Fluency, and Accuracy. *International Review of Research in Open and Distributed Learning*, *25*(3 Special Issue), 37–56. https://doi.org/10.19173/irrodl.v25i3.7646

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

