

SCAFFOLDING INDEPENDENT EFL WRITING THROUGH AI CHATBOT PROMPTING GUIDANCE: A MIXED-METHODS STUDY ON ANALYTICAL EXPOSITION WRITING

Novan Kurnia Rahardika¹, Buntara Adi Purwanto²
dhickhatemangung@gmail.com¹, buntara@unsig.ac.id²
Universitas Sains Al-Qur'an

ABSTRAK

Menulis teks Analytical Exposition secara mandiri merupakan tugas yang berat bagi siswa Sekolah Menengah Atas yang belajar Bahasa Inggris sebagai bahasa asing (EFL), karena menuntut kemampuan argumentasi, ketepatan tata bahasa, dan pengaturan diri yang belum banyak dikuasai siswa. Meskipun chatbot AI generatif semakin banyak digunakan untuk mendukung pembelajaran menulis EFL, penggunaan tanpa panduan berisiko mengalihkan upaya kognitif siswa alih-alih mengembangkannya. Penelitian ini mengembangkan dan menguji teknik Prompting Guidance, sebuah pendekatan pembelajaran yang melatih siswa menyusun prompt terstruktur sehingga chatbot AI berfungsi sebagai perancah (scaffold), bukan sebagai penghasil teks, sekaligus menggali persepsi siswa terhadap teknik tersebut. Penelitian menggunakan metode campuran dengan desain Non-Equivalent Control Group Pretest-Posttest yang melibatkan 50 siswa kelas XI (25 kelas eksperimen, 25 kelas kontrol); data dikumpulkan melalui tes menulis, kuesioner persepsi, wawancara semi terstruktur, dan observasi kelas. Skor menulis kelas eksperimen meningkat signifikan dari pre-test ke post-test (selisih rata-rata = 6,81; $t = 13,643$; $p < 0,001$) dan mengungguli kelas kontrol pada post-test (selisih rata-rata = 4,18; $t = 6,963$; $p < 0,001$). Siswa menilai teknik ini sangat positif ($M = 4,27$) dan menunjukkan peningkatan perencanaan serta sikap kritis terhadap umpan balik AI, meskipun satu dari empat siswa yang diwawancarai masih sangat bergantung pada AI. Temuan ini memposisikan prompting terstruktur sebagai cara mengubah chatbot AI menjadi "More Knowledgeable Other" yang menumbuhkan, bukan menggantikan, kompetensi menulis mandiri siswa.

Kata Kunci: Chatbot Ai, Panduan Prompting, Pembelajaran Mandiri, Pembelajaran Berbasis Regulasi Diri, Penulisan Analytical Exposition.

ABSTRACT

Writing an Analytical Exposition text independently is a demanding task for senior high school English as a Foreign Language (EFL) learners, since it requires argumentative reasoning, grammatical accuracy, and self-regulation that many students have not yet mastered. While generative AI chatbots are increasingly used to support EFL writing, unguided use risks shifting cognitive effort away from students rather than developing it. This study developed and tested a Prompting Guidance technique, an instructional approach that trains students to compose structured prompts so that AI chatbots function as a scaffold rather than a text generator, and examined students' perceptions of the technique. A mixed-methods, Non-Equivalent Control Group Pretest-Posttest design was employed involving fifty eleventh-grade students (25 experimental, 25 control), with data drawn from writing tests, a perception questionnaire, semi-structured interviews, and classroom observation. The experimental group's writing scores improved significantly from pre- to post-test (mean gain = 6.81, $t = 13.643$, $p < .001$) and outperformed the control group on the post-test (mean difference = 4.18, $t = 6.963$, $p < .001$). Students rated the technique very positively ($M = 4.27$) and showed increased planning and critical evaluation of AI feedback, although one of the four interviewed students remained strongly AI-dependent. The findings position structured prompting as a way of turning AI chatbots into a "More Knowledgeable Other" that fosters, rather than replaces, independent writing competence.

Keywords: Ai Chatbot, Prompting Guidance, Independent Learning, Self-Regulated Learning, Analytical Exposition Writing.

INTRODUCTION

Writing remains one of the most cognitively demanding skills for English as a Foreign Language (EFL) learners, requiring a higher level of language accuracy, structural organization, and self-regulation than oral communication (Richards & Renandya, 2002). This challenge is amplified when secondary school curricula, such as Indonesia's Kurikulum Merdeka, require students to produce argumentative genres like the Analytical Exposition, a text type that demands not only accurate vocabulary and syntax but also the ability to construct logical, persuasive arguments (Hyland, 2019). Classroom observation in a senior high school setting indicated that many eleventh-grade students struggled to compose these texts independently: they frequently experienced writer's block, had difficulty organizing the generic structure of thesis, arguments, and reiteration, and showed little intrinsic motivation to revise their drafts. Such difficulties point to weak self-regulated learning (SRL) strategies, which Teng and Zhang (2022) identify as a critical predictor of success in second-language writing.

Generative Artificial Intelligence (AI), particularly large language model-based chatbots such as ChatGPT and Gemini, has emerged as a potential tool to bridge this gap. Kuhail et al. (2023) argue that AI chatbots can reduce students' cognitive load by functioning as an on-demand virtual tutor, while Waziana (2024) reports that appropriately integrated chatbots can enhance EFL students' writing proficiency through adaptive, real-time scaffolding. At the same time, recent scholarship cautions that AI-assisted writing is a double-edged sword. Baidoo-Anu and Owusu-Ansah (2023) warn that unguided reliance on AI can undermine learning when students use chatbots merely to generate finished text, and Sharples (2022) similarly notes that without pedagogical structure, students risk submitting AI-generated essays as their own, bypassing the reflective thinking that writing is meant to cultivate. Systematic reviews reinforce this concern: Huang, Hew, and Fryer (2022) found that although chatbots reduce learners' anxiety and support access to learning resources, most conventional chatbot applications lack the pedagogical design needed to build long-term self-regulated learning habits, while Su, Lin, and Lai (2023) observed that EFL students collaborating with ChatGPT on argumentative writing frequently struggled to compose effective prompts and to critically evaluate the AI's output. Barrot (2023) likewise concludes that ChatGPT poses genuine risks to academic integrity and cognitive development unless students possess adequate prompt engineering skills.

These findings converge on a specific gap: existing research has established both the promise and the risk of AI-assisted EFL writing, but has offered comparatively little guidance on how to structure students' interaction with AI so that it scaffolds rather than substitutes for independent thinking. Encouragingly, emerging evidence suggests that this gap is addressable through prompt engineering instruction. Lee, Teo, and Tan (2024) demonstrate that specific, student-generated prompts can shift an AI's output from passive generation toward active pedagogical scaffolding, and Liu, Sihes, and Lu (2025) similarly report that EFL students trained to construct structured prompts engage with AI more reflectively and critically than those who interact with it passively. Building on this line of work, the present study developed and tested a Prompting Guidance technique: rather than allowing students to issue open commands such as "write an essay for me," the technique trains them to compose targeted prompts that position the AI as an iterative guide through the stages of planning, drafting, and revising an Analytical Exposition text. Framed within Vygotsky's (1978) sociocultural theory, the technique repositions the AI chatbot as a "More Knowledgeable Other" operating within students' Zone of Proximal Development, providing contingent support precisely where students need it while leaving the intellectual

work of composing the essay to the students themselves.

Accordingly, this study aimed to evaluate the effectiveness of integrating AI chatbots through the Prompting Guidance technique in improving eleventh-grade students' independent Analytical Exposition writing, to determine whether this technique produced statistically significant gains compared with conventional instruction, and to explore students' perceptions and lived experiences of learning to write with AI in this structured manner. Specifically, the study addressed three questions: (1) how effective was the integration of AI chatbots through the Prompting Guidance technique in improving students' independent ability to write Analytical Exposition texts; (2) was there a statistically significant difference in writing achievement between students taught with this technique and those taught conventionally; and (3) what were students' perceptions of using AI as a learning tool for independent writing? In answering these questions, the study sought to contribute both a practical instructional model and empirical evidence on how generative AI can be integrated into EFL writing classrooms responsibly, extending scaffolding theory and self-regulated learning frameworks to a technological context that remains under-examined at the senior high school level.

METHOD

This study employed a mixed-methods design in which quantitative data from a true experiment served as the primary source of evidence, supplemented by qualitative data used to explain and contextualize the statistical results (Creswell & Plano Clark, 2018). The quantitative component followed a Randomized Pretest-Posttest Control Group design, described by Fraenkel, Wallen, and Hyun (2012) as one of the most rigorous designs for establishing causal relationships because random assignment controls for major threats to internal validity. In this design, both the experimental and control groups received a pre-test and a post-test, with only the experimental group receiving the Prompting Guidance intervention between the two measurements.

The study was conducted in a senior high school in Sukorejo, Central Java, Indonesia, over approximately six weeks during the odd semester of the 2025/2026 academic year, in a special after-school program that did not interfere with regular instructional schedules. The population comprised the entire eleventh-grade cohort of the school, totaling approximately 164 students enrolled in several parallel classes; this grade level was selected because Indonesia's Kurikulum Merdeka requires eleventh graders to master complex argumentative genres, including Analytical Exposition. A sample of 50 students was selected using simple random sampling from a complete sampling frame of the population, employing a computer-assisted random number generator, in line with Cohen, Manion, and Morrison's (2018) recommendation that every population member have an equal and independent chance of selection. These 50 students were then randomly assigned into an experimental group ($n = 25$) and a control group ($n = 25$), a procedure intended to produce statistically equivalent baseline groups and eliminate intact-class bias.

Over the six-week implementation, the experimental group was taught prompt engineering techniques that trained them to request brainstorming assistance, targeted grammar explanations, and structural feedback from an AI chatbot rather than a finished essay, while the control group received conventional, teacher-centered instruction with feedback limited by classroom time constraints. Implementation proceeded across four weekly phases: instrument validation and random sampling with pre-test administration; introduction of AI prompting or conventional explanation with initial drafting; revision using AI feedback or limited teacher feedback with text finalization; and, finally, post-test administration followed by distribution of the perception questionnaire and semi-structured

interviews.

Data were collected using four instruments to enable methodological triangulation (Creswell, 2014). First, a writing test administered before and after the intervention asked students to compose an Analytical Exposition text on a specified topic within 60 minutes, without access to AI or the internet; the pre-test and post-test topics were matched in difficulty but differed in content to control for practice effects. Essays were scored using an analytic rubric adapted from Brown and Abeywickrama (2010) covering content (30%), organization (20%), vocabulary (20%), grammar (25%), and mechanics (5%), and were independently scored by two raters; Cronbach's alpha values of .944 (pre-test) and .971 (post-test) indicated high inter-rater reliability. Second, a 25-item closed-ended perception questionnaire, organized into four subscales, usability and interaction, writing process facilitation, overcoming writing challenges, and independent learning and self-regulation, was administered to the experimental group after the post-test using a five-point scale. Third, semi-structured interviews were conducted with four purposively selected experimental-group students representing a range of improvement levels (two with the greatest gains, one with typical gains, and one with the smallest gain), following a protocol validated by an expert judge. Fourth, two observation instruments, a general behavioral checklist and a writing-process checklist, each comprising 20 indicators, were used to record students' visible engagement, difficulty, and writing behaviors during the second and third treatment meetings.

Prior to hypothesis testing, the writing test data were checked for normality using the Shapiro-Wilk test and for homogeneity of variance using Levene's test; both assumptions were satisfied ($p > .05$), justifying the use of parametric statistics. A paired-samples t-test was then used to examine within-group change in students' combined pre- to post-test scores, and an independent-samples t-test was used to compare post-test scores between the experimental and control groups, both analyzed in SPSS. Qualitative data from the questionnaire, interviews, and observation instruments were analyzed thematically, following the six-phase approach of Braun and Clarke (2006), and were subsequently triangulated with the quantitative results to explain not only whether students' writing improved, but how and for whom that improvement occurred.

RESULTS AND DISCUSSION

1. Baseline Equivalence and Assumption Checks

Before the intervention, the experimental and control groups shared a statistically comparable starting point: the pre-test mean for the experimental group was 12.98 (SD = 2.14) and for the control group was 12.70 (SD = 2.33), indicating that random assignment had produced two initially equivalent groups (Table 1). The Shapiro-Wilk test confirmed that pre-test and post-test scores in both groups were normally distributed ($p > .05$ in all four comparisons), and Levene's test confirmed homogeneity of variance for both the pre-test ($p = .840$) and post-test ($p = .070$) data, satisfying the assumptions required for parametric hypothesis testing. Inter-rater reliability was also high, with Cronbach's alpha values of .944 for the pre-test and .971 for the post-test, indicating that the two raters scored the essays consistently using the analytic rubric.

Table 1. Descriptive Statistics of Writing Scores

Group	N	Pre-test Mean (SD)	Post-test Mean (SD)
Experimental	25	12.98 (2.14)	21.74 (1.67)
Control	25	12.70 (2.33)	17.56 (2.50)

2. Effectiveness of the Prompting Guidance Technique on Writing Achievement

By the end of the intervention, the two groups had diverged substantially. The experimental group's post-test mean rose to 21.74 (SD = 1.67), compared with 17.56 (SD = 2.50) for the control group. A paired-samples t-test on the combined pre- and post-test scores showed a statistically significant within-subject gain from 12.84 to 19.65, a mean difference of 6.81 points, $t(49) = 13.64$, $p < .001$, confirming that students' writing improved significantly after the treatment period regardless of group. More critically for the study's central hypothesis, an independent-samples t-test comparing the two groups' post-test scores showed that the experimental group significantly outperformed the control group, with a mean difference of 4.18 points, $t(48) = 6.96$, $p < .001$. In percentage terms, the experimental group's writing scores improved by 40.9% from pre- to post-test, compared with a 27.7% improvement in the control group, a 13.2 percentage-point gap that is difficult to attribute to instructional time alone, since both groups received the same six-week period of instruction. On this basis, the null hypothesis was rejected, and the alternative hypothesis, that AI chatbots used through the Prompting Guidance technique produce significantly better independent writing outcomes than conventional instruction, was supported.

Table 2. Summary of Hypothesis Testing

Test	Comparison	Mean Diff.	t	df	p
Paired-samples t-test	Pre- vs Post-test (combined)	6.81	13.643	49	< .001
Independent-samples t-test	Experimental vs Control (Post-test)	4.18	6.963	48	< .001

These results are consistent with a growing body of experimental research on ChatGPT-assisted EFL writing instruction. Song and Song (2023) and Boudouaia, Mouas, and Kouider (2024) both report significant gains in organization, coherence, and linguistic accuracy among AI-assisted groups relative to conventionally taught groups, while Woo (2024) attributes such gains to the lower cognitive load and higher motivation that ChatGPT-assisted writing affords learners. The present findings extend this evidence to a senior high school EFL context in Indonesia that has received comparatively little empirical attention, and further suggest that the specific mechanism driving the gain was not exposure to AI per se, but the structured manner in which students were trained to interact with it.

3. Students' Perceptions of the AI Prompting Guidance Technique

The perception questionnaire produced an overall mean score of 4.27, placing students' attitudes toward the technique in the "Strongly Agree" category (Table 3). Among the four subscales, Independent Learning and Self-Regulation received the highest rating ($M = 4.33$), followed closely by Writing Process Facilitation ($M = 4.31$) and Overcoming Writing Challenges ($M = 4.28$), while Usability and Interaction received the lowest, though still positive, rating ($M = 4.12$). Within the Writing Process Facilitation subscale, the highest-rated items concerned idea generation ($M = 4.48$) and the relevance of AI suggestions to the assigned topic ($M = 4.44$); within the Independent Learning subscale, students most strongly endorsed their intention to continue using the technique in future writing tasks ($M = 4.52$) and reported heightened motivation to write in English ($M = 4.48$). Importantly, the item most directly addressing authorship, "I feel like I am actively writing the essay, not just copying it from an AI," still scored strongly ($M = 4.28$), suggesting that most students experienced the intervention as scaffolding rather than as a substitute for their own thinking. The only area of comparative concern was technical reliability: the lowest-rated item in the entire questionnaire concerned the frequency of technical difficulties ($M =$

3.76), even though the same subscale's item on the AI's speed and clarity of response scored among the highest overall ($M = 4.48$), indicating that the friction students experienced was largely infrastructural rather than pedagogical.

Table 3. Mean Scores of Perception Questionnaire Subscales

Subscale	Items	Mean	Category
Usability and Interaction	5	4.12	Agree
Writing Process Facilitation	7	4.31	Strongly Agree
Overcoming Writing Challenges	6	4.28	Strongly Agree
Independent Learning and Self-Regulation	7	4.33	Strongly Agree
Overall	25	4.27	Strongly Agree

This pattern of results converges with the semi-structured interviews conducted with four purposively selected students. All four reported initial confusion when first using the AI chatbot with the prompting technique, but three found it progressively easier as they practiced; the fourth (Student D) continued to struggle with composing prompts that produced the intended results. Three of the four students identified idea generation as the stage where AI assistance was most valuable, consistent with the questionnaire's highest-rated item, while the fourth relied on AI more heavily during revision, for grammar correction and word choice. All four students reported that the AI helped them organize their essays' generic structure more clearly, and, notably, all four described actively evaluating rather than passively accepting AI suggestions: when the AI proposed vocabulary that felt too formal or stylistically unsuitable, students either requested a simpler alternative or replaced the suggestion themselves. Three of the four continued to view their final essays as their own work, describing the AI's role as limited to feedback, grammar checking, and clarifying sentence structure, while the fourth, the interviewee with the smallest test-score gain, admitted feeling less confident writing without AI assistance.

These findings align closely with Vygotsky's (1978) Zone of Proximal Development and the scaffolding concept articulated by Wood, Bruner, and Ross (1976), in which a more capable other, in this case the AI chatbot, provides contingent support precisely where a learner struggles, offering prompts and alternatives rather than a finished product. They also corroborate Song and Song's (2023) observation that AI-assisted EFL writers can experience meaningful support without a uniform loss of ownership over their texts, while nonetheless echoing Cotton, Cotton, and Shipway's (2023) caution that unregulated AI use carries a real risk of dependency for at least some learners, a risk visibly instantiated by the profile of the least-improved interviewee in the present study.

4. Behavioral Change during the Writing Process

Classroom observation data, collected at the second and third treatment meetings, provided independent behavioral evidence that corroborated the self-reported perceptions above. On the general behavioral checklist, compliance with the AI-assisted writing routine rose from 60% (12 of 20 indicators observed) at Meeting 2 to 80% (16 of 20) at Meeting 3. A more striking shift appeared on the writing-process checklist, where engagement with a structured composing sequence rose from 50% (10 of 20 indicators) at Meeting 2 to 95% (19 of 20) at Meeting 3. Critically, the indicators that emerged only at Meeting 3, manually outlining ideas before drafting, determining a thesis statement before developing arguments, requesting brief hints rather than full text from the AI, developing body paragraphs systematically, and, perhaps most importantly, examining rather than immediately copying the AI's explanations of grammatical errors and rejecting suggestions that altered intended

meaning, indicate a shift from passive consumption of AI output toward active planning and critical evaluation. This pattern is consistent with Zimmerman's (1990, 2000) cyclical model of self-regulated learning, in which learners progress from weak forethought and monitoring toward a fuller cycle of planning, monitoring, and evaluation, and with the "fading" component of scaffolding described by Wood, Bruner, and Ross (1976), whereby learners require progressively less explicit structuring as they internalize a skill.

Not every behavior improved evenly, however. Signs of difficulty, staring at a blank screen for an extended period and visible expressions of confusion or frustration, were observed at both meetings, indicating that some linguistic and cognitive challenges persisted even as overall compliance increased. Notably, the single indicator never observed at either meeting was students comparing their initial draft with a revised draft, a specific and addressable gap suggesting that certain higher-order revision behaviors may require a longer intervention period than the present six-week design allowed.

5. Integration of Findings and Discussion

Triangulating these three data sources, statistical test scores, self-reported perceptions, and independent behavioral observation, produces a coherent account of both what changed and how. The quantitative gain of 6.81 points documented by the paired-samples t-test explains what changed at the level of measured writing quality; the interview finding that three of four students found AI most useful for idea generation, together with the emergence of planning behaviors only at Meeting 3, explains how that change occurred, namely through a shift toward more deliberate structuring of the writing process rather than through direct copying of AI-generated text. Similarly, the strongly positive questionnaire mean ($M = 4.27$) is corroborated by the independent rise in observed behavioral compliance and structured engagement, supporting the inference that the positive self-report reflects genuine change in classroom behavior rather than a socially desirable response. At the same time, the qualitative data explain the spread evident in the quantitative results: the least-improved interviewee's persistent difficulty, together with the complete absence of draft-comparison behavior across both observed meetings, show that the benefits of the Prompting Guidance technique, although substantial at the group level, were neither automatic nor evenly distributed across all students.

Taken together, these findings support the conclusion that the AI Prompting Guidance technique functioned as an effective, Zone of Proximal Development-consistent scaffold for improving students' independent Analytical Exposition writing, a conclusion broadly consistent with recent experimental research on ChatGPT-assisted EFL writing instruction (Boudouaia et al., 2024; Song & Song, 2023; Woo, 2024). At the same time, the qualification introduced by the least-improved student's profile and by the absent draft-comparison behavior points to a clear direction for refining the technique: pairing structured AI prompting with more explicit, sustained instruction in independent revision, so that students who need more support in building confidence apart from the AI, rather than only those who adapt quickly, can benefit equally from the intervention.

CONCLUSION

This study set out to develop and evaluate a Prompting Guidance technique for using AI chatbots to support independent Analytical Exposition writing among eleventh-grade EFL students, and to understand how students themselves experienced this way of learning. The findings show that students who learned to compose structured, purposeful prompts, rather than simply asking an AI chatbot to produce a finished essay, made clearly greater gains in their writing ability than students who received conventional instruction over the same period, even though both groups began at a comparable level of ability. This

improvement was accompanied by a visible shift in how students approached the writing task itself: over the course of the intervention, students moved away from passively waiting for the chatbot to supply content and increasingly began to plan their essays, draft a thesis statement, and build arguments on their own initiative before turning to the AI for feedback, guidance, or clarification. All of the students interviewed described coming to understand writing as a multi-stage process rather than a single act of production, and most continued to regard their finished essays as genuinely their own work, with the AI serving mainly as a source of feedback rather than as a ghostwriter.

At the same time, this positive outcome was not experienced uniformly by every student. One of the four students interviewed continued to rely heavily on the AI to compose usable prompts and admitted to feeling less confident writing without its assistance, a reminder that structured AI use can support independence for most learners while still leaving a smaller group in need of further, more individualized guidance. Likewise, while students' overall engagement with the AI-assisted writing routine became markedly more independent and reflective over time, the habit of directly comparing an early draft with a revised one never fully took hold during the study period, suggesting that this particular self-regulation skill may need more sustained practice than a short intervention can provide.

Taken as a whole, these findings suggest that AI chatbots can meaningfully support the development of independent EFL writing when their use is deliberately structured around guiding questions rather than left open for students to request a finished product, and that it is this structure, more than the technology itself, that appears to encourage genuine planning, revision, and critical thinking. For students, the practical implication is to continue treating the AI as a source of ideas and feedback rather than a shortcut, and to make a habit of comparing draft versions of their own writing before finalizing it. For teachers wishing to adopt a similar approach, the experience of this study suggests providing clear, written prompting examples during the early stages of the program and arranging brief individual check-ins for students who continue to lean heavily on AI assistance, so that they receive additional support rather than being left to catch up unassisted. For schools, the experience also points to the importance of preparing clear guidelines for responsible AI use together with reliable technical infrastructure, since the most common source of student frustration in this study concerned access and connectivity rather than the learning approach itself. Finally, for future researchers, extending this kind of study to a larger number of schools and a longer period of implementation would help establish how consistently these benefits hold, and closer attention to which specific kinds of prompts are most strongly associated with writing improvement would help refine the technique further. In closing, despite its limited scale and duration, this study demonstrates that when AI chatbots are guided by a structured prompting strategy, they can support rather than replace the development of independent writing skills, offering a practical and encouraging model for bringing generative AI into the EFL writing classroom responsibly.

Acknowledgements

The author would like to express sincere gratitude to the school leadership, English teachers, and eleventh-grade students of the participating senior high school for their cooperation throughout the six-week intervention, and to the academic supervisors at the English Education Program, Faculty of Language and Literature, Universitas Sains Al-Qur'an Jawa Tengah, for their guidance during the design and execution of this study. No external funding was received for this research.

REFERENCES

- Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI*, 7(1), 52–62. <https://doi.org/10.61969/jai.1337500>
- Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, 57. <https://doi.org/10.1016/j.asw.2023.100745>
- Boudouaia, A., Mouas, S., & Kouider, B. (2024). A study on ChatGPT-4 as an innovative approach to enhancing English as a Foreign Language writing learning. *SAGE Open*, 14(2). <https://doi.org/10.1177/07356331241247465>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices* (2nd ed.). Pearson Education.
- Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). Routledge.
- Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61(2), 228–239. <https://doi.org/10.1080/14703297.2023.2190148>
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). SAGE Publications.
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.
- Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to design and evaluate research in education* (8th ed.). McGraw-Hill.
- Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning*, 38(1). <https://doi.org/10.1111/jcal.12610>
- Hyland, K. (2019). *Second language writing* (2nd ed.). Cambridge University Press.
- Kohnke, L. (2023). L2 learners' perceptions of a chatbot as a potential independent language learning tool. *International Journal of Mobile Learning and Organisation*.
- Kuhail, M. A., Alturki, N., Alamlawi, S., & Alhejori, K. (2023). Interacting with educational chatbots: A systematic review. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-022-11177-3>
- Lee, A. V. Y., Teo, C. L., & Tan, S. C. (2024). Prompt engineering for knowledge creation: Using chain-of-thought to support students' improvable ideas. *AI*, 5(3). <https://doi.org/10.3390/ai5030069>
- Liu, J., Sihes, A. J. B., & Lu, Y. (2025). How do generative artificial intelligence (AI) tools and large language models (LLMs) influence language learners' critical thinking in EFL education? A systematic review. *Smart Learning Environments*, 12(1).
- Richards, J. C., & Renandya, W. A. (2002). *Methodology in language teaching: An anthology of current practice*. Cambridge University Press.
- Sharpley, M. (2022). Automated essay writing: An AIED opinion. *International Journal of Artificial Intelligence in Education*. <https://doi.org/10.1007/s40593-022-00300-7>
- Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2023.1260843>
- Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing*. <https://doi.org/10.1016/j.asw.2023.100752>
- Teng, L. S., & Zhang, L. J. (2022). Can self-regulation be transferred to second/foreign language learning and teaching? Current status, controversies, and future directions. *Applied Linguistics*. <https://doi.org/10.1093/applin/amab032>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Waziana, W. (2024). Students' perceptions about the impact of AI chatbots on their vocabulary and

- grammar in EFL writing. *Register Journal*. <https://doi.org/10.18326/register.v17i2.352-382>
- Woo, D. J. (2024). Teaching EFL students to write with ChatGPT: Students' motivation to learn, cognitive load, and satisfaction with the learning process. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-024-12819-4>
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*. <https://doi.org/10.1111/j.1469-7610.1976.tb00381.x>
- Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist*. https://doi.org/10.1207/s15326985ep2501_2
- Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In *Handbook of self-regulation*. Academic Press.
- Zimmerman, B. J. (2008). Investigating self-regulation and motivation. *American Educational Research Journal*. <https://doi.org/10.3102/0002831207312909>