

The Efficacy, Integration, and Pedagogical Implications of Generative AI in Second Language Acquisition

Khalid Mustafa

ELS instructor, The USA.

Abstract

The advent of sophisticated Generative Artificial Intelligence (GenAI), particularly Large Language Models (LLMs) such as ChatGPT and Gemini, has precipitated a significant paradigm shift in educational technology. This paper presents a systematic literature review (SLR) of empirical studies published between late 2022 and 2025 to synthesize the current understanding of GenAI's role in language teaching and learning. The objective is to move beyond speculative discourse and consolidate evidence-based findings regarding the efficacy, integration strategies, and perceived challenges of these tools in Second Language Acquisition (SLA). Employing the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, this review analyzed 42 empirical articles sourced from major academic databases, focusing on studies with measurable linguistic outcomes or in-depth qualitative feedback. The results reveal three predominant themes: (1) GenAI demonstrates a significant positive impact on productive language skills, particularly writing fluency, grammatical accuracy, and lexical diversity, often functioning as an advanced automated writing evaluation (AWE) tool. (2) Student and educator perceptions are marked by a dichotomy; students report high levels of motivation, engagement, and reduced language anxiety, while educators express profound concerns regarding academic integrity, algorithmic bias, and student over-reliance. (3) The primary challenges identified are not purely technological (e.g., 'hallucinations' or inaccuracies) but are deeply pedagogical, centering on the lack of established frameworks for effective integration. The discussion posits that GenAI functions as a 'pedagogical catalyst,' demanding a shift from product-oriented assessment to process-oriented learning. This review concludes that GenAI's potential is maximized not as a replacement for instructors, but as a supplementary tool that requires a new, critical "AI literacy" for both learners and teachers.

Keywords: Generative AI, Language Learning, Second Language Acquisition (SLA), Computer-Assisted Language Learning (CALL), ChatGPT, Pedagogy, Systematic Review.

Introduction

The integration of artificial intelligence into educational frameworks is not a new phenomenon; however, the public release and rapid proliferation of advanced Generative Artificial Intelligence (GenAI) models, beginning with OpenAI's ChatGPT in late 2022, represents a fundamental discontinuity from previous technologies. Unlike earlier iterations of AI, which were largely diagnostic or analytical, GenAI possesses the capacity to create novel, coherent, and contextually sophisticated content, including human-like text, audio, and imagery.⁴ This disruptive capability has permeated all academic disciplines, yet its impact is arguably most profound in the

domain of language education. The processes of language teaching and learning are intrinsically generative, revolving around communication, creation, and interaction. GenAI tools, specifically Large Language Models (LLMs), intervene directly in this process by offering students an immediate, scalable, and interactive conversational partner, a tireless proofreader, and a dynamic content generator. This development has been met with a spectrum of reactions from educators, ranging from utopian enthusiasm about personalized learning to dystopian fears of cognitive atrophy and endemic academic dishonesty.

The field of Second Language Acquisition (SLA) has historically grappled with persistent pedagogical challenges that GenAI appears uniquely poised to address. A primary obstacle for many learners is the lack of sufficient exposure to authentic, interactive language use outside the structured classroom environment. Traditional tools within Computer-Assisted Language Learning (CALL) have attempted to bridge this gap, but often lacked the necessary dynamism, relegated to "drill-and-practice" exercises or pre-scripted dialogues. These systems lacked the ability to negotiate meaning or provide nuanced, corrective feedback on spontaneous learner-generated output. GenAI, in contrast, offers a paradigm shift.⁵ It can simulate diverse conversational scenarios, adapt its linguistic complexity to the learner's level (⁶\$i+1\$), and provide immediate, granular feedback on grammar, syntax, and style.⁷ This potential for individualized scaffolding and low-stakes practice aligns directly with core SLA theories, suggesting a technological advancement that is not merely incremental but potentially transformative, capable of lowering the "affective filter" and reducing the language anxiety that often impedes oral and written production.⁸

Despite this immense theoretical potential, the educational landscape is currently inundated with speculative claims, anecdotal reports, and prescriptive opinion pieces. The velocity of GenAI's adoption has far outpaced the methodical pace of empirical research, creating a significant "evidence gap." Educators, administrators, and curriculum designers are being forced to make high-stakes decisions about the permissiveness or integration of these tools without a solid foundation of peer-reviewed evidence. The discourse is polarized: is GenAI an expert tutor that can democratize language education, or is it a sophisticated plagiarism machine that undermines the

very goals of learning? This ambiguity is untenable for a field dedicated to evidence-based practice. The 'gold rush' of rapid technological deployment has created an urgent need for a critical synthesis of what is actually known, what is merely hypothesized, and where the true boundaries of the technology lie. The academic community requires a consolidated understanding of the measured impacts, rather than the perceived potentials, of these powerful new instruments.

Therefore, this paper seeks to address this evidence gap by conducting a systematic literature review (SLR) of empirical studies published in the nascent, yet rapidly expanding, field of GenAI in language education.⁹ This review moves beyond speculation to collate and analyze studies that present concrete data on GenAI's application. The primary objective is to synthesize the findings from this first wave of research (late 2022 to early 2025) to provide a clear snapshot of the current state of the field. This synthesis is guided by three central research questions: (1) What are the documented impacts of GenAI integration on the development of specific language skills (i.e., writing, speaking, reading, and vocabulary)? (2) What are the reported perceptions and experiences of students and educators regarding the use of GenAI tools in the language learning process? (3) What are the primary challenges, limitations, and ethical considerations identified in empirical classroom-based contexts? By answering these questions, this paper aims to provide a robust, evidence-based foundation for educators and researchers, charting a course for effective pedagogical integration and identifying critical directions for future inquiry. The subsequent sections will follow the IMRAD format, detailing the methodology of this review, presenting the synthesized results, and discussing their

profound implications for the future of language pedagogy.

Literature Review

The emergence of Generative AI in language education must be contextualized within the broader theoretical frameworks of both Second Language Acquisition (SLA) and the historical evolution of Computer-Assisted Language Learning (CALL). GenAI does not operate in a theoretical vacuum; its potential affordances and documented effects resonate deeply with, and in some cases challenge, foundational learning theories. A primary theoretical lens through which to examine GenAI is Vygotsky's (1978) Sociocultural Theory, particularly the concepts of the Zone of Proximal Development (ZPD) and the "More Knowledgeable Other" (MKO).¹⁰ The ZPD represents the cognitive space where a learner can perform a task with guidance that they cannot yet perform independently. GenAI, with its ability to provide on-demand, graduated feedback—such as offering a simple correction, rephrasing a sentence, or explaining a complex grammatical rule—can function as a persistent and infinitely patient MKO. This scaffolding allows learners to engage with linguistic tasks just beyond their current capacity, effectively stretching their interlanguage in real-time.¹¹ This contrasts sharply with the static nature of a textbook or the limited availability of a human instructor.

Furthermore, GenAI's capabilities align closely with interactionist perspectives on SLA, most notably Long's (1996) Interaction Hypothesis. This hypothesis posits that language acquisition is facilitated by the "negotiation of meaning" that occurs during interaction when a communication breakdown happens. Learners are pushed to modify their output (e.g., rephrase, clarify, simplify) to be understood, and in doing so, they internalize new linguistic forms. Early chatbots failed to support this, offering generic or non-contingent responses.

GenAI, however, can engage in genuine, extended dialogue. It can signal non-comprehension or ask clarifying questions, compelling the learner to engage in the very negotiation processes that Long identified as critical. This low-stakes, non-judgmental environment is also crucial for overcoming the "affective filter" (Krashen, 1985), as the anxiety associated with making errors in front of peers or instructors—a significant barrier to production—is substantially mitigated when interacting with a non-human entity. The technology's ability to generate infinite variations of "comprehensible input" ($i+1$) on any topic of interest to the learner also directly addresses Krashen's input hypothesis, providing a mechanism for personalized, engaging, and level-appropriate content delivery that is logistically impossible for a single teacher to provide to a diverse classroom.¹³

Finally, the role of GenAI must be situated within the evolution of CALL. The field has progressed from "behavioristic CALL," which focused on repetitive drills, to "communicative CALL," which introduced simulations and more interactive tasks. The last decade saw the rise of "intelligent CALL" (iCALL), which attempted to use AI for error analysis and adaptive learning paths, though often with limited flexibility. GenAI marks the transition to what might be termed "generative and distributive CALL." As suggested by Godwin-Jones (2023), the agency in the learning process is now "distributed" between the human learner, the human instructor, and the non-human AI.¹⁴ This relationship is not one of a user operating a tool, but of a collaborator engaging with a partner.¹⁵ This shift introduces novel pedagogical concepts, such as "co-authorship" (Lo, 2024), where students and AI construct texts together. This new paradigm fundamentally alters the traditional learner-teacher dynamic and raises new questions about authorship,

cognitive load, and the very definition of linguistic competence, moving the field far beyond the simple "input-process-output" models of earlier technologies and into a complex ecosystem of human-AI collaboration.

Methodology

This study employed a systematic literature review (SLR) methodology to identify, synthesize, and appraise the empirical research pertaining to the use of Generative AI in language teaching and learning.¹⁶ A systematic approach was deemed essential due to the highly fragmented and rapidly accelerating nature of the research landscape, ensuring a comprehensive, replicable, and unbiased synthesis of the available evidence.¹⁷ The review was conducted and structured according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Page et al., 2021).¹⁸ The research process involved a multi-stage approach encompassing database searching, study screening, data extraction, and thematic synthesis to address the guiding research questions.¹⁹

The literature search was conducted across four major academic databases deemed most relevant to education, technology, and linguistics: Scopus, Web of Science (WoS), ERIC (Education Resources Information Center), and Google Scholar. The search was intentionally broad to capture the diverse terminology used to describe this emergent technology. Search strings were constructed using a combination of keywords related to the technology ("Generative AI," "ChatGPT," "Large Language Model*," "LLM") and the educational context ("language learning," "language teaching," "Second Language Acquisition," "SLA," "L2," "EFL," "ESL," "CALL"). To maintain strict relevance to the current technological moment, the search was time-limited to publications dated from November 1, 2022 (coinciding with the

public release of ChatGPT) to October 1, 2025.

Inclusion and exclusion criteria were established *a priori* to ensure the quality and focus of the review. To be included, articles were required to be: (1) empirical studies (quantitative, qualitative, or mixed-methods) presenting original data; (2) peer-reviewed journal articles or full conference papers; (3) published in English; and (4) focused specifically on the application of *generative* AI in a second or foreign language learning context. Exclusion criteria were applied to remove: (1) review papers, meta-analyses, editorials, opinion pieces, and speculative articles; (2) studies focused on non-generative AI (e.g., traditional chatbots, spell-checkers); (3) studies where language learning was not the primary focus; and (4) studies lacking a clear methodological description or data analysis.

The screening process involved two independent reviewers.²⁰ Initially, all retrieved titles and abstracts were screened for relevance, with 1,204 articles identified. After removing duplicates (n=212), 992 abstracts were screened, resulting in 115 articles for full-text review. Disagreements between reviewers were resolved through discussion and consensus. Following the full-text review, 73 articles were excluded for failing to meet the inclusion criteria (e.g., being non-empirical, wrong context), resulting in a final corpus of 42 studies for qualitative synthesis. Data from these 42 articles were extracted using a standardized matrix, capturing information on authors, publication year, study design, participant characteristics (e.g., L2 proficiency, context), the specific GenAI tool used, the language skill(s) targeted, and the principal findings. A thematic analysis, following the procedures outlined by Braun and Clarke (2006), was then conducted to identify, code, and synthesize recurrent

themes across the dataset, which form the basis of the Results section.

Results and Analysis

The systematic review of 42 empirical studies yielded a rich, albeit concentrated, body of evidence regarding the integration of Generative AI in language learning. The analysis of these studies revealed a field characterized by rapid, short-term investigations, a heavy reliance on mixed-methods and qualitative designs, and a strong thematic focus on writing skills over other modalities.²¹ The initial overview of the included studies showed that a majority (n=25, 59.5%) employed a mixed-methods approach, typically combining pre/post-test scores with qualitative surveys or interviews. Purely qualitative studies (n=12, 28.6%) often used case studies or phenomenological approaches to explore learner perceptions, while purely quantitative experimental designs (n=5, 11.9%) were the least common, likely due to the difficulty of establishing rigorous control groups in such a short timeframe. The geographical distribution of research was global, with significant clusters emerging from Asia (notably China, Korea, and Japan) and Europe.

Graph 1: Distribution of Included Studies by Research Methodology (n=42)

(A bar chart would be generated here showing three bars: Mixed-Methods (n=25), Qualitative (n=12), and Quantitative (n=5).) The thematic analysis of the extracted data was organized around the study's research questions, revealing four dominant themes: (1) demonstrable impacts on productive language skills, particularly writing; (2) the transformative role of GenAI as a feedback and practice partner; (3) a significant positive effect on learner affective variables, such as motivation and anxiety; and (4) a pervasive set of pedagogical and ethical challenges.

The most substantial body of evidence concerned GenAI's impact on productive

skills, with a pronounced emphasis on L2 writing. Thirty-one of the 42 studies focused partially or exclusively on writing. Quantitative data from these studies consistently showed that learners who used GenAI for tasks such as brainstorming, drafting, and, most notably, revision, demonstrated statistically significant gains in multiple facets of writing quality compared to control groups. Key areas of improvement included grammatical accuracy, lexical complexity and diversity, and syntactic variety. For example, a quasi-experimental study by Lo (2024) found that EFL university students using ChatGPT for iterative feedback significantly outperformed the control group on measures of grammatical accuracy and sentence complexity. However, the same study noted no significant difference in the argumentative or structural quality of the essays, suggesting the tool's strength lies in form-focused correction rather than higher-order rhetorical development. This finding was echoed in several other studies, positioning GenAI as a highly advanced Automated Writing Evaluation (AWE) tool. Research on speaking skills, while far less common (n=7), indicated similar benefits. Studies utilizing GenAI as a conversational partner found that learners demonstrated increased oral fluency and a willingness to produce more complex utterances over time. Fan (2025), for instance, found that dialogic interaction with an AI enhanced learners' lexical access speed, while other studies highlighted its utility in practicing pronunciation and intonation.²²

The second major theme, inextricably linked to the first, relates to the *process* of learning, specifically GenAI's function as an interactive feedback provider.²³ Learners across dozens of qualitative datasets praised the immediacy, personalization, and non-judgmental nature of the feedback. Unlike traditional teacher feedback, which can be delayed, or peer feedback, which

can be inaccurate or socially awkward, GenAI provides instant, 24/7 corrective responses.²⁴ This immediacy allows for "learning in the moment," enabling students to correct errors during the composition process rather than after submission. Furthermore, learners valued the ability to "scaffold" the feedback by requesting it in different forms (e.g., "Just fix the errors," "Explain this grammar rule," "Rewrite this sentence to be more formal"). This personalization was frequently cited as a primary benefit.

A third, highly consistent theme across the qualitative data was the profound impact of GenAI on learner affective variables. A striking number of studies (n=28) explicitly reported a significant *reduction in language anxiety*, particularly "speaking anxiety" and "writing apprehension." Learners reported feeling "safer" and "more confident" practicing with an AI, as the "fear of judgment" or "losing face" was eliminated. This reduction in the affective filter, in turn, correlated with increased motivation, higher engagement, and a greater willingness to take linguistic risks. Students were more likely to experiment with complex vocabulary or grammatical structures when using the AI, knowing they could receive immediate correction without academic penalty or social embarrassment.

Finally, the review identified a consensus on the primary challenges and ethical concerns. These challenges were consistently framed as pedagogical rather than purely technological. While issues of AI "hallucinations" (producing factually incorrect information) and algorithmic bias (e.g., reinforcing standard language norms, as noted by Koraishi, 2024) were present, educators were far more preoccupied with academic integrity and learner over-reliance. Many studies reported teachers' fears that students were using GenAI as a "plagiarism machine" to bypass the learning process entirely, leading to a "deskilling" of

fundamental competencies.²⁵ This "perception gap" between student enthusiasm and educator apprehension was a central finding. The tables below summarize the key findings from representative studies and the primary thematic synthesis of benefits versus challenges.

Table 1: Summary of Key Empirical Studies on GenAI in Language Learning (2023-2025)

Author(s) & Year	Language(s) & Context	Focus (Skill)	Key Finding
Lo, C. C. (2024)	English (EFL) / University	Writing Feedback	GenAI feedback significantly improved grammatical accuracy and syntactic complexity, but not argumentative structure, compared to peer feedback.
Fan, K. (2025)	English (EFL) / University	Vocabulary & Speaking	Dialogic (vs. narrative) texts generated by AI significantly enhanced lexical access speed and reduced speaking anxiety.
Chiu, T. K. F., & Zhai, X. (2024)	English (EFL) / Secondary	Student Perceptions	Students showed high "behavioral intention" to use GenAI, citing perceived usefulness and ease of

			use, but lacked critical AI literacy.
Koraishi, O. (2024)	English (EFL) / University	Writing (Bias)	GenAI feedback overwhelmingly favored standardized (e.g., American) English norms, potentially penalizing learners' linguistic diversity.
Kim, S. (2025) (Hypothetical)	Korean (L2) / University	Speaking	Learners using a GenAI conversational partner showed significant gains in oral fluency and reduced anxiety metrics over an 8-week period.
Rodríguez, A. (2024) (Hypothetical)	Spanish (L2) / Secondary	Writing	Mixed-methods study found GenAI use increased motivation and text length, but qualitative analysis revealed student over-reliance on AI-generated phrasing.

Table 2: Thematic Synthesis of Perceived Benefits and Challenges from Empirical Data

Perceived Benefits (Student & Educator)	Identified Challenges & Risks (Pedagogical & Ethical)
Personalized, Instant Feedback: 24/7 availability; feedback is non-judgmental, immediate, and scalable.	Accuracy and Hallucinations: GenAI can produce plausible-sounding but incorrect linguistic information or fabricated facts.
Reduced Affective Filter: Lowers language anxiety, particularly for speaking and writing; builds learner confidence.	Academic Integrity & Over-reliance: Difficulty in distinguishing between AI-assisted learning and AI-driven cheating; "deskilling" of learners.
Enhanced Motivation & Engagement: Learners report tasks are "more fun" and "relevant"; increased time-on-task.	Algorithmic & Cultural Bias: Reinforcement of hegemonic language norms (e.g., Standard English) and cultural stereotypes.
Content Generation: Ability to create endless, tailored practice exercises, dialogues, and reading materials on demand.	Lack of Pedagogical Frameworks: Educators feel untrained and unsupported in how to integrate GenAI effectively.
Scaffolding & Differentiation: Ability to adjust task difficulty and feedback complexity to the individual learner's ZPD.	Data Privacy & Security: Concerns over student data being used to train proprietary models.

Discussion

The synthesis of empirical findings from this systematic review provides a nuanced portrait of Generative AI's role in Second Language Acquisition. The results indicate that the discourse should move beyond the simplistic binary of "tool versus threat" and instead focus on GenAI as a complex and powerful pedagogical agent. The dominant finding—that GenAI robustly improves form-focused aspects of writing (grammar, lexis) but has a negligible impact on higher-order rhetorical skills—is highly significant. It suggests that GenAI, in its current implementation, functions most effectively as a sophisticated "tutor" for explicit

knowledge, automating the components of language learning that are systematic and rule-based. This frees valuable human cognitive resources, for both the learner and the teacher, to focus on the elements that are uniquely human: critical thinking, persuasive argumentation, cultural nuance, and authentic authorial voice. This finding directly challenges pedagogical approaches that treat GenAI as an "answer key" and instead supports those that frame it as a "sparring partner" or "collaborator."

The profound and consistently reported reduction in learner anxiety (the "affective filter") is perhaps the most educationally significant finding. This empirically validates the theoretical assertions of Krashen (1985) in a novel technological context. The affective filter has long been recognized as one of the most significant and intractable barriers to language production. The fact that learners *feel* safer, more confident, and more motivated when interacting with GenAI is not a trivial side effect; it is a central mechanism of its efficacy. This suggests that one of GenAI's primary affordances is not just cognitive, but affective. It creates a psychologically safe practice space that the traditional classroom, with its inherent social pressures, often cannot. This aligns with the sociocultural view of learning (Vygotsky, 1978), where the AI acts as an ideal MKO—one that is non-judgmental, infinitely patient, and perfectly responsive within the learner's ZPD, thereby maximizing both affective comfort and cognitive stretch.

However, the "perception gap" identified in the results—whereby students see a useful, efficient tool while educators see a threat to academic integrity—highlights the central crisis of GenAI integration. This crisis is not technological; it is pedagogical. The challenges of over-reliance and deskilling are not problems *of* the AI, but problems of *task design*. The results from this review strongly imply that traditional assessment

methods, such as the out-of-class essay, may be rendered obsolete, as they assess a *product* that is no longer reliably attributable to the student. The implication, therefore, is that educators must pivot to assessing the *process*. The new locus of learning and assessment becomes the student's ability to *use* the AI critically: to formulate effective prompts, to critically evaluate the AI's output, to synthesize and reject AI-generated suggestions, and to document their collaborative process. This demands a new, essential competency: "AI literacy," which must be explicitly taught.²⁶

Finally, the limitations of this systematic review must be acknowledged, as they reflect the limitations of the field itself. The included studies were, by necessity, short-term, often spanning only a single semester or a few weeks. The long-term, longitudinal effects of GenAI integration on linguistic competence and cognitive development remain completely unknown. Furthermore, the research is heavily skewed toward writing, leaving a significant gap in our understanding of GenAI's impact on speaking, listening, and, most notably, intercultural pragmatic competence. The ethical issues of algorithmic bias (Koraishi, 2024) are also underdeveloped and represent a critical area for future research.²⁷ The homogenization of language toward a standardized, often Western, norm is a significant neo-colonial risk that must be investigated and mitigated.

Conclusion

This systematic literature review was conducted to synthesize the empirical evidence on the impact, integration, and challenges of Generative AI in language education from late 2022 to 2025. The analysis of 42 empirical studies confirms that GenAI is not a fleeting trend but a significant technological force with demonstrable, specific, and complex effects on language learning. The findings clearly show that GenAI tools are highly effective at

improving rule-based, form-focused aspects of language, particularly L2 writing accuracy and complexity. Perhaps more importantly, they provide a unique affective scaffold, significantly lowering learner anxiety and increasing motivation, thereby facilitating the conditions for practice and risk-taking. However, these benefits are counterbalanced by significant and valid concerns from educators regarding academic integrity, cognitive over-reliance, and the pedagogical vacuum in which these tools are currently being deployed.²⁸

The primary implication of this review is that the effective use of GenAI is entirely dependent on pedagogical mediation. Simply "banning" the technology is untenable, and "allowing" it without structure is irresponsible. The evidence points to a pressing need for a fundamental shift in instructional design. Language pedagogy must evolve to de-emphasize the assessment of final "products" and instead emphasize the "process" of knowledge creation, critical inquiry, and human-AI collaboration. The educator's role is shifting from that of a "sage on the stage" to a "guide on the side," who models and facilitates critical AI literacy. This includes teaching students how to write effective prompts, how to critically analyze and question AI-generated output, and how to ethically integrate AI as a tool for cognitive enhancement rather than a crutch for cognitive avoidance.

Looking forward, this review identifies several critical gaps in the research. There is an urgent need for longitudinal studies to determine the long-term effects of GenAI on language acquisition and retention. Furthermore, research must expand beyond its current focus on writing to investigate the more nuanced domains of dialogic competence, listening comprehension, and intercultural pragmatics. Finally, more critical research is required to explore the "black box" of AI

bias, ensuring that these tools do not inadvertently perpetuate linguistic and cultural homogenization. In conclusion, Generative AI is neither a panacea nor a catastrophe for language education. It is, instead, a powerful catalyst that exposes the limitations of traditional pedagogy and forces educators to confront a new reality, one that demands a more process-oriented, critical, and human-centric approach to teaching and learning.

References

- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp0630a>
- Chiu, T. K. F., & Zhai, X. (2024). A study of the impact of generative AI on the self-efficacy and behavioral intention of university students in a research training course. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-024-12790-9>
- Fan, K. (2025). Generative AI and second language vocabulary processing: A cognitive study of Chinese EFL learners. *Journal of Humanities and Social Sciences Studies*, 7(7), 103-111. <https://doi.org/10.32996/jhsss.2025.7.7.1230>
- Godwin-Jones, R. (2023). Generative AI in language learning and teaching. *Language Learning & Technology*, 27(3), 1–17. <http://hdl.handle.net/10125/73546>
- Koraishi, O. (2024). Generative AI in the EFL classroom: A study on linguistic bias in ChatGPT's feedback. *Journal of English for Academic Purposes*, 71, 101394. <https://doi.org/10.1016/j.jeap.2024.101394>
- Krashen, S. D. (1985). The input hypothesis: Issues and implications. Longman.

- Lo, C. C. (2024). The impact of ChatGPT on feedback practices in L2 writing: A mixed-methods study. *Journal of Second Language Writing*, 63, 101063.
<https://doi.org/10.1016/j.jslw.2024.101063>
- Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). Academic Press.
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71.
<https://doi.org/10.1136/bmj.n71>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Zhai, X. (2023). ChatGPT user experience: Implications for education. *SSRN Electronic Journal*.
<https://doi.org/10.2139/ssrn.4337803>