International Journal of Learning, Teaching and Educational Research
Vol. 24, No. 10, pp. 509-526, October 2025
https://doi.org/10.26803/ijlter.24.10.24
Received Aug 25, 2025; Revised Sept 1, 2025; Accepted Sept 30, 2025

Artificial Intelligence in ESL/EFL Education: Evidence from Recent Reviews (2024–2025)

Eun-Young Jeon*
Yonsei University, Seoul, Korea
Jeonbuk National University, Jeonju, Korea

Abstract. The advent of generative artificial intelligence (AI), exemplified by ChatGPT, has triggered a paradigm shift in English language education. This study adopts a "review of reviews" design, synthesizing 14 systematic reviews, meta-analyses, and meta-syntheses (2024–2025). Articles were identified through Scopus and Web of Science using a five-stage selection procedure and analyzed through thematic synthesis. A central contribution is the articulation of a pedagogy-first, ethically grounded research agenda, which distinguishes this study from prior work. The analysis highlights three dimensions: (1) the overall effectiveness of AI-enhanced instruction, (2) applications in writing and speaking, and (3) evolving learner roles in engagement, self-regulation, and emotional experiences. The findings confirm AI's strong potential to enhance productive skills: automated feedback systems improve accuracy, cohesion, and revision processes in writing, while dialogue-based chatbots strengthen speaking fluency, confidence, and willingness to communicate. These tools serve not only as technological aids but also as cognitive scaffolds and interactive partners that reshape how learners engage with language. At the same time, the benefits are conditional, shaped by factors such as intervention duration, interface design, learner characteristics, and educational context. Risks—including overreliance, plagiarism, privacy concerns, and uneven attention to the K-12 contexts—temper these gains. Current research remains concentrated in higher education, limiting cross-level generalizability. Unlike earlier reviews that cataloged applications or emphasized single-skill outcomes, this synthesis integrates effectiveness data with pedagogical, affective, and ethical perspectives. It underscores the need for future research that is longitudinal, context-sensitive, and multi-skilled, while ensuring innovation aligns with pedagogy and robust ethical safeguards.

Keywords: artificial intelligence; ChatGPT; English as a second and foreign language education; educational technology; review of reviews

*Corresponding author: Eun-Young Jeon; *joyjeon@hanmail.net*

1. Introduction

The integration of technology into language education has long been shaped by the field of Computer-Assisted Language Learning (CALL). However, since the release of generative artificial intelligence (AI) tools in late 2022, the field has undergone a paradigm shift of unprecedented scale. The emergence of models such as ChatGPT has fundamentally reconfigured English as a second and foreign language (ESL/EFL) education, triggering a rapid surge in scholarly interest (Wahyuni et al., 2024; Wang et al., 2025a). These developments have not only expanded the scope of research to include diverse pedagogical applications—from the creation of global Englishes materials (Lo, 2025) to the reimagining of pre-service teacher training (Zhang et al., 2025)—but also revealed striking contrasts in scope and depth.

For instance, Lo (2025) highlights how AI reshapes global Englishes pedagogy, while Zhang et al. (2025) illustrate its potential in teacher education. Taken together, these studies underscore AI's versatility yet also expose fragmentation: Lo emphasizes macro-level curriculum design, whereas Zhang centers on micro-level teacher preparation. Such contrasts underscore the absence of an integrated perspective and highlight the need for a review that critically synthesizes these parallel strands of research.

Recent research has investigated the role of AI across school-based contexts (primary and secondary education) and higher education (e.g., universities), emphasizing the importance of distinguishing how these populations differentially experience and benefit from AI-enhanced language learning. This underscores the urgency for a higher-order synthesis that critically evaluates, rather than simply catalogs, prior findings, while addressing deeper theoretical foundations, unresolved methodological weaknesses (e.g., small samples, short interventions), and socio-ethical risks (e.g., plagiarism, cognitive overreliance, data surveillance).

Importantly, the integration of AI into language education is not a recent phenomenon. A bibliometric review by Kartal and Yeşilyurt (2024) of research spanning 1995 to 2022 identified four major clusters: general AI, Natural Language Processing (NLP), Robot-Assisted Language Learning, and chatbots. These clusters demonstrate that foundational ideas such as automated feedback and dialogue systems have long informed language education.

Yet the capabilities of modern generative AI, particularly its ability to generate human-like language, sustain coherent interaction, and tailor content to individual learners, represent a qualitative leap over earlier systems. This shift has compelled educators to reconsider long-standing pedagogical models (Yang & Li, 2024) and has sparked research across varied instructional contexts (Lee et al., 2025) and goals (Daud et al., 2025).

The historical evolution and recent paradigm shift are visually summarized in Figure 1, which traces the progression of AI-powered tools in language education, from foundational CALL to Intelligent Tutoring Systems, Automated Writing

Evaluation, and early chatbots, culminating in the current era of generative AI. Dominated by large language models such as ChatGPT, this stage is marked by unprecedented versatility and multimodal interaction (Li et al., 2025a).

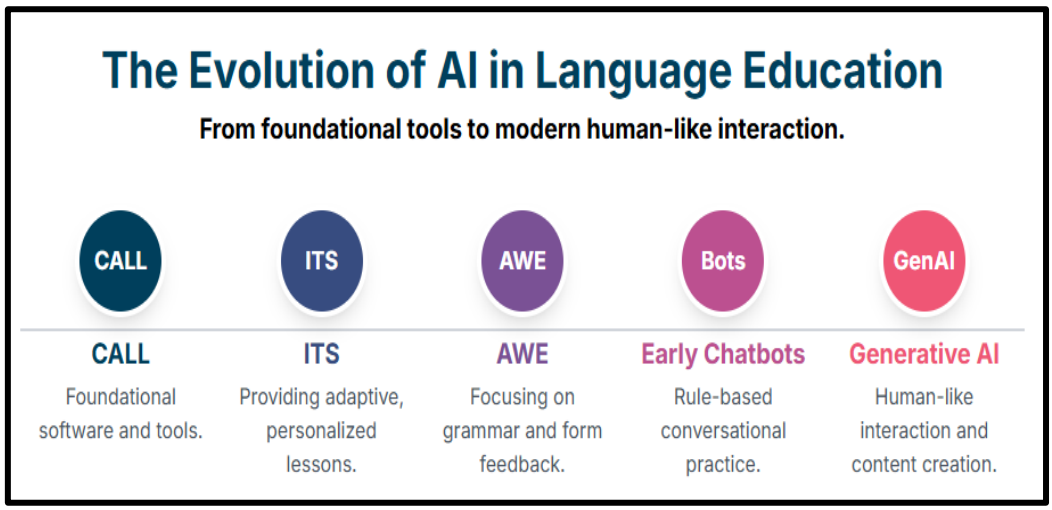

Figure 1: The evolution of AI in language education

Building on this evolution, a growing number of systematic reviews and meta-analyses have recently been published, reflecting the field's rapid expansion and the increasing need for knowledge consolidation. Yet despite their contributions, many of these reviews remain limited to reporting effectiveness metrics or listing applications, often neglecting how such findings complicate interpretations of pedagogical impact. As a result, there remains a lack of clarity about how to interpret AI's pedagogical potential alongside its limitations.

The distinctive contribution of this article is that it moves beyond descriptive cataloging to offer an integrative synthesis: it not only aggregates quantitative effectiveness data (e.g., g-ranges) but also connects these outcomes to pedagogical design, affective learner experiences, and ethical implications. Extending this line of inquiry, the present study offers a "review of reviews" that critically synthesizes major systematic reviews and meta-analyses from 2024–2025, to develop a comprehensive, evidence-informed understanding of AI in ESL/EFL education.

By situating its contribution within these research gaps, this review not only consolidates findings but also highlights these broader concerns. This sharper focus underscores the urgency of a pedagogy-first, ethically grounded framework. Importantly, the review narrows its scope to writing and speaking, domains where AI tools (e.g., automated feedback systems, chatbots) have been most extensively studied and where their pedagogical potential and risks are most evident. Accordingly, this study seeks to provide a nuanced synthesis of recent findings by examining AI's effectiveness, its differentiated contributions to skill-specific instruction, and its influence on learner engagement and affective experiences.

It addresses the following research questions:
1. How effective is AI overall in enhancing English language learning outcomes?
2. How do AI applications differ in their effectiveness between writing proficiency and speaking competence?
3. In what ways do AI systems influence learner engagement, self-regulation, and emotional experiences in language learning contexts?

2. Methodology

Relevant research articles on AI in ESL/EFL education were identified through a systematic search of Scopus and Web of Science to ensure comprehensive coverage. Scopus is recognized as the "largest abstract and citation database of peer-reviewed literature" (Schotten et al., 2018, p. 31). To address the research questions, a systematic screening process was applied using the following search terms: "artificial intelligence" AND ("EFL" OR "ESL" OR "English language learning") AND ("review" OR "meta-analysis"). This initial search yielded 259 articles.

Data extraction was guided by a coding protocol that included study characteristics (year, type, scope), educational level, targeted skills, AI tool type, methodological design, and reported effect sizes. First, only peer-reviewed journal publications from 2024 to 2025 were included, capturing the most recent methodological trends, emerging tools, and pedagogical debates. This step reduced the sample to 140 articles. Next, disciplinary relevance was assessed, restricting the corpus to social sciences, computer science, and arts and humanities, which narrowed the set to 63 articles.

Subsequently, only non-empirical articles, specifically reviews, meta-analyses, and meta-syntheses, were retained, reducing the pool to 56. The relevance of the topic was then evaluated through careful reading of titles and abstracts, leading to the exclusion of studies not directly related to AI in English language education, leaving 24 articles. Finally, to ensure methodological rigor and research quality, only articles published in Social Science Citation Index (SSCI)-listed journals were retained. SSCI journals adhere to rigorous peer review standards, providing confidence in the credibility of the selected studies (Duman et al., 2015). This criterion narrowed the final corpus to 14 articles.

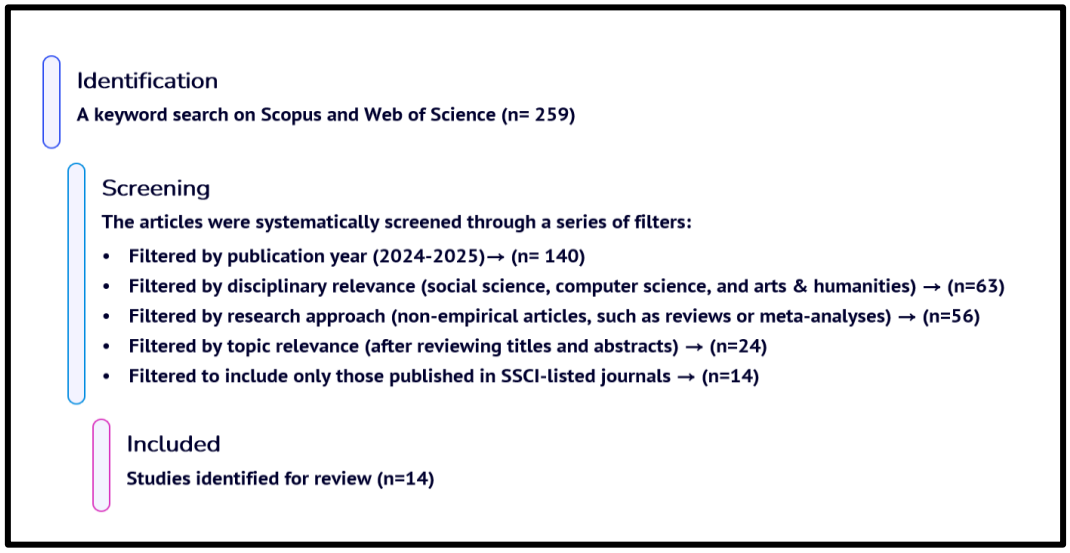

Figure 2: The process of the literature search and selection

Thematic synthesis was then conducted on the final set of 14 articles to identify recurring patterns and gaps. This process involved initial open coding of key findings, grouping related codes into descriptive categories, and refining them into higher-order analytical themes. The entire selection and screening procedure is visually represented in Figure 2, adhering to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Page et al., 2021).

3. Results

This section reviews recent meta-analyses, systematic reviews, and meta-syntheses to examine three key areas: (1) the effectiveness of AI in ESL/EFL education, (2) its varied applications across language skills (e.g., speaking, writing), and (3) learners' roles in AI-mediated environments.

3.1 Quantifying the Impact: Overall Effectiveness

The cumulative evidence from recent meta-analyses and systematic reviews confirms the overall effectiveness of AI in ESL/EFL learning. Xu and Wang (2024) reported a large positive effect (g = 0.812) across 40 studies, while Wu (2024), analyzing 49 studies, also found significant benefits across listening, speaking, writing, and vocabulary, moderated by factors such as context and intervention design. Taken together, Xu and Wang (2024) and Wu (2024) converge on the conclusion that AI integration yields strong benefits, but they emphasize different levers of impact: Xu and Wang foreground learner demographics, while Wu stresses contextual and instructional moderators. Their findings are complementary yet expose a gap in cross-level synthesis that considers how individual and contextual factors interact.

Koç and Savaş (2025) further showed that voice-based chatbots are especially effective for speaking and listening, reinforcing Wu's emphasis on interactivity but contrasting with Xu and Wang's focus on demographic moderators. Collectively, these findings underscore the potential of adaptive AI tools, particularly those combining automated and human feedback, to optimize

outcomes across diverse learner profiles. Table 1 provides an overview of 14 recent reviews, summarizing study types, scopes, foci, and outcomes.

Table 1: Summary of recent reviews on AI in ESL/EFL education (2024–2025)

Author(s) & Year	Study Type & Scope	Primary Focus	Key Findings
Chang & Sun (2024)	Systematic Review (22 studies, 2000–2022)	AI's impact on self-regulated language learning	AI fosters autonomy and cyclical improvement in higher education learners, but two-way interactivity remains limited.
Feng et al. (2025)	Systematic Review (112 studies, 2014–2024)	AI-integrated L2 writing research	Writing tasks are diverse, but argumentative essays dominate; research was short-term, often without a clear theory; need for longitudinal studies on AI literacy.
Hou & Min (2025)	Meta-analysis (16 studies, 89 effect sizes)	Dialogue-based CALL and L2 speaking	Dialogue systems moderately improved speaking; effectiveness varied by system type, meaning constraint, and modality.
Kartal & Yeşilyurt (2024)	Bibliometric Analysis (185 SSCI articles, 1995–2022)	AI applications in L2 teaching & applied linguistics	Four clusters emerged (AI, NLP, robots, chatbots); highlighted trends in tutoring systems, gamification, and personalized feedback.
Koç & Savaş (2025)	Qualitative Systematic Meta-synthesis (57 studies, 2010–2024)	Voice-based AI chatbots for English learning	Chatbots enhance communication but are limited by technical and interactional constraints.
Li & Zhao (2025)	Systematic Review (28 studies, 2020–2024)	AI for oral communication skills in EFL	AI tools improved fluency, confidence, and reduced anxiety; challenges remain in emotional interaction, technical support, and data privacy.
Li et al. (2025b)	Meta-analysis (41 studies, 2023 to date)	Generative AI (GenAI) chatbots in SLA	Generative AI chatbots had moderate-to-large positive effects; effects varied by language, task (especially vocabulary), and intervention duration.
Liu et al. (2024)	Systematic Review (20 studies, 2001–2023)	Emotional AI applications	Identified five affordances (e.g., human-like conversations, personalized feedback); emotional and cognitive support improved outcomes more than emotional support alone.

Author(s) & Year	Study Type & Scope	Primary Focus	Key Findings
Shi & Aryadoust (2024)	Systematic Review (83 SSCI articles, 1993–2022)	Automated Written Feedback (AWF)	AWF studies show heterogeneous contexts and mixed results, highlighting the need for clearer validation and role definition.
Wang et al. (2025b)	Scoping Review (43 SSCI studies, 2022–2024)	Generative AI (GAI) in language education	Progress noted but gaps remain overreliance on ChatGPT, lack of theory, and perception research dominated.
Wu (2024)	Meta-analysis (49 studies, 2013-2024)	Effectiveness of AI in language learning	Significant positive effects across skills; boosted motivation and reduced anxiety; medium-duration interventions are most effective; interactivity and feedback crucial; need to address learner differences.
Wu & Li (2024)	Meta-analysis (21 studies, 2008–2023)	AI chatbot effectiveness in EFL learning and its moderators	Moderate-to-large positive effect, shaped by duration and interface design.
Xu & Wang (2024)	Meta-analysis (40 studies, 2022–2023)	Overall effectiveness of AI in English learning	Large positive effect; moderated by sample size, learning phase, and students' majors.
Yang & Li (2024)	Systematic Review (44 studies, 2022–2023)	ChatGPT for L2 learning	ChatGPT is used for content generation, feedback, and teaching support; focused on EFL college learners; writing skills and perceptions are central.

Importantly, while all reviews confirmed positive effects, only a subset reported moderator analyses with effect sizes. Table 2, therefore, narrows the focus to five studies that provide quantitative moderator-level data, highlighting how intervention duration, learner demographics, or tool design shaped outcomes.

Table 2: Comparative moderator effects across reviews (2024–2025)

Author(s) & Year	Education Level	Key Moderators	Reported Effect Size	Notes
Wu (2024)	Mainly higher education, mixed contexts	Intervention duration (6 weeks–6 months most effective); interactivity types; feedback sources; learner motivation, anxiety, willingness to communicate	Moderate-to-large: varied across skills	Strongest for listening and speaking; medium-duration optimal; both automated and human feedback beneficial.
Wu & Li (2024)	Mixed	Intervention duration; Interface design	g = 0.648	Outcomes shaped by duration and design.
Xu & Wang (2024)	Mixed	Sample size; learning phase; students' majors	g = 0.812	Moderators significantly influenced effectiveness.
Hou & Min (2025)	Mixed	Type of system; meaning constraint; modality	g = 0.61	No significant moderation by proficiency, education stage, location, or feedback.
Li et al. (2025b)	Mixed	Language type (L1/L2); GenAI tool (ChatGPT); learning task (esp. vocabulary); intervention duration (1–7 days)	*ES* = 0.576	No differences by education stage/environment; strongest effects for vocabulary tasks.

Note. g = Hedges' g; ES = effect size.

However, several critical limitations temper these findings. Shi and Aryadoust (2024) revealed that while AWF systems are generally perceived positively, concerns persist regarding accuracy, reliability, and limited alignment with human feedback. Similarly, Wang et al. (2025b) and Yang and Li (2024) observed that many recent studies on generative AI, such as ChatGPT, lack theoretical grounding, employ short-term experimental designs, and rarely include longitudinal data, thus limiting the generalizability of findings. The heterogeneity of AI tools and research methodologies further complicates meta-level interpretations, highlighting the need for standardized and theory-informed approaches.

3.2 Differentiated Applications: AI's Role Across Writing Proficiency and Speaking Competence

AI's affordances in EFL education vary significantly depending on the linguistic domain targeted. Recent reviews reveal that the most notable gains occur in

productive skills, particularly writing and speaking, where AI tools function not only as instructional aids but also as interactive partners.

3.2.1 Enhancing writing proficiency through automated feedback

Feng et al. (2025) provides strong evidence for the effectiveness of AI-based tools in improving L2 writing, with benefits observed in vocabulary use, grammatical accuracy, and textual cohesion. Similarly, Shi and Aryadoust (2024) found that AWF systems offer valuable feedback that enhances learners' revision behaviors.

However, both studies caution that such tools may lead to overreliance, plagiarism, and a reduction in critical thinking, especially when learners unquestioningly accept machine-generated suggestions. Feng et al. (2025) emphasize skill gains and learner perceptions, while Shi and Aryadoust (2024) focus more on tool performance and system design. Their findings are mutually reinforcing in highlighting benefits but diverge in diagnosing the source of risks, whether learner behavior or technological limitations, indicating that pedagogy and tool development must advance in tandem.

Wang et al. (2025b) complement these findings by pointing to broader shortcomings in generative AI research on writing. Technologically, prior studies often exhibited an overreliance on ChatGPT, lacked detail on the versions employed, and failed to report prompt designs. Theoretically, nearly half of the reviewed studies did not specify a guiding framework, and those that did relied primarily on psychological, technological, or social theories. In terms of research aims, most studies (41.9%) focused on user perceptions, with fewer examining AI's impact, exploring pedagogical practices, or benchmarking performance against other tools. Taking together, these patterns suggest that while AWF systems hold considerable promise, the evidence base remains fragmented, requiring more theory-driven and methodologically rigorous inquiry to establish how such tools can best support sustained L2 writing development.

3.2.2 Developing speaking competence with dialogue-based systems

Dialogue-based CALL systems have shown promise in promoting speaking fluency, interactional competence, and learner confidence. Hou and Min (2025) found a moderate effect ($g = 0.61$) with stronger outcomes in goal-oriented, multimodal designs. Wu and Li (2024) reported a slightly higher effect ($g = 0.648$), showing that sustained and well-designed chatbot interactions yield consistent speaking gains.

Li and Zhao (2025) highlighted improvements in pronunciation, fluency, and anxiety reduction, but also noted barriers such as limited emotional responsiveness, technical challenges, and privacy concerns, pointing to the need for hybrid strategies that combine AI with teacher mediation. Similarly, Koç and Savaş (2025) observed limitations in spontaneity, underscoring the need for more naturalistic design. Overall, these studies suggest that while AI-driven speaking tools are effective, their success depends on thoughtful integration with human facilitation.

3.3 The Learner in the AI Ecosystem: Self-Regulation, Engagement, and Affect

While cognitive gains are well-documented, recent studies have increasingly emphasized the learner's role in navigating AI-based environments, particularly in self-regulation, emotional response, and engagement. Chang and Sun (2024) examined the role of AI as a metacognitive scaffold in supporting self-regulated language learning. They concluded that AI tools can enhance goal setting, strategic planning, progress monitoring, and autonomy, particularly when integrated with personalized feedback mechanisms. These findings align with broader calls for AI to be conceptualized not merely as a delivery mechanism but as a catalyst for learner agency in digital language learning ecosystems.

Chatbots and generative AI tools have been credited with improving student motivation, attention, and interaction (Koç & Savaş, 2025; Li et al., 2025b). However, several reviews caution against uncritical optimism. Yang and Li (2024) and Wang et al. (2025b) noted that engagement may be surface-level or short-lived, particularly in studies that lack sustained pedagogical integration or fail to consider learners' prior digital literacies.

Moreover, the novelty effect of AI can diminish over time, raising concerns about retention and transfer. Wu (2024) similarly observed that while AI-supported learning can enhance motivation, reduce anxiety, and increase willingness to communicate, learner engagement and satisfaction remain inconsistent. Liu et al. (2024) introduced the concept of emotional AI—systems that detect and respond to learners' emotions in real time to foster humanized and affect-aware learning environments. While still in its infancy, emotional AI promises to address the affective dimension often overlooked in traditional language instruction. It holds potential for reducing foreign language anxiety, enhancing perceived support, and facilitating emotionally adaptive feedback. However, ethical considerations regarding privacy, consent, and the interpretability of affective data remain unresolved.

Figure 3 illustrates contrasting pedagogical functions of AI, showing its corrective role in writing versus its interactive role in speaking. In writing, AI primarily functions as a corrective tool, offering form-focused feedback, grammar and spelling correction, vocabulary suggestions, and coherence analysis. In contrast, in speaking, AI serves as an interactive partner, reducing anxiety, supporting fluency development, and providing immediate responses.

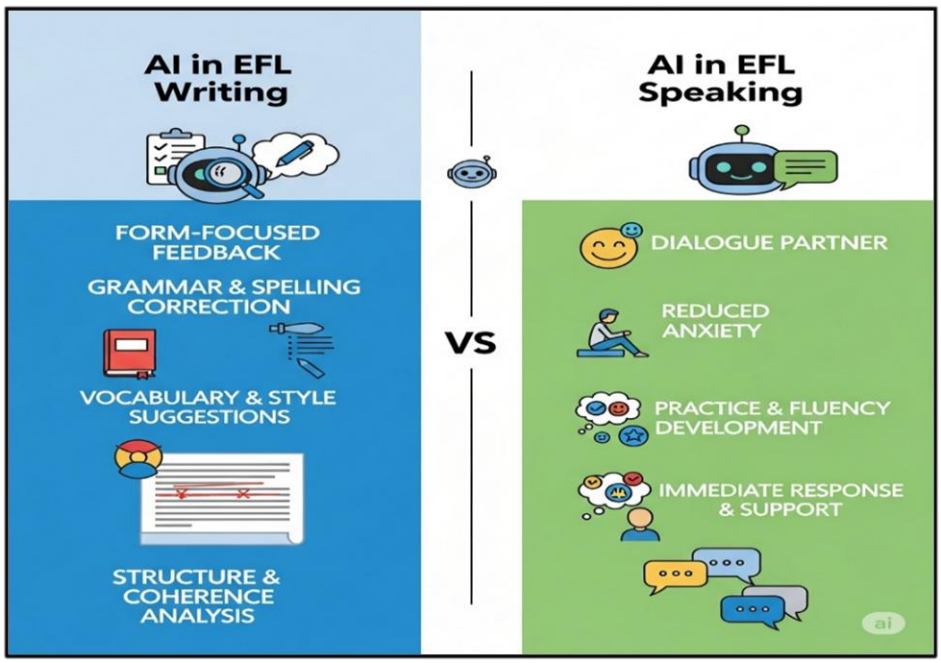

Figure 3: Contrasting pedagogical functions of AI in ESL/EFL writing and speaking

The contrast underscores the need for pedagogy-sensitive integration strategies. Importantly, these functional distinctions also intersect with educational contexts. As summarized in Table 3, reviews comparing K–12 and higher education reveal asymmetries: while both levels report affective gains and some benefits in oral communication, writing and self-regulated learning research remain predominantly focused on higher education. This imbalance highlights the need for more balanced and inclusive research across age groups.

Table 3: AI research in K–12 vs. Higher education

Theme / Skill	K-12	Higher Education	Overlap	Divergence
Writing	Limited coverage (few studies in K-12; AWF occasionally tested)	Strong focus (Feng et al., 2025; Shi & Aryadoust, 2024; Wang et al., 2025b)	n/a	Writing research is overwhelmingly higher education-centric, especially on AWF and argumentative essays.
Speaking / Oral Communication	Rare (some dialogue systems piloted in secondary school contexts)	Frequent focus (Hou & Min, 2025; Koç & Savaş, 2025; Li & Zhao, 2025)	Both levels benefit from AI speaking tools.	Empirical evidence is stronger in higher education.
Vocabulary / Reading / Listening	Minimal reporting (occasional chatbot or	Frequently targeted (Wu, 2024; Li et al., 2025b; Xu & Wang, 2024)	n/a	Vocabulary gains and multimodal listening tasks emphasized

Theme / Skill	K-12	Higher Education	Overlap	Divergence
	game-based applications)			mainly in higher education.
Learner Perceptions / Motivation / Affective Outcomes	Often emphasized in chatbots and emotional AI (Liu et al., 2024)	Motivation, anxiety reduction, and autonomy were widely reported (Chang & Sun, 2024; Wu, 2024)	Both levels highlight affective factors.	Higher education links affect self-regulated learning and autonomy, K–12 links affect engagement/fun.
Self-Regulated Learning)	Rare in K-12	Central in higher education (Chang & Sun, 2024; Feng et al., 2025)	n/a	SRL was explicitly theorized only in higher education.
Technology Orientation (E.g., ChatGPT, Emotional AI, AWF)	Limited integration, often exploratory (chatbots in class, emotional AI support)	Strong adoption and critique (overreliance on ChatGPT; AWF validation; diverse tools)	Chatbot trials in both.	Higher education dominates systematic adoption and critique.

Note. n/a = not applicable.

4. Discussion

The integration of AI into ESL/EFL education is no longer a speculative possibility but an empirical reality. As the findings in Sections 4.1 to 4.3. demonstrate, AI technologies have shown robust potential to enhance language learning outcomes, particularly in writing and speaking, while also reshaping the learner's experience in terms of self-regulation, engagement, and affective dynamics. Rather than reiterating specific results, this section extends their meaning by focusing on interpretation, synthesis, and the broader theoretical implications of AI integration.

4.1 From Effectiveness to Conditional Validity

While recent meta-analyses (e.g., Xu & Wang, 2024; Wu, 2024) report large positive effects of AI across multiple skills, the methodological underpinnings of many primary studies are cause for concern. As noted, heterogeneity in AI tool types, short intervention durations, and overreliance on self-reported measures undermine the internal and external validity of the aggregated results (Jeon et al., 2023; Lo et al., 2024b; Yiling et al., 2025).

The current study demonstrates that intervention duration, interface design, and learner characteristics systematically shape the strength of these effects,

underscoring that general effect size reporting is less informative than moderator-sensitive analyses. This contribution highlights the conditional validity of AI's effectiveness and reinforces the importance of contextual and pedagogical factors. Researchers must also be mindful of potential biases, as illustrated by Farangi and Nejadghanbar (2024), who investigated questionable research practices in AI-related applied linguistics.

4.2 AI's Pedagogical Function: Skill-Specific Contributions

AI's impact is not evenly distributed across all language domains. Writing and speaking, particularly when supported by automated feedback systems (Shi & Aryadoust, 2024) and dialogue-based chatbots (Du & Daniel, 2024; Jeon et al, 2024), respectively, stand out as the most thoroughly studied and pedagogically promising areas. In writing, form-focused feedback dominates at the expense of higher-order thinking and discourse coherence (Zhou et al., 2023).

In speaking, the affective benefits, particularly the reduction of language anxiety and enhancement of willingness to communicate, are significant (Fathi et al., 2024; Ghafouri, 2024; Li & Zhao, 2025; Zhang et al., 2024), but dialogue spontaneity and speech recognition limitations persist (Hou & Min, 2025; Koç & Savaş, 2025; Liu et al, 2025). This review interprets these patterns as evidence that AI plays differentiated pedagogical functions: corrective and form-focused in writing, confidence-building and interactive in speaking. The challenge lies in ensuring that such functions complement, rather than constrain, learner creativity and critical engagement.

Moreover, the present synthesis reveals clear cross-contextual imbalances: while affective gains and oral communication improvements are consistently reported across both school and university learners, research on writing and self-regulated learning remains heavily concentrated in higher education. This suggests that insights from university contexts cannot be uncritically generalized to K–12 learners, whose developmental needs and motivational profiles differ.

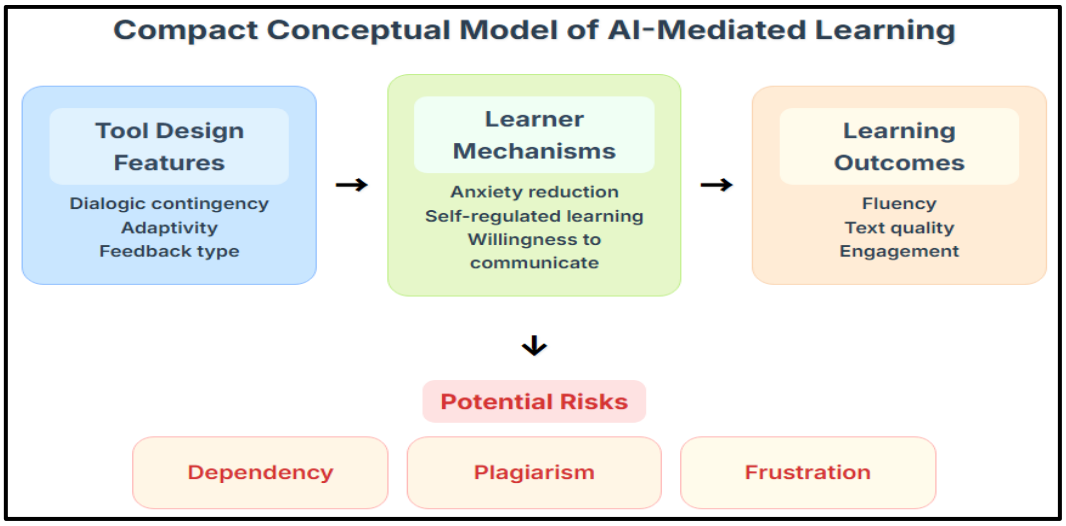

Figure 4: Compact conceptual model of AI-mediated learning

To integrate these insights, Figure 4 presents a conceptual model linking three core dimensions: (a) tool design features, (b) learner mechanisms, and (c) learning outcomes. Tool design features include dialogic contingency, adaptivity, and feedback type, which shape the ways AI systems interact with learners. These features activate mechanisms such as anxiety reduction, enhanced self-regulated learning, and increased willingness to communicate.

In turn, these mechanisms lead to outcomes such as improved fluency, greater text quality, and heightened engagement. Importantly, the model also acknowledges feedback loops: poorly designed tools may generate opposite effects, such as dependency, plagiarism, or learner frustration. By foregrounding this pathway from design → mechanism → outcome, the model clarifies how AI's pedagogical potential is contingent not on technology alone, but on its thoughtful integration into instructional contexts.

4.3 Rethinking the Role of the Learner

A key insight from the reviewed studies is the shift from viewing learners as passive recipients of AI-generated feedback to active participants navigating AI-mediated environments (Kundu & Bej, 2025). AI systems, particularly those designed for personalization and scaffolding, hold the potential to enhance self-regulated learning and foster metacognitive awareness (Chang & Sun, 2024; Pan et al, 2025; Xiao et al., 2025). However, when overused or poorly designed, these same systems risk undermining autonomy, critical thinking, and deep engagement, transforming helpful scaffolds into cognitive crutches (Feng et al., 2025; Lo et al, 2024a).

AI-mediated engagement produces both enabling and constraining effects. While AI tools can generate short-term motivation (Liu & Reinders, 2025) and participation (Guan et al., 2024), they may also elicit frustration, promote plagiarism, or erode intrinsic motivation if not thoughtfully integrated (Yang & Li, 2024). Moreover, emerging emotional AI systems offer exciting potential to humanize learning by responding to affective cues, yet their ethical implications—ranging from emotional surveillance to data privacy—require urgent attention (Liu et al., 2024).

4.4 Toward Pedagogy-Driven, Ethically Grounded AI Integration

The field is at a crossroads. While technological capabilities are rapidly expanding, the pedagogical and ethical frameworks lag behind (Kartal & Yeşilyurt, 2024; Li et al, 2025c; Zhang et al, 2025). A more cohesive trajectory between subtopics reveals that methodological validity, skill-specific contributions, and learner roles converge in one key imperative: pedagogy-first integration. Teachers must learn to leverage AI's strengths while mitigating risks such as bias, dependency, or academic dishonesty—for example, pairing AI feedback with peer review, or balancing chatbot use with human interaction. Moving forward, the field must adopt frameworks that define clear learning objectives, consider learner variability, and ensure AI complements rather than replaces human instruction.

4.5 Implications and Limitations

The current synthesis contributes to both theory and practice by clarifying the differentiated pedagogical functions of AI: as a corrective scaffold in writing and as a conversational partner in speaking. Theoretically, it highlights the conditional validity of AI effectiveness and underscores the need for skill-sensitive, context-dependent pedagogical frameworks. Practically, it links ethical concerns to classroom strategies, emphasizing how educators can balance AI's benefits with its risks through transparent policies, critical digital literacy training, and scaffolded integration that fosters autonomy rather than dependency.

At the same time, several limitations must be acknowledged. This study is constrained by its narrow time frame (2024–2025) and reliance on SSCI-listed journals, which may have excluded valuable perspectives from other contexts or gray literature. In addition, the qualitative synthesis approach did not permit recalculation of effect sizes. Best practices in systematic reviewing, such as employing two independent coders with inter-rater reliability, pre-registration in Operations Feedback Systems, and applying appraisal tools (e.g., AMSTAR 2 or ROBIS), were not feasible due to practical constraints. Restricting the dataset to SSCI-indexed journals enhanced the credibility of the review; however, this criterion may also have excluded high-quality non-SSCI research, thereby introducing potential bias.

5. Conclusion

The rapid proliferation of AI in ESL/EFL education presents both opportunities and challenges. This review synthesized recent meta-analyses, systematic reviews, and meta-syntheses (2024–2025) to evaluate AI's effectiveness, applications, and learner-centered dynamics. Across studies, AI integration consistently enhanced language learning, particularly in writing and speaking. Automated Writing Feedback (AWF) systems improved accuracy and cohesion, while dialogue-based chatbots supported fluency, confidence, and reduced speaking anxiety.

These benefits, however, must be interpreted cautiously. Reported effect sizes may be inflated by methodological limitations such as short interventions, small samples, and reliance on self-reports. Moreover, tool quality and function vary: AWF often focuses narrowly on form-based feedback, while chatbots remain constrained by conversational spontaneity and speech recognition. This review emphasizes that AI's effectiveness is conditional, moderated by intervention duration, interface design, learner characteristics, and educational context.

Beyond cognitive gains, AI reshapes learner roles. It can scaffold self-regulated learning by providing feedback, task management, and motivational support. Yet indiscriminate use risks undermining autonomy, encouraging shallow engagement, or fostering overreliance. Emerging emotional AI highlights both promise for affective support and urgent ethical concerns regarding privacy, surveillance, and data security. Furthermore, research remains disproportionately concentrated in higher education, leaving K–12 contexts underexplored and limiting cross-level generalizability.

A pedagogy-first and ethically grounded approach is therefore imperative. Future research should adopt longitudinal and mixed-methods designs, sharpen research agendas with cross-skill integration and multi-site studies, and incorporate learning analytics with robust safeguards. AI must be designed with usability and contextual relevance, centering both learners and teachers. In sum, AI holds significant promise, but its success depends not on technology alone but on alignment with pedagogy, methodological rigor, ethical responsibility, and learner empowerment. Responsible, pedagogy-driven integration will be essential to balance opportunity and risk in the evolving AI-enhanced educational landscape.

6. Acknowledgments

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2022S1A5B5A16051353).

7. References

Chang, W. L., & Sun, J. C. Y. (2024). Evaluating AI's impact on self-regulated language learning: A systematic review. *System, 126*, 103484. https://doi.org/10.1016/j.system.2024.103484

Daud, A., Aulia, A. F., Muryanti, Harfal, Z., Nabilla, O., & Ali, H. S. (2025). Integrating artificial intelligence into English language teaching: A systematic review. *European Journal of Educational Research, 14*(2), 677–691. https://doi.org/10.12973/eu-jer.14.2.677

Du, J., & Daniel, B. K. (2024). Transforming language education: A systematic review of AI-powered chatbots for English as a foreign language speaking practice. *Computers and Education: Artificial Intelligence, 6*, 100230. https://doi.org/10.1016/j.caeai.2024.100230

Duman, G., Orhon, G., & Gedik, N. (2015). Research trends in mobile assisted language learning from 2000 to 2012. *ReCALL, 27*(2), 197–216. https://doi.org/10.1017/S0958344014000287

Farangi, M. R., & Nejadghanbar, H. (2024). Investigating questionable research practices among Iranian applied linguists: Prevalence, severity, and the role of artificial intelligence tools. *System, 125*, 103427. https://doi.org/10.1016/j.system.2024.103427

Fathi, J., Rahimi, M., & Derakhshan, A. (2024). Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions. *System, 121*, 103254. https://doi.org/10.1016/j.system.2024.103254

Feng, H., Li, K., & Zhang, L. J. (2025). What does AI bring to second language writing? A systematic review (2014-2024). *Language Learning & Technology, 29*(1), 1–27. https://doi.org/10.64152/10125/73629

Ghafouri, M. (2024). ChatGPT: The catalyst for teacher-student rapport and grit development in L2 class. *System, 120*, 103209. https://doi.org/10.1016/j.system.2023.103209

Guan, L., Li, S., & Gu, M. M. (2024). AI in informal digital English learning: A meta-analysis of its effectiveness on proficiency, motivation, and self-regulation. *Computers and Education: Artificial Intelligence, 7*, 100323. https://doi.org/10.1016/j.caeai.2024.100323

Hou, Z., & Min, S. (2025). Dialogue-based computer-assisted language learning systems for second language speaking development: A three-level meta-analysis. *ReCALL*, 1–17. https://doi.org/10.1017/S0958344025100268

Jeon, J., Lee, S., & Choe, H. (2023). Beyond ChatGPT: A conceptual framework and systematic review of speech-recognition chatbots for language learning.

Computers & Education, 206, 104898. https://doi.org/10.1016/j.compedu.2023.104898

Jeon, J., Lee, S., & Choi, S. (2024). A systematic review of research on speech-recognition chatbots for language learning: Implications for future directions in the era of large language models. *Interactive Learning Environments, 32*(8), 4613–4631. https://doi.org/10.1080/10494820.2023.2204343

Kartal, G., & Yeşilyurt, Y. E. (2024). A bibliometric analysis of artificial intelligence in L2 teaching and applied linguistics between 1995 and 2022. *ReCALL, 36*(3), 359–375. https://doi.org/10.1017/S0958344024000284

Koç, F. Ş., & Savaş, P. (2025). The use of artificially intelligent chatbots in English language learning: A systematic meta-synthesis study of articles published between 2010 and 2024. *ReCALL, 37*(1), 4–21. https://doi.org/10.1017/S0958344024000168

Kundu, A., & Bej, T. (2025). Transforming EFL teaching with AI: A systematic review of empirical studies. *International Journal of Artificial Intelligence in Education*. Advance online publication. https://doi.org/10.1007/s40593-025-00470-0

Lee, S., Choe, H., Zou, D., & Jeon, J. (2025). Generative AI (GenAI) in the language classroom: A systematic review. *Interactive Learning Environments.* Advance online publication. https://doi.org/10.1080/10494820.2025.2498537

Li, B., Tan, Y. L., Wang, C., & Lowell, V. (2025a). Two years of innovation: A systematic review of empirical generative AI research in language learning and teaching. *Computers and Education: Artificial Intelligence, 9*, 100226. https://doi.org/10.1016/j.caeai.2025.100445

Li, M., Wang, Y., & Yang, X. (2025b). Can generative AI chatbots promote second language acquisition? A meta-analysis. *Journal of Computer Assisted Learning, 41*(4), e70060. https://doi.org/10.1111/jcal.70060

Li, Y., Zhou, X., & Chiu, T. K. F. (2025c). Systematics review on artificial intelligence chatbots and ChatGPT for language learning and research from self-determination theory (SDT): What are the roles of teachers? *Interactive Learning Environments, 33*(3), 1850–1864. https://doi.org/10.1080/10494820.2024.2400090

Li, D., & Zhao, Y. (2025). Artificial intelligence applications for oral communication skills in EFL contexts: A systematic review. *Asia-Pacific Education Researcher*. Advance online publication. https://doi.org/10.1007/s40299-025-01023-8

Liu, M., & Reinders, H. (2025). Do AI chatbots impact motivation? Insights from a preliminary longitudinal study. *System, 128*, 103544. https://doi.org/10.1016/j.system.2024.103544

Liu, Y., binti Ab Rahman, F., & binti Mohamad Zain, F. (2025). A systematic literature review of research on automatic speech recognition in EFL pronunciation. *Cogent Education, 12*(1), 2466288. https://doi.org/10.1080/2331186X.2025.2466288

Liu, Y., Zhang, H., Jiang, M., Chen, J., & Wang, M. (2024). A systematic review of research on emotional artificial intelligence in English language education. *System, 126*, 103478. https://doi.org/10.1016/j.system.2024.103478

Lo, C. K., Hew, K. F., & Jong, M. S. Y. (2024a). The influence of ChatGPT on student engagement: A systematic review and future research agenda. Computers & Education, 219, 105100. https://doi.org/10.1016/j.compedu.2024.105100

Lo, C. K., Yu, P. L. H., Xu, S., Ng, D. T. K., & Jong, M. S.-Y. (2024b). Exploring the application of ChatGPT in ESL/EFL education and related research issues: A systematic review of empirical studies. *Smart Learning Environments, 11*, Article 21. https://doi.org/10.1186/s40561-024-00342-5

Lo, A. W. T. (2025). The educational affordances and challenges of generative AI in global Englishes-oriented materials development and implementation: A critical ecological perspective. *System, 130*, 103610. https://doi.org/10.1016/j.system.2025.103610

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffman, T. C., Mulrow, C. D.,... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Systematic Reviews, 10*, Article 89. https://doi.org/10.1186/s13643-021-01626-4

Pan, M., Lai, C., & Guo, K. (2025). Technology-enhanced self-regulated second language learning interventions: A systematic review. *System, 131*, 103675. https://doi.org/10.1016/j.system.2025.103675

Schotten, M., Aisati, M., Meester, W. J. N., Steigninga, S., & Ross, C. A. (2018). A brief history of Scopus: The world's largest abstract and citation database of scientific literature. In F. J. Cantu-Ortiz (Ed.), *Research analytics: Boosting university productivity and competitiveness through scientometrics* (pp. 33–57). Taylor & Francis.

Shi, H., & Aryadoust, V. (2024). A systematic review of AI-based automated written feedback research. *ReCALL, 36*(2), 187–209. https://doi.org/10.1017/S0958344023000265

Wahyuni, S., Putro, N. H. P. S., & Efendi, A. (2024). Trends in artificial intelligence-infused English language learning: A comprehensive bibliometric and content review. *Advanced Education, 2024*(25), 162–178. https://doi.org/10.20535/2410-8286.315035

Wang, H., Abdul Aziz, A., & Kutty, F. M. (2025a). Integrating AI into Asian tertiary EFL learners' speaking instruction: A systematic literature review. *Forum for Linguistic Studies, 7*(3), 104–119. https://doi.org/10.30564/fls.v7i3.8263

Wang, Y., Zhang, T., Yao, L., & Seedhouse, P. (2025b). A scoping review of empirical studies on generative artificial intelligence in language education. *Innovation in Language Learning and Teaching*. Advance online publication. https://doi.org/10.1080/17501229.2025.2509759

Wu, X., & Li, R. (2024). Unraveling effects of AI chatbots on EFL learners' language skill development: A meta-analysis. *Asia-Pacific Education Researcher*. Advance online publication. https://doi.org/10.1007/s40299-024-00853-2

Wu, X.-Y. (2024). Artificial intelligence in L2 learning: A meta-analysis of contextual, instructional, and social-emotional moderators. *System, 126*, 103498. https://doi.org/10.1016/j.system.2024.103498

Xiao, Y., Liu, X., & Yao, Y. (2025). Students' development of AI metacognitive awareness in an EAP course: A qualitative inspection through the experiential learning theory. *System, 132*, 103790. https://doi.org/10.1016/j.system.2025.103790

Xu, T., & Wang, H. (2024). The effectiveness of artificial intelligence on English language learning achievement. *System, 125*, 103428. https://doi.org/10.1016/j.system.2024.103428

Yang, L., & Li, R. (2024). ChatGPT for L2 learning: Current status and implications. *System, 124*, 103351. https://doi.org/10.1016/j.system.2024.103351

Yiling, J., Omar, M., & Kamaruzaman, F. M. (2025). Exploring the AI-enhanced project-based learning for English language acquisition: A systematic review of the key elements and emerging technology trends. *International Journal of Learning, Teaching and Educational Research, 24*(2), 636–652. https://doi.org/10.26803/ijlter.24.2.31

Zhang, D., Wu, J. G., & Fu, Z. (2024). From shy to fly: Facilitating EFL learners' willingness to communicate with an AI chatbot and an intelligent tutoring system. *System, 127*, 103501. https://doi.org/10.1016/j.system.2024.103501

Zhang, Y., Lai, C., & Gu, M. M. Y. (2025). Becoming a teacher in the era of AI: A multiple-case study of pre-service teachers' investment in AI-facilitated learning-to-teach practices. *System, 133*, 103746. https://doi.org/10.1016/j.system.2025.103746

Zhou, T., Cao, S., Zhou, S., Zhang, Y., & He, A. (2023). Chinese intermediate English learners outdid ChatGPT in deep cohesion: Evidence from English narrative writing. *System, 118*, 103141. https://doi.org/10.1016/j.system.2023.103141