

## Generative AI (GenAI) in the language classroom: A systematic review

Seongyong Lee, Hohsung Choe, Di Zou & Jaeho Jeon

**To cite this article:** Seongyong Lee, Hohsung Choe, Di Zou & Jaeho Jeon (22 May 2025):  
Generative AI (GenAI) in the language classroom: A systematic review, Interactive Learning  
Environments, DOI: [10.1080/10494820.2025.2498537](https://doi.org/10.1080/10494820.2025.2498537)

**To link to this article:** <https://doi.org/10.1080/10494820.2025.2498537>



© 2025 The Author(s). Published by Informa  
UK Limited, trading as Taylor & Francis  
Group



Published online: 22 May 2025.



Submit your article to this journal [↗](#)



Article views: 11017



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 18 View citing articles [↗](#)

REVIEW ARTICLE



# Generative AI (GenAI) in the language classroom: A systematic review

Seongyong Lee <sup>a</sup>, Hohsung Choe <sup>b</sup>, Di Zou <sup>c</sup> and Jaeho Jeon <sup>d</sup>

<sup>a</sup>School of Education and English, University of Nottingham Ningbo China, Ningbo, People's Republic of China; <sup>b</sup>Department of English Linguistics, Hankuk University of Foreign Studies, Seoul, Republic of Korea; <sup>c</sup>Department of English and Communication, The Hong Kong Polytechnic University, Kowloon, Hong Kong; <sup>d</sup>Department of Instruction and Curriculum, Indiana University Bloomington, Bloomington, USA

## ABSTRACT

The integration of generative AI (GenAI) into language education has attracted significant attention due to its versatile capabilities. However, previous reviews rarely focused on classroom-based research with language learners, a crucial aspect for understanding GenAI's practical application. Therefore, this systematic review addresses this gap by examining 49 empirical studies published between January 2023 and December 2024, focusing on research design, research focus, and the roles and challenges of GenAI in language learning. Findings show that higher education was the dominant setting and English was the primary target language. Although the studies employed diverse research methods, they largely relied on self-reported data. They mainly investigated learner perceptions, such as attitude, self-efficacy, and motivation, with language acquisition studies focusing primarily on writing skills. GenAI served diverse roles, including feedback provider, learning tutor, cognitive stimulator, interaction facilitator, and conversation partner. However, its integration revealed technological challenges, such as content quality and querying issues, as well as educational challenges, including overreliance, pedagogical limitations, and concerns over academic integrity. The review suggests that future research should diversify its research focus and tools, extend studies into K-12 contexts, employ longitudinal designs, explore behavioral interactions with GenAI, and develop customized GenAI applications for personalized learning.

## ARTICLE HISTORY

Received 12 August 2024  
Accepted 21 April 2025

## KEYWORDS

Generative AI; ChatGPT; artificial intelligence; large language model; chatbot; language learning

## 1. Introduction

The launch of ChatGPT has marked a significant shift in language learning, as GenAI, leveraging large language models (LLMs), offers a substantial leap in its capabilities beyond previous AI chatbots (W. Huang et al., 2022; Ji et al., 2023). GenAI tools, such as ChatGPT, Gemini, and Claude, enable more natural, contextually rich interactions and enhance learner experiences across diverse language learning contexts (Chiu, 2023; W. Huang et al., 2022; Ji et al., 2023; Schneider, 2024). The literature highlights GenAI's diverse pedagogical advantages for language education through personalized learning with immediate and targeted feedback that improves writing, reading, and speaking skills (Karataş et al., 2024; Pan et al., 2025; Tai & Chen, 2024). It also fosters learner autonomy and

**CONTACT** Jaeho Jeon  jaehojeon21@gmail.com

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group  
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

reduces anxiety, supporting diverse proficiency levels and individual differences, including grit, motivation, and self-efficacy (Ghafouri, 2024; J. H. Lee et al., 2023; X. Li et al., 2023). However, concerns regarding academic integrity, content bias, feedback quality, and data security necessitate careful consideration in its actual integration into classrooms (Abdelhalim, 2024; Long, 2024; Zou & Huang, 2023).

Recent reviews have explored GenAI's affordances and ethical implications in language education (Law, 2024; Lai & Lee, 2024; B. Li et al., 2024; Teng, 2024b; Yang & Li, 2024). Despite this growing attention, significant gaps remain in literature. First, there is a lack of classroom-based empirical research involving learners, which limits our understanding of GenAI's pedagogical impact. Second, many reviews combine teacher and learner perspectives, potentially masking unique insights from learners. Finally, several reviews either adopt a broad scope or focus on only one language skill, restricting a full view of GenAI's holistic application.

This systematic review addresses these gaps by focusing on the classroom-based GenAI usage in second (L2) and foreign language (FL) learning. It aims to clarify how evidence-based GenAI integration impacts language learning practices, while identifying and addressing ethical and instructional challenges.

## 2. Literature review

### 2.1. Impact of GenAI in language learning: benefits and challenges

Since ChatGPT's introduction in late 2022, GenAI tools have revolutionized language education (Chiu, 2023; Karataş et al., 2024). While previous AI chatbots craft through Dialogflow or intelligent assistants offered language learning benefits for skill development (Ji et al., 2023), affective aspects (Ji et al., 2023), and intercultural communication (Lee et al., 2025a), these earlier models experienced communication breakdowns due to their reliance on pre-defined dialogue scenarios and limited corpus data (W. Huang et al., 2022). Their rule-based systems failed to adapt to unexpected responses beyond scripted interactions, resulting in unsatisfactory learner experiences (Lee & Jeon, 2024b). Their restricted ability to process natural language and contexts limited capacity to generate diverse, natural, and contextually appropriate responses, ultimately hindering effective language learning (Ji et al., 2023).

In contrast, LLM-based GenAI tools offer advanced capabilities for human-like interaction, as they are pre-trained on substantially larger language datasets to produce optimal multimodal outputs (Chiu, 2023). This expansion in training data, along with the use of transformer architectures featuring parallel processing and attention mechanisms, enables LLMs to better understand context, generate more diverse and spontaneous responses, and manage open-ended conversations (Schneider, 2024). In language learning, these features allow GenAI to produce context-aware, natural responses to unpredictable learner input, fostering more immersive and interactive learning experiences (Jeon & Lee, 2023; Kohnke et al., 2023). Accordingly, GenAI has sparked active debate among scholars regarding its pedagogical benefits and challenges (Mai et al., 2024).

In language education, GenAI tools have shown a potential to support language skills through immediate feedback and personalized instruction across different levels (Pan et al., 2025; Tai & Chen, 2024). Empirical studies indicate that, when effectively integrated into classroom settings, these tools can support language acquisition. For example, Tai and Chen (2024), using a mixed-methods design with 85 Taiwanese elementary students, found that both individual and paired interactions with ChatGPT improved speaking performance. Using sociocultural theory, their study highlighted how GenAI-mediated interactions lowered anxiety for authentic communication by bridging classroom learning and real-world language use. Similarly, Boudouaia et al. (2024), in a 10-week quasi-experimental study with 76 Algerian undergraduates, examined ChatGPT's impact on EFL writing using the Technology Acceptance Model. They reported that

immediate, detailed feedback facilitated students' writing performance in task achievement, coherence, and grammatical accuracy.

Despite its advantages, research raises concerns about GenAI utilization in language learning, particularly regarding its authenticity and reliability in classroom contexts (Tseng & Lin, 2024; Vaccino-Salvadore, 2023). Studies have highlighted issues such as academic dishonesty (Boudouaia et al., 2024), content bias (Wiboolyasarin et al., 2024), unreliable information sources (Vaccino-Salvadore, 2023), and data security risks (Zou & Huang, 2023). In technological aspects, GenAI's content generation can compromise academic integrity by producing inaccurate information and fabricated citations, which learners often fail to recognize (Abdelhalim, 2024). Its inconsistent feedback quality may confuse lower-proficiency learners and hinder autonomous language development (Long, 2024). Moreover, GenAI may reinforce stereotypical cultural representations and favors standardized language varieties, potentially disadvantaging culturally marginalized groups (Payne et al., 2024). These biases often stem from training data dominated by mainstream voices and algorithm designs based on statistical probabilities (Navigli et al., 2023), which can limit the development of intercultural competence among multilingual learners (Dai et al., 2025). Thus, scholars recommend designing language courses that foster critical thinking (Xiao & Zhi, 2023), leverage the GenAI-based formative feedback (Banihashem et al., 2024), and enhance teachers' digital literacy (Jeon et al., 2024).

The review of literature highlights diverse topics related to GenAI usage in language learning classrooms. Evidence on GenAI's effectiveness across language learning areas shows that its benefits largely depend on factors such as learner proficiency, instructional design, and implementation strategy (Yang & Li, 2024). To balance GenAI's potential with its limitations, a comprehensive approach is necessary, including systematic reviews that synthesize existing evidence for effective practices and offer clear guidance for addressing concerns, including academic integrity, content reliability, and sociocultural representation (Lai & Lee, 2024).

## ***2.2. Previous systematic reviews on the use of GenAI for language learning***

To date, five systematic reviews have examined the use of GenAI in language teaching and learning, with three focusing on its overall benefits and limitations. Law (2024) analyzed 41 articles published between 2017 and July 2023, exploring key research areas, educational opportunities, and limitations of GenAI. The methodological challenge was the inclusion of both empirical research and non-empirical research (e.g. review papers offering critical technology analysis and position papers presenting researchers' specific viewpoints). It may blur the distinction between classroom applications and conceptual insights, making it harder to assess GenAI's direct educational impact. B. Li et al. (2024) reviewed 36 empirical studies from November 2022 to November 2023, focusing on ChatGPT's roles in language education and future directions. It identified focal areas like ChatGPT's benefits, perceived roles, learner attitudes, and instructional design, presenting ChatGPT's diverse functions such as conversation practice, writing assistance, teaching support, information gathering, and assessment with feedback. Yang and Li (2024) examined 44 studies from 2022 to 2023, highlighting ChatGPT's roles as a content generator, feedback provider, teaching aid, and assessment tool, alongside potential challenges. However, the three reviews combined teacher and student perspectives, as well as formal and informal learning contexts, which limits knowledge of ChatGPT's practical implications for learners in classrooms (Walter, 2024).

Two review studies take a broader perspective beyond GenAI or only focus on one skill. Lai and Lee (2024) reviewed 32 empirical studies from 2013 to 2023, examining conversational AI tools in ELT, including ChatGPT. The review identified affective, cognitive, and behavioral benefits, alongside ecological factors influencing chatbot integration. The authors called for more comprehensive research into long-term impacts and broader educational contexts beyond language-specific use. Teng (2024b) examined 20 empirical studies published between 2023 and early 2024, focusing on ChatGPT's effective role in EFL writing instruction, while raising ethical concerns such as plagiarism, overreliance, and reduced critical thinking. Both reviews revealed methodological limitations,

including the exclusion of recent studies beyond early 2024 and a lack of classroom-based empirical research and student-centered inquiry.

Although the five review studies offer insights into GenAI usage for language education from different angles, three research gaps should be identified. First, they lack a strong focus on classroom-based research involving student participants. Empirical data from real classroom settings – whether online or offline – provides critical insights into students’ experiences with GenAI and helps deepen our understanding of AI-student interaction in authentic learning environments (Chiu, 2023; Liang et al., 2023). Examining how GenAI tools are integrated into instructional practices in classrooms can offer evidence-based pedagogical implications for effective language teaching (Jeon & Lee, 2023; Walter, 2024). Second, the reviews addressed both teachers and learners broadly, rather than centering specifically on learners. This limits our understanding of learner behavior, perspectives, and experiences with GenAI, even though teaching and learning are interdependent processes (Chiu, 2023; Jeon et al., 2024). As noted in the literature, learners may interact with GenAI through self-regulated learning (Wiboolyasarini et al., 2024), while teachers focus on orchestrating its use for instruction (Jeon et al., 2024). A learner-centered review is thus important to uncover insights specific to student engagement (J. H. Lee et al., 2024). Third, several studies (e.g. Lai & Lee, 2024; Teng, 2024b) did not focus exclusively on GenAI or language education. Lai and Lee (2024) examined both GenAI and non-GenAI tools, while Teng (2024b) addressed ChatGPT’s impact on EFL writing only. As B. Li et al. (2024) suggest, one reason for these gaps may be the publication timelines – most reviewed studies were published before early 2024 – leading to limited classroom-based, learner-focused empirical research and broader inclusion of teachers, informal contexts, and non-GenAI tools.

Therefore, this systematic review examines how GenAI tools are used for language learning in classroom settings with a particular focus on L2 or FL learners’ experiences. We analyze classroom-based empirical studies in three domains: research design (Liang et al., 2023; Mai et al., 2024), research focus (Jeon & Lee, 2023; Ji et al., 2023; Law, 2024), and roles and challenges (Jeon & Lee, 2023; B. Li et al., 2024; L. Yan et al., 2024). The following research questions (RQs) guided our study.

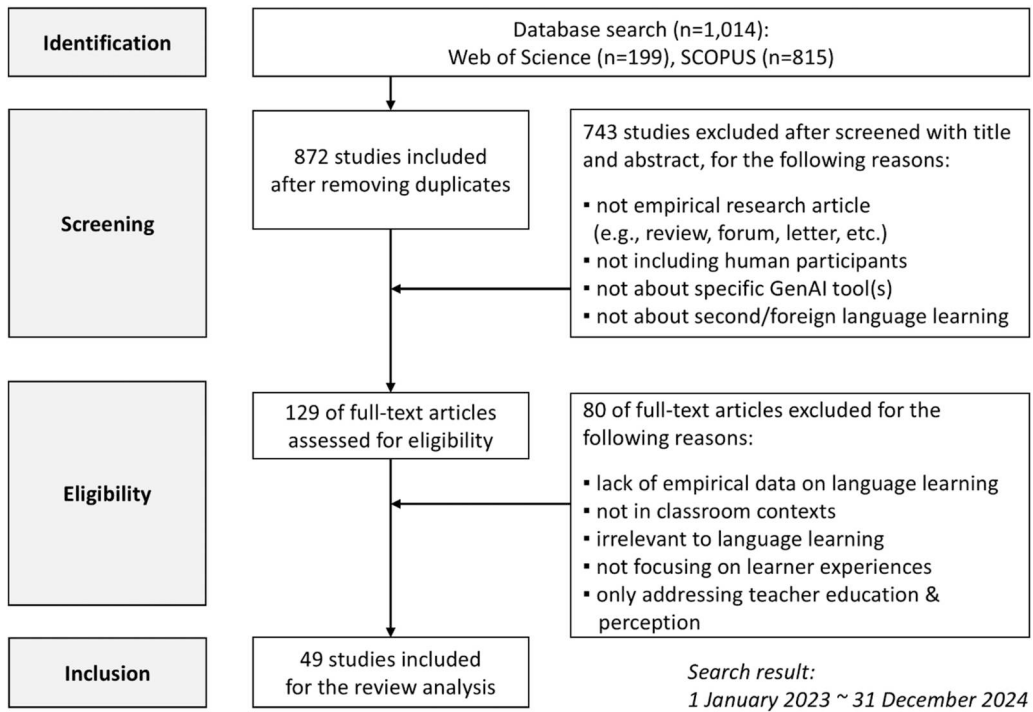
1. In what research design are GenAI tools utilized for classroom-based language learning?
2. What are the focuses of research regarding the use of GenAI in language learning?
3. What are the roles and challenges of GenAI in language learning?

### 3. Methods

#### 3.1. Data collection process

We used the PRISMA guidelines to collect data for current research (Moher et al., 2009; Page et al., 2021). Drawing on previous studies on AI chatbots and GenAI (e.g. Ji et al., 2023; Mai et al., 2024), we used Web of Science and SCOPUS to “find high-quality peer-reviewed publications” (L. Yan et al., 2024, p. 96). To identify search terms relevant to current research, we reviewed highly cited peer-reviewed journals across various education fields, including language education, computer-assisted language learning, general education, and educational technology. We determined search keywords on GenAI (“Generative AI” OR “Generative artificial intelligence” OR “ChatGPT” OR “GPT” OR “Large language model” OR “DALL·E” OR “Bing” OR “Copilot” OR “Claude” OR “Gemini” OR “Midjourney” OR “Bard”) and language learning (“language learning” OR “language learner” OR “language education” OR “language acquisition” OR “second language” OR “foreign language” OR “EFL” OR “ESL”).

Our search was limited to research articles published between January 2023 and December 2024 (see Figure 1), considering the public release of ChatGPT in late 2022. Using “title,” “abstract,” and “keywords” as filtering criteria, we initially found 1014 articles written in English from the two



**Figure 1.** Article selection flow.

databases. After excluding duplicates, we included studies that (1) adopted an empirical research design, (2) included human participants, (3) used specific GenAI tools, and (4) were about L2 or FL learning. On the other hand, we excluded articles that belong to non-empirical research (e.g. review, editorial, position paper, conference paper, etc.), only focused on GenAI's functionalities without human participants, and addressed general education or first language acquisition. Accordingly, we obtained 129 articles at this screening stage.

Next, through a full-text search, we filtered articles that were not closely relevant to our RQs by the following excluding criteria: (1) lacking empirical data on language learning; (2) not in classroom contexts; (3) irrelevant to language learning; (4) not focusing on learner experience; (5) only addressing teacher education and perceptions. As a result, a total of 49 studies were included in the final set for data analysis.

### 3.2. Data analysis

We adopted a hybrid thematic analysis combining deductive and inductive methods (Fereday & Muir-Cochrane, 2006). First, we developed a coding scheme with various research dimensions and categories based on previous reviews of AI chatbots and ChatGPT (Table 1). For RQ1, we incorporated categories including location (Mai et al., 2024), target language (Ji et al., 2023), educational level (Jeon & Lee, 2023; Ji et al., 2023), sample size and language proficiency (Liang et al., 2023), research method (W. Huang et al., 2022; Ji et al., 2023), data source (Law, 2024), intervention duration (Liang et al., 2023), and specific GenAI tool (W. Huang et al., 2022). Regarding RQ2, we used three research focus categories: (1) perception (Jeon & Lee, 2023; Liang et al., 2023); (2) acquisition (Ji et al., 2023; Liang et al., 2023); and (3) behavior (Liang et al., 2023). In relation to RQ3, we included various GenAI roles such as evaluator, resource provider, and feedback provider (Jeon & Lee, 2023; Liang et al., 2023), and referred to previous review articles to identify challenges associated with GenAI use, including technological and pedagogical aspects (Mai et al., 2024; L. Yan et al., 2024).

**Table 1.** Coding scheme regarding research dimensions and categories.

Dimension	Category	Subcategory	Reference
Research design	Location	Countries	Mai et al. (2024)
	Target language	Different languages	Ji et al. (2023)
	Education level	<ul style="list-style-type: none"> <li>• Preschool or kindergarten</li> <li>• Elementary education</li> <li>• Secondary education</li> <li>• Higher education</li> <li>• Adult workers</li> </ul>	Jeon and Lee (2023)
	Sample size	<ul style="list-style-type: none"> <li>• Small (S): <math>\leq 10</math></li> <li>• Medium (M): 11–30</li> <li>• Large (L): 31–100</li> <li>• Extra-large (XL): <math>\geq 101</math></li> </ul>	Liang et al. (2023)
	Language proficiency	<ul style="list-style-type: none"> <li>• Beginner</li> <li>• Intermediate</li> <li>• Advanced</li> <li>• Mixed</li> <li>• Not specified</li> </ul>	Liang et al. (2023)
	Method	<ul style="list-style-type: none"> <li>• Quantitative</li> <li>• Qualitative</li> <li>• Mixed-methods</li> </ul>	W. Huang et al. (2022)
	Data source	<ul style="list-style-type: none"> <li>• Learning artifacts</li> <li>• Questionnaire</li> <li>• Interview</li> <li>• Observation</li> <li>• Reflection paper</li> <li>• Group discussion</li> <li>• Study log</li> </ul>	Law (2024)
	Intervention duration	<ul style="list-style-type: none"> <li>■ One session</li> <li>• Short-term: <math>\leq 10</math> weeks</li> <li>• Intermediate-term: 11–16 weeks</li> <li>• Long-term: <math>&gt; 16</math> weeks</li> <li>• Not specified</li> </ul>	Liang et al. (2023)
	GenAI tool	<ul style="list-style-type: none"> <li>• ChatGPT: ChatGPT3, ChatGPT3.5, ChatGPT4.0</li> <li>• Bing AI</li> <li>• Midjourney</li> <li>• Others: Claude, Bard, Gemini</li> <li>• Not specified</li> </ul>	Jeon and Lee (2023)
			W. Huang et al. (2022)
Research focus	Perception	<ul style="list-style-type: none"> <li>• Learning benefit (e.g. writing, grammar, vocabulary)</li> <li>• Affective aspect (e.g. attitude, grit, self-efficacy, WTC, motivation, acceptance)</li> <li>• Knowledge (e.g. culture, subject, digital literacy)</li> <li>• Competence (e.g. communicative)</li> </ul>	Jeon and Lee (2023)
	Acquisition	<ul style="list-style-type: none"> <li>• Reading</li> <li>• Writing</li> <li>• Listening</li> <li>• Speaking</li> <li>• Grammar</li> <li>• Vocabulary</li> </ul>	Liang et al. (2023)

(Continued)

**Table 1.** Continued.

Dimension	Category	Subcategory	Reference
Pedagogical application	Behavior	<ul style="list-style-type: none"><li>• Classroom interaction</li><li>• Linguistic interaction</li><li>• Culture or linguistic ecology</li><li>• Individual difference</li></ul>	Liang et al. (2023)
	Role of GenAI	<ul style="list-style-type: none"><li>• Evaluator</li><li>• Interviewer</li><li>• Resource provider</li><li>• Feedback provider</li><li>• Conversation partner</li></ul>	Jeon and Lee (2023) Ji et al. (2023) B. Li et al. (2024)
	Challenge	<ul style="list-style-type: none"><li>• Technological aspect (e.g. content quality, bias, privacy)</li><li>• Pedagogical challenge (e.g. ethical issue, academic integrity, overreliance)</li></ul>	Mai et al. (2024) L. Yan et al. (2024)

Next, we conducted a thematic analysis using a hybrid approach that combined deductive and inductive coding, described as “the data-driven inductive approach” (Fereday & Muir-Cochrane, 2006, p. 83). We first applied the coding scheme from previous review studies (Table 1) to the 49 selected studies. While our initial categories (i.e. research focus, GenAI role, challenges) covered most data, we identified gaps where important themes were missing. To address these gaps, we introduced an inductive phase using iterative cycles of open coding and constant comparison. Two authors independently reviewed the datasets, assigning initial codes to content that did not fit the existing scheme. Emergent codes included “student interaction” (e.g. Abdelhalim, 2024) and “teacher-student rapport” (Ghafouri, 2024) within the broader category of classroom interaction in Table 1. We also identified new GenAI roles such as “cognitive stimulator” and “affective supporter” (Ji et al., 2023; Liang et al., 2023; Mai et al., 2024).

We refined and validated these emerging codes through an iterative process, merging overlapping codes and developing clear operational definitions. The refined codes were then applied to the entire dataset to ensure analysis consistency. For validation, two researchers coded overlapping data subsets. The inter-coder reliability for the final coding results, calculated using Cohen’s kappa, showed a high agreement level of 0.82. Any remaining discrepancies were resolved through discussion until consensus was reached on the final codes, which are reflected in our findings. This systematic approach allowed us to integrate coded categories (e.g. querying challenge, pedagogical limitation, assessment disruption) with the original deductive framework following the coding scheme.

## 4. Results

### 4.1. RQ1: research design

#### 4.1.1. Location

The results (Table 2) show that most of the studies were conducted in Asia ( $n = 39$ ) that includes 13 locations, such as China ( $n = 16$ ), including mainland China ( $n = 7$ ), Hong Kong ( $n = 4$ ), Taiwan ( $n = 3$ ), and Macau ( $n = 2$ ), Saudi Arabia ( $n = 5$ ), South Korea ( $n = 4$ ), Japan ( $n = 3$ ), Türkiye ( $n = 3$ ), Iran ( $n = 2$ ), Thailand ( $n = 2$ ), Indonesia ( $n = 2$ ), Oman ( $n = 1$ ), and India ( $n = 1$ ). Four studies were done in Europe, including the Netherlands ( $n = 1$ ), Germany ( $n = 1$ ), Czech ( $n = 1$ ), and Norway ( $n = 1$ ). Three studies were conducted in Oceania (New Zealand,  $n = 3$ ), while two studies were done in the USA and one study in Africa (Algeria,  $n = 1$ ). This underscores the diverse range of locations and the global interest in using GenAI for language learning.

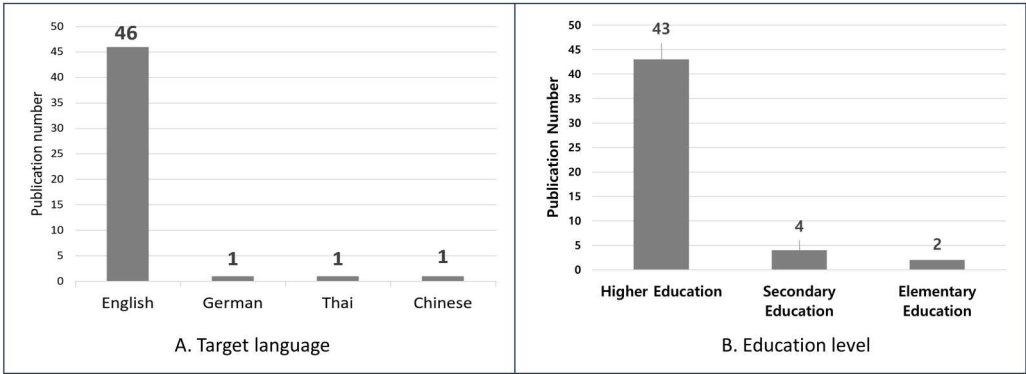
The concentration of research from Asia indicates regional research priorities and existing structural advantages. However, this geographical imbalance limits the generalizability of findings, as noted in the literature (Chiu et al., 2023; Liang et al., 2023). It also creates gaps in understanding how GenAI is integrated across diverse educational contexts worldwide, given

**Table 2.** Summary of research design.

Author	Location (country)	Target language	Edu. level	Sample Size	Language proficiency	Method	Data collection method	Duration (week)	Tool
Abdelhalim (2024)	Saudi Arabia	English	HE	27	Intermediate	Mixed	Questionnaire, interview, study log	13	ChatGPT
Alm (2024)	New Zealand	German	HE	4	Not specified	Quantitative	Reflection, portfolio, interview	13	ChatGPT3.5, DeepL
Almanea (2024)	Saudi Arabia	English	HE	135	Not specified	Mixed	Questionnaire, interview	15	ChatGPT3.5
Alsager (2024)	Saudi Arabia	English	HE	83	Intermediate	Quantitative	Questionnaire	Not specified	ChatGPT3.5
Alsofyani and Barzanji (2025)	Saudi Arabia	English	HE	102	Intermediate	Mixed	Writing sample, questionnaire	8	ChatGPT3.5
Bai and Wei (2024)	China	English	HE	34	Not specified	Mixed	Writing sample, questionnaire	2	ChatGPT3.5
Banihashem et al. (2024)	Netherlands	English	HE	74	Not specified	Quantitative	Writing sample, student work	2	ChatGPT
Behforouz and Ghaithi (2024)	Oman	English	HE	60	Intermediate	Quantitative	Test item	5	ChatGPT
Boudouaia et al. (2024)	Algeria	English	HE	76	Intermediate	Quantitative	Writing sample, questionnaire	10	ChatGPT4
Darmawansah et al. (2025)	Indonesia	English	HE	67	Mixed	Quantitative	Speaking recording, questionnaire, worksheet	6	ChatGPT
Du and Alm (2024)	New Zealand	English	HE	24	Beginner	Qualitative	Interview	Not specified	ChatGPT4
Foung et al. (2024)	China (HK)	English	HE	74	Intermediate	Qualitative	Reflection, interview	13	ChatGPT
Ghafouri (2024)	Iran	English	SE	30	Intermediate	Quantitative	Questionnaire	16	ChatGPT
Ghafouri et al. (2024)	Iran	English	HE	60	Intermediate	Quantitative	Writing sample, questionnaire	10	ChatGPT3.5
Gozali et al. (2024)	Indonesia	English	HE	18	Mixed	Qualitative	Writing sample, reflection, interview	16	ChatGPT
Hiniz (2024)	Türkiye	English	HE	27	Intermediate	Qualitative	Interview	8	ChatGPT, Bing Chat
Heygen, Synthesia J. Huang and Mizumoto (2025)	Japan	English	HE	80	Not specified	Quantitative	Questionnaire	5	ChatGPT
Ismail and Alharkan (2024)	Saudi Arabia	English	HE	70	Intermediate	Quantitative	Questionnaire	Not specified	ChatGPT
Karataş et al. (2024)	Türkiye	English	HE	13	Mixed	Qualitative	Interview	4	ChatGPT3.5
Kohnke (2024)	China (HK)	English	HE	14	Not specified	Qualitative	Interview	13	ChatGPT
Y. J. Lee & Davis (2024)	S. Korea	English	HE	89	Not specified	Quantitative	Questionnaire	16	ChatGPT
J. H. Lee et al. (2023)	S. Korea	English	EE	121	Not specified	Mixed	Questionnaire, interview	4	CopyAI
J. H. Lee et al. (2024)	S. Korea	English	HE	22	Not specified	Mixed	Questionnaire, dialogue log	Not specified	ChatGPT
X. Li et al. (2023)	USA	Chinese	SE	4	Mixed	Mixed	Writing sample, reflection	3	ChatGPT3.5
Liu et al. (2024)	New Zealand	English	HE	8	Mixed	Qualitative	Multimodal text, interview	2	ChatGPT3.5, Bing AI
Long (2024)	China (HK)	English	SE	4	Intermediate	Qualitative	Interview	5	ChatGPT
Mahapatra (2024)	India	English	HE	72	Not specified	Mixed	Writing sample, group discussion	16	ChatGPT
Meyer et al. (2024)	Germany	English	SE	459	Not specified	Quantitative	Writing sample, questionnaire	One day	ChatGPT3.5
Pan et al. (2025)	China	English	HE	61	Not specified	Mixed	Questionnaire, study log, interview	12	ChatGPT (ReadMate)

Polakova and Ivenz (2024)	Czech	English	HE	110	Not specified	Mixed	Writing sample, Questionnaire, interview	10	ChatGPT
Punar Özçelik and Ekşi (2024)	Türkiye	English	HE	11	Intermediate	Qualitative	Observation, interview	2	ChatGPT3.5
Robillos (2024)	Thailand	English	HE	30	Not specified	Mixed	Writing sample, interview	Not specified	ChatGPT
Shaikh et al. (2023)	Norway	English	HE	10	Mixed	Quantitative	Questionnaire	One day	ChatGPT3.5
Tai and Chen (2024)	China (Taiwan)	English	EE	85	Beginner	Mixed	Test item, interview	3	ChatGPT (CoolE)
Teng (2024a)	China (Macau)	English	HE	45	Mixed	Mixed	Questionnaire, interview	15	ChatGPT4
Teng (2025)	China (Macau)	English	HE	40	Mixed	Mixed	Questionnaire, interview	15	ChatGPT
Tsai et al. (2024)	China (Taiwan)	English	HE	52	Intermediate	Quantitative	Writing sample, questionnaire	Not specified	ChatGPT3.5
Tseng and Lin (2024)	China (Taiwan)	English	HE	15	Not specified	Qualitative	Writing sample, reflection	16	ChatGPT3.5
Van Horn (2024)	S. Korea	English	HE	120	Intermediate	Mixed	Questionnaire, interview, observation	15	ChatGPT3.5
C. Wang (2024)	USA	English	HE	6	Not specified	Qualitative	Writing sample, reflection, interview	14	ChatGPT3.5
Y. Wang (2024)	Japan	English	HE	61	Intermediate	Mixed	Questionnaire, writing sample, reflection	14	ChatGPT, DeepL, Bing AI
C. Wang et al. (2024)	China	English	HE	99	Intermediate	Mixed	Questionnaire, interview	6	Typebot, D-ID, EAP Talk
Wiboolyasarin et al. (2024)	Thailand	Thai	HE	39	Not specified	Quantitative	Writing sample	10	ChatGPT
Xiao and Zhi (2023)	China	English	HE	5	Not specified	Qualitative	Interview	Not specified	ChatGPT3.5
Yamaoka (2024)	Japan	English	HE	33	Intermediate	Qualitative	Interview	15	ChatGPT
D. Yan (2023)	China	English	HE	35	Not specified	Qualitative	Writing sample, interview	1	ChatGPT3.5
Zhang et al. (2025)	China (HK)	English	HE	35	Not specified	Qualitative	Questionnaire, writing sample, interview, eye-tracking	3	ChatGPT4.0 (LogicalHamster)
Zheng (2024)	China	English	HE	40	Not specified	Mixed	Questionnaire, interview	Not specified	Reading Bot
Zou and Huang (2023)	China	English	HE	125	Intermediate	Qualitative	Reflection, interview	1	ChatGPT

Note: HE, higher education; SE, secondary education; EE, elementary education.



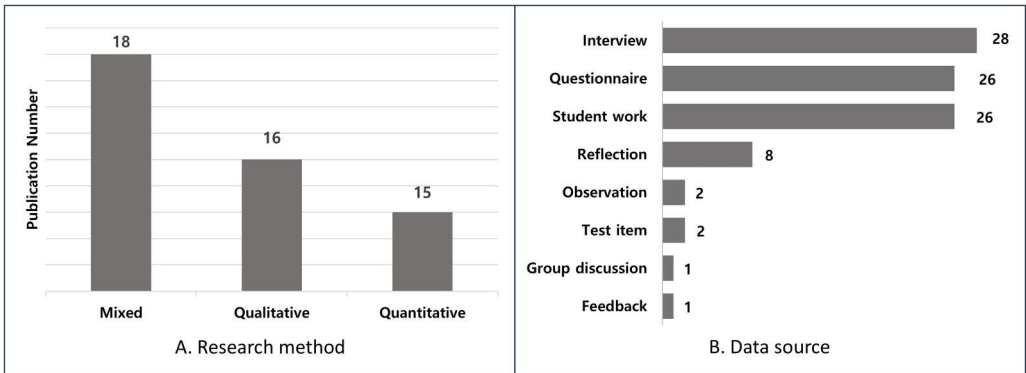
**Figure 2.** Target language and education level.

differences in their infrastructure, pedagogical approaches, and attitudes toward GenAI usage (Chiu et al., 2023; Teng, 2024b). For example, insights from well-resourced Chinese universities may not be readily transferred to under-resourced regions like Africa (Zawacki-Richter et al., 2019).

#### 4.1.2. Target language and education level

As Figure 2(A) shows, most research ( $n = 46$ ) addressed English as the target language, followed by German ( $n = 1$ ), Thai ( $n = 1$ ), and Chinese ( $n = 1$ ). Regarding the education level (Figure 2(B)), most of the studies were conducted in higher education contexts ( $n = 43$ ). Four studies explored GenAI usage in secondary education, with the other studies focused on the elementary education context ( $n = 2$ ).

The focus on English as the primary target language may limit our understanding of GenAI's potential across diverse language contexts (Teng, 2024b). Similarly, the concentration on higher education leaves an unknown area regarding GenAI application across K-12 (Liang et al., 2023). In fact, higher education students' advanced digital literacy and institutional resources may inflate GenAI's perceived effectiveness across all education levels. As a result, pedagogical concerns relevant to elementary and secondary education – such as developmentally appropriate content, scaffolding strategies, and play-based learning – are likely overlooked (J. H. Lee et al., 2023).



**Figure 3.** Research method and data source.

#### 4.1.3. Sample size and language proficiency

Table 2 shows that a large sample size ( $n = 24$ ) was the most common one, followed by medium ( $n = 11$ ), small ( $n = 7$ ), and extra-large ( $n = 7$ ). Regarding language proficiency, 18 studies involved intermediate-level students, eight studies included participants with different language levels, two studies with beginners, and the remaining 21 studies did not specify this information. Using large sample sizes in many studies strengthens the statistical power of their findings, while the inclusion of intermediate-level students provides insights into a core stage of language development (W. Huang et al., 2022). However, the studies show methodological limitations in sampling transparency, given that 21 studies did not report language proficiency levels. This potentially undermines the generalizability of findings and limits implementing appropriate GenAI strategies for different learner groups, as beginners and advanced learners may require distinct approaches (Lai & Lee, 2024).

#### 4.1.4. Research method and data source

Most studies adopted a mixed-methods approach ( $n = 18$ ), followed by a qualitative approach ( $n = 16$ ) and a quantitative research method ( $n = 15$ ) (Figure 3(A)). The exploration of data sources (Figure 3(B)) demonstrates that most of the studies used interviews ( $n = 28$ ), followed by students' work ( $n = 26$ ), including writing samples ( $n = 20$ ), study log ( $n = 2$ ), and multimodal texts ( $n = 1$ ), speaking recording ( $n = 1$ ), eye-tracking data ( $n = 1$ ), dialogue with GenAI ( $n = 1$ ), and questionnaires ( $n = 26$ ). Students' reflective journals were used in eight articles (e.g. X. Li et al., 2023), observations in two papers (e.g. Tseng & Lin, 2024), test items in two papers (e.g. Tai & Chen, 2024), group discussion in one study (e.g. Mahapatra, 2024), and ChatGPT's feedback in the other study.

A balanced use of qualitative, quantitative, and mixed-methods approaches reflects methodological diversity, enhancing our understanding of GenAI's impact on language learning outcomes (B. Li et al., 2024). However, the heavy reliance on self-reported data from interviews and questionnaires may raise concerns about subjective bias – especially due to the novelty effect of GenAI usage (Chiu et al., 2023; W. Huang et al., 2022). In addition, current research often lacks evidence-based methodologies that can offer insights into actual language development (Jeon & Lee, 2023). The limited use of observational studies also restricts knowledge of authentic classroom practices with GenAI tools (Lee & Jeon, 2024a). This gap likely undermines the credibility of reported language gains (Chiu et al., 2023; Ji et al., 2023).

#### 4.1.5. Research duration and specific tool

Most of the studies were conducted over a short-term period ( $n = 22$ , one to 10 weeks) and an intermediate-term period ( $n = 17$ , 11–16 weeks) (Figure 4). Two studies were conducted using a one-shot

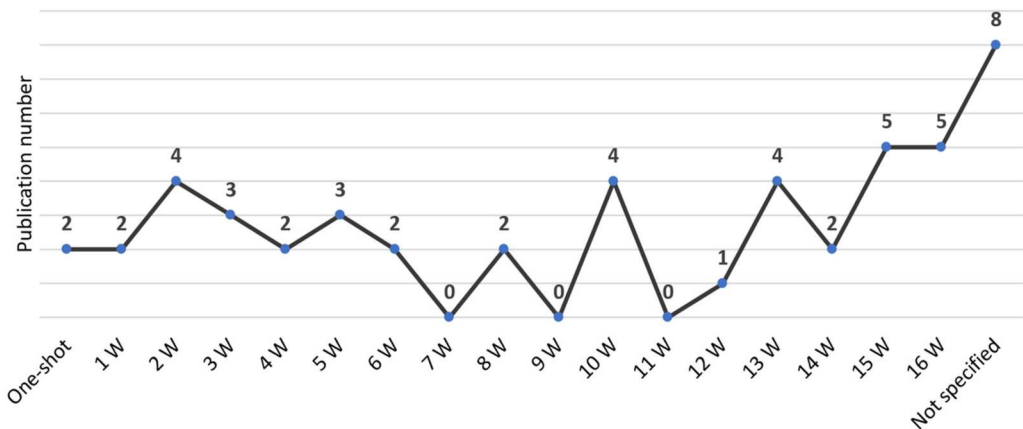


Figure 4. Study duration.

design, and eight studies did not specify the research duration, while no study was done over a long-term period (longer than 16 weeks). The dominance of short- and intermediate-term studies may pose methodological constraints. Given no studies conducted beyond 16 weeks, little is known about GenAI's long-term effects on language development or changes in usage as its novelty fades (Ji et al., 2023). This concern is heightened by the fact that 13 studies lasted only 4 weeks or less, including two one-day interventions and eight with unspecified durations. Such brief time-frames are likely to capture only initial effects rather than the sustained integration of GenAI into language learning (W. Huang et al., 2022).

Table 2 shows that most of the studies ( $n = 46$ ) adopted ChatGPT as a GenAI tool, including some ChatGPT-based customized bots such as CopyAI in J. H. Lee et al. (2023). Three studies utilized Bing AI, such as Liu et al. (2024), which adopted Bing Chat and Bing Image Creator in addition to ChatGPT to engage Chinese undergraduate students in producing multimodal texts for a multimodal Power-Point project. Other GenAI tools include DeepL ( $n = 2$ ), Heygen ( $n = 1$ ), Synthesia ( $n = 1$ ), Typebot ( $n = 1$ ), D = ID ( $n = 1$ ), Reading Bot ( $n = 1$ ), and EAP Talk ( $n = 1$ ). The results indicate that ChatGPT dominates the research landscape, yet many studies did not specify the model version. Given the significant performance differences between versions (e.g. GPT-3.5 vs. GPT-4o), this may hinder replicability and complicate cross-study comparisons (Jeon et al., 2024). Likewise, the limited exploration of other GenAI tools (e.g. Bing AI, DeepL, Claude) can leave the broader GenAI usage in language learning unknown (B. Li et al., 2024).

**Table 3.** Research focus of the selected articles.

Category	Subcategory	Item	Frequency
Perception	• Language skill	Writing	25
		Reading	4
		Speaking	3
		Grammar	3
		Vocabulary	2
		Listening	1
		<i>Total</i>	38
	• Affective aspect	Attitude	21
		Self-efficacy	9
		Motivation	7
		Anxiety	6
		Acceptance	3
		Critical thinking	3
		Engagement	3
		Autonomy	3
		Grit	1
		Willingness to communicate	1
		<i>Total</i>	57
	• Knowledge	Digital literacy	3
		Feedback literacy	1
		Register knowledge	1
		Research competence	1
		<i>Total</i>	6
		<i>Perception total</i>	101
Acquisition	• Writing		15
		• Speaking	1
		• Reading	1
		• Grammar	1
		<i>Acquisition total</i>	18
Behavior	• Classroom interaction	Students' collaboration	5
		Teacher-student rapport	1
	• Interaction with GenAI	Learning strategy	2
		Prompting strategy	1
		<i>Behavior total</i>	9
		<i>Total of all focuses</i>	128

## 4.2. RQ2: research focus

Most studies were centered on topics related to learner perception ( $n = 101$ ), whereas acquisition ( $n = 18$ ) and behavior ( $n = 9$ ) gained significantly less attention (Table 3). Notably, many studies ( $n = 18$ ) addressed multiple topics regarding GenAI in the language classroom. For example, Tsai et al. (2024) explored 44 Taiwanese undergraduate students' writing skill enhancement and their attitudes after using ChatGPT for revision.

### 4.2.1. Perception

Learner perception ( $n = 101$ ) was the most common focus of research in the selected papers. Our analysis identified three areas of research: language skill ( $n = 38$ ), affective aspect ( $n = 57$ ), and knowledge ( $n = 6$ ). Among the studies examining language skills, writing ( $n = 25$ ) was the most frequently addressed topic, followed by reading ( $n = 4$ ), speaking ( $n = 3$ ), grammar ( $n = 3$ ), vocabulary ( $n = 2$ ), and listening ( $n = 1$ ). Some studies addressed multiple language skills: for example, Karataş et al. (2024) examined three skills, including writing, grammar, and vocabulary, regarding ChatGPT's impact on college students' English learning. 57 cases were found to examine the affective aspects that involve attitudes ( $n = 21$ ), self-efficacy ( $n = 9$ ), motivation ( $n = 7$ ), anxiety ( $n = 6$ ), critical thinking ( $n = 3$ ), engagement ( $n = 3$ ), acceptance ( $n = 3$ ), autonomy ( $n = 3$ ), grit ( $n = 1$ ), and willingness to communicate ( $n = 1$ ). J. H. Lee et al. (2023) explored 121 primary students' enjoyment, engagement, and motivation after GenAI-assisted reading activities. Finally, knowledge gain ( $n = 6$ ) was identified in four areas: digital literacy ( $n = 3$ ), feedback literacy ( $n = 1$ ), register knowledge ( $n = 1$ ), and research competence ( $n = 1$ ). Gozali et al. (2024) reported 18 Indonesian college students' feedback literacy development after EFL writing courses with ChatGPT.

The strong emphasis on perception-based studies represents a mismatch between perceived and actual learning outcomes, with limited consideration of contextual factors. Banihashem et al. (2024) highlight that learners' self-perceptions often do not align with measurable performance, while Yang and Li (2024) stress the need to move beyond perceptual data to assess actual proficiency gains. In addition, the lack of contextual sensitivity in current research limits the generalizability of findings across diverse educational settings, institutional situations, and cultural contexts (Crompton et al., 2024).

### 4.2.2. Acquisition

Most reviewed studies examined writing ( $n = 15$ ) as the primary language skill for acquisition. The positive effects of GenAI usage on enhancing writing skills were examined on diverse topics, involving feedback provision (Banihashem et al., 2024; Mahapatra, 2024; Meyer et al., 2024; Tsai et al., 2024; Wiboolyasarin et al., 2024), revision and uptake (Boudouaia et al., 2024; Tseng & Lin, 2024), writing tutoring (Ghafouri et al., 2024; X. Li et al., 2023), and multimodal composition (Liu et al., 2024). In addition, speaking ( $n = 1$ ), reading ( $n = 1$ ), and grammar ( $n = 1$ ) were addressed as topics related to language acquisition (Table 3).

The findings reveal several limitations in research on GenAI-assisted language acquisition. First, a bias is identified toward writing skills, leaving a significant gap in understanding GenAI's effectiveness on speaking, listening, and reading skills. This focus may stem from the text-based nature of LLMs and methodological convenience (B. Li et al., 2024). Second, although seven of the 18 studies employed quasi-experimental pre- and post-test designs, many of them lack the specificity needed to identify which aspects of language development were measured (Boudouaia et al., 2024). Moreover, the absence of proper control conditions or baseline comparisons undermines the credibility of their claims, making it difficult to attribute the outcomes specifically to GenAI rather than other factors like increased practice or instructor input (Meyer et al., 2024). Finally, despite the significance of learner characteristics in L2 or FL acquisition, current studies have largely ignored how individual differences mediate GenAI's impact on language development.

4.2.3. Behavior

Learner behaviors ( $n=9$ ) were examined in two categories: Classroom interaction ( $n=6$ ), which encompassed students' collaboration ( $n=5$ ) and teacher-student rapport ( $n=1$ ), and interaction with GenAI ( $n=3$ ), which included learning strategy ( $n=2$ ) and prompting strategy ( $n=1$ ). For example, Abdelhalim (2024) and Van Horn (2024) reported on ChatGPT's positive impact on enhancing student collaboration, whereas Liu et al. (2024) observed that ChatGPT facilitated students' interaction with audience. Ghafouri (2024) found that ChatGPT assisted in designing a teacher-student rapport system in which teachers could provide L2 students with "self-expression promotion, empathy, group learning, and assessment" (p. 5). Regarding learner interaction with GenAI, Pan et al. (2025) investigated EFL students' self-regulated reading strategy use with ChatGPT, focusing on metacognitive, cognitive, motivational, and behavioral strategies. Zhang et al. (2025) examined behavioral development by measuring students' engagement behaviors, including total learning time and the number of prompts produced.

This analysis reveals that learner behavior remains underexplored, despite its importance in GenAI-integrated language classrooms (Jeon et al., 2024). The gaps include concerns about the ecological validity of GenAI use, as overreliance on controlled, task-based experimental designs may disconnect research from real-world classroom practice (Van Horn, 2024; Zawacki-Richter et al., 2019). Little is known about how learners engage with GenAI tools over time, including the development of prompt engineering skills and response evaluation strategies (Zhang et al., 2025). Although some evidence suggests that learners' strategic GenAI use is influenced by metacognitive awareness, empirical research on how these strategies evolve remains limited (Du & Alm, 2024). In particular, the cognitive processes involved in navigating and coordinating multiple GenAI tools alongside traditional learning resources have been largely overlooked (Jeon & Lee, 2023). In fact, research shows that effective learning outcomes depend on the strategic integration of diverse resources rather than reliance on isolated tools (Foung et al., 2024).

4.3. RQ3: roles and challenges of GenAI

4.3.1. Roles of GenAI

Table 4 summarizes nine roles of GenAI in language learning: feedback provider ( $n=42$ ), language learning tutor ( $n=26$ ), resource provider ( $n=22$ ), content generator ( $n=20$ ), evaluator ( $n=14$ ), affective supporter ( $n=12$ ), conversation partner ( $n=12$ ), interaction facilitator ( $n=10$ ), and cognitive stimulator ( $n=8$ ). In most of the studies ( $n=48$ ), notably, multiple roles of GenAI tools were reported. For example, three roles, including feedback provider, content generator, and language learning tutor, were identified in C. Wang (2024).

Feedback provider ( $n=42$ ) was the most common role. In all studies, feedback for language learners was linked to their writing outcomes. For example, Banihashem et al. (2024) compared ChatGPT-assisted feedback with peer feedback on argumentative essays, finding that ChatGPT provided general, descriptive feedback, while peer feedback identified specific issues for improvement. In

Table 4. Roles of GenAI.

Category	Subcategory	Frequency
GenAI role	Feedback provider	42
	Language learning tutor	26
	Resource provider	22
	Content generator	20
	Evaluator	14
	Affective supporter	12
	Conversation partner	12
	Interaction facilitator	10
	Cognitive stimulator	8
Role total		166

contrast, Ghafouri (2024, p. 8) found that, with well-constructed prompts, ChatGPT can deliver “action-based, explicit, and personalized feedback” on written work.

Language learning tutor ( $n = 26$ ) was the second most common role with GenAI tools aiding in managing a holistic learning process. GenAI was used as writing assistants ( $n = 19$ ) to guide learners through drafting and revision (e.g. Boudouaia et al., 2024; Punar Özçelik & Ekşi, 2024) and supported general language learning ( $n = 7$ ) by offering pedagogical benefits such as teaching support (Ghafouri et al., 2024; Gozali et al., 2024), formal language skills development (Shaikh et al., 2023), and self-directed learning (Xiao & Zhi, 2023).

Resource provider ( $n = 22$ ) and content generator ( $n = 20$ ) were the third and fourth most common roles. As a content generator, GenAI helped create learning materials or tasks, such as when Karataş et al. (2024) encouraged EFL students to use ChatGPT in creating short stories, and Ghafouri et al. (2024) highlighted its role in lesson design. As a resource provider, GenAI assisted language learners in enriching writing topics (Mahapatra, 2024), seeking additional information (Punar Özçelik & Ekşi, 2024), learning technical terms (Abdelhalim, 2024), and improving grammar or vocabulary (C. Wang 2024). As a cognitive stimulator ( $n = 8$ ), GenAI supported brainstorming (Karataş et al., 2024; Zou & Huang, 2023), developing outlines (Ghafouri et al., 2024), and generating new research ideas (Abdelhalim, 2024). Zou and Huang highlighted its capacity to inspire multi-perspective thinking, while Abdelhalim noted its role in helping graduate students refine their research arguments.

The roles of Evaluator ( $n = 14$ ) and conversation partner ( $n = 12$ ) were identified in some of the studies reviewed. GenAI engages in conversations with learners to provide additional information (Abdelhalim, 2024; Mahapatra, 2024), offer consultation on language tasks (Ghafouri et al., 2024), and enhance writing and speaking skills (Shaikh et al., 2023; Van Horn, 2024). GenAI was also used as an evaluator across various contexts, including high-stake exams (Ghafouri, 2024), self-evaluation and assessment literacy (Gozali et al., 2024; Shaikh et al., 2023), and assessment for learning (Tseng & Lin, 2024; Wiboolyasarin et al., 2024).

The roles of affective supporter ( $n = 12$ ), interaction facilitator ( $n = 10$ ), and cognitive stimulator ( $n = 8$ ) were also identified. As an interaction facilitator, GenAI facilitated collaborative learning (Abdelhalim, 2024; Mahapatra, 2024; Van Horn, 2024; Zou & Huang, 2023), rapport-building (Ghafouri, 2024), and peer interaction during presentations (Liu et al., 2024; Zou & Huang, 2023). It also provided support in affective aspects such as enjoyment (Abdelhalim, 2024; J. H. Lee et al., 2023), self-efficacy (X. Li et al., 2023), motivation (Abdelhalim, 2024; Ghafouri, 2024), L2 grit (Ghafouri, 2024), and autonomy (Mahapatra, 2024). As a cognitive stimulator, GenAI supported brainstorming (Zou & Huang, 2023), developing outlines (Ghafouri et al., 2024), and generating new research ideas (Abdelhalim, 2024).

**Table 5.** Challenges of GenAI.

Subcategory	Subcategory	Frequency
Technological aspect	Content quality	25
	Querying challenge	8
	Information bias	8
	Technical issue	5
	Privacy and security	5
	Information overload	2
	<i>Technology total</i>	53
Educational aspect	Overreliance	22
	Pedagogical limitation	21
	Academic integrity	15
	Learning inequality	4
	Assessment disruption	2
	<i>Education total</i>	64
	<i>Challenge total</i>	117

While studies introduced diverse GenAI roles, significant research gaps persist. First, the longitudinal evaluation of these roles remains limited, neglecting their evolution over learner progress (Jeon et al., 2024; Lee & Jeon, 2024a). Furthermore, the roles are often examined in isolation, overlooking their potential interactions, such as tensions between GenAI's feedback and assessment functions (Jeon & Lee, 2023, 2024; B. Li et al., 2024). Second, research lacks investigation into optimal customization of GenAI roles for diverse learner characteristics and objectives (Ji et al., 2023; Teng, 2024). Finally, there is insufficient understanding of how GenAI roles integrate with human stakeholders in the language educational ecosystem, leaving the optimal distribution of responsibilities between human and GenAI unexplored despite its importance for effective GenAI implementation (Gozali et al., 2024).

#### 4.3.2. Challenges from using GenAI

Our review identified 11 themes of GenAI challenges in language learning, categorized into two main areas: technology and education (Table 5). Nineteen studies reported at least one challenge associated with GenAI in language learning, while four studies focused solely on its benefits. In the technology domain, the most prevalent concern was content quality ( $n = 25$ ), followed by querying challenge ( $n = 8$ ), information bias ( $n = 8$ ), technical issues ( $n = 5$ ), privacy and security ( $n = 5$ ), and information overload ( $n = 2$ ).

Regarding content quality, diverse concerns were highlighted, including low information accuracy and authenticity (Abdelhalim, 2024; Fount et al., 2024; Gozali et al., 2024; Liu et al., 2024), non-specific content (J. H. Lee et al., 2023; Meyer et al., 2024; Punar Özçelik & Ekşi, 2024; Tseng & Lin, 2024; Van Horn, 2024), and lack of originality due to information duplication (Abdelhalim, 2024; Xiao & Zhi, 2023). Querying challenges (e.g. J. H. Lee et al., 2023; Liu et al., 2024; Punar Özçelik & Ekşi, 2024; Xiao & Zhi, 2023) were also reported: for instance, Liu et al. (2024) noted the difficulty students faced in crafting effective prompts to generate realistic AI images for multimodal composition. Technical issues, the third most common challenge, included occasional disconnection (Karataş et al., 2024), login failures (Punar Özçelik & Ekşi, 2024), and error messages (Van Horn, 2024). Information bias was another concern: GenAI's preference for standard language (Punar Özçelik & Ekşi, 2024) and biased feedback resulting from its training data (Fount et al., 2024; Wiboolyasarín et al., 2024). Privacy and security issues were raised on the risk of personal data leakage (Ghafouri et al., 2024; Zou & Huang, 2023). Finally, research indicated that GenAI tools provided students with potentially excessive information relative to their proficiency levels (Zheng, 2024).

Concerning educational challenges, overreliance ( $n = 22$ ) emerged as the most significant issue in relation to ethical issues, particularly when students focused on GenAI outcomes rather than the learning process (e.g. Abdelhalim, 2024; Mahapatra, 2024; Wiboolyasarín et al., 2024). Pedagogical limitations ( $n = 21$ ) included reduced motivation and engagement (Karataş et al., 2024; Mahapatra, 2024; Wiboolyasarín et al., 2024), fewer hands-on practice opportunities (Karataş et al., 2024; D. Yan, 2023), diminished creativity (Mahapatra, 2024; Zou & Huang, 2023), and limited proficiency development (Van Horn, 2024; Zou & Huang, 2023). Academic integrity ( $n = 15$ ) also emerged as a significant issue, which involves plagiarism and cheating (Abdelhalim, 2024; Boudouaia et al., 2024; Gozali et al., 2024; Xiao & Zhi, 2023; C. Wang 2024; D. Yan, 2023; Zou & Huang, 2023), misuse (Tsai et al., 2024), and concerns about ownership (Tseng & Lin, 2024). Learning inequality ( $n = 4$ ) was reported when students lacked equal access to GenAI (Boudouaia et al., 2024; Zou & Huang, 2023). Finally, assessment disruption was noted in two studies in relation to challenging traditional assessment methods (Tsai et al., 2024) and creating unfair advantages for users (D. Yan, 2023).

This review on technological and educational barriers presents some limitations. First, a disconnect exists between identified GenAI-related challenges and the evaluation of potential solutions. Although many studies raised concerns about GenAI in language learning, few proposed or assessed evidence-based strategies, thereby limiting their practical value for educators who seek actionable guidance (Ji et al., 2023). Second, current research overlooked how

challenges differ across educational settings, learner proficiency levels, and access to technology (Crompton et al., 2024). This lack of contextual factors may restrict the generalizability of their findings (Chiu et al., 2023). Third, the dominant focus on technological issues often adopts a technocentric perspective, separating technical problems from broader educational contexts such as learner agency and institutional dynamics (Jeon & Lee, 2023; Lee & Jeon, 2024a). Yet, as research shows, technological challenges are embedded within social, institutional, and pedagogical systems (Bower et al., 2024; Chiu, 2023). Finally, minimal attention has been paid to learning inequality – an important oversight given GenAI’s potential to widen educational disparities among learners (Zou & Huang, 2023).

## 5. Discussion: Gaps and future research

This review of 49 studies on language learning systematically examined GenAI usage, focusing exclusively on students’ experiences in classroom settings, unlike previous comprehensive reviews that include non-empirical studies, non-classroom contexts, and teachers’ perspectives. This section discusses research gaps and offers suggestions for future research.

### 5.1. The need for diverse research designs

GenAI tools in language learning require research in diverse contexts to validate their pedagogical benefits in actual classrooms. Most studies have centered on English, likely due to earlier ChatGPT models’ stronger English proficiency (Kohnke et al., 2023). This linguistic bias restricts our understanding of GenAI’s effectiveness across multilingual contexts and underrepresented languages (Dai et al., 2025; Lee et al., 2025b). The English-centric approach limits knowledge about how GenAI addresses languages with different orthographic and grammatical systems and cultural representations, potentially neglecting important lingual-cultural aspects of language learning (Crompton et al., 2024). Although newer GenAI tools, such as ChatGPT-4o, demonstrate improved multilingual capabilities (OpenAI, 2024a), future research should address these limitations by explicitly exploring GenAI usage for underrepresented languages with fewer digital resources (Jeon et al. 2025b).

Most studies focused on higher education, possibly due to easier data collection and institutional support (Abdelhalim, 2024; Liu et al., 2024). This limits our understanding of GenAI’s effectiveness during early language acquisition and in less structured environments typical for younger learners (Jeon et al., 2024). Primary and secondary students need different pedagogical strategies and developmental considerations than those in higher education (J. H. Lee et al., 2023). Differences in digital literacy, learner agency, and institutional infrastructure between K-12 and higher education further complicate applying findings from higher education across contexts (Lee & Jeon, 2024a). Thus, future studies should address GenAI’s impacts in primary and secondary education for comprehensive understanding across all educational levels (Liang et al., 2023).

This review highlights the need for more diverse research approaches. While classroom-based GenAI studies used quantitative, qualitative, and mixed-methods approaches in a balanced manner, data collection mainly relied on student work, interviews, and questionnaires. Furthermore, qualitative studies typically used self-reported data, such as interviews, reflections, and essays, to explore writing revisions and students’ perceptions. Therefore, future research should expand the data sources to capture students’ dynamic interactions with GenAI across language learning modes, including observational field notes of real-time classroom behaviors, video recordings for detailed analysis of learner-technology interactions, group discussion transcripts highlighting collaborative processes, reflective journals providing insights into learner autonomy, and multimodal artifacts like digital portfolios from in-class activities (Du & Alm, 2024; Ji et al., 2023; Van Horn, 2024). Scholars should also conduct longitudinal studies beyond one semester to investigate long-term effects of GenAI tools on language acquisition (Jeon & Lee, 2023).

Finally, we noted the need to adopt diverse GenAI tools to fully harness their capabilities to enhance student learning (Chiu, 2023). All reviewed studies mainly used ChatGPT, with only a few additionally utilizing other tools (e.g. Alm, 2024; Hınız, 2024). This ChatGPT-centric approach limits research findings by restricting our knowledge of multimodal learning capabilities and overlooking unique affordances of other GenAI tools, such as image creation and specialized interaction methods (Jeon et al., 2024; Kohnke et al., 2023). While diverse tools like Gemini, Claude, and Midjourney have been discussed in conceptual research (e.g. Kohnke et al., 2023), their practical classroom application with students remains underexplored. Future research should expand the range of GenAI tools investigated, exploring their specific affordances to maximize educational outcomes and multimodal learning opportunities (Crompton et al., 2024; Jeon & Lee, 2023). Furthermore, integrating these tools across various devices – mobile phones, smart speakers, educational robots, and tablets – can create more authentic contexts for language acquisition (Jeon et al., 2023, 2024). Similarly, incorporating GenAI into virtual reality could enrich learning by enhancing engagement, self-efficacy, and communication skills (Bower et al., 2024; Lee et al., 2025a).

### **5.2. The need for more diversified research topics**

Regarding research topics, learner perceptions of GenAI usage were the primary focus, whereas only some examined language acquisition. This imbalance suggests a need for more quasi-experimental studies that include diverse groups and employ pre-tests, post-tests, and delayed post-tests to directly measure language skill development from using GenAI and to assess its positive or negative impacts on specific aspects of language learning (Jeon & Lee, 2023; Liang et al., 2023). In addition, learners' classroom behaviors regarding GenAI use received minimal attention. To address this gap, future research could adopt in-depth qualitative methods, such as multimodal conversation analysis using video recordings, to closely monitor student interactions with GenAI.

This review emphasizes the need to diversify research across all language skills. Most studies examined GenAI's benefits for writing performance, while only a few exclusively targeted other skills. This writing emphasis likely arises from GenAI's text-based interactions, as major platforms primarily process textual data (Kohnke et al., 2023; D. Yan, 2023). However, this focus potentially limits our understanding of GenAI's effectiveness for other critical skills. Specifically, text-based models may struggle with oral skills such as natural speech flow, precise pronunciation feedback, and conversational pragmatics (Crompton et al., 2024; Lee et al., 2025b; B. Li et al., 2024). Recent GenAI tools (e.g. ChatGPT-4) offer advanced multimodal features, handling text, audio, and images, and generating both text and speech (OpenAI, 2024a). Leveraging these multimodal technologies could broaden GenAI's applicability beyond writing, supporting multiple language skills (Jeon & Lee, 2023; Van Horn, 2024), and enhancing students' communicative competence in authentic learning environments (Karataş et al., 2024).

### **5.3. Applying multifaceted roles of GenAI with instructional design**

A wide range of GenAI roles in language learning, as identified in this review, highlights its high versatility and potential as a comprehensive language learning tool (Kohnke et al., 2023; Yang & Li, 2024). For example, ChatGPT, the most salient GenAI tool in the studies, engages in human-like conversations based on its extensive pre-trained knowledge to interpret inputs, reason contextually, and generate appropriate responses (Chiu, 2023). This capability allows it to perform multiple functions in various domains, including language education, unlike earlier chatbot models whose capacity was relatively limited on prescribed scenarios and restricted text data (W. Huang et al., 2022; Ji et al., 2023).

However, the results suggest that the potential of GenAI has not yet been fully realized in all areas of language learning. One likely reason is that the reviewed studies focused on written communication, such as providing feedback on essays, generating writing tasks and reading materials, and

tutoring revision processes. Its role as a conversation partner was primarily explored in learners' text-based communication with GenAI, rather than in enhancing their language skills. That is, in most studies, the focus of its conversational role was on obtaining information or feedback, rather than on developing or assessing students' language performance. Therefore, with the recent development of multimodal GenAI (OpenAI, 2024a), future studies should explore how students' interactions with GenAI (e.g. speech- and image-recognition ChatGPT) can positively impact their language development. To validate the adaptability of GenAI for actual language acquisition, we recommend employing GenAI-enhanced instructional designs that offer comprehensive learning experiences (Bower et al., 2024).

#### ***5.4. Redefining teaching and assessment paradigms in the GenAI age***

The current review identified technological and educational challenges regarding GenAI use in language learning classrooms. Most technological issues – low content quality, information bias, and privacy concerns – may arise from GenAI's reliance on statistical probability and its training datasets (Dai et al., 2025). One potential solution is to use customized GenAI such as My GPTs, allowing for tailoring GPT results to specific educational purposes (Lee et al., 2025b; OpenAI, 2024a). Teachers can customize these platforms with their own data, enhancing GenAI's specificity in language learning tasks. However, practical barriers such as potential costs, the significant time investment, and the need for teacher- and learner-training required for data preparation and model fine-tuning may limit teachers' ability to utilize such tools efficiently in the classroom environments (Bower et al., 2024; Chiu, 2023). Also, personalization functions like "Manage Memory" enable users to control ChatGPT's conversation archiving, allowing learners to personalize outputs and address privacy concerns by anchoring interactions solely within their prompts (OpenAI, 2024b).

Regarding educational challenges, the most frequent issues – academic integrity, overreliance, and pedagogical limitations – relate directly to GenAI's impact on student learning. Academic integrity issues and overdependence on GenAI may hinder language development and restrict critical thinking skills (C. Wang 2024). To address these issues, future research should develop a critical literacy framework that integrates digital, AI, and GenAI literacies (Chiu, 2023). This framework may include competencies for critically evaluating GenAI-generated content in terms of accuracy, authenticity, and bias, recognizing ethical considerations and potential misuse of AI tools, understanding different GenAI models' limitations and affordances, and effectively combining human judgment with AI outputs in education (Chiu, 2023; Crompton et al., 2024; Jeon & Lee, 2023; Lee et al., 2025b). Systematic training informed by this framework could help learners thoughtfully evaluate and effectively leverage GenAI's pedagogical benefits (Vaccino-Salvadore, 2023). Similarly, exploring GenAI-student collaboration may provide insights into how students actively exercise agency when integrating GenAI into their learning processes, rather than passively adopting it for predetermined outcomes (Bower et al., 2024; Chiu, 2023).

Finally, learning inequity and assessment disruption require pedagogical and institutional solutions. Instead of viewing GenAI as a panacea for future education, a critical research approach is needed (Vaccino-Salvadore, 2023). On the one hand, teaching paradigms should shift from knowledge acquisition to "learning processes that included creativity, collaboration, and multimedia, and to achieve efficiencies" (Bower et al., 2024, p. 18). Transitioning to these innovative teaching paradigms may present practical barriers. First, GenAI literacy should be addressed in teacher training, as current teachers often lack adequate technological preparation to fully integrate GenAI's potential into classroom practices (Bower et al., 2024; Chiu, 2023). Institutional readiness also widely varies when many educational institutions lack necessary infrastructure and administrative support to sustainably embed AI technologies (Chiu, 2023). Moreover, unequal student access to cutting-edge technologies exacerbates existing disparities, and it may potentially widen educational gaps among socioeconomically diverse learner groups (Li et al., 2023). On the other hand,

assessment frameworks should also expand to evaluate students' critical thinking, GenAI literacy, and AI collaborative skills (Chiu, 2023; Jeon et al., 2024). Successful implementation of these teaching and assessment methods necessitates collaboration among students, educators, developers, and policymakers to establish GenAI integration guidelines (L. Yan et al., 2024).

## 6. Conclusion and implications

Although many studies have explored the potential benefits of GenAI in language education, its empirical evidence in classroom contexts has not been systematically reviewed. In this study, we reviewed 49 empirical studies, focusing on research trends, research focuses, and the roles and limitations of GenAI. Our findings highlight research gaps and suggest future directions, including the adoption of diverse research designs, the use of different GenAI tools, the expansion of research topics, and the redefinition of teaching and assessment paradigms in the GenAI era.

The review suggests six key research implications for using GenAI in language learning. First, with the current advancement of GenAI technology, scholars should explore the combined use of multiple GenAI tools across offline and virtual environments, including virtual reality, while utilizing diverse devices such as smart speakers, mobile phones, and educational robots. Second, given the multimodal capabilities of recent GenAI models like ChatGPT-4o, research should expand beyond writing skills to encompass multiple language skills (Jeon & Lee, 2023; Van Horn, 2024). Third, the multifaceted roles of GenAI in language learning should be further explored to understand its potential as a comprehensive learning tool, possibly through GenAI-assisted instructional design (Shaikh et al., 2023). Fourth, the effectiveness of customized GenAI platforms such as My GPTs should be investigated in supporting personalized and targeted language learning tasks (OpenAI, 2024a). Fifth, future research should focus on developing students' critical GenAI literacy to address ethical concerns in academic contexts (C. Wang 2024). Finally, as GenAI begins to be integrated into educational systems, researchers should explore shifts in teaching paradigms toward creativity, collaboration, and critical AI literacy, as well as the redefinition of assessment frameworks to evaluate these competencies necessary in the GenAI era (Bower et al., 2024; Chiu, 2023).

This review has limitations: first, its focus on L2/FL learning and learner perspectives may limit its generalizability to broader educational context. Thus, future research should include language teaching, especially teacher perspectives. Second, the review investigated GenAI use only in a formal language learning classroom. Follow-up studies may explore both formal and informal contexts to better understand GenAI application to language learning. Specifically, informal contexts might offer different learning contexts such as flexible and diverse interaction with GenAI, greater opportunities for self-directed learning, reduced performance anxiety, and enhanced learner autonomy, which are often structured in formal settings (Bower et al., 2024; Chiu, 2023). This review focused on "learning," not "assessment," suggesting research on GenAI's role in L2/FL assessment. GenAI offers new assessment opportunities, such as automated feedback, adaptive testing, and AI-based formative assessment (Bower et al., 2024; Teng, 2024b). Future studies should address academic integrity, assessment equity, and evaluation of higher-order thinking in GenAI contexts (Chiu, 2023).

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This work was supported by the Hankuk University of Foreign Studies Research Fund.

## Notes on contributors

**Seongyong Lee** is an Assistant Professor in School of Education and English at University of Nottingham Ningbo China. His research addresses AI in L2 education and Global Englishes. His articles have appeared in prestigious international journals, such as *Computer Assisted Language Learning*, *Computers & Education*, *Education & Information Technologies*, *Interactive Learning Environments*, *Applied Linguistics Review*, *ELT Journal*, *TESOL Quarterly*, and *System*, among others.

**Hohsung Choe** is Dean of the Office of Administrative Support and Professor of English Linguistics at Hankuk University of Foreign Studies (HUFS), South Korea. He teaches various courses in foreign and second language education from the undergraduate to the doctoral level. He holds a B.A. in English Education, an M.A. in English Linguistics from HUFS, and an M.S. and a Ph.D. in Language Education from Indiana University Bloomington. His research interests include language teacher and learner identity, world Englishes, multilingualism and multiculturalism, and computer-mediated language learning. His articles have appeared in *English Today*, *Applied Linguistics Review*, *Lingua*, *ELT Journal*, *System*, and *Computers and Education*, among others.

**Dr. Di Zou** is an Associate Professor at The Hong Kong Polytechnic University. Her research interests include AI in language education, technology-enhanced and game-based language learning. She has published more than 150 papers in international journals. Stanford University listed her as one of the World's Top 2% Most-cited Scientists for four consecutive years.

**Jaeho Jeon** is a Ph.D. Candidate in the Department of Curriculum and Instruction at Indiana University Bloomington. His research interests include computer-assisted language learning, dynamic assessment, and teacher education. His articles have been published in *Computer Assisted Language Learning*, *Computers & Education*, *Interactive Learning Environments*, *Education and Information Technologies*, *System*, *TESOL Quarterly*, and *ELT Journal*, among others.

## Ethics approval and consent to participate

Not applicable.

## Data availability statement

Data and materials collected in this study are available upon reasonable request.

## ORCID

Seongyong Lee  <http://orcid.org/0000-0002-9436-4272>

Hohsung Choe  <http://orcid.org/0000-0002-4355-595X>

Di Zou  <http://orcid.org/0000-0001-8435-9739>

Jaeho Jeon  <http://orcid.org/0000-0002-1161-3676>

## References

\*Indicates the studies included in the review.

\*Abdelhalim, S. M. (2024). Using ChatGPT to promote research competency: English as a Foreign Language undergraduates' perceptions and practices across varied metacognitive awareness levels. *Journal of Computer Assisted Learning*, 40(3), 1261–1275. <https://doi.org/10.1111/jcal.12948>

\*Alm, A. (2024). Exploring autonomy in the AI wilderness: Learner challenges and choices. *Education Sciences*, 14(12), 1369. <https://doi.org/10.3390/educsci14121369>

\*Almanea, M. (2024). Instructors' and learners' perspectives on using ChatGPT in English as a foreign language courses and its effect on academic integrity. *Computer Assisted Language Learning*, 1–26. <https://doi.org/10.1080/09588221.2024.2410158>

\*Alsager, H. N. (2024). To look from another window in education: Artificial intelligence assisted language learning and its reflections on academic demotivation, foreign language learning anxiety and autonomy. *Computer Assisted Language Learning Electronic Journal*, 25(4), 124–147.

\*Alsofyani, A. H., & Barzanji, A. M. (2025). The effects of ChatGPT-generated feedback on Saudi EFL learners' writing skills and perception at the tertiary level: A mixed-methods study. *Journal of Educational Computing Research*, 63(2), 431–463. <https://doi.org/10.1177/07356331241307297>.

\*Bai, L., & Wei, Y. (2024). Exploring EFL learners' integration and perceptions of ChatGPT's text revisions: A three-stage writing task study. *IEEE Transactions on Learning Technologies*, 17, 2161–2172. <https://doi.org/10.1109/TLT.2024.3491864>

- \*Banihashem, S. K., Kerman, N. T., Noroozi, O., Moon, J., & Drachsler, H. (2024). Feedback sources in essay writing: Peer-generated or AI-generated feedback? *International Journal of Educational Technology in Higher Education*, 21(1), Article 23. <https://doi.org/10.1186/s41239-024-00455-4>
- \*Behforouz, B., & Ghaithi, A. A. (2024). Grammar gains: Transforming EFL learning with ChatGPT. *Educational Process International Journal*, 13, 4. <https://doi.org/10.22521/edupij.2024.134.2>
- \*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. *Journal of Educational Computing Research*. Advance online publication. <https://doi.org/10.1177/07356331241247465>
- Bower, M., Torrington, J., Lai, J. W. M., Petocz, P., & Alfano, M. (2024). How should we change teaching and assessment in response to increasingly powerful generative Artificial Intelligence? Outcomes of the ChatGPT teacher survey. *Education and Information Technologies*. Advance online publication. <https://doi.org/10.1007/s10639-023-12405-0>
- Chiu, T. K. F. (2023). The impact of Generative AI (GenAI) on practices, policies and research direction in education: A case of ChatGPT and Midjourney. *Interactive Learning Environments*. Advance online publication. <https://doi.org/10.1080/10494820.2023.2253861>
- Chiu, T. K. F., Xia, Q., Zhou, X., Chai, C. S., & Cheng, M. (2023). Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 4, 100118. <https://doi.org/10.1016/j.caeai.2022.100118>
- Crompton, H., Edmett, A., Ichaporia, N., & Burke, D. (2024). AI and English language teaching: Affordances and challenges. *British Journal of Educational Technology*, 55(6), 2503–2529. <https://doi.org/10.1111/bjet.13460>
- Dai, D. W., Suzuki, S., & Chen, G. (2025). Generative AI for professional communication training in intercultural contexts: Where are we now and where are we heading? *Applied Linguistics Review*, 16(2), 763–774. <https://doi.org/10.1515/applirev-2024-0184>
- \*Darmawansah, D., Rachman, D., Febiyani, F., & Hwang, G.-J. (2025). ChatGPT-supported collaborative argumentation: Integrating collaboration script and argument mapping to enhance EFL students' argumentation skills. *Education and Information Technologies*, 30(3), 3803–3827. <https://doi.org/10.1007/s10639-024-12986-4>
- \*Du, J., & Alm, A. (2024). The impact of ChatGPT on English for academic purposes (EAP) students' language learning experience: A self-determination theory perspective. *Education Sciences*, 14, 726. <https://doi.org/10.3390/educsci14070726>
- Fereday, J., & Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International Journal of Qualitative Methods*, 5(1), 80–92. <https://doi.org/10.1177/160940690600500107>
- \*Foung, D., Lin, L., & Chen, J. (2024). Reinventing assessments with ChatGPT and other online tools: Opportunities for GenAI-empowered assessment practices. *Computers and Education: Artificial Intelligence*, 6, Article 100250. <https://doi.org/10.1016/j.caeai.2024.100250>
- \*Ghafouri, M. (2024). ChatGPT: The catalyst for teacher-student rapport and grit development in L2 class. *System*, 120, Article 103209. <https://doi.org/10.1016/j.system.2023.103209>
- \*Ghafouri, M., Hassaskhah, J., & Mahdavi-Zafarghandi, A. (2024). From virtual assistant to writing mentor: Exploring the impact of a ChatGPT-based writing instruction protocol on EFL teachers' self-efficacy and learners' writing skill. *Language Teaching Research*. Advance online publication. <https://doi.org/10.1177/13621688241239764>
- \*Gozali, I., Wijaya, A. R. T., Lie, A., Cahyono, B. Y., & Suryati, N. (2024). Leveraging the potential of ChatGPT as an automated writing evaluation (AWE) tool: Students' feedback literacy development and AWE tools integration framework. *The JALT CALL Journal*, 20(1), 1–22. <https://doi.org/10.29140/jaltcall.v20n1.1200>
- \*Hiniz, G. (2024). A year of generative AI in English language teaching and learning: A case study. *Journal of Research on Technology in Education*, 1–21. <https://doi.org/10.1080/15391523.2024.2404132>
- 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), 237–257. <https://doi.org/10.1111/jcal.12610>
- \*Huang, J., & Mizumoto, A. (2025). The effects of generative AI usage in EFL classrooms on the I2 motivational self system. *Education and Information Technologies*, 30, 6435–6454. <https://doi.org/10.1007/s10639-024-13071-6>
- \*Ismail, S. M., & Alharkan, A. (2024). EFL learners' positive emotions in the era of technology: Unpacking the effects of artificial intelligence on learning enjoyment, self-efficacy, and resilience. *Computer Assisted Language Learning Electronic Journal*, 25(4), 526–551.
- Jeon, J., & Lee, S. (2023). Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. *Education and Information Technologies*, 28, 15873–15892. <https://doi.org/10.1007/s10639-023-11834-1>
- 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>

- Ji, H., Han, I., & Ko, Y. (2023). A systematic review of conversational AI in language education: Focusing on the collaboration with human teachers. *Journal of Research on Technology in Education*, 55(1), 48–63. <https://doi.org/10.1080/15391523.2022.2142873>
- \*Karataş, F., Abedi, F. Y., Ozek Gunyel, F., Karadeniz, D., & Kuzgun, Y. (2024). Incorporating AI in foreign language education: An investigation into ChatGPT's effect on foreign language learners. *Education and Information Technologies*. Advance online publication. <https://doi.org/10.1007/s10639-024-12574-6>
- \*Kohnke, L. (2024). Exploring EAP students' perceptions of GenAI and traditional grammar-checking tools for language learning. *Computers and Education: Artificial Intelligence*, 7, 100279. <https://doi.org/10.1016/j.caeai.2024.100279>
- Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, 54(2), 537–550. <https://doi.org/10.1177/00336882231162868>
- Lai, W. Y. W., & Lee, J. S. (2024). A systematic review of conversational AI tools in ELT: Publication trends, tools, research methods, learning outcomes, and antecedents. *Computers and Education: Artificial Intelligence*, 7, 100291. <https://doi.org/10.1016/j.caeai.2024.100291>
- Law, L. (2024). Application of generative artificial intelligence (GenAI) in language teaching and learning: A scoping literature review. *Computers and Education Open*, 6, Article 100174. <https://doi.org/10.1016/j.caeo.2024.100174>
- \*Lee, Y. J., & Davis, R. O. (2024). A case study of implementing generative AI in university's general English courses. *Contemporary Educational Technology*, 16(4), ep533. <https://doi.org/10.30935/cedtech/15218>
- \*Lee, J. H., Shin, D., & Hwang, Y. (2024). Investigating the capabilities of large language model-based task-oriented dialogue chatbots from a learner's perspective. *System*, 127, 103538. <https://doi.org/10.1016/j.system.2024.103538>
- \*Lee, J. H., Shin, D., & Noh, W. (2023). Artificial intelligence-based content generator technology for young English-as-a-foreign-language learners' reading enjoyment. *RELC Journal*, 54(2), 508–516. <https://doi.org/10.1177/00336882231165060>
- Lee, S., & Jeon, J. (2024a). Teacher agency and ICT affordances in classroom-based language assessment: The return to face-to-face classes after online teaching. *System*, 121, 103218. <https://doi.org/10.1016/j.system.2023.103218>
- Lee, S., & Jeon, J. (2024b). Visualizing a disembodied agent: Young EFL learners' perceptions of voice-controlled conversational agents as language partners. *Computer Assisted Language Learning*, 37(5–6), 1048–1073. <https://doi.org/10.1080/09588221.2022.2067182>
- Lee, S., Jeon, J., & Choe, H. (2025a). Enhancing pre-service teachers' Global Englishes awareness with technology: A focus on AI chatbots in 3D metaverse environments. *TESOL Quarterly*, 59(1), 49–74. <https://doi.org/10.1002/tesq.3300>
- Lee, S., Jeon, J., & Choe, H. (2025b). Generative AI (GenAI) and pre-service teacher agency in ELT. *ELT Journal*. <https://doi.org/10.1093/elt/ccaf005>
- Li, B., Lowell, V. L., Wang, C., & Li, X. (2024). A systematic review of the first year of publications on ChatGPT and language education: Examining research on ChatGPT's use in language learning and teaching. *Computers and Education: Artificial Intelligence*, 7, Article 100266. <https://doi.org/10.1016/j.caeai.2024.100266>
- \*Li, X., Li, B., & Cho, S.-J. (2023). Empowering Chinese language learners from low-income families to improve their Chinese writing with ChatGPT's assistance afterschool. *Languages*, 8(4), Article 238. <https://doi.org/10.3390/languages8040238>
- Liang, J.-C., Hwang, G.-J., Chen, M.-R. A., & Darmawansah, D. (2023). Roles and research foci of artificial intelligence in language education: An integrated bibliographic analysis and systematic review approach. *Interactive Learning Environments*, 31(7), 4270–4296. <http://doi.org/10.1080/10494820.2021.1958348>
- \*Liu, M., Zhang, L. J., & Biebricher, C. (2024). Investigating students' cognitive processes in generative AI-assisted digital multimodal composing and traditional writing. *Computers & Education*, 211, Article 104977. <https://doi.org/10.1016/j.compedu.2023.104977>
- \*Long, H. S. (2024). Exploring the use of ChatGPT as a tool for written corrective feedback in an EFL classroom. *The Journal of Asia TEFL*, 21(2), 397–412. <https://doi.org/10.18823/asiatefl.2024.21.2.8.397>
- \*Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments*, 11, Article 9. <https://doi.org/10.1186/s40561-024-00295-9>
- Mai, D. T. T., Da, C. V., & Hanh, N. V. (2024). The use of ChatGPT in teaching and learning: A systematic review through SWOT analysis approach. *Frontiers in Education*, 9, Article 1328769. <https://doi.org/10.3389/feduc.2024.1328769>
- \*Meyer, J., Jansen, T., Schiller, R., Liebenow, L. W., Steinbach, M., Horbach, A., & Fleckenstein, J. (2024). Using LLMs to bring evidence-based feedback into the classroom: AI-generated feedback increases secondary students' text revision, motivation, and positive emotions. *Computers and Education: Artificial Intelligence*, 6, Article 100199. <https://doi.org/10.1016/j.caeai.2023.100199>
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, 6(7), e1000097. <http://doi.org/10.1371/journal.pmed.1000097>
- Navigli, R., Conia, S., & Ross, B. (2023). Biases in large language models: Origins, inventory, and discussion. *Journal of Data and Information Quality*, 15(2), 1–21. <https://doi.org/10.1145/3597307>
- OpenAI. (2024a, May 13). *Hello GPT-4o*. <https://openai.com/index/hello-gpt-4o/>
- OpenAI. (2024b, February 13). *Memory and new controls for ChatGPT*. <https://openai.com/index/memory-and-new-controls-for-chatgpt/>

- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). *The PRISMA 2020 statement: An updated guideline for reporting systematic reviews*. *BMJ*, 372, Article n71. <https://doi.org/10.1136/bmj.n71>
- \*Pan, M., Lai, C., & Guo, K. (2025). Effects of GenAI-empowered interactive support on university EFL students' self-regulated strategy use and engagement in reading. *The Internet and Higher Education*, 65, 100991. <https://doi.org/10.1016/j.iheduc.2024.100991>
- Payne, A. L., Austin, T., & Clemons, A. M. (2024). Beyond the front yard: The dehumanizing message of accent-altering technology. *Applied Linguistics*, 45(3), 553–560. <https://doi.org/10.1093/applin/amae002>
- \*Polakova, P., & Ivenz, P. (2024). The impact of ChatGPT feedback on the development of EFL students' writing skills. *Cogent Education*, 11(1), 2410101. <https://doi.org/10.1080/2331186X.2024.2410101>
- \*Punar Özçelik, N., & Ekşi, Y. G. (2024). Cultivating writing skills: The role of ChatGPT as a learning assistant—a case study. *Smart Learning Environments*, 11, Article 10. <https://doi.org/10.1186/s40561-024-00296-8>
- \*Robillos, R. (2024). Synergizing generative pre-trained transformer (GPT) chatbots in a process-based writing paradigm to enhance university students' writing skills. *Journal of Language and Education*, 10(3), 79–94. <https://doi.org/10.17323/jle.2024.18708>
- Schneider, J. (2024). *What comes after transformers? A Selective Survey Connecting Ideas in Deep Learning*. ArXiv. <https://doi.org/10.48550/arXiv.2408.00386>
- \*Shaikh, S., Yayilgan, S. Y., Klimova, B., & Pikhart, M. (2023). Assessing the usability of ChatGPT for formal English language learning. *European Journal of Investigation in Health, Psychology and Education*, 13(9), 1937–1960. <https://doi.org/10.3390/ejihpe13090140>
- \*Tai, T.-Y., & Chen, H. H.-J. (2024). Improving elementary EFL speaking skills with generative AI chatbots: Exploring individual and paired interactions. *Computers & Education*, 220, 105112. <https://doi.org/10.1016/j.compedu.2024.105112>
- \*Teng, M. F. (2024a). “ChatGPT is the companion, not enemies”: EFL learners' perceptions and experiences in using ChatGPT for feedback in writing. *Computers and Education: Artificial Intelligence*, 7, 100270. <https://doi.org/10.1016/j.caeai.2024.100270>
- \*Teng, M. F. (2025). Metacognitive awareness and EFL learners' perceptions and experiences in utilising ChatGPT for writing feedback. *European Journal of Education*, 60(1), e12811. <https://doi.org/10.1111/ejed.12811>
- Teng, M. F. (2024b). A systematic review of ChatGPT for English as a Foreign language writing: Opportunities, challenges, and recommendations. *International Journal of TESOL Studies*, 6(3), 36–57. <https://doi.org/10.58304/ijts.20240304>
- \*Tsai, C.-Y., Lin, Y.-T., & Brown, I. K. (2024). Impacts of ChatGPT-assisted writing for EFL English majors: Feasibility and challenges. *Education and Information Technologies*. Advance online publication. <https://doi.org/10.1007/s10639-024-12722-y>
- \*Tseng, Y.-C., & Lin, Y.-H. (2024). Enhancing English as a foreign language (EFL) learners' writing with ChatGPT: A university-level course design. *Electronic Journal of E-Learning*, 22(2), 78–97. <https://doi.org/10.34190/ejel.21.5.3329>
- Vaccino-Salvadore, S. (2023). Exploring the ethical dimensions of using ChatGPT in language learning and beyond. *Languages*, 8(3), 191. <https://doi.org/10.3390/languages8030191>
- \*Van Horn, K. (2024). ChatGPT in English language learning: Exploring perceptions and promoting autonomy in a university EFL context. *TESL-EJ*, 28, Article 1. <https://doi.org/10.55593/ej.28109a8>
- Walter, Y. (2024). Embracing the future of Artificial Intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education. *International Journal of Educational Technology in Higher Education*, 21(1), 15. <https://doi.org/10.1186/s41239-024-00448-3>
- \*Wang, Y. (2024). Cognitive and sociocultural dynamics of self-regulated use of machine translation and generative AI tools in academic EFL writing. *System*, 126, 103505. <https://doi.org/10.1016/j.system.2024.103505>
- \*Wang, C. (2024). Exploring students' generative AI-assisted writing processes: Perceptions and experiences from native and nonnative English speakers. *Technology, Knowledge and Learning*. Advance online publication. <https://doi.org/10.1007/s10758-024-09744-3>
- \*Wang, C., Zou, B., Du, Y., & Wang, Z. (2024). The impact of different conversational generative AI chatbots on EFL learners: An analysis of willingness to communicate, foreign language speaking anxiety, and self-perceived communicative competence. *System*, 127, 103533. <https://doi.org/10.1016/j.system.2024.103533>
- \*Wiboolyasarin, W., Wiboolyasarin, K., Suwanwihok, K., Jinowat, N., & Muenjanchoey, R. (2024). Synergizing collaborative writing and AI feedback: An investigation into enhancing L2 writing proficiency in wiki-based environments. *Computers and Education: Artificial Intelligence*, 6, Article 100228. <https://doi.org/10.1016/j.caeai.2024.100228>
- \*Xiao, Y., & Zhi, Y. (2023). An exploratory study of EFL learners' use of ChatGPT for language learning tasks: Experience and perceptions. *Languages*, 8, Article 212. <https://doi.org/10.3390/languages8030212>
- \*Yamaoka, K. (2024). ChatGPT's motivational effects on Japanese university EFL learners: A qualitative analysis. *International Journal of TESOL Studies*, 6(3), 24–35. <https://doi.org/10.58304/ijts.20240303>
- \*Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Education and Information Technologies*, 28(11), 13943–13967. <https://doi.org/10.1007/s10639-023-11742-4>

- Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., & Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, 55(1), 90–112. <https://doi.org/10.1111/bjet.13370>
- 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>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, 16(1), 39. <https://doi.org/10.1186/s41239-019-0171-0>
- \*Zhang, R., Zou, D., Cheng, G., & Xie, H. (2025). Flow in ChatGPT-based logic learning and its influences on logic and self-efficacy in English argumentative writing. *Computers in Human Behavior*, 162, 108457. <https://doi.org/10.1016/j.chb.2024.108457>
- \*Zheng, S. (2024). The effects of chatbot use on foreign language reading anxiety and reading performance among Chinese secondary school students. *Computers and Education: Artificial Intelligence*, 7, 100271. <https://doi.org/10.1016/j.caeai.2024.100271>
- \*Zou, M., & Huang, L. (2023). The impact of ChatGPT on L2 writing and expected responses: Voice from doctoral students. *Education and Information Technologies*. Advance online publication. <https://doi.org/10.1007/s10639-023-12397-x>