Enhancing L2 Writing with Generative AI: A Systematic Review of Pedagogical Integration and Outcomes

2 authors, including:

Hui Wang
University of Arizona
9 PUBLICATIONS   **17** CITATIONS

SEE PROFILE

Enhancing L2 Writing with Generative AI: A Systematic Review of Pedagogical Integration and Outcomes

Hui Wang[a], Anh Dang[a]

[a]*Second Language Acquisition and Teaching, University of Arizona, Tucson, USA*

Abstract

The rapid development of Generative Artificial Intelligence (GenAI) opens up new possibilities for the fields of education. GenAI tools can serve as writing assistants, supporting student writers by producing human-like text and suggestions. With the growing adoption of these tools, this study examines how instructors have integrated GenAI in second language (L2) writing classes to support students' writing skills. We focus on the AI-assisted learning experiences, including student population, teaching contexts, pedagogical design, measurements, and learning outcomes. 21 empirical studies were identified for this systematic review. Results indicate that GenAI tools are primarily applied for brainstorming ideas, assessing and providing feedback, generating writing samples for analysis, scaffolding writing process, raising GenAI literacy, and developing critical thinking skills. The research also emphasizes learning challenges and critical factors by reviewing the post-implementation assessment, including reliance on GenAI tools, inaccurate information, cultural bias, and language assimilation for L2 students. The findings inform future AI-assisted design by highlighting teacher mediation and supervision as well as teacher training on the use of GenAI in L2 writing activities. Future practices and research should be encouraged to further explore the impact and potential of AI-assisted writing instruction across a broader range of academic levels and domains.

Keywords: Generative AI, Artificial Intelligence, ChatGPT, L2 writing, AI-assisted instruction, AI-enhanced education

1. Introduction

The rise and advancement of Generative Artificial Intelligence (GenAI), such as ChatGPT, has achieved profound impacts on educational fields. This powerful virtual assistant can enable human-like conversations and provides thorough writing support, including generating ideas, composing essays, offering personalized feedback, and polishing language to meet a range of needs from writers [1, 2, 3]. Previous research has highlighted the inevitable presence of GenAI in students' writing practices [4], although its benefits on teaching and learning writing remain inconclusive [5]. For second language (L2) writing, the absence of cultural diversity and inclusion in GenAI's responses may exacerbate biases to the underrepresented L2 student group [6]. This begs a question of what it means for L2 students to integrate GenAI in their writing processes, given the different positions that they're being put in. This complex nature of GenAI has led to a need for a systematic review of actual AI-assisted L2 writing instruction implementation and its impacts on students' learning. However, limited systematic reviews have focused specifically on the application of GenAI in writing education, particularly in the context of L2 writing.

Driven by the current research gap, this systematic review focuses on the integration of GenAI technologies in L2 writing courses to support L2 students' writing skills. We will delve into the application of GenAI tools through an examination of 21 empirical studies in diverse L2 teaching contexts. This study aims to explore the current application of GenAI in L2 writing pedagogy and inform future AI-assisted L2 writing instructional designs.

2. Literature review

Generative Artificial Intelligence (GenAI) has experienced rapid development and brought profound impacts in the field of education. AI-powered chatbots, such as ChatGPT, are designed to provide immediate human-like conversations to help people solve problems in writing tasks ranging from linguistic errors to genre-specific patterns [7, 8]. Numerous studies have been conducted on the implications of GenAI in education. For writing-intensive courses, GenAI has been employed to achieve a range of goals, including evaluating student writing [9, 10], offering corrective feedback [11, 12], generating samples for analysis [13, 14], fostering student collaboration [15], and

enhancing students' learning experiences [16]. On the other hand, due to the nature of these GenAI tools, there are also discussions regarding the risks of integrating them in writing courses. As banning GenAI from schools is not an option [17, 18], ethical concerns regarding plagiarism and academic integrity have become central of these discussions [19, 20, 21], along with other unignorable concerns, such as overreliance and inaccuracy [22, 23]. When it comes to the context of L2 writing, more complex perspectives and nuanced voices tend to emerge, highlighting the additional challenges faced by L2 student writers. For example, according to Bender et al. [6], GenAI models were primarily trained on the internet-based language dataset, where white supremacist and misogynistic, ageist, etc. views are often overrepresented. Training on these datasets worsens diversity and inclusion in their responses and in turn enforces White Mainstream English and amplifies biases and harms to the underrepresented L2 student population [6]. Additionally, current GenAI detectors often misclassify L2 writing as AI-generated texts, raising bias, fairness, and robustness concerns for L2 students [24, 25]. Overall, GenAI's impacts on L2 writing education remain inconclusive and complicated. Such diverse discussion leads to a pressing need for a systematic review of previous empirical research on AI-assisted L2 writing instruction with a focus on the diverse student populations, guiding concepts, GenAI application approaches, and learning outcomes.

Several existing systematic review studies have explored GenAI's impacts in the educational area. For instance, Dempere et al. [26] reviewed 143 studies on the use of ChatGPT in higher education, identifying significant benefits and risks, and recommended thoughtful and responsible integration of GenAI within the higher education context. Li et al. [27] analyzed the design of learning tasks for AI education in K-12 contexts and revealed its effectiveness in fostering understanding and engagement as well as learning challenges in developing AI literacy and solving technical issues. Labadze et al. [28] investigated the ways in which teachers and students used AI chatbots in their teaching and learning and addressed its important benefits in providing grammatical corrections and syntactic suggestions.

Given the potential of GenAI, educators have expressed interest and likelihood to use GenAI in writing courses while acknowledging that there will be hesitation, and uneasiness regarding AI-assisted writing instruction [29]. However, there have been few systematic reviews focusing on how GenAI was applied in writing education specifically. This scarcity is even more prominent in the context of L2 writing. Dang & Wang [17] examine over

3

100 universities' publicly available statements on the use of GenAI for the teaching of writing, and suggest educators to consider diverse student populations in their GenAI policies and advocate for more related research on the relationship between L2 writing and GenAI to amplify the voices of L2 students in the era of GenAI. Pedagogical implications and accommodation strategies for L2 students should receive more attention in discussions about the integration of GenAI in education.

Therefore, driven by the existing research gap and the need for a comprehensive understanding of AI-assisted L2 writing instruction, our systematic review analyzes both pedagogical approaches and implications of integrating GenAI technologies in L2 writing courses. We aim to explore student populations and contexts, pedagogical strategies, learning focuses, measurements, and learning outcomes. Our research questions are as follows:

1. What were the student population and teaching contexts in the empirical studies of integrating GenAI into L2 writing classes?
2. In what ways did instructors incorporate GenAI tools into their L2 writing classes? What AI-assisted pedagogical strategies were applied to teach L2 writing?
3. What were the learning focuses that guide AI-assisted teaching practices in L2 writing contexts?
4. How did researchers measure AI-assisted L2 writing instruction, and what were the learning outcomes and perceptions of L2 students after engaging with AI-assisted L2 writing instruction?

3. Method

3.1. Data collection

Data collection was conducted in May 2024 by using educational databases including Google Scholar, ProQuest, Scopus, Science Direct, Web of Science, and JSTOR. To align with the research questions, we used a set of keywords as search terms, such as "GenAI", "AI", "L2 writing", "education", and etc. The search results were evaluated by the research team based on the inclusion criteria. Eligible papers should meet the inclusion criteria: 1) empirical studies, 2) focus on the application of GenAI in writing education, 3) implemented in L2 contexts, 4) written in English and full-text available. We considered both published and unpublished studies to reduce publication

Table 1: Student population and contexts.

Parent code	Child code
Population information	Sample size
	Student original country
	First language
	Age
	Proficiency
	Academic levels
	Target language
Context information	Duration
	Implementation country
	Class mode
	GenAI tools
	Writing task genres

bias [30]. Participants cover a range of age groups from K-12 to graduate students. After an examination with the research team, twenty one studies met the inclusion criteria, including 17 published journal articles, 2 proceeding articles, and 2 articles from conference books (see Data availability).

Table 2: Pedagogical application of GenAI in L2 writing.

Code	Explanation
Evaluation and feedback	GenAI was used as a resource for students to receive feedback, and conduct automatic assessments on their drafts.
Brainstorming	GenAI was used to help students brainstorm initial writing ideas and outlines.
Predicting	GenAI was used for students to: 1. Predict topics; 2. Continue writing after the given partial sentences/paragraphs/drafts; 3. Continue developing ideas.
Revision	Instructors asked students to revise and edit AI-generated information.
Modeling	GenAI was used for students to: 1. Generate writing samples as models for analysis and evaluation; 2. Generate writing samples and compare them with human writing.
Summarizing	GenAI was used for students to summarize long or difficult texts.
Content generation	GenAI was used for students to: 1. Generate longer texts or first drafts; 2. Produce images for multimodal writing.
Intellectual writing support	GenAI was used for students to: 1. Scaffolding the writing process; 2. Analyze data; 3. Search for more resources.
Teaching support	GenAI was used for teachers to: 1. Generate writing prompts, and other teaching materials; 2. Conduct learning analytics; 3. Identify learning patterns.

3.2. Data analysis

To answer the research questions, we developed a coding scheme focusing on four dimensions of the studies: 1) student population and contexts (see Table 1); 2) pedagogical application of GenAI in L2 writing (see Table 2); 3) learning focuses and guiding concepts in AI-assisted classes (see Table 3); 4) measurements and learning outcomes (see Table 4). The first dimension was designed to answer the first research question about student population and contexts. It included sample size, student original country, first language, age groups, proficiency, academic levels, target language, duration, implementation country, class mode, GenAI tools, and writing task genres. Following Wang et al. [18], the second dimension explored how GenAI tools were applied in L2 writing classes. We categorized the applications into evaluation and feedback, brainstorming, predicting, revision, modeling, summarizing, content generation, intellectual writing support, and teaching support. The third dimension was informed by Hyland [31] and Long & Magerko [32] on learning focuses and guiding concepts in AI-assisted L2 writing classes. It covered language structure, text functions, creative expression, writing process, content knowledge, genre and contexts of writing, GenAI literacy, and critical thinking. The fourth dimension investigated measurements and learning outcomes, including several measurement methods, writing performances, and students' affective responses.

Thematic analysis [33] was conducted to identify the main themes in writing performances and affective responses. Specifically, we found that the emerging themes can be categorized into linguistic accuracy, genre conventions, content, and structure. Regarding students' affective responses, we considered the primary emerging themes, such as writing support, usefulness, motivation, confidence, GenAI features, reliance on AI, creativity, and critical thinking.

Table 3: Learning focuses and guiding concepts in AI-assisted classes.

Code	Explanation
Language structure	The implementation focused on essential linguistic knowledge and vocabulary choices.
Text functions	The implementation focused on relating structures to meanings and using language to express target functions.
Creative expression	The implementation focused on encouraging students to express their personal experiences and opinions. Writing was considered a creative act of self-discovery.
Writing process	The implementation focused on developing students' metacognitive awareness of their writing process, helping them reflect on the strategies they use to write.
Content knowledge	The implementation involved subject matter themes and topics, such as pollution and health.
Genre and contexts of writing	The implementation focused on genre knowledge, including awareness of audience, adaptation to contexts, and clarity of purposes to successfully meet the expectations of specific text types.
GenAI literacy	The implementation involved: 1. Play around and get familiar with tools; 2. Understand the strengths and weaknesses of GenAI tools; 3. Define prompts, learn prompt engineering.
Critical thinking	The implementation included developing students' critical thinking abilities.

Table 4: Measurements and learning outcomes.

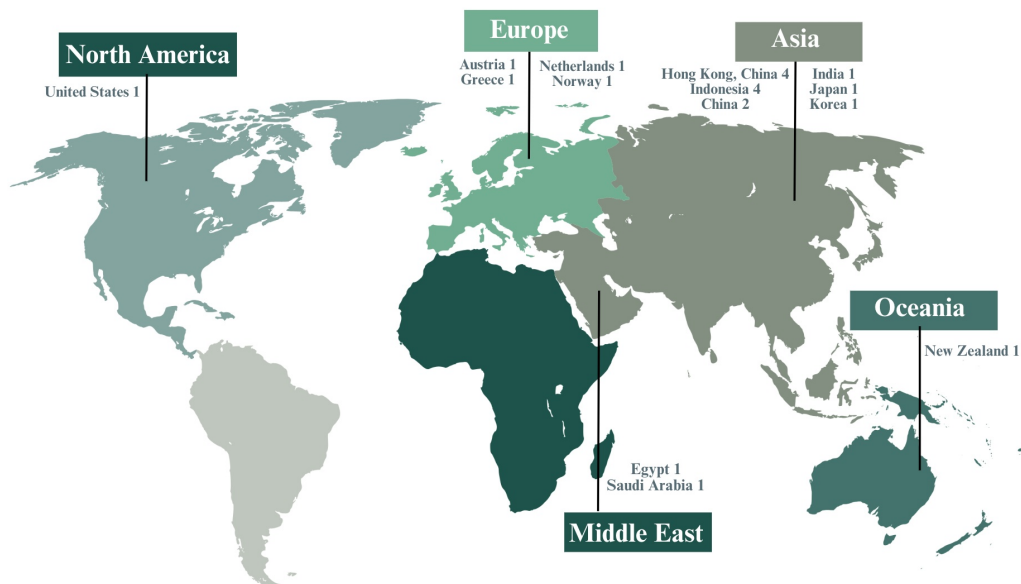| Parent code	Child code

Measurements	Surveys
	 Interviews
	 Students' productions and artifacts:
	     1. Implementation vs. Control
	     2. Diagnostic writing tests
	 Journals, think-aloud, stimulated recalls
	 Observation
	 Student reflection

Writing performances	Linguistic accuracy Genre conventions Content Structure
Context information	Duration Implementation country Class mode GenAI tools Writing task genres

Affective responses	Writing support
	 Writing support
	 Usefulness
	 Motivation
	 Confidence
	 GenAI features:
	     1. Quick response
	     2. Ease of use
	     3. Convenience
	     4. User-friendly
	 Reliance on AI
	 Creativity
	 Critical thinking

Figure 1: Countries and the number of studies.

4. Results

4.1. Results for RQ1: student population and teaching contexts

To answer RQ1, this section introduces student demographics and teaching contexts, including duration, technologies, and writing tasks of the 21 empirical studies. On average, the included empirical studies of AI-assisted L2 writing instruction had a mean student sample size of 31 but with a large variance (SD = 27.29), which suggests the sample sizes vary widely from each other. Eleven studies were small-scale exploratory research, with enrollment of no more than 30 participants. Two studies did not report the number of participants. The remaining studies included fewer than 100 student participants. Regarding student demographics, as shown in Table 5, more studies aimed at teaching advanced adult learners, who came from diverse L1 backgrounds. Notably, four studies (19%) did not report students' L1 background, although it may be related to the country where the implementation was conducted. Ten studies (48%) did not disclose the students' language proficiency. Twelve studies (57%) did not provide information on their students' ages.

Table 5: Student demographics.

Theme	Code	Count	Percentage
Proficiency	Low	1	5
	Low-intermediate	2	10
	Intermediate	1	5
	High-intermediate	2	10
	Advanced	5	24
	Not reported	10	48
First language (L1)	Mandarin	3	14
	Cantonese	2	10
	German	1	5
	Indonesian	3	14
	Dutch	1	5
	Korean	1	5
	Japanese	1	5
	Arabic	2	10
	English	1	5
	Mixed	2	10
	Not reported	4	19
Age	>=18	6	29
	13-17	3	14
	<=12	0	0
	Not reported	12	57

Despite the small number of empirical studies in this emerging AI-assisted instruction field, research has been conducted across 13 countries in Europe, Asia, the Middle East, Oceania, and North America (see Fig. 1). Asia leads with the highest number of studies (n = 13, 62%) with Hong Kong and Indonesia each conducting four studies (19%). Europe follows, with four studies (19%) conducted across four different countries. Two studies were from the Middle East countries. North America and Oceania each had one study in this domain. This distribution reflected a notable concentration of interest and research activity in AI-assisted L2 writing instruction in Asia, alongside a growing global engagement with this educational technology.

Table 6: Teaching contexts.

Theme	Code	Count	Percentage
Institution	Elementary school	0	0
	Middle school	2	10
	High school	4	19
	Undergraduate	12	57
	Graduate and post-graduate	2	10
	Language learning center	1	5
Language specialty	Foreign language	19	90
	Second language	2	10
Target language	English	18	86
	Chinese	1	5
	German	2	10
Duration	Less than 1 week	2	10
	2-4 weeks	7	33
	5-8 weeks	3	14
	More than 8 weeks	2	10
	Not reported	7	33
GenAI tools	ChatGPT	21	100
	Mixed	4	19

In terms of more details on teaching contexts, the institutions examined span from middle school to graduate levels, most (n=12, 57%) were situated in the undergraduate programs, while K-12 setting came in second with 4 studies (19%) in high school context and two studies in middle school context (10%) (see Table 6). Most studies (n=19, 90%) were conducted in foreign language learning settings where English was the target language for learning (n=18, 86%). As to the duration of the studies, seven studies (33%) were short-term implementations that lasted 2-4 weeks. Meanwhile, seven studies (33%) did not report their length of teaching.

As revealed at the bottom of Table 6, we found that so far ChatGPT served as the prevalent GenAI tool in L2 writing classrooms. Almost all selected studies used ChatGPT as their main GenAI tool for teaching and learning L2 writing. Four studies (19%) also combined with other GenAI tools, such as Bing Chat [14] and Microsoft Aim writing [34].

Regarding the writing tasks that these studies focused on (see Fig. 2),
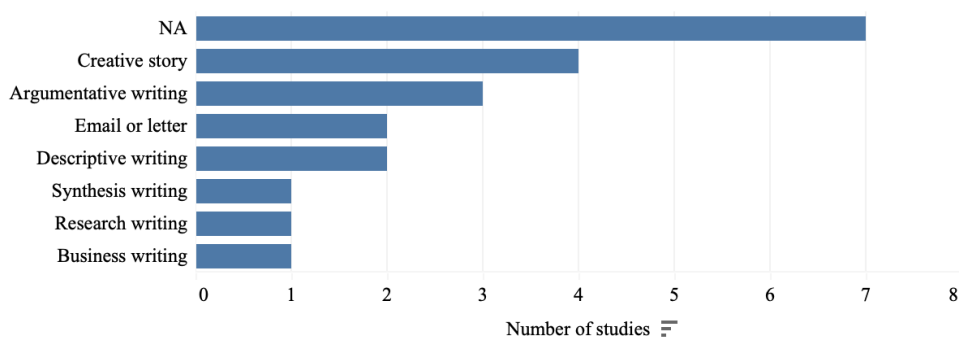
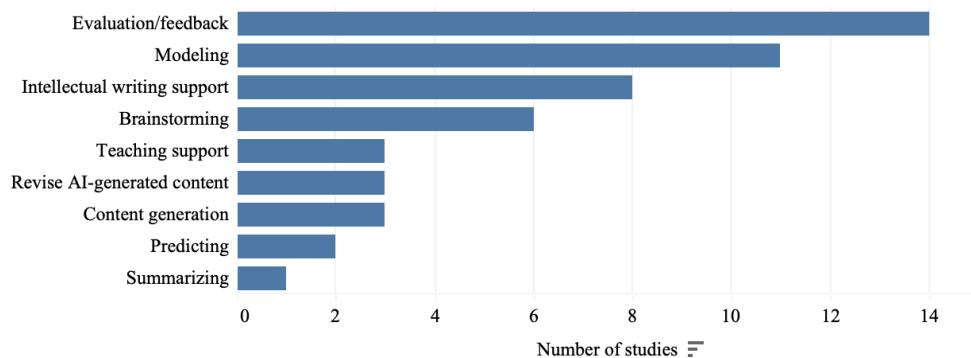Figure 2: Writing task genres and the number of studies.

Figure 3: AI-assisted pedagogical applications in L2 writing contexts.

we notice a wide range of genres, with creative story (n=4, 19%) and argumentative writing (n=3, 14%) being used the most, with a learning focus on forms and functions. We used the term "genre" here to indicate. However, seven studies did not specify the genres of the writing tasks they had, instead categorizing them as academic writing or L2 writing in general.

4.2. Results for RQ2: AI-assisted pedagogical applications in L2 writing contexts

Research question 2 examines the variety of AI-assisted pedagogical applications that were applied to L2 writing classes. The 21 studies included comprehensive descriptions of AI-assisted language learning activities as well as the roles GenAI performed during the implementation process to support

L2 students in developing their writing and writing-related skills. We categorized them into 9 categories developed based on Crompton & Burke [35] and Wang et al. [18]. Fig. 3 reflects the analysis of AI-assisted pedagogical applications.

The result shows that evaluation and feedback were the most frequent application of GenAI in L2 writing classrooms (n=14, 67%). This category includes using GenAI tools as a resource for students to evaluate their drafts and receive feedback. For L2 student writers who face extra language difficulties, numerous studies have employed GenAI as a tool to facilitate students' self-evaluation and provide instant corrective feedback. It helped L2 students avoid long-time waiting for teacher feedback, keeping them engaged and motivated throughout the writing process, as well as promoting self-regulated learning [35, 36]. Regarding the content of evaluation and feedback, the majority of studies (n=14, 67%) focused on correcting linguistic errors in L2 writing [37, 12, 38, 39], with a few studies (n=2, 10%) also mentioned feedback on genre features and conventions using proper prompts [36, 39]. In addition, it is worth noting that compared with traditional teacher feedback, which often highlights language errors in texts and allows students to make changes accordingly, GenAI usually automatically corrects and edits original texts and presents rewritten texts without further explanation [40].

Modeling was in second place as a common application of GenAI in L2 writing classrooms (n=11, 52%). It refers to 1) generating writing samples as models for analysis and evaluation; and 2) generating writing samples and comparing them with human writing. Considering the L2 learning background, L2 students often face extra disadvantages in some unfamiliar cultural contexts, exposure to a variety of genres can accelerate the processes of situated acquisition [41]. Leveraging its access to an extensive corpus of data, GenAI can generate writing samples of different genres, which can serve as writing samples for L2 students to analyze genre and language conventions. For example, in the study conducted by Liu et al. [14], the drafts produced by ChatGPT have functioned as model texts to improve L2 students' writing skills regarding organizational structures, language and content. However, Kim et al. [42] observed that AI-generated model texts may often provide rigid, machine-like templates that led L2 students to passively follow these structures and write with strict organization without independent thoughts and varieties.

Using GenAI as intellectual writing support was identified as another useful AI-assisted pedagogical application (n=8, 38%). Three subcodes were

recognized in how GenAI supported the L2 students' writing process: scaffolding the writing process, analyzing data, and searching for resources. Our extant reviews of the selected studies highlighted the use of GenAI for scaffolding (n=3, 14%). Scaffolding support is defined as a strategy that prompts students to start, connect, and develop their thoughts step by step [41], guiding them to construct their texts through "gradual approximation" [43]. This approach helps L2 learners understand writing as a process and develop greater confidence [41]. An example of this is the study by Harunasari [44], students were assigned to write a short story where ChatGPT broke down the whole task into smaller manageable pieces, guiding them through developing an initial storyline, identifying key plot points, and composing narrative effectively. Besides scaffolding guidance on the writing process, in a study by Liu et al. [14], students interacted with Bing Chat to search for academic references, reflecting the application of GenAI in resource searching for academic writing.

Another code that emerged in data analysis was applying GenAI for brainstorming (n=6, 29%). This includes brainstorming initial writing ideas and outlines at the beginning of the writing process. Generating creative and unique ideas can be some students' writing blocks. The integration of ChatGPT can spark students' creativity and enhance their idea generation abilities [13, 44, 37]. Instead of automatically generating writing content, this approach aimed at fostering students' own critical thinking, planning and creativity at the initial stages of the writing process.

Aside from the four most frequent applications discussed above, a smaller portion of studies explored other pedagogical uses of GenAI, as illustrated in Fig. 3. These applications included integrating GenAI for students to generate longer texts for their first drafts and to produce images for multimodal writing projects, documented in three studies (14%). Another three studies (14%) addressed how students revised and edited AI-generated content to improve their language awareness and writing skills. Additionally, three studies (14%) focused on GenAI's role in teaching preparation and management. These studies found that GenAI can serve as a teaching assistant to generate writing prompts, create teaching materials, analyze learning data, and identify learning patterns. Two studies (10%) introduced that students can use GenAI to predict and continue their writing from partial sentences, paragraphs, or drafts. One study (5%) reported that they used GenAI as a learning tool to help L2 students summarize long and difficult texts.
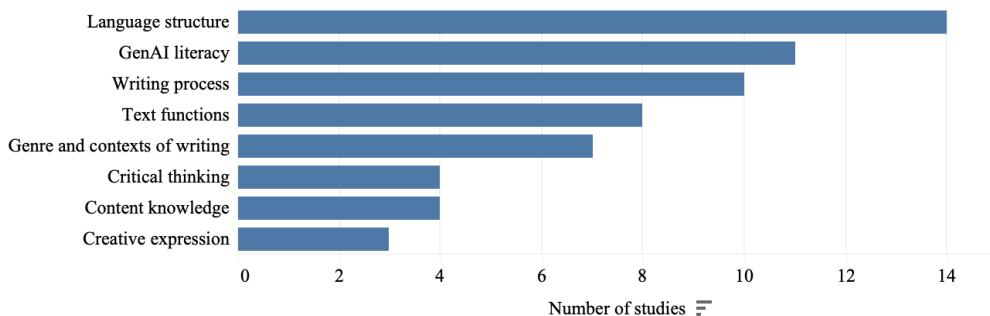
Figure 4: Learning focuses in AI-assisted L2 writing teaching.

4.3. Results for RQ3: Learning focuses in AI-assisted L2 writing teaching

RQ3 explores different learning focuses that guided AI-assisted teaching practices in L2 writing contexts. Learning focuses support teachers in deciding what students need to learn and what they need to provide for an effective AI-assisted writing class. According to Hyland [31], L2 writing teaching is often guided by six focuses: language structures, text functions, creative expression, composing processes, content, and genre and contexts of writing. In the context of integrating GenAI into L2 writing pedagogy, we have added two pertinent themes: GenAI literacy and critical thinking. According to Long and Magerko [32], GenAI literacy refers to a set of competencies that support us to understand, engage, and evaluate GenAI applications. It is worth noting that teachers often draw on multiple focuses rather than strictly adhering to one of those, but it is common for one focus to prevail in shaping their teaching design and implementation [45, 31]. Therefore, although these focuses are seldom reflected as distinct teaching approaches, examining the diversity and disparity of learning focuses on AI-assisted L2 writing classrooms can shed light on the features of AI-assisted L2 writing instruction and provide insights into some potential ways to further develop our pedagogical approaches.

As revealed in Fig. 4, language structure emerged as a prominent learning focus, with 14 studies (67%) emphasizing this goal [40, 13, 37, 14, 34]. A big portion of AI-assisted L2 writing instruction involved developing L2 students' linguistic knowledge and grammar competence which were the most immediately obvious learning obstacles of L2 student writers [46, 31, 47]. GenAI literacy came as the second prevalent learning focus, appearing in 11 studies (53%), as observed in addressing the strengths and weaknesses of AI-
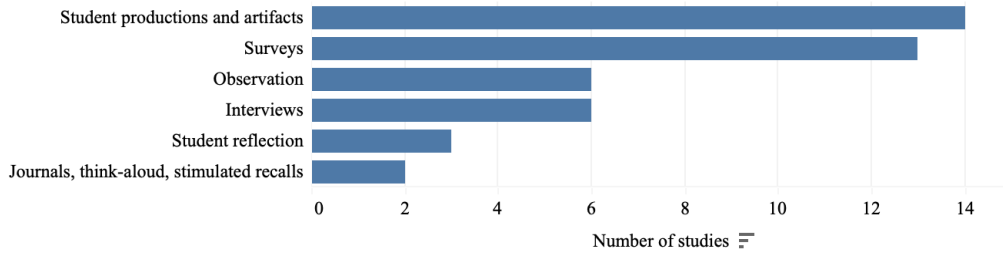
Figure 5: Measurement methods in AI-assisted L2 writing teaching.

generated texts and helping L2 students enhance their prompt engineering skills [13, 48, 39, 49, 34]. Writing process was another prevalent learning focus (n = 10, 48%) which fostered a process-based writing approach to emphasize developing students' awareness and abilities to review, revise, and reflect in the process of writing [10, 12, 50]. Text functions were seen in 8 studies (38%). They focused on teaching structures and relating structures to meanings. Some relevant learning activities included using GenAI to generate an outline [44, 42, 14] and comment on paragraph structure to better scaffold L2 students in writing [10]. Seven studies (33%) focused on genre and contexts of writing, drawing L2 students' attention to awareness of audience, adaptation to contexts, and clarity of purposes [14, 38, 50, 51]. These areas might be the most deficiencies of L2 students due to cultural differences in rhetoric and the unfamiliarity of rhetorical conventions in their second language [52, 31, 53, 54]. Additionally, both critical thinking and content knowledge have been the learning focus in 4 studies (19%). Creative expression was addressed in 3 studies (14%). Four studies (19%) did not present specific learning focuses or objectives in their AI-assisted L2 writing instruction.

4.4. Results for RQ4: Measurements, learning outcomes and perceptions

To address RQ4, we noticed six types of measurement methods in the collected articles: student artifact evaluation, surveys, observations, interviews, student reflection, and journal reports (see Fig. 5). These measurements were used to analyze two dimensions of the effects of AI-assisted L2 writing instruction: (1) students' writing performances, such as linguistic and genre-specific features; and (2) students' affective responses to AI-assisted L2 writing instruction, including perceptions and feedback.

17

For the first dimension, most of the studies (n = 14, 67%) included a mixed analysis of students' writing performances through collected student productions and artifacts [40, 55, 51]. Among these studies, eleven studies (52%) applied a pre-test and post-test design to assess students' writing performances [44, 37, 48, 38], and 3 studies (14%) compared the drafts from implementation and control groups [10, 13, 39]. To understand more about the effects of GenAI on students' writing performances, we delved deep into their findings in terms of artifact analysis. The results widely acknowledged that GenAI was effective in correcting linguistic errors and thus improving the accuracy of student writing [13, 49]. Some studies also addressed that GenAI helped L2 students notice and adopt appropriate styles and registers in their language use, such as converting informal expressions into formal academic language in scientific writing genres [50]. However, when it came to the use of examples, results became inconsistent in the collected studies. Some studies highlighted that GenAI can provide a variety of examples during writing process [14, 34], however, the provided examples might not always be concrete, specific, or relevant enough to support the claims effectively [13, 14]. As many studies were focused on teaching academic writing, citations were often one of the major elements evaluated in student texts. The results of citations tended to be negative, as they can be missing or inaccurate when students engaged with GenAI tools. Studies also revealed that the structure of students' essays and paragraphs did not significantly improve and even declined in some cases [13].

Regarding examining students' affective responses to AI-assisted L2 writing instruction (the second dimension), self-reported evaluation was often used with surveys as the most frequent data collection instrument (n = 13, 62%). Other instruments included observation (n = 6, 29%), interviews (n = 6, 29%), student reflection (n = 3, 14%), and journals (n = 2, 10%). As for perceptions of L2 students after engaging with AI-assisted L2 writing instruction, many students across different studies affirmed the effectiveness of using GenAI in lexicogrammatical revision, which greatly alleviated their anxiety and boosted their confidence and motivation in English writing as foreign language learners [37, 55, 36, 39]. Some students expressed a sense of empowerment in their second language acquisition attributed to GenAI's feature of quick response, ease of use, convenience [56, 48, 12, 39]. Additionally, the ease of understanding the responses was highlighted as user-friendly for L2 learners [10, 14].

The most frequent concern that emerged from student responses and ob-

servations was overreliance on GenAI. With the adoption of GenAI in teaching and learning, a number of students became excessively dependent on it, which hindered their brainstorming ability, critical thinking and creativity [13, 55, 51]. Since the default tone of ChatGPT was usually White Mainstream English [6], L2 students in the study of Song & Song [39] realized that the suggestions and revised writing did not always align with their own writing style or were not contextually accurate. Moreover, the individuality, human nuances, and authenticity in writing were reported reduced or eliminated after AI revision, resulting in writing that tended to be homogeneous and standardized, lacking stances, originality, and creativity [10, 50].

5. Discussion

In the following sections, we highlight the main findings and trends in the collected studies, discuss pedagogical implications on both the design and implementation of AI-assisted L2 writing instruction, and explore potential gaps and opportunities for future research in this area.

5.1. Contexts of studies

The systematic review of the 21 studies highlights a prevalent focus on AI-assisted L2 writing instruction for adult English language learners in foreign language contexts (EFL). Together with the development of GenAI, this focus raises concerns about the underrepresentation of K-12 students, ESL learners, and learners of languages other than English. This phenomenon calls for future empirical research to explore a wider range of student demographics in various contexts, including non-English learners and countries where the target language is the dominant language, to address the needs of L2 writing educators who work with diverse student populations globally.

Over half of the collected studies had less than 30 participants. The implementation was frequently in small-scale short-term contexts. More empirical research is needed to explore the design and implications of AI-assisted L2 writing instruction with a larger participant group and a longer duration.

The predominance of using ChatGPT as their GenAI tool may be attributed to its ease of use and stability across education contexts [57]. Fewer studies have tested other GenAI tools or combined ChatGPT with other digital tools [14, 34]. As large language models continue to evolve, we expect more opportunities to explore and compare various tools in future educational practices and research.

19

5.2. AI-assisted pedagogical approaches

With regard to how teachers applied GenAI to assist their L2 writing classes, we identified key pedagogical approaches emerging from the studies, including evaluation and feedback, modeling, intellectual writing support, brainstorming, etc. A salient pattern from the findings was the prevalent use of GenAI to address linguistic errors, which was an efficient and reliable approach for L2 students who need extra language support for writing. Beyond linguistic errors, GenAI was also useful in generating sample texts for L2 students to observe and identify genre conventions [14, 55, 38]. For L2 students who are unfamiliar with naturalistic language settings, learning about genres can be a foundational approach to engaging with cultural and textual practices [41]. Furthermore, GenAI was integrated to scaffold the writing process by breaking large projects into smaller pieces and providing ideas to inspire students' thoughts.

While we learned various approaches to integrate GenAI in L2 writing education, it is important to highlight that GenAI often automatically edits original texts [40], allowing L2 students to directly copy and paste the revised text into their drafts. This should bring our attention to the question of whether it merely improves a single essay per se or genuinely enhances students' writing skills. Further teaching practices and research are needed to evaluate students' authentic self-writing skills without GenAI after engaging with AI-assisted instruction.

In addition, we argue that it is crucial to emphasize that the role of GenAI should be a writing assistant offering support for consideration rather than an alternative writing editor for students or an alternative teacher imparting writing knowledge. The findings in the studies present that texts generated by GenAI may contain fake citations and rigid, machine-like templates, which could be misleading and confusing for L2 student writers. Thus, it is essential for L2 writing teachers to explicitly discuss these risks and increase students' awareness of GenAI. Specifically, teachers can emphasize the assistant role of GenAI and remind students not to overrely on GenAI. Students should be cautious not to accept all AI-generated content as accurate and should avoid treating AI-generated texts as standards or requirements. Exemplary designs can be found in some studies, such as Dwina [58], Harunasari [44], and Strobl et al. [50]. They had students analyze and compare AI-generated texts with human writing with prompts guiding students to assess content. This approach can promote students' abilities of critical thinking, evaluation,

fact-checking, and verifying information when using GenAI in L2 writing contexts.

5.3. Learning content in GenAI-based class

In terms of learning content in AI-assisted L2 writing instruction, studies in this domain widely focused on using GenAI to improve L2 students' language structure, which signaled its effectiveness in addressing the foundational aspect of L2 writing [40, 13, 37, 14, 34]. The second salient trait was developing students' GenAI literacy, including understanding GenAI capabilities, delving into ethical considerations, and enhancing prompt engineering skills [13, 48, 39, 49, 34]. It indicated that building GenAI awareness and literacy has been listed as one of the learning objectives in some classes and tended to become a core component of AI-assisted L2 writing instruction. This pattern resonates with the discussion in Chiu and Sanusi [59] and Stolpe and Hallström [60], emphasizing cultivating GenAI literacy across broader educational domains and preparing students for a future where GenAI permeates everyday and professional environments.

While it is important to explore the new technology, we suggest that the integration of GenAI in writing education should align with the learning objectives and curriculum content, recognize its pedagogical affordances, and ensure coherent and responsive courses. As a kind of digital tool, TPACK (Technological Pedagogical Content Knowledge) framework is also applicable to the use of GenAI and relevant teacher training. Mishra et al., [61] claim that "TPACK focuses on the integration of tools with content and pedagogy, not on the specific tools themselves" (p.239). In order to use GenAI effectively, teachers need to be equipped with the skills to integrate these tools with pedagogy knowledge (PK)—the understanding of teaching strategies and management—and content knowledge (CK)—the understanding of content and learning objectives to be taught. In other words, this integration should go beyond only using GenAI for innovation per se; it requires an intentional focus on the pedagogical and content objectives that the technology is meant to support [62, 63, 18]. For example, in the context of L2 writing, teachers may ask students to observe and analyze the language patterns used in AI-generated texts with prompts leading their attention to authorial voice and audience awareness. This allows students to concentrate on the content learning of the rhetorical situation and writing techniques. As this example indicates, responsive teacher training is essential to ensure that instructors

can strategically and purposefully leverage GenAI to enhance students' writing skills.

5.4. Measurement and outcomes

The results of this systematic review reveal a wide range of measurement methods used in research on AI-assisted L2 writing instruction, including qualitative and quantitative analysis of student productions and artifacts, surveys, observations, interviews, reflections, and journals. Student work revealed that GenAI was effective in improving the accuracy of student writing as well as raising their genre awareness. However, noteworthy is the fact that in some cases, the structure of students' essays and paragraphs did not evidently improve and even declined after engaging with GenAI. This might be due to some students submitting drafts generated exclusively by ChatGPT with minimal self input [13]. AI-generated writing often lacked concrete examples and citations [13] and tended to share a uniform template-like structure which made students' writing lack effectiveness and creativity [49]. Additionally, since GenAI often adopts White Mainstream English in its responses, using it to assist L2 writing education may reinforce the idea of standard English. This approach can assimilate L2 student writing and exacerbate hegemonic biases against L2 students [6]. These findings suggest greater teacher mediation and supervision on the use of GenAI in AI-assisted writing practices. In teaching practices, teachers can encourage L2 students to value linguistic diversity, understand diverse linguistic structures, and critically engage with AI-generated texts.

Analyzing students' perceptions and experiences, we found that GenAI can increase L2 students' confidence and reduce anxiety in writing tasks, which suggested that GenAI may be more effective for facilitating the initial stages of language acquisition, particularly in lexicogrammatical revision. However, dependence on GenAI raises concerns about the development of critical thinking and creative writing skills. For L2 students, GenAI may pose bias and language assimilation for L2 students. According to Hawanti and Zubaydulloevna [37], ChatGPT learns from text data found on the internet, which can include biases from biased or offensive sources. This may lead the model to show cultural or regional biases toward certain norms or languages and therefore assimilate L2 students in writing practices. The results from students' perceptions and experiences once again emphasize the importance of teacher supervision and mediation on GenAI in L2 writing classes. Effective strategies may include establishing clear policies and guidelines for

GenAI use and requiring students to submit their chat history of GenAI, which could help manage if it is used effectively and ethically [44, 18]. More research is needed to continue to explore the broader implications of GenAI integration from the perspective of instructors who are implementing AI-assisted activities in L2 writing courses.

6. Conclusion

By examining 21 studies that integrated GenAI in L2 writing classrooms across various language proficiencies and countries, the study mapped out the current student demographics, pedagogical applications, learning focuses, measurements, and learning outcomes on AI-assisted L2 writing courses. The findings showed that a primary focus on integrating ChatGPT for adult English language learners in EFL contexts, reflecting the need for attention to more diverse student groups and broader educational contexts. The reviewed studies presented diverse pedagogical applications of GenAI in language learning, with instant corrective feedback for linguistic errors emerging as the most prevalent application, alongside other applications such as evaluation, modeling, intellectual writing support, and brainstorming. This highlighted GenAI's effectiveness in assisting L2 students who require additional linguistic support. In AI-assisted L2 writing classrooms, the learning objectives also included exploring the new technology and cultivating GenAI awareness and literacy. Understanding strengths and weaknesses of GenAI tools as well as prompt engineering skills would help students learn how to interact with GenAI tools to enhance their learning without limiting their creative thinking. However, the review also found important limitations of integrating GenAI in L2 writing education. From the reviewed studies, we noticed that automatic edits from GenAI may contain fake and unnatural content, which could be misleading and confusing for L2 students. Additionally, while qualitative and quantitative analysis revealed that GenAI was generally effective in improving student writing and boosting L2 students' confidence, a few reviewed studies reported that there were student works that did not improve or even declined. This phenomenon might be attributed to the excessive reliance on GenAI tools and the biases and language assimilation that these tools may present to L2 students. These findings reflected the importance of developing students' critical thinking skills to evaluate AI-generated content and cultivating students' awareness to avoid over-reliance.

The findings of this study carried pedagogical and research implications

for the future integration of GenAI in L2 writing courses. For teaching practices, this systematic review summarizes a variety of pedagogical approaches and learning focuses that could serve as valuable resources for teacher scholar who are interested in implementing AI-assisted L2 writing classes. In addition, the study emphasizes the important role of teacher training on this new technology, advocating for frameworks such as the TPACK model to ensure that GenAI integration aligns with specific learning objectives and course content. Moreover, the findings also address the importance of teacher supervision and mediation in AI-assisted L2 writing environments to guide students toward the ethical and effective use of GenAI tools. Specifically, teachers' responsibilities in this new era may extend to raising L2 students' awareness about potential pitfalls of AI-generated content, such as misinformation, cultural bias, and language assimilation. Establishing clear policies and guidelines may also become essential to regulate students' GenAI usage, and thereby building a responsible and effective L2 writing classroom.

For a research-oriented perspective, this study contributed to a comprehensive overview of GenAI's role in L2 writing education. As this technology rapidly evolves and becomes more widely used, more empirical studies are needed to explore its impact across diverse student demographics and educational contexts. For example, future studies could explore the engagement and benefits of AI-assisted writing instruction for learners of different language proficiencies and L1 backgrounds. Additionally, more research should further explore student performance in AI-assisted L2 writing classes and teachers' role in scaffolding these activities. For instance, students' authentic independent writing skills without GenAI after participating in AI-assisted instruction might be a worthy area of investigation, as it could provide insights into the transferability of skills acquired through AI-assisted instruction. Finally, while most of the reviewed studies implemented GenAI over short periods (within two months), it would be beneficial to examine its long-term effects on L2 students' writing skills and language development.

Declarations

References

[1] M. Abdullayeva, Z. M. Musayeva, The impact of chat gpt on student's writing skills: An exploration of ai-assisted writing tools, in: International Conference of Education, Research and Innovation, Vol. 1, 2023, pp. 61–66.

[2] K. Fuchs, Exploring the opportunities and challenges of nlp models in higher education: is chat gpt a blessing or a curse?, in: Frontiers in Education, Vol. 8, Frontiers Media SA, 2023, p. 1166682.

[3] J. Rudolph, S. Tan, S. Tan, Chatgpt: Bullshit spewer or the end of traditional assessments in higher education?, Journal of applied learning and teaching 6 (1) (2023) 342–363.

[4] M. Sullivan, A. Kelly, P. McLaughlan, Chatgpt in higher education: Considerations for academic integrity and student learning (2023).

[5] M. M. Rahman, Y. Watanobe, Chatgpt for education and research: Opportunities, threats, and strategies, Applied Sciences 13 (9) (2023) 5783.

[6] E. M. Bender, T. Gebru, A. McMillan-Major, S. Shmitchell, On the dangers of stochastic parrots: Can language models be too big?, in: Proceedings of the 2021 ACM conference on fairness, accountability, and transparency, 2021, pp. 610–623.

[7] C. K. Y. Chan, W. Hu, Students' voices on generative ai: Perceptions, benefits, and challenges in higher education, International Journal of Educational Technology in Higher Education 20 (1) (2023) 43.

[8] G.-G. Lee, L. Shi, E. Latif, Y. Gao, A. Bewersdorf, M. Nyaaba, S. Guo, Z. Wu, Z. Liu, H. Wang, et al., Multimodality of ai for education: Towards artificial general intelligence, arXiv preprint arXiv:2312.06037 (2023).

[9] A. Barrett, A. Pack, Not quite eye to ai: student and teacher perspectives on the use of generative artificial intelligence in the writing process, International Journal of Educational Technology in Higher Education 20 (1) (2023) 59.

[10] J. Escalante, A. Pack, A. Barrett, Ai-generated feedback on writing: insights into efficacy and enl student preference, International Journal of Educational Technology in Higher Education 20 (1) (2023) 57.

[11] A. McGuire, W. Qureshi, M. Saad, A constructivist model for leveraging genai tools for individualized, peer-simulated feedback on student writing., International Journal of Technology in Education 7 (2) (2024) 326–352.

[12] R. Schmidt-Fajlik, Chatgpt as a grammar checker for japanese english language learners: A comparison with grammarly and prowritingaid, AsiaCALL Online Journal 14 (1) (2023) 105–119.

[13] K. Glahn, Using chatgpt to teach english for academic purposes: Perspectives from the classroom, Journal of Education, Innovation and Communication 5 (2) (2024) 63–75.

[14] M. Liu, L. J. Zhang, C. Biebricher, Investigating students' cognitive processes in generative ai-assisted digital multimodal composing and traditional writing, Computers & Education 211 (2024) 104977.

[15] L. Law, Application of generative artificial intelligence (genai) in language teaching and learning: A scoping literature review, Computers and Education Open (2024) 100174.

[16] H. Li, Y. Wang, S. Luo, C. Huang, The influence of genai on the effectiveness of argumentative writing in higher education: evidence from a quasi-experimental study in china, Journal of Asian Public Policy (2024) 1–26.

[17] A. Dang, H. Wang, Ethical use of generative ai for writing practices: Addressing linguistically diverse students in u.s. universities' ai statements, Journal of Second Language Writing (2024).

[18] H. Wang, A. Dang, Z. Wu, S. Mac, Seeing chatgpt through universities' policies, resources and guidelines, arXiv preprint arXiv:2312.05235 (2023).

[19] J. Luo, A critical review of genai policies in higher education assessment: A call to reconsider the "originality" of students' work, Assessment & Evaluation in Higher Education (2024) 1–14.

[20] M. A. Yeo, Academic integrity in the age of artificial intelligence (ai) authoring apps, Tesol Journal 14 (3) (2023) e716.

[21] A. Yusuf, N. Pervin, M. Román-González, Generative ai and the future of higher education: a threat to academic integrity or reformation? evidence from multicultural perspectives, International Journal of Educational Technology in Higher Education 21 (1) (2024) 21.

[22] W. Alharbi, Ai in the foreign language classroom: A pedagogical overview of automated writing assistance tools, Education Research International 2023 (1) (2023) 4253331.

[23] C. K. Y. Chan, K. K. Lee, The ai generation gap: Are gen z students more interested in adopting generative ai such as chatgpt in teaching and learning than their gen x and millennial generation teachers?, Smart learning environments 10 (1) (2023) 60.

[24] J. Campino, Unleashing the transformers: Nlp models detect ai writing in education, Journal of Computers in Education (2024) 1–29.

[25] W. Liang, M. Yuksekgonul, Y. Mao, E. Wu, J. Zou, Gpt detectors are biased against non-native english writers, Patterns 4 (7) (2023).

[26] J. Dempere, K. Modugu, A. Hesham, L. K. Ramasamy, The impact of chatgpt on higher education, in: Frontiers in Education, Vol. 8, Frontiers Media SA, 2023, p. 1206936.

[27] L. Li, Y. Fengchao, E. Zhang, A systematic review of learning task design for k-12 ai education: Trends, challenges, and opportunities, Computers and Education: Artificial Intelligence (2024) 100217.

[28] L. Labadze, M. Grigolia, L. Machaidze, Role of ai chatbots in education: systematic literature review, International Journal of Educational Technology in Higher Education 20 (1) (2023) 56.

[29] H. Nguyen Thi Thu, Efl teachers' perspectives toward the use of chatgpt in writing classes: A case study at van lang university, Nguyen, TTH (2023). EFL Teachers' Perspectives toward the Use of ChatGPT in Writing Classes: A Case Study at Van Lang University. International Journal of Language Instruction 2 (3) (2023) 1–47.

[30] R. Rosenthal, The file drawer problem and tolerance for null results., Psychological bulletin 86 (3) (1979) 638.

[31] K. Hyland, Second language writing, Cambridge university press, 2019.

[32] D. Long, B. Magerko, What is ai literacy? competencies and design considerations, in: Proceedings of the 2020 CHI conference on human factors in computing systems, 2020, pp. 1–16.

[33] V. Braun, V. Clarke, Using thematic analysis in psychology, Qualitative research in psychology 3 (2) (2006) 77–101.

[34] D. Yan, Impact of chatgpt on learners in a l2 writing practicum: An exploratory investigation, Education and Information Technologies 28 (11) (2023) 13943–13967.

[35] H. Crompton, D. Burke, Artificial intelligence in higher education: the state of the field, International Journal of Educational Technology in Higher Education 20 (1) (2023) 22.

[36] L. M. Silitonga, S. Hawanti, F. Aziez, M. Furqon, D. S. M. Zain, S. Anjarani, T.-T. Wu, The impact of ai chatbot-based learning on students' motivation in english writing classroom, in: International Conference on Innovative Technologies and Learning, Springer, 2023, pp. 542–549.

[37] S. Hawanti, K. M. Zubaydulloevna, Ai chatbot-based learning: alleviating students' anxiety in english writing classroom, Bulletin of Social Informatics Theory and Application 7 (2) (2023) 182–192.

[38] S. Shaikh, S. Y. Yayilgan, B. Klimova, M. Pikhart, Assessing the usability of chatgpt for formal english language learning, European Journal of Investigation in Health, Psychology and Education 13 (9) (2023) 1937–1960.

[39] C. Song, Y. Song, Enhancing academic writing skills and motivation: assessing the efficacy of chatgpt in ai-assisted language learning for efl students, Frontiers in Psychology 14 (2023) 1260843.

[40] S. Athanassopoulos, P. Manoli, M. Gouvi, K. Lavidas, V. Komis, The use of chatgpt as a learning tool to improve foreign language writing in a multilingual and multicultural classroom, Advances in Mobile Learning Educational Research 3 (2) (2023) 818–824.

[41] K. Hyland, Genre pedagogy: Language, literacy and l2 writing instruction, Journal of second language writing 16 (3) (2007) 148–164.

[42] S. Kim, J. Shim, J. Shim, et al., A study on the utilization of openai chatgpt as a second language learning tool, Journal of Multimedia Information System 10 (1) (2023) 79–88.

[43] H. G. Widdowson, Teaching language as communication, Oxford University Press, 1978.

[44] S. Y. Harunasari, Examining the effectiveness of ai-integrated approach in efl writing: A case of chatgpt, International Journal of Progressive Sciences and Technology (IJPSAT) 39 (2) (2023) 357–368.

[45] A. Cumming, Experienced esl/efl writing instructors' conceptualizations of their teaching: Curriculum options and implications, Exploring the dynamics of second language writing (2003) 71–92.

[46] E. Hinkel, Teaching academic ESL writing: Practical techniques in vocabulary and grammar, Routledge, 2003.

[47] A. C. Purves, Writing across languages and cultures: Issues in contrastive rhetoric, (No Title) (1988).

[48] X. Li, B. Li, S.-J. Cho, Empowering chinese language learners from low-income families to improve their chinese writing with chatgpt's assistance afterschool, Languages 8 (4) (2023) 238.

[49] D. J. Woo, Evaluating a learning design for efl writing using chatgpt.

[50] C. Strobl, I. Menke-Bazhutkina, N. Abel, M. Michel, Adopting chatgpt as a writing buddy in the advanced l2 writing class, Technology in Language Teaching & Learning 6 (1) (2024) 1–19.

[51] D. J. Woo, K. Guo, S. Z. Salas-Pilco, Writing creative stories with ai: learning designs for secondary school students, Innovation in Language Learning and Teaching (2024) 1–13.

[52] U. Connor, Contrastive rhetoric, Cross cultural aspects of second language writing/Cambridge University (1996).

[53] Y. Kachru, Cultural meaning and contrastive rhetoric in english education, World Englishes 16 (3) (1997) 337–350.

[54] R. B. Kaplan, Cultural thought patterns in inter-cultural education, Language learning 16 (1-2) (1966) 1–20.

[55] S. Mahapatra, Impact of chatgpt on esl students' academic writing skills: a mixed methods intervention study, Smart Learning Environments 11 (1) (2024) 9.

[56] M. A. Ahmed, Chatgpt and the efl classroom: Supplement or substitute in saudi arabia's eastern region, Information Sciences Letters 12 (7) (2023) 2727–2734.

[57] N. Saif, S. U. Khan, I. Shaheen, F. A. ALotaibi, M. M. Alnfiai, M. Arif, Chat-gpt; validating technology acceptance model (tam) in education sector via ubiquitous learning mechanism, Computers in Human Behavior 154 (2024) 108097.

[58] L. Dwina, Chatgpt: Empowering self-directed writing through mind mapping and ai-assisted composition, Nusantara Science and Technology Proceedings (2024) 33–45.

[59] T. K. Chiu, I. T. Sanusi, Define, foster, and assess student and teacher ai literacy and competency for all: Current status and future research direction (2024).

[60] K. Stolpe, J. Hallström, Artificial intelligence literacy for technology education, Computers and Education Open 6 (2024) 100159.

[61] P. Mishra, M. Warr, R. Islam, Tpack in the age of chatgpt and generative ai, Journal of Digital Learning in Teacher Education 39 (4) (2023) 235–251.

[62] S. Leahy, P. Mishra, Tpack and the cambrian explosion of ai, in: Society for Information Technology & Teacher Education International Conference, Association for the Advancement of Computing in Education (AACE), 2023, pp. 2465–2469.

[63] P. Perera, Preparing to revolutionize education with the multi-model genai tool google gemini? a journey towards effective policy making (2023).