Applications and Research Gaps of LLM-based English-as-a-foreign-language Education

Hengke Jiang

College of Letters & Science, University of Wisconsin-Madison, Madison, Wisconsin, 53703, The United States of America

hjiang296@wisc.edu

Abstract. English as a Foreign Language (EFL) education is crucial in today's globalized world. With the rapid development of Artificial Intelligence (AI) and Large Language Models (LLMs) over the past few years, educators continuously investigate how to use them to improve EFL educational methods. This article reviews the studies on integrating LLMs with EFL speaking and writing education from 2020 to 2024. It separately summarizes current LLMs' strengths and weaknesses in improving EFL learner's skills and learning motivation. This research demonstrates a promising future of applying LLMs in EFL education because of their strong abilities of immediate feedback generation and no-stress communication environment creation. However, they also find their shortages like superficial incorrect feedback. The research gaps in these areas are also analyzed based on these articles, which include uneven participants, the lack of differentiated investigation on samples with different ages and education levels, and the lack of research on various LLMs.

Keywords: English education; large language models; artificial intelligence.

1. Introduction

In an era of globalization, English language education is essential for personal growth, professional development, and international communication. As a globally used language, English allows learners with another native language to communicate with people from other countries with fewer barriers, enabling them to participate effectively and benefit from international activities or projects. Due to the value attached to language education, researchers are discovering methods of adopting technology into language education and using them to stimulate learner's studying interests [1].

In recent years, Large Language Models (LLM) and their fast development have drawn attention from educators. As a branch of Artificial Intelligence (AI), LLMs aim to train models that can understand and generate human-like language. These models are featured by their massive numbers of parameters and diverse training datasets, and they demonstrate good performance on human-level text generations and relative tasks [2].

Due to LLMs' strong functions, they are already used in education and have a profound impact. For instance, the integration with LLMs allows intelligent tutoring systems to understand students' questions and past performances better to provide students with more personalized learning support [3]. Similarly, the potential practicality is also shown on chatbots. Taking Chat Generative Pre-trained Transformer (ChatGPT) as an example, its feature of conversational interaction means it can encourage users to raise following reactions, which enables learners to personalize the knowledge they gain and encourages them to think thoroughly [4]. Moreover, LLMs demonstrate their potential abilities to assist language education and academic writing as powerful tools. They are already used in tasks such as creating questions, providing prompts, grading essays, and generating feedback [5].

Since LLM tools like ChatGPT have opportunities for enhancing teaching and learning experiences, it is essential to analyze their potential applications and future development trends. For all language learners, reading, writing, listening, and speaking skills are four significant skills [6]. This study focuses on English as a Foreign Language (EFL) speaking and writing education from 2020 to 2024, reviewing the research on applying LLMs on these two topics. For EFL teachers, their responsibilities include two sides: 1. Recognizing EFL learners' needs and adopting suitable techniques and methods to improve their English skills 2. Using learner-centered approaches and

applying attractive teaching materials to stimulate students' learning interests [6]. Therefore, this study will discuss the aspects of enhancing EFL learner's language skills and learning motivation separately. Based on these reviewed articles, it will analyze the strengths and shortages of LLMs-assisted EFL education and summarize the limitations of studies in this area. These pieces of information can help EFL educators plan the usage of current LLMs and point a direction for future research in this area.

2. LLM-based Application for EFL

This research's topic is understanding and summarizing the current research progress of applying chatbots and LLMs to enhance the English abilities of EFL learners. The relevant research papers published since 2020 were collected, and the influences of LLMs on EFL speaking and writing education will be analyzed from two aspects: (1) Using LLMs and chatbots to improve EFL learner's English level. (2) Using LLMs and chatbots to improve EFL learner's motivation and willingness

2.1. LLM-based Spoken English Education

Acquiring a good speaking skill means developing the ability to communicate effectively, clearly, and confidently, which is essential for personal, academic, and professional socializing. However, for EFL learners, especially those not living in a non-English-speaking environment, limited resources of reliable human experts are a common hindrance. Without expert feedback or personalized assistance, it would lead to an ineffective speaking learning process [7]. Another problem that EFL learners usually encounter is the anxiety of making mistakes when speaking English. Therefore, they need an environment where they can practice speaking skills without stress [8]. In recent years, chatbots' function of interacting with users through human-like languages promotes researchers to explore the methods of using them to meet these EFL learner's requirements.

The studies evaluating the effectiveness of providing personalized feedback to learners' speaking exercises through chatbots show optimistic results. In 2024, Mhasakar et al. developed an LLM-based application called Comuniqa to transcribe and evaluate user speech inputs. They observed that the output of Comuniqa is accurate, and it is more helpful than other traditional applications for learners who cannot access human instruction. However, compared with human experts, AIs cannot provide the same level of emotional support and personalized guidance [7]. Hsu, Chen, and Todd trained college students in Taiwan to practice speaking with Intelligent personal assistants (IPA) Amazon Alexa by asking questions, searching for word definitions, and playing interactive games. After seven Alexa courses, there was an obvious improvement in these students' Test of English for International Communication (TOEIC) speaking tests. Researchers attribute the improvement to reasons that these students are able to chat with IPA in an appropriate context and fix their faults based on IPA's reaction [9]. TPBOT, a task-oriented chatbot developed to help the preparation of TOEIC speaking tasks, also demonstrates its ability to enhance learner's scores. While the average speaking grade of the control group students who only self-learned with traditional TOEIC materials did not have an obvious change, the average speaking grades of the experimental group students increased by around 65 scores [10].

In addition to making up for the lack of human experts' assistance, another benefit of AI-assisted speaking learning is reducing learner's speaking anxiety. For example, Tai and Chen's study shows that eighth graders' speaking confidence, fluency, and willingness were enhanced by the interaction with IPA Google Assistance. Besides, the immediate feedback from IPA can encourage learners to self-correct and repeat exercises [8]. Similarly, EFL students who learned with Amazon Alexa reflected that this learning process is less stressful than practicing with classmates or instructors [9], and those who learned with Tian Peng (TPBot) reflected that this tool could increase their self-assurance than directly communicating with foreigners [10].

To sum up, there are sufficient articles demonstrating the positive effects of LLMs on improving EFL learners' speaking skills and alleviating their anxiety, which is enough to prove the potential of

introducing LLMs into this area. Despite these studies showing many faults of LLMs that make them inferior to experienced human experts, LLMs' advantages of high accessibility and interactivity would enable them to create a conversation environment for EFL learners. This is rare for EFL learners with limited resources of experienced instructors, and it can let LLMs become useful tools for self-studying.

2.2. LLM-based Written English Education

For EFL students, good writing skill refers to the ability to effectively communicate ideas, thoughts, and information in written English, which plays an important role in EFL students' academic achievements. For EFL teachers, giving effective feedback on student writing is time-consuming, labor-intensive, and subjective. Furthermore, English language learners also face motivation issues and time constraints, limiting their ability to improve their writing skills [11]. MMLs, especially ChatGPT, are seen as a possible solution to these problems by many educators.

To investigate the influences of applying ChatGPT on EFL writing, much research has been done in recent years. The study from Song and Song investigated the effectiveness of ChatGPT writing aids by training Chinese EFL learners to pose inquiries, request clarification, or ask for additional examples and explanations from ChatGPT. They compared these students' performance with other EFL students at similar English writing levels and found that after a 12-week experiment, the average International English Language Testing System (IELTS) writing score of students assisted by ChatGPT increased by about one-third compared to those assisted by human experts, with scores of 59.12 and 45.18 respectively [11]. Boudouaia, Mouas, and Kouider's research demonstrated a similar result. They found that students who revised their papers based on ChatGPT's suggestions performed better in the post-test than those who did not and showed significant improvement compared to their pre-test scores [12]. However, Algaraady and Mahyoob discovered ChatGPT's shortages in recognizing writing errors. They found that ChatGPT can efficiently catch surface-level mistakes like grammar, spelling, and punctuation errors. Still, it has difficulty identifying issues related to deeper structural and pragmatic aspects like human experts, and it also sometimes fabricates errors when the input is correct [13].

However, although LLMs' assistance would be helpful for EFL learners to improve their writing skills, whether they can enhance student's writing motivation remains uncertain. The participants in Song and Song's study reflected that the interaction with ChatGPT made them more confident and engaged during the writing process. [11] Similarly, through Guo and Li's research on Chinese undergraduate students, creating and using personalized retrieval augmented generation (RAG) chatbots enables augmentative writing to not only address their problems, but also increase their confidence and motivate them to improve [14]. Conversely, a study conducted on Hong Kong Secondary school students showed that teaching and training them to write with ChatGPT enhanced learning motivation insignificantly, and this increase might be attributed to the "novelty effect". In addition, crafting appropriate prompts for ChatGPT has the risk of increasing students' cognitive load, which could overwhelm them [15].

All in all, there are plenty of articles about enhancing EFL learners' writing skills with LLMs as assistive or correction tools, and they show consistently positive results. However, the number of studies on applying LLMs to enhance EFL learners' writing motivation is limited, and their effects remain debatable. For these studies with different conclusions, age and education level seem to be the main differences among their samples. College students seem to gain confidence from writing with LLMs more significantly, but secondary school students would be overwhelmed for undetermined reasons. Therefore, more studies to discover which factors increase these students' cognitive loads to develop more suitable LLMs are needed.

3. Discussion

The reviewed articles in this paper demonstrate the practicable methods of applying LLMs to EFL speaking and writing education. For English oral instructions, LLMs, especially IPAs like Google Assistant and Amazon Alexa, demonstrate the ability to help EFL learners overcome speaking anxiety by supporting them with encouraging conversation and immediate feedback [8, 9]. Besides, even though human experts are more professional and more empathetic than LLM-based applications [7], they can provide EFL learners with powerful assistance for self-studying when human experts are not affordable or available [7, 9, 10]. For English writing instructions, researchers mainly focused on exploring the usage of ChatGPT. They proved ChatGPT's abilities to answer writing questions, provide suggestions and feedback, and detect surface mistakes [11-13], but whether MMLs are able to improve EFL learner's writing motivation requires further research. In general, these studies show a bright prospect of integrating LLMs into EFL education, which would enhance the efficiency and performance of traditional education.

However, while the power of LLMs is proven, these articles also exhibit some similar concerns about LLMs' performances and reliability at present. According to Mhasakar et al.'s interview conducted on participants, sometimes they doubt that the reflection from AI is wrong, and a more transparent speaking score-calculating process is needed [7]. Likewise, Algaraady and Mahyoob indicated ChatGPT's behavior of making up mistakes when the original text was correct [13]. These show the deficiency of applying current LLMs to EFL learning and emphasize the value of enhancing their accuracy. Due to the overall good performances of LLMs in EFL education, training EFL educators to teach and supervise students' usages of AIs is also important, because they need to avoid EFL learners being misled by LLMs.

Furthermore, to confirm the future research and development directions of applying LLMs in EFL education, the research gaps and limitations of reviewed papers should also be identified.

Whether in the speaking or writing education field, the sampling range of these studies is limited to only a few countries, which leads to an unbalanced situation in that most of the learners have the same native language. Specifically, among these 9 studies, 6 of them sampled students whose native language is Chinese [8-13], two did not specify students' native languages [12,13], and only one specified that their samples have 8 different native languages [7]. This would indicate that Chinese native speakers have a relatively high demand for EFL education. However, it also means that more research needs to be conducted focusing on speakers of other languages to evaluate the LLMs' effectiveness.

Another concern related to sampling is the different ages and education levels of EFL learners. The participants of these studies can be divided into college and middle school students, and they usually have varying learning abilities, learning motivations, and English levels. In reviewed articles, most of them are focusing on college education [9-14], and only two studied middle students [8, 15]. According to these articles, these two groups' reactions to LLMs are different. Taking writing education part as an example, college students in Guo and Li's research reflected that chatbots could enhance their learning motivation [14], but Woo et. al.'s research shows that middle students cooperating with ChatGPT will not inspire them but could lead to higher cognitive load [15]. Considering the differences between college students and middle school students, researchers can explore more on the application methods that are more suitable for each of these groups separately, as well as focus on groups of other age ranges.

Lastly, the choices of LLMs in the studies on writing education are ChatGPT series, which cannot represent the overall performances of LLMs. Except for two studies that used Platform for Open Exploration (Poe) to access other LLMs or customize chatbots [14, 15], all other studies focused on writing use ChatGPT directly [11-13]. Even though ChatGPT is the most powerful LLM for now, the customizing chatbots function on Poe platforms also has the special advantage of allowing chatbots to meet students' specific requirements, improving learners' writing skills and motivations [14]. This indicates that other LLMs, especially RAG chatbots, should receive more attention from researchers in the future.

4. Conclusion

To conclude, this literature review collected and summarized the research results of studies in recent years that focused on the application of LLMs in EFL speaking or writing education. These studies demonstrated the strength of LLMs in interaction activities but also showed their defects of slight inaccuracy. Generally, they depict an optimistic prospect of integrating them into many aspects of English language education. The quality of accessibility would enable LLMs to become a strong tool to assist EFL learners' self-study on speaking, and the relaxed and meaningful conversation created by IPAs can eliminate EFL learners' anxiety effectively. These features show the probability of developing LLMs into educational software or platforms for speaking practice, especially for EFL learners who find it difficult to access experienced person instructors. In addition, to assist English writing, LLMs can answer questions, generate useful suggestions, and detect surface errors quickly, which means it can be an assistant to promote instructor's work efficiency.

However, although most of these studies have a positive result, there are many limitations. The research gaps of these studies include inadequate linguistic diversity and age diversity of participants, the lack of investigating the differences between middle school students and college students, and a lack of analysis of other LLMs's abilities besides ChatGPT. This research provides a reference for future research directions. To advance further, it is recommended to increase the samples' diversity, investigate diverse methods to assist EFL learners of various ages or English levels, and explore different LLMs' capabilities, so that educators can maximize LLMs' advantage and develop teaching strategies suitable for multiple students.

While this review provides a comprehensive analysis of the current literature on integrating LLMs into EFL speaking and writing education, it is important to acknowledge several limitations. First, this review may not cover all relevant studies due to the vast amount of available research or language barriers so that some important contributions might be ignored inadvertently. Also, considering the rapid evolution of the LLMs field, the relevance between this review's conclusion and current LLMs development would be restricted.

5. References

[1] Iqbal S., S. Niazi, and M. Hafeez. Developments on technology integration in language teaching and learning. Global Educational Studies Review, 2021, 6: 21-28.

[2] Chu Zhibo, Ni Shiwen, Wang Zichong, et al. History, Development, and Principles of Large Language Models-An Introductory Survey. arXiv preprint, 2024: 2402.06853 (2024).

[3] Shen Shiyi. Application of large language models in the field of education. Theoretical and Natural Science, 2024, 34: 147-154.

[4] Rospigliosi Pericles 'asher. Artificial intelligence in teaching and learning: what questions should we ask of ChatGPT?. Interactive Learning Environments, 2023, 31(1): 1-3.

[5] Caines Andrew, Benedetto Luca, Taslimipoor Shiva, et al. On the application of large language models for language teaching and assessment technology. arXiv preprint, 2023: 2307.08393.

[6] Rao Parupalli Srinivas. The importance of teaching language skills to the second or foreign language learners of english: A comprehensive study. ACADEMICIA: An International Multidisciplinary Research Journal, 2019, 9(4): 6-19.

[7] Mhasakar Manas, Sharma Shikhar, Mehra Apurv, et al. Comuniqa: Exploring Large Language Models for improving speaking skills. arXiv preprint, 2024: 2401.15595.

[8] Tai Tzu-Yu, Howard Hao-Jan Chen. The impact of Google Assistant on adolescent EFL learners' willingness to communicate. Interactive Learning Environments, 2023, 31(3): 1485-1502.

[9] Hsu Hsiao-Ling, Howard Hao-Jan Chen, Andrew G. Todd. Investigating the impact of the Amazon Alexa on the development of L2 listening and speaking skills. Interactive Learning Environments, 2023, 31(9): 5732-5745.

[10] Hsu Mei-Hua, Pei-Shih Chen, Chi-Shun Yu. Proposing a task-oriented chatbot system for EFL learners speaking practice. Interactive Learning Environments, 2023, 31(7): 4297-4308.

[11] Song Cuiping, Yanping Song. Enhancing academic writing skills and motivation: assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. Frontiers in Psychology, 2023, 14: 1260843.

[12] Boudouaia Azzeddine, Samia Mouas, Bochra Kouider. A study on ChatGPT-4 as an innovative approach to enhancing English as a foreign language writing learning. Journal of Educational Computing Research, 2024: 07356331241247465.

[13] Al-Garaady Jeehaan, Mohammad Mahyoob. ChatGPT's capabilities in spotting and analyzing writing errors experienced by EFL learners. Arab World English Journals, Special Issue on CALL, 2023, 9: 3-17.

[14] Guo Kai, Danling Li. Understanding EFL students' use of self-made AI chatbots as personalized writing assistance tools: A mixed methods study. System, 2024, 124: 103362.

[15] Woo David James, Wang Deliang, Guo Kai, et al. Teaching EFL students to write with ChatGPT: Students' motivation to learn, cognitive load, and satisfaction with the learning process. Education and Information Technologies, 2024: 1-28.