
RESEARCH ARTICLE

Leveraging Large Language Models to Enhance Learner Agency in Foreign Language Education

Xiang Chen

Undergraduate Student, Foreign Studies College, Hunan Normal University, Changsha, China

Corresponding Author: Xiang Chen, **E-mail:** 1987787870@qq.com

ABSTRACT

As a prominent manifestation of artificial intelligence, large language models (LLMs) have profoundly influenced foreign language education with their exceptional capabilities in natural language understanding and generation. To fully realize their potential in education, a growing body of research and policy documents has focused on the empowering role of LLMs in foreign language teaching, particularly regarding innovations in instructional models and the professional development of instructors. However, existing studies predominantly emphasize the dimension of “teaching”, highlighting how instructors can leverage technology to optimize teaching, while comparatively less attention has been paid to the dimension of “learning”. In fact, as the most important role in the process of foreign language teaching, learners’ initiative and autonomy should be further enhanced through technological support. This paper explores how students can leverage LLMs to strengthen their learner agency by examining their core functions, practical strategies for use, and the importance of cultivating critical awareness in the learning process. Through a mixed-methods questionnaire survey, this study reveals current usage trends, learner perceptions, and strategies to empower language learners in the age of artificial intelligence.

KEYWORDS

Foreign Language Education; Large Language Models; Learner Agency; Output Hypothesis; Input Hypothesis

ARTICLE INFORMATION

ACCEPTED: 01 June 2025

PUBLISHED: 30 June 2025

DOI: 10.32996/jhsss.2025.7.7.1

1. Introduction

The rapid development of artificial intelligence is profoundly reshaping education. As a significant achievement in the field of artificial intelligence, large language models (LLMs), with their powerful capabilities in natural language understanding and generation, have come to play a vital role in foreign language instruction, continuously driving its transformation. Against this backdrop, a growing body of research has focused on the integration of LLMs into foreign language education. However, the majority of existing studies adopt a teacher-centered perspective, concentrating on the transformation of instructional models, the enhancement of teachers’ professional competencies, and the opportunities and challenges brought by artificial intelligence. At the same time, national-level and education governing bodies have issued numerous policy documents that explicitly outline the requirements for instructional tasks and teacher qualifications. Nevertheless, from the perspective of foreign language learners, research and guidance on how learners themselves can harness LLMs to empower their language acquisition remain relatively limited.

The prevailing instructional models are undergoing a transformation, shifting from the traditional “teacher-student” binary structure to a “teacher-student-machine” ternary structure (Zhang, 2023: 6). Both traditional and emerging instructional models emphasize the central role of students in the teaching process, recognizing them as pivotal agents in foreign language learning. Likewise, the *Compulsory Education English Curriculum Standards* (2022 Edition) also highlights the necessity of enhancing students’ agency in the teaching process (Ministry of Education of the People’s Republic of China, 2022: ii). The rapid development of artificial intelligence in recent years has further expanded the possibilities for strengthening learners’

autonomy and initiative. This paper adopts a learner-centered perspective to analyze how foreign language learners can leverage large language models to empower their learning and enhance their autonomy.

2. Literature Review

As digital natives, contemporary students have been exposed to a variety of smart devices from an early age and possess relatively strong information technology literacy, which provides them with certain advantages in understanding and utilizing large language models to support foreign language learning. However, despite the increasing prevalence of LLMs, many students' understanding of these models remains superficial, often perceiving them merely as conversational tools. To fully realize their potential, it is essential for students to develop a clear understanding of the fundamental functions of large language models and their advantages in foreign language acquisition.

2.1 Understanding Learner Agency

Learner agency refers to students' capacity to act purposefully and constructively in their learning processes. It encompasses autonomy, self-regulation, and critical thinking. While often discussed in tandem with learner autonomy, agency goes beyond independence by emphasizing learners' active participation and influence over learning conditions (Bandura, 2001). In technology-enhanced environments, learner agency is vital for meaningful engagement.

2.2 Providing Comprehensible Input

The Input Hypothesis posits that “people acquire second languages only if they obtain comprehensible input, and if their affective filters are low enough to let the input in.” (Krashen, 1985: 4). However, in the context of large-class instruction in China, the high number of students often makes it difficult for teachers to provide individualized instruction. As a result, both classroom input and after-class assignments frequently fail to align with the actual needs of each learner and struggle to effectively target every student's “zone of proximal development”.

The application of large language models offers a potential solution to the challenges mentioned above. LLMs are not only capable of rapidly generating learning materials, but the materials they produce also possess qualities such as comprehensibility, multimodality, and customizability (Yang, 2024: 580). Comprehensibility ensures that the generated content remains within students' zones of proximal development, providing effective input. It addresses the problem of the uniformity of instructional materials often found in large-class settings. Multimodality accommodates different learning styles: for auditory learners, LLMs can generate more auditory information, while for visual learners, they can produce richer visual representations. Customizability enables personalized support, catering to the individual learning needs of students. LLMs can tailor input based on learners' language proficiency, cognitive styles, and personal interests, thereby lowering affective filters and facilitating second language acquisition. Another notable advantage of LLM-generated materials lies in their authenticity. Here, authenticity refers not to factual accuracy but to greater alignment with native speaker usage. For instance, ChatGPT demonstrates highly proficient English language abilities, sometimes surpassing non-native English-speaking teachers in this regard.

2.3 Providing Opportunities for Meaningful Output

The Output Hypothesis posits that language production facilitates the internalization of linguistic knowledge and constitutes a crucial component of second language acquisition (Swain, 2000: 99). This theory emphasizes that language learning relies not only on abundant input but also on the output. Output is not merely mechanical imitation; rather, it involves the active construction of language within a given context, fostering communicative needs and helping learners identify errors in their language use. Therefore, it promotes linguistic development and internalization. In China, in traditional English teaching models, students have had very limited opportunities for output both in and outside the classroom, leading to the phenomenon of “mute English”. This highlights the critical role of language output in second language acquisition.

Although foreign language teaching is undergoing a transformation that increasingly emphasizes the importance of output, several issues persist. The first is insufficient output. Second language acquisition demands a substantial amount of output while opportunities provided during class are far from adequate; outside the classroom, students often lack a supportive environment for language production. The second issue concerns meaningless output. Mechanical imitation and context-detached production easily lead to meaningless output. Even when efforts are made to artificially situate output within a context, problems may still arise. For instance, if an English writing task requires students to compose a letter to describe a visit to a museum, students who are uninterested in museums or have never visited one may diminish motivation. Consequently, it negatively impacts second language acquisition.

Large language models can offer learners more opportunities for interactive output in context. They are capable of generating language closely aligned with authentic contexts based on specific prompts and can simulate real-world communicative

environments, enabling learners to engage in dynamic dialogue. In doing so, they effectively address the lack of supportive environments for after-class output. With the aid of large language models, each student can have access to a personalized language practice partner.

2.4 Timely Correction of Language Errors

Mistakes and errors are two distinct concepts in language learning. Mistakes can generally be corrected by learners themselves through their own efforts, whereas errors, resulting from insufficient mastery of knowledge, are difficult for learners to recognize and correct without external intervention. If errors are not promptly addressed, they may become fossilized, negatively impacting the process of second language acquisition. Fossilization leads learners to develop incorrect linguistic understandings and hinders their overall language development.

In traditional foreign language instruction, due to the large number of learners, it is often difficult for teachers to provide timely and individualized feedback. Typically, teachers can only address common errors observed in classroom activities or assignments, leaving individual learners' specific needs unattended. As a result, many learners fail to realize the errors they are making. Even when students are aware that something is wrong, they may still struggle to correct complex errors without the teacher's direct intervention.

Large language models can effectively bridge this gap. Through real-time interaction, they are capable of providing personalized and immediate feedback based on each learner's output. They not only assist students in identifying and correcting individual errors but also offer targeted improvement suggestions according to learners' language proficiency levels and specific needs, greatly enhancing both the efficiency and quality of learning.

3. Methodology

3.1 Questionnaire design rationale

This paper employed the method of questionnaire to investigate the basic information, the use of large language models, perceptions and critical Awareness. Here is the questionnaire:

Part I: Basic Information

1.Your year of study:

Freshman

Sophomore

Junior

Senior

2.Your major:

English

Translation

Other foreign language major (please specify: _____)

Non-foreign language major

Part II: Usage

3.Have you ever used large language models (such as ChatGPT, ERNIE Bot, etc.) to assist in foreign language learning?

Yes

No

4.How often do you use large language models for language learning?

Almost every day

Several times a week

Several times a month

Rarely

5. What are your main purposes for using large language models in language learning? (Select all that apply)

Completing assignments

Editing compositions

Improving listening skills

Practicing speaking

Others (please specify: _____)

6. Have you encountered any of the following problems with the content generated by large language models? (Select all that apply)

Inaccurate content

Fabricated information

Illogical reasoning

Never encountered problems

7. How important do you think prompts are for content generation?

(Scale of 1 to 5, where 1 = Not important at all, 5 = Extremely important)

1 2 3 4 5

8. Have you systematically learned how to use large language models?

Never

Occasionally watch short videos

Frequently search for tutorials

Received systematic training

Part III: Perceptions and Critical Awareness

9. Do you think large language models can completely replace professional guidance from teachers?

Yes

No

Not sure

10. When using large language models, you usually:

Directly accept the generated content

Partially refer to it and combine with other resources

Carefully verify the authenticity of the generated content

11. Are you aware of the limitations of large language models (e.g., factual errors, knowledge gaps, etc.)?

Fully aware

Somewhat aware

Slightly aware

Not aware at all

3.2 The detailed information of questionnaire

The second section of the questionnaire investigates students' use of large language models (LLMs) in foreign language learning, focusing on usage frequency, learning tasks, and problems encountered. These aspects correspond to three key theories discussed in the literature review: Input Hypothesis, Output Hypothesis, and the role of corrective feedback. Krashen's Input Hypothesis emphasizes the importance of comprehensible input that matches the learner's developmental level. In large EFL classes, personalized input is often lacking. This section seeks to determine whether students are using LLMs to compensate for this gap by accessing customized and understandable learning content. Swain's Output Hypothesis underlines the value of meaningful language production. Given the limited speaking and writing opportunities in traditional classrooms, the questionnaire explores whether LLMs are being used to generate output in more interactive and contextualized ways. Finally, the section also investigates whether learners are aware of and capable of identifying inaccuracies in model-generated content. This relates to the role of timely and individualized corrective feedback in preventing error fossilization and promoting language development.

The third part of the questionnaire integrates two originally separate dimensions: learners' attitudes toward LLMs and their critical awareness when using them. These dimensions are included to gain a fuller picture of how students position LLMs in relation to traditional teacher roles and how critically they evaluate LLM outputs. This section addresses whether students perceive LLMs as potential replacements for expert guidance, and to what extent they understand the models' limitations—such as factual inaccuracies, hallucinated content, or gaps in knowledge. It also examines whether learners verify LLM-generated content or blindly accept it. This aligns with the discussion in the literature review that highlights the necessity of cultivating user agency and digital literacy to ensure that LLMs support rather than hinder effective language learning.

3.3 Data analysis

3.3.1 Quantitative Analysis

3.3.1.1 General Information

Participants

A total of 44 university students participated in the questionnaire. The majority of participants are third-year students (61.36%), indicating that the sample group generally possesses sufficient learning experience to evaluate the role of LLMs in language learning objectively.

Academic background: Students majoring in foreign languages (including English, translation, and other foreign language majors) account for 79.54%, demonstrating a high level of relevance between respondents and the research topic.

Prevalence and Frequency of Use

Usage rate: A total of 95.45% of students reported having used large language models (LLMs) to support foreign language learning, suggesting widespread adoption and high acceptance among university students.

Usage frequency: Approximately 69.05% of students use LLMs at least several times per week, with 30.95% using them almost daily. This indicates that LLMs have become an integral part of students' daily study routines.

Functional Use Analysis

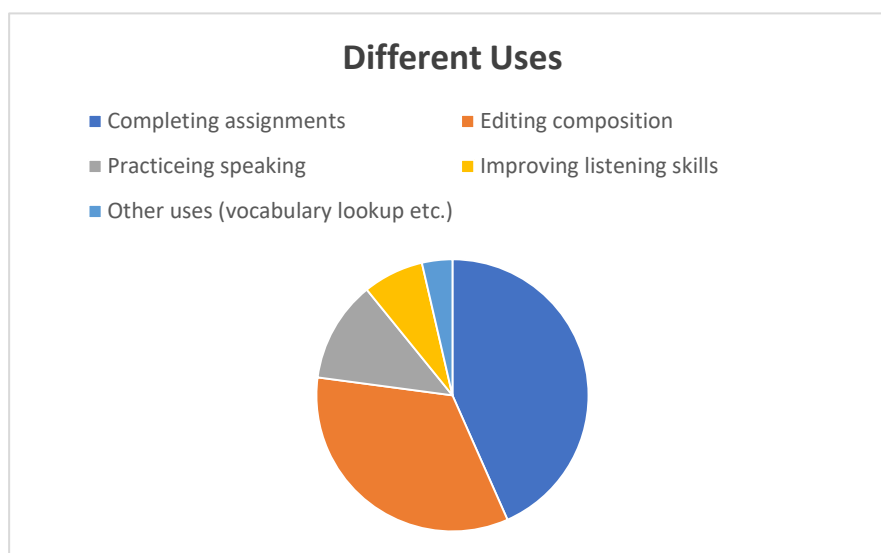


Figure 1 Different Uses

Figure 1 shows that students primarily use LLMs for written output and input-related tasks, reflecting their practical applications in providing comprehensible input and supporting meaningful output, as emphasized in SLA theories.

Output Quality and Feedback Awareness

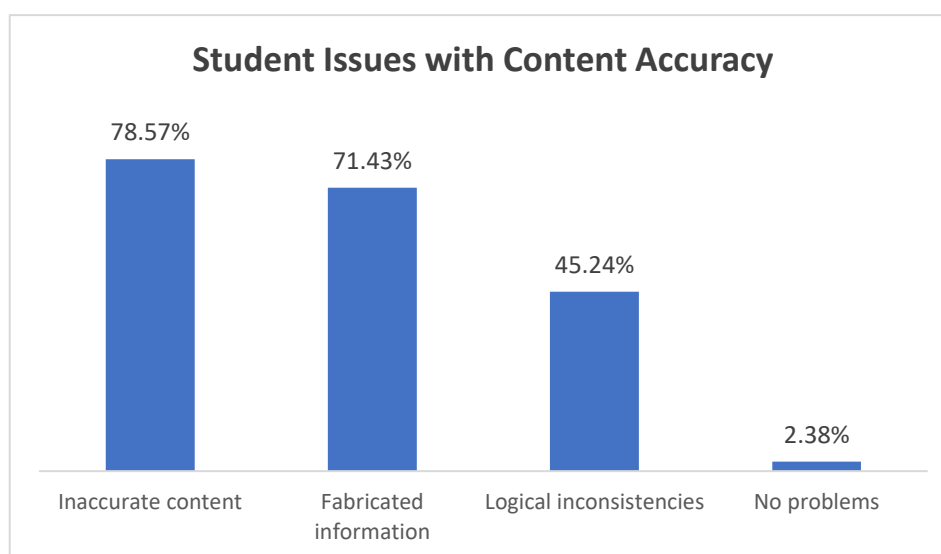


Figure 2 Student Issues with Content Accuracy

Nevertheless, 61.9% of students "partially refer and verify with other resources," while 35.71% "carefully fact-check," suggesting a relatively strong level of critical awareness among users.

Prompt Design and Learning Strategies

The average score for the importance of prompt design was 4.52 out of 5, indicating widespread recognition of the role of well-crafted prompts. However, only 14.28% of students have received systematic training (through formal courses or regular tutorial searches), revealing that most students learn prompt design through self-directed, fragmented means and lack structured support.

Attitudes Toward Replacing Teachers

A total of 90.48% of students disagreed that LLMs can fully replace professional teacher guidance, with only one student responding “yes.” This indicates that students maintain a rational and cautious attitude toward the capabilities of AI.

3.3.2 Qualitative Analysis

Practical-Oriented Use: Focus on Written Tasks

Students most frequently use LLMs to complete assignments and edit compositions. This suggests that many perceive AI mainly as a functional writing assistant, rather than an interactive language partner. Although LLMs can assist in listening and speaking tasks, limitations in platform design (e.g., lack of voice interaction) and habitual user behavior mean multi-modal output functions remain underdeveloped.

General Critical Awareness, but Behavioral Differences

While most students reported encountering issues such as inaccuracy and hallucination, they generally responded with partial trust and cautious verification. This reflects a critical acceptance of AI-generated content. However, a small number of students directly accept outputs without verification, revealing that some users may still lack mature media literacy and exhibit overreliance on technology.

High Recognition of Prompt Design, but Lack of Systematic Strategy Learning

Although most students acknowledged the importance of prompt design (with an average rating of 4.52), very few received formal training. Most rely on watching short videos or informal online tips. This shows that students’ prompt design abilities are still in a trial-and-error stage, lacking conversion into strategic competence or cognitive tools.

4. Results

The survey results show that most students have started using large language models to enhance their foreign language learning. However, the majority of students have not received systematic training. Using large language models effectively has become the key to improving learning outcomes.

Specifically, factors such as the setting of learning goals, the selection of large language models, and the design of prompts directly affect the relevance and quality of the generated content, thereby influencing the effectiveness of large language models in foreign language learning. Therefore, students should learn how to use large language models effectively, so that these models can empower foreign language teaching and increase student engagement in foreign language education.

4.1 Clarifying the Purpose of Using Large Language Models

With the rapid development of artificial intelligence, some foreign language learners have begun using large language models to assist their studies and improve their learning outcomes. However, in practice, learners often make the mistake: they focus primarily on completing tasks while neglecting the improvement of their language abilities. A typical example is that many students rely on large language models to generate essays directly for their writing assignments, overlooking the essential processes of developing writing skills and cultivating critical thinking. Large language models should serve as auxiliary tools for learning rather than replacements for genuine engagement. Their true value lies in supporting learners’ continuous reflection and improvement through practice.

How, then, should foreign language learners set reasonable learning goals? The most fundamental goal should be the improvement of language skills. As the foundation of foreign language learning, language skills, including listening, speaking, reading, writing, and translating, are prerequisites for a deeper understanding of the culture, literature, and broader knowledge embedded within a language. Only by consolidating these basic skills can learners effectively expand their knowledge and achieve deeper levels of foreign language learning.

Another equally important objective is the cultivation of cultural literacy. “With the increasingly widespread application of artificial intelligence technologies, if foreign language education remains limited to the cultivation of language skills, it risks

being replaced by AI, which already possesses varying degrees of comprehensive language ability" (Hu, 2024: 12). Enhancing cultural literacy requires students to develop abilities such as appreciating literary works, mastering area and country-specific knowledge, and understanding cross-cultural differences. Both language proficiency and cultural literacy cannot be achieved overnight; they demand consistent accumulation and long-term engagement. By leveraging large language models, learners can accelerate this process and improve the overall quality of their learning.

4.2 Selecting the Appropriate Large Language Model

Currently, there are various large language models on the market, such as DeepSeek, ERNIE Bot, and ChatGPT. Due to differences in processors and databases, these models display distinct styles. Feng & Zhang (2024: 6) conducted a comparison between ChatGPT-4o and ERNIE Bot 4.0, analyzing their abilities in handling tasks. Their study found that ChatGPT-4o offers more precise, comprehensive, and detailed analyses of lower-scoring essays compared to ERNIE Bot 4.0. Therefore, for foreign language learners seeking to improve their writing skills, ChatGPT-4o may provide a clearer and more thorough understanding of their shortcomings and help them make targeted improvements.

This highlights that learners at different language proficiency levels and with different learning needs should carefully assess large language models and select the one that best suits their specific circumstances. "Performance evaluation of large language models refers to the quantitative and qualitative measurement and assessment of a model's performance on specific tasks. Such evaluations can be categorized into four types: NLG (Natural Language Generation) task evaluation, NLU (Natural Language Understanding) task evaluation, evaluation of other tasks (such as multimodal tasks), and comprehensive task evaluation" (Song et al., 2025: 4). Although foreign language learners may not be able to conduct full quantitative and qualitative assessments, this research provides a useful framework. Learners can classify their tasks into these four categories and select different models accordingly, rather than relying on a single model for all tasks.

4.3 Learning to Optimize Prompts

In actual practice, many learners encounter a common problem: the content generated by large language models often deviates from their expectations. One major reason for this discrepancy lies in the quality of the prompts. "Different prompts yield different answers; more specific prompts are easier for the system to understand and are more likely to produce the desired responses" (Qin, 2023: 27). High-quality prompts should be clear and specific. The more precise and explicit the prompt, the more likely the generated content will align with the learner's actual needs.

Crafting high-quality prompts requires a high level of critical thinking and cognitive ability, as learners with stronger thinking skills are better equipped to pose insightful and targeted questions, guiding the model to generate more valuable content. Developing this ability demands long-term training and practice. A more direct strategy is to increase the specificity of prompts. Even with the same task goal, learners can refine prompts by setting a role (e.g., "You are an IELTS speaking examiner"), specifying task requirements (e.g., "Please summarize the main ideas of this article"), defining a style (e.g., "Please use formal, academic language"), providing a context (e.g., "You are attending an academic lecture"), or adding additional instructions (e.g., "Please point out grammatical errors and explain the reasons"). These techniques make prompts more targeted and actionable, resulting in answers that better meet learners' expectations.

Figure 3 illustrates response generated by the model Doubao based on different prompts. When the prompt was simply "Practice oral English with me," Doubao selected topics more related to daily life. However, when the prompt was refined by specifying a role as an IELTS speaking examiner, clarifying the task for IELTS Part 3, and adding more specific requirements, the answers were much closer to the IELTS speaking exam, with a stronger emphasis on logical coherence.

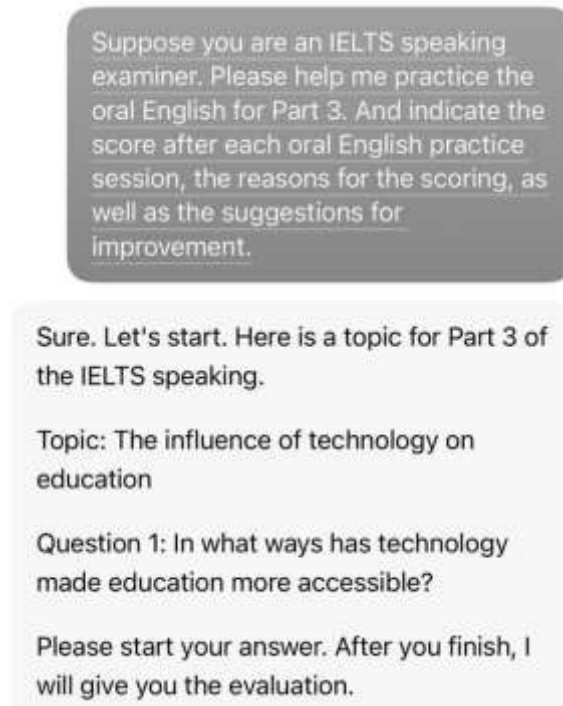


Figure 3 Response Generated by Doubao

5. Conclusion

The rapid development of large language models has created new possibilities and opportunities for enhancing student autonomy. In applying large language models to foreign language education, students should not only understand their basic functions and advantages but also master effective usage strategies to improve learning outcomes. However, the use of technology should never replace the cultivation of critical thinking. While leveraging large language models to support their learning, students must maintain independent thought and rational judgment. By combining AI-generated insights with teacher guidance and diverse perspectives, students can truly master large language models and empower their foreign language learning.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

- [1] Feng, Q., & Zhang, K. (2024). Exploring the Capabilities of Artificial Intelligence in Assisting Foreign Language Teaching and Research: A Case Study of ChatGPT-4o and Wenxin Model 4.0. *Technology Enhanced Foreign Language Education*, (03), 3-12+109. doi:10.20139/j.issn.1001-5795.20240301.
- [2] Hu, K. (2024). Artificial Intelligence and Foreign Language and Literature Discipline. *Foreign Languages in China*, 21 (06), 1+12-16.
- [3] Krashen, S. D. (1985). *The Input Hypothesis: Issues and Implications*. Longman.
- [4] Qin, Y. (2023). Exploration of Foreign Language Education Methods in the Context of Human-Machine Symbiosis: A Case Study on ChatGPT. *Technology Enhanced Foreign Language Education*, (02), 24-29+108.
- [5] Song, J., Zuo, X., Zhang, X., & Huang, H. (2025). A Review of Large Language Model Evaluation Methods. *Journal of Astronautic Metrology and Measurement*, 45 (02), 1-30.
- [6] Swain, M. (2000). The Output Hypothesis and Beyond: Mediating Acquisition through Collaborative Dialogue. In James, P. L. (Ed), *Sociocultural Theory and Second Language Learning* (pp. 97–114). Oxford University Press.
- [7] Yang, L. (2024). Second Language Acquisition in the Context of Large Language Model of ChatGPT. *Modern Foreign Languages*, 47 (04), 578-585.
- [8] Zhang, R. (2023). The Impact of Generative Artificial Intelligence Technology on Education: An Interview on ChatGPT. *e-Education Research* 44 (02), 5-14.
- [9] Zhong, B., Shang, J., Wang, J., Han, Y., Liu, J., Zou, H., & Wang, Z. (2023). ChatGPT's Challenges to Education (Written Talk). *Chongqing Higher Education Research*, 11 (03), 3-25. doi:10.15998/j.cnki.issn1673-8012.2023.03.001.
- [10] Ministry of Education of the People's Republic of China. (2022). *Compulsory Education English Curriculum Standards* (2022 ed.). Beijing Normal University Press.