Computer Life
ISSN: 1819-4818 | Vol. 12, No. 3, 2024

Exploring the Efficacy of a ChatGPT-Based Application in Enhancing Oral English Proficiency for Chinese ESL Students

Weiyi Guo *, Difei Li

School of Education, Johns Hopkins University, Maryland, USA
* **Corresponding author**: Weiyi Guo (Email: wguo22@jh.edu)

Abstract: This research explores the impact of Chat GPT AI Characters on ESL learning, comparing them to traditional peer practice in language education. It investigates AI's influence on language proficiency, engagement, and response quality, drawing from literature on ESL challenges. Using mixed methods—pre- and post-test questionnaires, classroom experiments—the study evaluates two groups' language fluency, response speed, and answer quality. Thematic analysis uncovers student perspectives, while statistical methods like t-tests and correlation studies highlight group differences. Ethical considerations prioritize privacy and consent. Anticipated outcomes include improved language skills and engagement. Acknowledging limitations in sample size and cultural biases, this study illuminates AI's role in ESL education, informing future teaching practices.

Keywords: ESL; AI Integration; Language Learning.

1. Introduction

English as a Second Language (ESL) education has witnessed a transformative evolution with the integration of Artificial Intelligence (AI) technology. In recent years, AI has emerged as a powerful tool, offering innovative solutions to enhance language learning experiences [5]. The intersection of AI and ESL instruction has sparked considerable interest among educators, researchers, and learners alike. This introduction provides a brief background on the integration of AI technology in ESL teaching, highlighting the potential benefits and challenges it brings to the forefront.

Traditionally, ESL instruction has relied on conventional teaching methods, emphasizing face-to-face interactions, textbooks, and language laboratories. However, the dynamic landscape of education is now marked by the proliferation of AI technologies, reshaping the way language learning is approached [6]. The integration of AI in ESL classrooms is driven by the recognition of its ability to cater to diverse learning styles, offer personalized feedback, and provide immersive language experiences [4].

One significant application of AI in ESL education is the utilization of language processing models, such as Chat GPT, to simulate real-life conversations. These models leverage natural language understanding and generation, allowing learners to engage in interactive dialogues that mimic authentic communication scenarios. The adaptability of AI technologies enables tailored learning experiences, catering to the unique needs and proficiency levels of individual students.

As educators explore the potential of AI in ESL teaching, questions arise regarding its efficacy in comparison to traditional methods. This research delves into the specific inquiry: Is English-speaking practice with Chat GPT AI Characters more effective through peer practice in ESL-speaking classes? To address this question, the study incorporates theoretical frameworks focused on engagement, language proficiency, and response quality, aiming to contribute valuable insights to the ongoing discourse on the role of AI in ESL education.

2. Research Aims and Objectives

The primary aim of this research is to investigate the effectiveness of English-speaking practice with Chat GPT AI Characters in ESL-speaking classes, comparing it with traditional peer practice methods. The study seeks to delve into the impact of Chat GPT-based activities on students' engagement in ESL speaking activities, examining whether the novel technological approach fosters increased interest and participation. Furthermore, the research aims to assess the influence of Chat GPT AI Characters on language proficiency development in ESL learners, considering both theoretical frameworks and practical outcomes. The study also aims to analyze the response rate and quality of feedback provided by Chat GPT in simulated oral communication scenarios, shedding light on the potential benefits and challenges associated with AI-assisted language learning. Finally, the research endeavors to explore the attitudes and perceptions of ESL students towards the integration of AI technology in language learning, providing valuable insights into the student experience and acceptance of innovative teaching methodologies in the ESL context. (Figure 1)

3. Literature Reviews

The integration of Artificial Intelligence (AI) technology into English as a Second Language (ESL) teaching has garnered considerable attention in recent years. As communication skills, especially spoken language proficiency, are integral components of language learning, this literature review aims to explore the existing research on the utilization of AI in enhancing ESL speaking skills.

The article “English speaking practice with conversational AI: Lower secondary students’ educational experiences over time” written by Elin Ericsson and Stefan Johansson [1] lists three dominant challenges for second language (L2) learners to improve ESL speaking skills: “the complexity of speaking skills, the practical and physical constraints of practicing

speaking with native speaker partners and receiving feedback in formal education." These three challenges not only illustrate the major dilemma faced by ESL learners in English-speaking environments but also implicitly address the question that will be elaborated upon in this paper—why it is necessary to introduce new technologies such as AI into ESL-speaking instruction. To be more specific, the article states that the introduction of AI technology into English-speaking teaching can activate students' both explicit and implicit knowledge. The explicit knowledge referred to here mainly points to grammar and vocabulary, and implicit knowledge refers to "the ability to interact, including planning, encoding, and producing utterances and simultaneously interpreting the interlocutor's utterances in real time" [1].

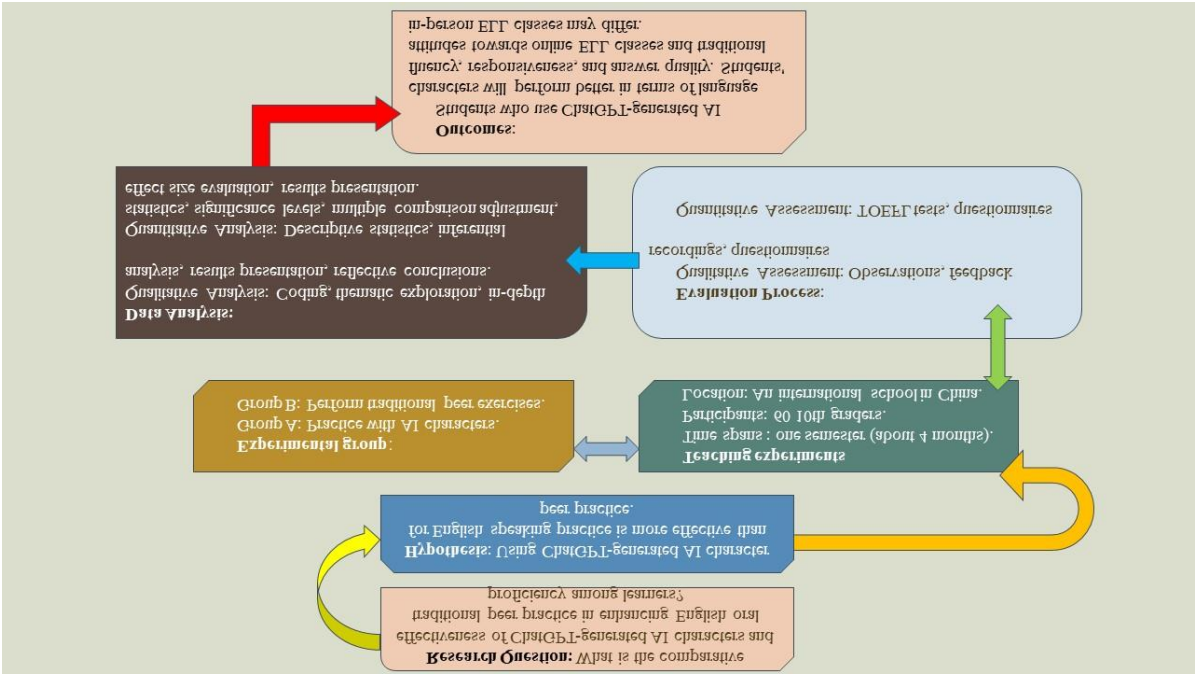

Figure 1. The Relationship among the Research Question, Hypothesis, Experiment Group, Teaching Experiment, Evaluation Process, Data Analysis, and Outcomes

Since one theory involved in this paper is associated with the higher-class engagement associated with AI technology-based ESL classrooms, previous studies associated with testing AI classroom engagement levels are critical to this paper's discussion. As evidenced by the rapid development of AI technology, there has been a growing interest in collaboration between AI technology and teaching. The creation of courseware is an important part of how class engagement is achieved in the collaboration between AI technology and teaching [4]. During the teaching experiment with the courseware designed by AI technology, the teachers kept adjusting the courseware designed by AI technology based on the latest student engagement data received. The result shows that the courseware created by AI technology does show a more consistent engagement among students and the graphs involved in the courseware created by AI technology has a crucial role in the improved class engagement.

Furthermore, recently China has invested a lot in inventing AI technology that can prominently facilitate education. Since English speaking is one of the most essential parts of English teaching, developing AI technology that can be used in English-speaking teaching is meaningful. Zoubin, a professor in Xi'an Jiaotong Liverpool University, has developed an artificial intelligence oral English evaluation system called EAP talk. According to the data provided by Wang, Zou, & Xue [5], most participants feel engaged and fresh about the new way of learning built on AI technology. The study testing the effectiveness of EAP talk utilizes the tracking experiment instead of controlling participants in a closed classroom environment. Therefore, most of the data cannot be supervised or collected directly due to the untraditional experiment environment. Under such an environment, systematic data analysis of questionnaires and interview results become an important resource for researchers, and researchers need to spend more time on getting data from participants instead of experiment observers or other recording methods.

4. Research Methodology

This section of our study comprises two essential components. The first, Research Design, emphasizes the distribution of questionnaires before and after the experiment. These questionnaires aim to gather both quantitative and qualitative data regarding participants' attitudes towards traditional English Language Learning (ELL) classrooms and the integration of Artificial Intelligence (AI) in language learning. The second component, Data Analysis, outlines the methods employed for interpreting this gathered data.

This includes statistical analysis to identify trends and thematic analysis for a deeper understanding of participant perspectives, thereby providing comprehensive insights into the effects of the teaching methodologies used.

4.1. Research Design

Research Design consists of two parts: questionnaire, which deals with the creation and application of pre-test and post-test questionnaires to gather both quantitative and qualitative data on student perceptions about traditional ELL classrooms and AI in language learning; and teaching experiment, outlining the practical execution of our study, including teaching methods, participant groups, and overall

implementation.

4.1.1. Questionnaire

Questionnaire includes three parts: the pre-test questionnaire to assess initial attitudes and expectations, the post-test questionnaire to evaluate changes in perceptions post-experiment, and questionnaire development.

Before the experiment, the pre-test questionnaire can be divided into different parts below:

(1) Quantitative Part: Utilizes Likert scale to assess participants, specifically students' initial attitudes and expectations towards traditional ELL classrooms, and their openness to using AI for language learning. Includes closed-ended questions on topics such as study habits, key elements of effective oral practice, frequency of previous technology use in language learning, and general attitudes and expectations towards language learning and technology application.

(2) Qualitative Part: Open-ended questions will explore participants' personal preferences for oral practice methods and their expectations and potential concerns about using AI characters as a tool for language learning.

(3) Question Design: In designing the questionnaire, we ensure that the questions are concise, clear, and easily understandable, avoiding technical jargon or complex structures. The design aims to facilitate participation from all students.

(4) Data Analysis: Quantitative data collected from the questionnaire will provide baseline data for subsequent analysis, while qualitative data will offer background information and a deeper understanding of the sample for the researcher's interventions in the study. By combining these qualitative and quantitative elements in the pre-test questionnaire, researchers can obtain a comprehensive view of the participants' initial attitudes and expectations, providing important background information for comparing changes before and after the experiment.

After the experiment, the post-test questionnaire can be divided into different parts below:

(1) Quantitative Part: Uses Likert scale to quantify participants' attitude changes post the teaching experiment, including aspects like language fluency, confidence in oral practice, response speed, classroom participation, and overall satisfaction with AI character practice generated by ChatGPT. Includes closed-ended questions, such as multiple-choice and true/false questions, to quantify participants' feedback.

(2) Qualitative Part: Consists of a series of open-ended questions asking participants about their experiences with using AI character and traditional peer practice, especially regarding their impact on language learning, perceived advantages and disadvantages, and any other detailed feedback. These questions will reference the structure of the Oxford Strategies for Language Learning (SILL) [3], but will be appropriately modified to fit the specific needs of this study.

(3) Question Design: Following best practice in questionnaire design [2], we will ensure that open-ended questions in the questionnaire are clear and specific, encourage detailed feedback, and avoid leading questions.

(4) Data Analysis: Quantitative data will be analyzed through statistical methods to identify significant trends and patterns. Qualitative data will be evaluated through thematic analysis to capture richer data and understand participants' personal experiences and viewpoints.

By combining quantitative and qualitative methods, the post-test questionnaire can provide not only statistical evidence about learning outcomes but also reveal deep insights into participants' feelings and suggestions regarding the experimental teaching methods, offering a more comprehensive perspective for the study. For the questionnaire development, the pre-test questionnaire can be divided into different parts below:

(1) Development of Questionnaire Items: In line with SILL and best practice in questionnaire design, we will develop questionnaire items and organize the questionnaire to ensure that each section has clear guiding objectives and structure.

(2) Pilot Testing: The questionnaire will undergo a pre-test to verify the validity of the questions. Necessary revisions will be made based on the feedback received.

(3) Use of Questionnaire Data: The data gathered from the questionnaire will be used to compare the impact of the teaching experiment on participants. It will also help to further document and demonstrate the effects of the two ELL classroom formats on participants.

4.1.2. Teaching Experiment

Purpose and hypothesis: This experiment aims to evaluate the difference in the effectiveness of English-speaking practice using AI characters generated by ChatGPT and traditional peer practice in ELL teaching. We hypothesized that Group A using AI characters would show more significant improvements in language fluency, response speed, and answer quality than Group B using traditional peer practice.

Participants: The experimental subjects were 60 grade 10 students in four classes, all from an international school in China. The four classes will be divided into two experimental groups, and the researchers will try to ensure that the average level of students in each group is similar in terms of language ability, age, and gender.

Experimental design: The experiment will be conducted over one semester, lasting approximately four months. Group A will use AI characters as auxiliary tools for English-speaking practice, while Group B will conduct traditional peer practice.

Interventions: Group A: Students will conduct regular speaking practice through AI characters under the guidance of teachers. After practicing, students will receive AI-based feedback to enhance learning. Group B: Students will engage in traditional role-playing and conversational exercises with peers, with feedback provided by the teacher.

Assessment method: Qualitative Assessment: The quality of student engagement and interaction during the exercise will be assessed through classroom observations and student feedback recordings.

Quantitative assessment: TOFEL standardized speaking tests and questionnaires will be used to quantify changes in students' language fluency, response speed and answer quality.

Blind assessment: When evaluating a student's performance, the rater does not know which experimental group the student belongs to. This helps ensure fair and consistent scoring and reduces subjective bias.

Data Collection: Group A: The computer automatically stores each exercise performed by students and performs certain automatic analysis to provide assistance for subsequent analysis. Group B: Teachers will record the length, frequency, and student performance of each exercise. Students' speaking performance will be audio-recorded for subsequent analysis. Before and after the experiment, students will fill out questionnaires to assess their attitudes toward various teaching methods and their personal language

learning progress.

4.2. Data Analysis

Data will be analyzed using a mixed methods approach, combining qualitative descriptions and quantitative statistics to provide a comprehensive view of the effectiveness of teaching methods. We will use appropriate statistical methods to analyze quantitative data, while qualitative feedback will be coded and thematically analyzed.

4.2.1. Qualitative Data Analysis

The data collection and organization can be divided into parts below:

Data Sources: We will ensure the collection of rich qualitative data from classroom observations, student feedback, and open-ended questions in questionnaires. Organization and Categorization: Collected qualitative data such as recordings of oral practices and student feedback notes will be organized and transcribed into an analyzable format.

The Coding and Thematic Analysis can be divided into parts below:

Preliminary Coding: We will read through the data, preliminarily marking or coding significant words, phrases, or sentences. Theme Identification: Through the coding process, we plan to identify important themes, patterns, or concepts. Iterative Review: We will repeatedly review and adjust the themes to ensure they accurately reflect the content of the data.

The In-depth analysis can be divided into parts below:

Detailed Interpretation: We will conduct a thorough analysis of each theme to understand its underlying meaning and context. Comparison and Contrast: We plan to compare responses from different student groups to identify varying perspectives and experiences.

The results presentation can be divided into parts below:

Citing Examples: We will cite specific examples in the analysis report to support the conclusions of thematic analysis. Visual Display: We intend to use charts or diagrams to aid in presenting the results of the qualitative analysis.

For conclusions and reflection, we will present the synthesized Insights and conclusions drawn from the qualitative data. Also, we will discuss the limitations of the methods and their potential implications for future research.

4.2.2. Quantitative Data Analysis

Quantitative data analysis covers a range of processes including data preparation and preprocessing for ensuring accuracy, descriptive statistics to provide an initial overview and measure tendencies, inferential statistics for drawing conclusions, significance level and multiple comparison correction to validate statistical significance, and effect size assessment and results presentation to quantify impacts and clearly convey findings.

Data Preparation and Preprocessing involves two parts:

Verifying Data Integrity: We will ensure there are no missing or erroneous data in the dataset, and perform necessary data cleaning. Data Transformation: We plan to convert the raw data into a format suitable for analysis, including standardization and normalization processes.

Descriptive Statistics involves four parts:

Dataset Overview: We will record the sample size and basic characteristics of the subjects, to assess the representativeness of the sample and fully understand its characteristics. Measuring Central Tendency: We will calculate average values, medians, and modes to understand the performance level of various indicators in the dataset. Measuring Variability: We will analyze standard deviation and coefficient of variation to understand the variability or consistency of data across different indicators. Data Distribution: We plan to use histograms, box plots, and quartiles to provide a visual representation of data distribution and to identify any skewness or outliers.

There are four methods are used in the Inferential Statistics:

T-test: We will compare the mean differences between Group A and Group B on various indicators, and clearly propose and verify related statistical hypotheses. Effect Size: We will calculate and interpret Cohen's d value to assess the actual importance of the effect. Correlation Analysis: We plan to use Spearman correlation coefficients to assess the degree of correlation between different variables, performing necessary hypothesis testing. Regression Analysis: We will use regression analysis to predict the impact of continuous predictive variables on key outcome variables, strictly following the related hypotheses.

For Significance Level and Multiple Comparison Correction, we will set an appropriate level of significance and apply methods like Bonferroni correction during multiple testing to reduce the risk of false positives.

While for Effect Size Assessment and Results Presentation, we will assess the actual importance of the relationships between variables based on the results of statistical tests, and we plan to use charts to clearly present key data and provide a detailed interpretation of the data analysis results. Finally, we will write conclusions and discuss the limitations of the methods and their impact on future research.

5. Discussion

In this study, we first ensured that the privacy of all participants was strictly protected, and participants' personal information would not be disclosed or used for purposes other than research without consent. Particularly when processing audio or text feedback, researchers will delete and mask any potentially personally identifiable information. In terms of data security, researchers will take strict measures to protect research data from unauthorized access or leakage.

Secondly, we adhere to the principle of informed consent to ensure that all participants (or their legal guardians) can fully understand the purpose, methods, potential risks and benefits of the study, and participate voluntarily. Each participant will sign a written consent form. Participants have the right to withdraw from the study at any time. The researchers ensured that participants' decisions would not result in any adverse consequences or penalties.

In this research, the participants are mainly students and teachers, including minors or other vulnerable groups. We will particularly protect their rights and interests to ensure fairness and respect in the research process. We will consider the impact of research interventions on students' daily learning, minimize disruption, and handle research results cautiously to avoid negative impacts on students' self-confidence or motivation to learn.

Research results will be disseminated responsibly, ensuring the accuracy and appropriateness of the information. In addition, we will deal transparently with any potential conflicts of interest to ensure the integrity of the research.

6. Conclusion

For the expected outcomes, we expect that this study will

reveal the differences in the English-speaking learning effects and students' attitudes using two different methods, ChatGPT and traditional peer practice. Specifically, we expected to find that practice via ChatGPT significantly improved students' language fluency, response speed, and answer quality. Furthermore, we anticipate that the use of this new technology will increase student engagement and motivation, thereby bringing about positive changes in language education. Ultimately, this study aims to explore the application value and potential limitations of ChatGPT as a new educational technology in educational environments, and to provide insights and directions for future educational technology practices.

Despite the comprehensive approach have used in this study, there are specific constraints we must recognize to better interpret our findings.

1) Narrow sample size and study scope

The study was conducted at a specific grade level and in a specific setting, which may limit the widespread application of the findings. In addition, we acknowledge that the sample size of this study was insufficient to cover the diversity of students.

2) Experiments are affected by background noise

In this study, the experimental design may not adequately consider all variables in the learning environment, and some experiment-induced variables may be difficult to eliminate. For example, the teaching style of different teachers, the different teaching methods caused by experiments, etc., may affect student engagement, etc., which may affect learning outcomes.

3) There is a subjective bias in the data collection methodology

The data collection methods we employ, such as self-report questionnaires, may be influenced by participant subjectivity, which can affect the objectivity and reliability of the data.

4) The interpretation of the results is not fully considered

Other possible interpretations of experimental results may not be fully considered when interpreting the data, such as a new teaching method based on AI characters, which may arouse students' interest and encourage them to learn actively.

5) Research is culturally and educationally dependent

The sample size of the studies was not diverse enough, which may have relied to some extent on specific cultural and educational backgrounds, limiting their universal applicability.

6) There are technical implementation challenges and technical dependencies

We recognize that the application of AI technology in practical teaching and learning may present challenges with resources and technical feasibility. In addition, there may be a risk of over-reliance on technology in actual teaching, ignoring the impact of individual differences between teachers and students.

7) Long-term effects and subsequent effects were not tracked

This study failed to assess the long-term impact of AI-assisted teaching on students' learning outcomes, nor did it assess students' knowledge retention and skill transfer after AI-assisted teaching.

References

[1] Ericsson, E., & Johansson, S. (2023). English speaking practice with conversational AI: Lower secondary students' educational experiences over time. Computers and Education. Artificial Intelligence, 5, 100164-.

https://doi.org/10.1016/j.caeai.2023.100164

[2] Imperial College London. (n.d.). Best practice in questionnaire design. Retrieved from https://www.imperial.ac.uk/education-research/evaluation/tools-and-resources-for-evaluation/questionnaires/best-practice-in-questionnaire-design/#:~:text=Best%20practice%20in%20questionnaire%20design,Presser%2C%202010%3B%20Schwarz%2C%201999

[3] Oxford, R. L. (1990). Language learning strategies and beyond: A look at strategies in the context of styles. Shifting the instructional focus to the learner, 35, 55.

[4] Schroeder, K. T., Hubertz, M., Van Campenhout, R., & Johnson, B. G. (2022). Teaching and Learning with AI-Generated Courseware: Lessons from the Classroom. Online Learning (Newburyport, Mass.), 26(3), 73-.

https://doi.org/10.24059/olj.v26i3.3370.

[5] Wang, W., Zou, B., & Xue, S. (2023). AI technology used as a tool for enhancing. university students' English speaking skills: perceptions and practices. 12779, 1277917-1277917–10.

https://doi.org/10.1117/12.2689728.

[6] Zafar, A. (2006). Traditional and Modern Approaches in ESL Teaching at Different. Levels in Pakistan. International Journal of Learning, 13(7), 61–66.

https://doi.org/10.18848/1447-9494/CGP/v13i07/44931.