

Enhancing English Proficiency Through AI Conversations: The Impact of ChatGPT on TOEIC and Speechace Performance in a Project-Based Learning Context

Jun Sakaue¹ & Tsukasa Yamanaka²

¹ Language Education Center, Ritsumeikan University, Kyoto, Japan

² College of Life Sciences, Ritsumeikan University, Kusatsu, Japan

Correspondence: Tsukasa Yamanaka, College of Life Sciences, Ritsumeikan University, Kusatsu, Shiga, 525-8577, Japan. E-mail: yaman@fc.ritsumei.ac.jp

Received: January 27, 2025

Accepted: February 25, 2025

Online Published: February 27, 2025

doi: 10.5539/elt.v18n3p70

URL: <https://doi.org/10.5539/elt.v18n3p70>

Abstract

This study examined the impact of integrating ChatGPT into a project-based English learning course on students' English proficiency, as assessed by TOEIC Listening and Reading scores and Speechace metrics. The participants, 110 first- and second-year university students, were divided into Homework and No-Homework groups, with the former required to engage in weekly English conversations with ChatGPT. The results revealed that while the Homework group demonstrated significant improvements in several Speechace categories, particularly among second-year students, no significant correlation was found between the number of completed assignments and score improvements. Additionally, TOEIC scores indicated greater progress in Listening and Reading among the Homework group compared to the No-Homework group. Notably, the reliance on text during ChatGPT conversations possibly explained the stronger impact on Reading and Vocabulary, while Fluency and Pronunciation also improved. However, an increase in unnatural pauses highlighted the need for further refinement in task design to encourage more natural conversational dynamics. This study underscored the potential of AI tools like ChatGPT to enhance English communication skills while identifying areas for improvement in their pedagogical application.

Keywords: ChatGPT, conversation, project-based English program, speechace, TOEIC

1. Introduction

1.1 Introduce the Problem

In pursuit of the goal of cultivating “English-speaking Japanese” (Eigo ga Tsukaeru Nihonjin), efforts are ongoing to enhance English education in Japan, particularly focusing on communication skill development. The new curriculum guidelines that the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT) introduced in April 2021 explicitly highlight the objective of fostering the ability to engage in actual communication in English (MEXT, 2018). However, the achievement of this objective remains elusive. For instance, a 2023 MEXT survey revealed that approximately 60% of third-year junior high school students scored zero on a speaking test, despite efforts to encourage active use of English in the classroom (MEXT, 2023). These findings underscored the insufficiency of classroom-based approaches alone in developing communication skills effectively.

Various initiatives, such as the project-based English program (PEP) the authors introduced, have been implemented to address these challenges. In PEP, students explore topics of personal interest, conduct research or experiments, and present their findings through academic presentations, aiming to enhance their English proficiency through active classroom engagement (Yamanaka et al., 2021). PEP actively integrates AI tools to achieve more efficient instruction. Studies have shown that this approach is effective in enhancing educational outcomes (Kimura, 2024; Yamashita et al., 2024). This aligns with global trends in leveraging technology for language education to enhance learning outcomes.

Another promising avenue involves leveraging artificial intelligence (AI) to support English learning. Advances in AI technology, particularly conversational voice interactions, have prompted researchers to investigate its potential for enhancing communication skills. While initial studies show encouraging results, their number remains limited, and definitive conclusions are yet to be drawn. In Japan, AI tools have primarily been utilized for writing instruction, with their impact on communication skills largely unexplored.

We aim to investigate how engaging in English conversation with ChatGPT affects students' English proficiency in the context of PEP classes. Specifically, we examine changes in students' performance using two evaluation metrics: TOEIC scores, which assess listening and reading proficiency; and Speechace scores, which evaluate speaking skills. By analyzing these metrics, we seek to provide insights into the effectiveness of AI-driven conversational activities in improving English communication skills. Ultimately, this research contributes to the broader discussion on leveraging AI in language education and aligns with ongoing national efforts to produce "English-speaking Japanese."

2. Literature Review

2.1 *ChatGPT and Its Role in Conversational Learning*

ChatGPT, developed by OpenAI, is a large language model designed to generate human-like text in response to user prompts (OpenAI, n.d.). Its advanced natural language processing capabilities enable it to engage users in interactive dialogues, making it a valuable conversational learning tool. By simulating human conversation, ChatGPT provides learners with opportunities to practice language skills in a controlled environment. This interaction can enhance learners' communicative competence by allowing them to experiment with language use, receive immediate feedback, and build confidence in their conversational abilities.

Recent research has examined ChatGPT's application in educational contexts, highlighting its potential to support language acquisition and conversational practice. ChatGPT's adaptability allows it to address diverse learning needs and proficiency levels, offering personalized support that traditional teaching methods may struggle with (Amin, 2023). Monica and Suganthan (2024) demonstrated that ChatGPT significantly enhances ESL learners' skills across multiple areas, including vocabulary, listening, speaking, reading, and writing, providing an efficient and time-saving tool for language acquisition in English classrooms. Similarly, Kasneci et al. (2023) discussed the opportunities for and challenges of integrating large language models into classroom settings, emphasizing the importance of pedagogical alignment and active teacher involvement to maximize AI's educational benefits.

Further research by Pratiwi et al. (2024) identified the strengths of ChatGPT's voice conversation features, noting its ability to provide personalized and interactive instruction through advanced natural language processing and speech recognition. However, the study also highlighted certain limitations, such as difficulties in accurately assessing pronunciation variability and interpreting culturally nuanced language use.

Although studies generally agree that ChatGPT can enhance learners' fluency (Hatmanto & Sari, 2023; Nhu, 2024; Zhou, 2023), findings about its impact on pronunciation remain inconsistent. Nhu (2024) found minimal improvement in pronunciation among EFL learners, whereas Hatmanto and Sari (2023) reported significant gains in pronunciation accuracy. These mixed results underscored the importance of leveraging AI tools like ChatGPT to enhance conversational learning experiences while acknowledging their limitations. Future research should explore ways to address these shortcomings, particularly in the areas of pronunciation and cultural context comprehension.

2.2 *Speechace*

Speechace is an automated speech evaluation tool widely used for assessing English-speaking proficiency. According to the official Speechace website (<https://www.speechace.com>), it employs advanced speech recognition technology to analyze various aspects of learners' oral performance, providing comprehensive feedback across multiple metrics.

Research highlights the effectiveness of Speechace in enhancing speaking skills. For instance, Ningsih (2024) found that Speechace significantly improved pronunciation, fluency, vocabulary, grammar, and overall speaking competence, with female students outperforming male students. Similarly, Mutiara et al. (2024) reported that the use of Speechace improved students' pronunciation through its feedback mechanism. Zainuddin and Mohamad (2024) further demonstrated that the tool's automated feedback supports the development of key speaking skills, including pronunciation and fluency, among preuniversity ESL students.

These studies underscore the potential of Speechace in advancing speaking proficiency, but limitations remain. The tool primarily focuses on lower-level components, such as pronunciation and grammar, and may not fully

address higher-order pragmatic skills or cultural nuances. Additionally, research on Speechace in Japan, particularly its integration with project-based English education or tools like ChatGPT, is scarce. Educators are encouraged to use it alongside other methods for a more holistic evaluation of learners' communicative abilities.

2.3 Research on English Communicative Competence in Japan and Its Challenges

In Japan, the development of English communicative competence has been a focal point of educational reform. Despite efforts to improve English education, challenges persist in achieving desired proficiency levels among learners. A MEXT report showed that while improvements have been observed, particularly in reading and listening skills, there is a need for further enhancement in speaking and writing abilities to foster comprehensive communicative competence (MEXT, 2019). Factors contributing to these challenges include limited opportunities for authentic language use, traditional grammar-translation teaching methods, and a lack of emphasis on interactive communication in the classroom.

Further, societal attitudes toward English learning in Japan may hinder the development of practical language skills. English is often viewed as an academic subject rather than a communicative tool, with emphasis on grammatical accuracy and test performance over real-world usage. This perception, reinforced by Japan's examination-oriented education system, prioritizes reading and writing for standardized tests, while speaking and listening receive comparatively less attention. To address these issues, researchers have advocated for the integration of communicative language teaching approaches, increased exposure to authentic language environments, and technology utilization to create interactive learning experiences. For instance, Sakai (2023) investigated ChatGPT's integration into language learning, finding that it improved learners' performance in tests and writing tasks. While highlighting its potential, Sakai noted limitations in addressing complex language acquisition and emphasized the need for educators to critically evaluate the tool's use and stay updated on technological advancements. Moreover, Butler and Iino (2005) analyzed Japan's English education reforms, identifying the structural challenges of large class sizes and limited teacher training as barriers to fostering communicative competence. These studies indicated that while Japan has made strides in improving English education, a more holistic approach incorporating pedagogical innovations and AI tools may better address existing challenges.

3. Method

3.1 Participants

The participants in this study were 110 students (33 first-year and 55 second-year students) enrolled at a large private university in Japan. The first-year students were divided into two classes, while the second-year students were divided into four classes, all taught by one of the authors. No specific criteria were applied to narrow the selection of participants; however, students who failed to complete any of the required tests during the study period were excluded from the analysis. The students attended English classes twice a week: one focused on the PEP and another dedicated to foundational skills workshops aimed at improving their reading, listening, speaking, writing, and grammar skills.

The PEP courses spanned from the spring semester of the first year to the fall semester of the second year. The instructor referred to the courses as Project English 1 through 4 (hereafter, P1 to P4) for each semester. For students in the College of Life Sciences and the College of Pharmaceutical Sciences, Junior Project 1 (hereafter, JP1) was also a required course in the first semester of their third year. In P1 and P2, students became familiar with academic presentations, while in P3, students participated in group activities such as debates and panel discussions. The goal of P4 was to enable students to write an academic paper of approximately 15 pages as well as give oral presentations. In JP1, students addressed specialized topics, received feedback from professors and instructors in their respective fields, and presented their work in the form of poster presentations modeled on international academic conferences.

3.2 Procedure

We conducted this experiment during the spring semester of the 2024 academic year. We divided first- and second-year students into two groups: a Homework Group and a No-Homework Group. Only one class from each year level was designated as the No-Homework Group. As a result, there were 16 first-year students in the Homework Group and 17 of them in the No-Homework Group, whereas there were 42 second-year students in the Homework Group and 13 of them in the No-Homework Group.

The Homework Group was assigned to engage in English conversation with ChatGPT for at least 1 hour per week, over 12 weeks (excluding the first orientation session, the final session, and one other week). To confirm the homework was completed, students were required to submit conversation scripts, which ChatGPT then

assessed to verify whether the conversations lasted for 60 minutes or more. If it determined that the conversations were too short, the homework was marked incomplete. Students were allowed to choose their conversation topics without restrictions. Currently, the voice conversation feature can be used for free by downloading the ChatGPT smartphone application. Therefore, students downloaded it and completed their homework.

We measured the effects on homework of using Speechace and the TOEIC Listening and Reading (TOEIC L&R) tests. For Speechace, students took the test at the beginning of the semester in April and at the end of the semester in July, receiving scores for each category: Overall, Pronunciation, Fluency, Vocabulary, Grammar, Average Words Per Minute (WPM), and Average Bad Pauses. The growth in scores from April to July was used to evaluate the effectiveness of the homework. The “Overall” score reflected the learner’s general speaking proficiency, while “Pronunciation” measured the accuracy of sounds and stress patterns. “Fluency” evaluated the natural flow of speech, including speed and smoothness, and “Vocabulary” assessed the appropriateness and variety of words used during the speech. Additionally, “Grammar” analyzed sentence structure and grammatical accuracy, “Average WPM” tracked speaking speed to ensure it fell within a natural range, and “Average Bad Pauses” identified interruptions or unnatural pauses that disrupted speech coherence. In Speechace, scores are presented in a format converted to other test standards. For example, they may be displayed as A1 or B1 levels according to the CEFR scale. In this study, we used the TOEIC-converted scores, where each category was shown on a 200-point scale, for all analyses.

Regarding TOEIC L&R, the university conducts institutional group testing in June and December. Because no pre-experiment TOEIC data were available for first-year students, the analysis was limited to second-year students. The relationship between homework and improvement in TOEIC scores was assessed by comparing the scores from December 2023 to June 2024.

More specifically, we first analyzed whether there were differences in the growth of Speechace scores between the Homework and No-Homework Groups. Additionally, within the Homework Group, we examined the correlation between the number of completed homework assignments and the growth in scores. Next, focusing on second-year students, we investigated whether there were differences in the growth of TOEIC scores between the Homework and No-Homework Groups. Finally, we divided participants into three groups—upper, middle, and lower—based on their TOEIC scores and explored the correlation between these categories and the growth in Speechace score for each.

4. Results

4.1 Results of Homework and TOEIC

Among the Homework Group, the average number of completed assignments was 10.4 (achievement rate: 86.7%) for the 16 first-year students and 10.3 (achievement rate: 85.8%) for the 42 second-year students, indicating that most students successfully completed their weekly homework.

Regarding TOEIC scores, for first-year students, the total Listening and Reading scores in the June 2024 test averaged 490.9 points for the Homework Group and 446.5 points for the No-Homework Group. For second-year students, the average total score in the December 2023 test was 525.6 points for the Homework Group and 454.4 points for the No-Homework Group. In the June 2024 test, the average total score for the Homework Group increased to 546.3 points, while that for the No-Homework Group remained at 452.5 points.

We calculated the growth for each of the seven categories by comparing the results of the Speechace speaking tests conducted in April and July. The growth in each category was then compared between the Homework and No-Homework Groups. Negative values indicated a decline in scores for the corresponding categories. In the overall Homework Group, statistically significant differences were observed in the categories of Overall, Vocabulary, Grammar, Average WPM, and Average Bad Pauses. Conversely, in the overall No-Homework Group, significant differences were found only in Vocabulary, Grammar, and Average Bad Pauses. When analyzing second-year students, the Homework Group demonstrated statistically significant differences across all categories, while the No-Homework Group showed no significant differences in Pronunciation and Fluency. For first-year students, both the Homework and No-Homework Groups exhibited significant differences only in Vocabulary, Grammar, and Average Bad Pauses. Table 1 illustrates the growth in each category for each group.

Table 1. The Growth in Individual Speechace Categories for Each Group

Group	Overall	Pronunciation	Fluency	Vocabulary	Grammar	Average WPM	Average Bad Pauses
1st- and 2nd-Year Homework	16.2	11.4	10.9	10.7	13.6	12.4	2.3
1st- and 2nd-Year No Homework	13.7	7.7	7.0	15.7	17.3	7.7	3.2
2nd-Year Homework	17.1	14.8	15.0	9.0	11.4	13.8	1.9
2nd-Year No Homework	29.2	27.7	26.9	21.5	20.0	11.9	2.8
1st-Year Homework	13.8	2.5	0.0	15.0	19.4	8.6	3.3
1st-Year No Homework	1.8	-7.6	-8.2	11.2	15.3	4.5	3.6

4.2 Correlation Between Homework and Speechace Scores

The analysis focused on the Homework group to examine the correlation between the number of times homework was completed and the improvement in each Speechace category. Table 2 presents the results. No correlation was observed for any of the categories, even when limiting the analysis to first-year or second-year students separately.

Table 2. Correlation Between Homework and Speechace Scores

	Overall	Pronunciation	Fluency	Vocabulary	Grammar	Average WPM	Average Bad Pauses
Number of Homework Assignments	-0.24	-0.18	-0.17	-0.14	-0.33	-0.04	-0.23

4.3 Correlation Between Homework and TOEIC Scores

Table 3 shows the results of comparing the average TOEIC scores between the Homework and No-Homework groups, focusing on second-year students. When comparing the average scores within each group between December and June, no statistically significant differences were observed for either group.

However, when comparing the average scores between the two groups, statistically significant differences were observed in Listening for both December 2023 and June 2024. In contrast, for Reading, no significant differences were observed in December, but statistically significant differences were found in June. While the average Reading score of the Homework group increased, the average score of the No-Homework group decreased.

Additionally, focusing on the Homework group, we examined the correlation between the number of times homework was completed and the improvement in TOEIC scores. However, no significant correlations were observed for any of the categories.

Table 3. Correlation Between Homework and TOEIC Scores

Group	Listening Score (Dec)	Listening Score (June)	Listening Growth	Reading Score (Dec)	Reading Score (June)	Reading Growth
Homework (HW)	287.5	296.5	9	238.1	249.7	11.6
No Homework (No HW)	235.8	246.9	11.1	218.6	205.6	-13
HW-No HW	51.7	49.6	-2.1	19.5	44.1	24.6

4.4 Analysis of Speechace Score Improvements Across Proficiency Groups

The participants were divided into three groups—upper, middle, and lower—based on their TOEIC L&R scores, and an analysis of their improvement was conducted using the growth in Speechace TOEIC-equivalent scores. The upper group consisted of 31 participants with an average TOEIC score of 649.7 points; there were 27 participants in the Homework group and four participants in the No-Homework group. The middle group consisted of 28 participants with an average TOEIC score of 496.4 points; there were 18 participants in the Homework group and 10 in the No-Homework group. The lower group consisted of 29 participants with an average TOEIC score of 342.6 points; there were 13 in the Homework group and 16 in the No-Homework group.

Table 4 shows the extent of score improvements in each Speechace category for each group. In the upper Homework group, significant differences were observed between the first and second test results in Overall, Vocabulary, Grammar, Average WPM, and Average Bad Pause. However, in the upper No-Homework group, no significant differences were found in any category. In the middle Homework group, significant differences were observed in Grammar and Average WPM, whereas in the middle No-Homework group, significant differences were found in four categories: Overall, Vocabulary, Grammar, and Average WPM. In the lower group, the Homework group showed significant differences in Overall, Vocabulary, Grammar, and Average Bad Pause, while the No-Homework group showed significant differences in Vocabulary, Grammar, and Average Bad Pause.

Table 4. Score Improvements Across Speechace Categories by Group

Group	Overall	Pronunciation	Fluency	Vocabulary	Grammar	Average WPM	Average Bad Pauses
Upper (Homework)	19.6	11.1	12.6	8.5	11.9	14.7	2.5
Upper (No Homework)	-10	-20	-25	17.5	17.5	-7.3	1.5
Middle (Homework)	6.1	6.7	5.6	8.3	9.4	10.9	0.9
Middle (No Homework)	25	20	20	15	8	14.5	2.1
Lower (Homework)	23.1	18.5	14.6	18.5	23.1	9.5	3.7
Lower (No Homework)	12.5	6.9	6.9	15.6	23.1	7.2	4.4

5. Discussion and Future Study

The findings of this study provided several important insights into the relationship among homework completion, TOEIC scores, and Speechace performance across different student groups. Below, we discuss their implications, drawing on relevant literature to contextualize them.

The results revealed that the Homework group demonstrated significant improvements in various Speechace categories compared to the No-Homework group, particularly among second-year students. For the second-year Homework group, significant differences were observed across all categories, indicating that consistent homework completion positively influenced speaking skills, as measured by Speechace. In contrast, the No-Homework group showed fewer significant improvements, with notable deficits in Pronunciation and Fluency. These findings aligned with previous research that emphasized the importance of deliberate practice in language learning (Ericsson et al., 1993).

However, the lack of significant correlation between the number of homework assignments completed and the improvement in Speechace scores was notable. Previous studies have suggested that the quality of tasks often outweighs their quantity in determining learning outcomes (Tomlinson, 2011). This finding suggested that merely assigning homework may not be sufficient; instead, the design and relevance of assignments should be carefully considered to maximize their impact. Some participants repeatedly completed their homework on similar topics. Moving forward, providing structured guidance, such as specifying topics or increasing task diversity, could help students practice a wider range of linguistic structures.

The Homework group consistently outperformed the No-Homework group in TOEIC L&R scores. This trend was evident among both first- and second-year students, with significant differences observed between the two groups in Listening scores for both December 2023 and June 2024. Additionally, a significant difference in Reading scores emerged in June 2024, with the Homework group showing a marked increase in average scores, while the No-Homework group experienced a decline. These results supported those from studies that highlighted the role of regular and structured homework in improving standardized test scores (Cooper et al., 2006).

The greater improvement in reading, vocabulary, and grammar may be attributed to the text-based nature of ChatGPT conversations, which encourages learners to focus on comprehension and written language structures. In contrast, speaking-related skills, such as Fluency and Pronunciation, may require auditory feedback and real-time correction, which text-based AI interactions do not provide. Additionally, the cognitive demands of formulating responses in English may have diverted attention from aspects like pronunciation accuracy and natural speech patterns. Auditory input's absence likely contributed to limited improvement in pronunciation because students lacked opportunities to imitate native pronunciation models. The lack of time constraints during the AI conversations allowed students to delay responses, possibly explaining the slower development of real-time fluency.

Despite these overall trends, no significant correlations were observed between the number of completed homework assignments and TOEIC score improvements. This finding was in line with those from prior studies, such as Lyddon and Kramer (2019), who found that the quantity of reading during extensive reading activities did not directly correlate with score improvements in TOEIC tests. These results underscored the complexity of language acquisition, where factors such as student motivation, task design, and individual learning differences mediate instructional practices' effectiveness.

The division of participants into upper, middle, and lower groups based on TOEIC L&R scores revealed nuanced differences in performance. In the upper group, significant improvements were observed in multiple Speechace categories within the Homework group, including Overall, Vocabulary, Grammar, Average WPM, and Average Bad Pauses. These findings matched Hattie and Yates (2014), who suggested that high-performing students benefit the most from structured homework because they are better equipped to handle independent learning tasks. In contrast, the No-Homework group in the upper tier exhibited no significant improvements and even declines in some categories, such as Pronunciation and Fluency.

While reducing unnatural pauses is generally considered preferable, the results showed an increase in the number of Bad Pauses. Although fluency improvements are typically associated with a reduction in such pauses, previous research by Sugiura et al. (2024) suggested that disfluency may temporarily increase during the early stages of language learning. This increase reflects trial-and-error processes as learners experiment with new language structures. In the current study, participants' interactions with ChatGPT likely represented such an initial learning stage. Further, the absence of time constraints during AI interactions may have allowed participants to pause longer, leading to increased Bad Pauses in formal testing. To mitigate this, future task designs should incorporate structured scaffolding, requiring participants to respond within set time limits, practice topic rotation, and engage in progressively complex conversational exchanges. Furthermore, explicitly teaching students how to use fillers appropriately could be beneficial because mastery of fillers reflects natural speech patterns commonly observed in real-life conversations.

Future research should investigate whether long-term engagement with AI-based conversations leads to transfer effects, such as improved spontaneous speaking in real-life interactions. Additionally, comparing different AI tools—such as ChatGPT and Google Assistant—could help determine which platform offers more effective conversational practice. Future studies should also incorporate human verification alongside AI for checking homework completion, including spot-checking transcripts to assess engagement depth. Researchers should address potential confounding variables, such as prior English proficiency and motivation levels, which could influence AI-assisted learning interventions' outcomes. Incorporating student interviews or qualitative surveys into the analysis could further clarify how learners perceive AI-driven conversation tasks and what their role is in improving speaking fluency. By refining task design and diversifying AI interactions, future studies can better address the complexities of language learning and enhance AI-assisted educational interventions' effectiveness.

6. Conclusion

In this study, we investigated how engaging in conversations with ChatGPT as homework in a project-based English learning course affected TOEIC L&R and Speechace scores. There were several areas where improvements were observed, but there were also many categories where no significant progress was made. Notably, the results indicated that AI-based conversations had a greater impact on Reading, and although Fluency and Pronunciation improved, the increase in Bad Pauses provided significant insights for future approaches. Moving forward, addressing this issue through more detailed and structured task design will be essential for exploring more effective methods to improve students' skills.

References

- Amin, M. Y. M. (2023). AI and ChatGPT in language teaching: Enhancing EFL classroom support and transforming assessment techniques. *International Journal of Higher Education Pedagogies*, 4(4), 1-15. <https://doi.org/10.33422/ijhep.v4i4.554>
- Butler, Y. G., & Iino, M. (2005). Current Japanese reforms in English language education: The 2003 “Action Plan.” *Language Policy*, 4(1), 25-45. <https://doi.org/10.1007/s10993-004-6563-5>
- Cooper, H., Robinson, J. C., & Patall, E. A. (2006). Does homework improve academic achievement? A synthesis of research, 1987-2003. *Review of Educational Research*, 76(1), 1-62. <https://doi.org/10.3102/00346543076001001>
- Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100(3), 363-406. <https://doi.org/10.1037/0033-295X.100.3.363>
- Hatmanto, E. D., & Sari, M. I. (2023). Aligning theory and practice: Leveraging ChatGPT for effective English language teaching and learning. *E3S Web of Conferences*, 440, 05001. <https://doi.org/10.1051/e3sconf/202344005001>
- Hattie, J., & Yates, G. (2014). *Visible learning and the science of how we learn*. Routledge. <https://doi.org/10.4324/9781315885025>
- Kasneeci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Kasneeci, E., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeiffer, J., Poquet, O., Sailer, M., Schmidt, A., ... Kasneeci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Kimura, S. (2024). AI jidai no eigo kyouiku wo kangaeru: Kikai hon'yaku nado wo seikajugyou ni dounyu shite mite (Rethinking English education in the AI era: Integrating machine translation into formal instruction). *e-Learning Kyouiku Kenkyuu*, 18, 1-12. https://doi.org/10.20623/well.18.0_1
- Lyddon, P. A., & Kramer, B. (2019). Connecting extensive reading to TOEIC performance. In F. Meunier, J. Van de Vyver, L. Bradley, & S. Thouéšny (Eds.), *CALL and complexity: Short papers from EUROCALL 2019* (pp. 257–262). Research-publishing.net. <https://doi.org/10.14705/rpnet.2019.38.1019>
- Monica, M., & Suganthan, C. (2024). A study on analyzing the role of ChatGPT in English acquisition among ESL learners during English language classroom. *Bodhi International Journal of Research in Humanities, Arts and Science*, 8(2), 75-84. <https://doi.org/10.13140/RG.2.2.28252.56961>
- Ministry of Education, Culture, Sports, Science and Technology of Japan. (2018, July 31). *Koutou gakkou gakusyu shidou youryou kaisetsu (Guidelines for high school curriculum instruction: Commentary)* [in Japanese]. https://www.mext.go.jp/a_menu/shotou/new-cs/1407074.htm
- Ministry of Education, Culture, Sports, Science and Technology of Japan. (2019, July 1). *Reiwa gannendo eigokyouiku kaizen puran (The 2019 English education improvement plan)* [in Japanese]. https://www.mext.go.jp/a_menu/kokusai/gaikokugo/1418086.htm
- Ministry of Education, Culture, Sports, Science and Technology of Japan. (2023, July 31). *Reiwa 5nendo zenkoku gakuryoku gakusyuujoukyou chousa (Report on the 2023 nationwide academic achievement and learning survey)* [in Japanese]. https://www.nier.go.jp/23chousakekkahoukou/report/middle_eng.html
- Mutiara, A., Wakhda, S. C., Alfidariyani, I. M., & Indriani, L. (2024). The Effectiveness of Speech-Ace Website on Students' Pronunciation. *English Education: Journal of English Teaching and Research*, 9(1), 92-104. <https://doi.org/10.29407/jetar.v9i1.22372>
- Nhu, P. T. T. (2024). Investigating the effectiveness of practicing English conversation with ChatGPT in improving non-English majored students' English speaking skills at Nguyen Tat Thanh University. *Journal of Science and Technology*, 7(4). <https://doi.org/10.55401/b1rnw21>
- Ningsih, F. (2024). Analyzing students' English-speaking skills using SpeechAce: Insights from an AI-powered assessment tool. *Paraph Journal*, 1(3), 111-125. <https://doi.org/10.70574/9w2prx09>
- OpenAI. (n.d.). *ChatGPT: Optimizing language models for dialogue*. <https://openai.com/blog/chatgpt/>
- Pratiwi, N., Efendy, A. G., Rini, H. C., & Ahmed, N. A. (2024). Speaking Practice using ChatGPT's Voice Conversation: A Review on Potentials and Concerns. *Journal of Language Intelligence and Culture*, 6(1),

59-72. <https://doi.org/10.35719/jlic.v6i1.149>

- Sakai, N. (2023). *Investigating the feasibility of ChatGPT for personalized English language learning: A case study on its applicability to Japanese students*. <https://doi.org/10.31219/osf.io/cv9f2>
- Sugiura, M., Eguchi, A., Abe, M., Muraio, R., Koizumi, T., & Abe, D. (2024). Tyugakusei no supi-kingu ni okeru ryuutyousa no hattatsuteki henka (Developmental changes in speaking fluency among junior high school students). *Learner Corpus Studies in Asia and the World*, 6, 169-184. <https://doi.org/10.24546/0100487721>
- Tomlinson, B. (Ed.). (2011). *Materials development in language teaching*. Cambridge University Press. <https://doi.org/10.1017/9781139042789>
- Yamanaka, T., Kimura, S., Yamashita, M., & Kondo, Y. (2021). *Purojekuto hasshingata eigo puroguramu: jibun jikuwo kitaeru "oshienai" kyouiku (Project-based English program: Train your own axis "without teaching")* [in Japanese]. Kitaohji Shobo.
- Yamashita, M., Yamanaka, T., & Sugiyama, K. (2024). AI tsu-ru wo ikashita eigo writing jugyou: Eibun sakusei shien tsu-ru Transable wo dounyu shite (English writing classes enhanced with AI tools: Implementing the Transable writing support tool). *Ritsumeikan Koutou Kyouiku Kenkyuu*, 24, 75-87. <https://doi.org/10.34382/0002000853>
- Zainuddin, S. Z., & Mohamad, M. (2024). Utilizing SpeechAce to enhance speaking skills among English as a second language pre-university students. *International Journal of Academic Research in Business and Social Sciences*, 14(1), 258-273.
- Zhou, W. (2023). *ChatGPT integrated with voice assistant as learning oral chat-based constructive communication to improve communicative competence for EFL learners*. arXiv. <https://doi.org/10.48550/arxiv.2311.00718>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).