Improving Language Learning with AI: Insights from Speaking and Writing Studies

Hisami Tsuichibaru[1*], Yuko Ito[2], Hiroyuki Obari[3]

[1]The University of Shiga Prefecture, Japan
[2]University of Tsukuba, Japan
[3]Globiz Professional University, Japan
[*]Corresponding author's email: kenchami2@gmail.com
https://orcid.org/0000-0002-2196-7881
https://doi.org/10.54855/979-8-9870112-8-7_7

Received: 09/06/2025 Revision: 27/05/2026 Accepted: 29/05/2026 Online: 18/06/2026

ABSTRACT

The purpose of this study is to highlights innovative applications of artificial intelligence (AI) in enhancing English language proficiency among Japanese learners. The first study investigated how non-English major university students improved speaking skills through smartphone recordings, AI analysis using ChatGPT, and peer assessment over ten weeks. Results revealed enhanced fluency and motivation, with AI feedback and collaborative evaluations fostering critical self-reflection and autonomy. The second study examined Transable (Tr), an AI tool for technical college students, which evaluated and revised essays according to CEFR standards. Tr notably increased unique vocabulary usage, improved fluency, and raised CEFR scores by offering detailed, objective feedback and targeted corrections, supporting independent learning and writing habit development. The third presentation explored AI-based tools Scribo (writing) and Progos (speaking) at Japanese universities, demonstrating that AI facilitates timely feedback, boosts motivation, and allows educators to focus on advanced skills, while human oversight ensures fairness. Overall, the symposium advocates for a blended approach, combining AI's efficiency with human expertise, to optimize language learning in the digital era.

Keywords: AI tools, Transable, ChatGPT, Scribo, Progos, Speaking, Writing

Introduction

In recent years, the advancement of artificial intelligence (AI) in language education has been remarkable, especially for STEM students in Japan, where the ability to communicate and express oneself in English has become more crucial than ever. Against the backdrop of globalization and the increasing pressure to enhance industry's international competitiveness, English language education has shifted from a traditional focus on grammar and reading comprehension to the development of comprehensive practical skills. These now include speaking, writing, and peer assessment, which are increasingly emphasized in curricula. This shift is consistent with task-based and communicative approaches to second language learning,

CITATION | Tsuichibaru, H., Ito, Y., & Obari, H. (2026). Improving Language Learning with AI: Insights from Speaking and Writing Studies. *ICTE Conference Proceedings, 7*, 95-119. ISSN: 2834-0000, ISBN: 979-8-9870112-8-7. DOI: https://doi.org/10.54855/979-8-9870112-8-7_7

which emphasize learners' engagement in meaningful tasks and the ways in which planning, interaction, and feedback can shape language performance (Foster, 1996; Foster & Skehan, 1996; Ellis, 2003). Previous research on technology-supported collaborative writing has also shown the pedagogical value of peer interaction and digital tools in supporting writing development (Suwantarathip & Wichadee, 2014).

This paper addresses these contemporary demands by synthesizing three empirical studies conducted among Japanese technical college, undergraduate, and graduate students. The studies collectively examine the effectiveness and challenges of integrating AI tools—namely Transable, Scribo/Progos, and ChatGPT—into English language education. These tools support a wide range of learning activities, such as automated essay correction and scoring, automated speaking evaluation, and smartphone-based speaking practice with AI feedback

---

Effectiveness of Speaking Learning with ChatGPT and Smartphones: Cooperative Learning through Mutual Grading and Independent Support Learning

Hisami Tsuichibaru

Introduction

English language education is undergoing a major transformation thanks to rapidly evolving digital technologies. In particular, the development of information and communication technologies (ICT) has opened new possibilities for English learning methodologies by enabling learners to learn without being constrained by location or time. This study examines a new educational approach that uses artificial intelligence, such as ChatGPT, to improve speaking skills. Compared to the old style of teaching, AI-based teaching methods are said to facilitate learners' self-learning and provide feedback tailored to their individual needs.

Literature Review

Overview of Speaking Research

In the 19th and early 20th centuries, the grammar-translation method dominated speaking instruction, emphasizing grammatical rules and translation rather than conversational skills. The Audiolingual Method, grounded in behaviorist psychology, shifted focus to repetitive practice and imitation, effectively developing early speaking abilities (Doughty & Long, 2003).

With the rise of the Internet and digital technologies in the 21st century, speaking instruction has undergone a significant transformation. Online practice and virtual classrooms enable learners to engage in speaking activities anytime and anywhere, enhancing convenience and real-time interaction (Godwin-Jones, 2015). Task-Based Learning (TBL) has gained prominence, promoting natural language use through meaningful tasks that reflect real-life communication (Skehan, 1996).

Technological advances further enrich speaking instruction. Virtual reality (VR) offers immersive environments for practicing self-regulated learning strategies, while speech recognition and analysis tools provide immediate feedback to improve pronunciation (Chun, 2016). Moreover, AI-driven chatbots facilitate interactive speaking practice with real-time corrective feedback, supporting learner autonomy (Stockwell, 2012). These innovations continue to reshape speaking pedagogy, highlighting the need for effective integration of technology in future research and practice.

Moreover, AI-driven chatbots facilitate interactive speaking practice with real-time corrective feedback, supporting learner autonomy (Stockwell, 2012). Recent studies in AI-assisted

English learning have also reported that ChatGPT can increase learners' engagement, confidence, and willingness to communicate when used as a supplementary support tool in language learning environments (Nhan, 2025; Phuong, 2024).

Mobile Assisted Language Learning (MALL)

Mobile Assisted Language Learning (MALL) leverages mobile devices to deliver educational content flexibly, supporting autonomous and personalized learning processes (Heil et al., 2016). Mobile technologies offer advantages such as mobility, social interaction, and adaptability, enabling learners to engage without constraints of time or place (Sung et al., 2015). The prevalence and user-friendliness of smartphones make them particularly effective MALL tools.

Research confirms that MALL enhances learning outcomes, notably in listening skills, through situational activities, collaboration, and social engagement. Educational systems must rapidly adapt to evolving technologies and pedagogical innovations to meet learners' needs. MALL also fosters critical thinking, problem-solving, communication, and teamwork skills, contributing to academic achievement and lifelong learning competencies (Kim, 2013).

Cooperative Learning

Cooperative learning, where students collaborate to solve shared intellectual problems, is widely recognized for its educational value (Laal & Ghodsi, 2012). Particularly in online environments, it promotes active learning, knowledge sharing, innovation, and problem-solving. Socially, it cultivates student voice and positive attitudes toward learning. Common cooperative practices include group projects, discussions, peer review, and feedback, all supported by digital tools such as Google Docs, Microsoft Teams, and video conferencing platforms (e.g., Zoom) that facilitate collaboration

Outcomes of cooperative learning include deeper understanding, critical thinking, improved communication, and learner autonomy. Its success depends on effective teacher guidance, thoughtful environment design, appropriate tool selection, and active student participation. Research underscores that when these elements align, cooperative learning significantly enhances educational experiences across diverse contexts (Suwantarathip and Wichadee, 2014).

Method

Purpose of this study

The purpose of this study is to verify the effectiveness of smartphone-based recording technology for English speech practice and to analyze recordings using ChatGPT, an AI tool, to clarify its effects on speech quality and motivation to learn. Specifically, the effectiveness of speech transcription and analysis using ChatGPT, as well as the effects of peer scoring and peer feedback on learners' motivation to learn, is determined.

Research Questions

1. Is smartphone recording effective in English speech practice?

2. How effective is speech transcription and Analysis using ChatGPT?

3. What are the effects of peer scoring and peer feedback on motivation to learn?

Through these research questions, we aim to examine the impact of smartphone recording and AI technology on English speech education from various perspectives and propose effective ways to use ICT tools in language education.

Research methodology

The study adopted a quasi-experimental design centered on experimental lessons and was conducted over a 10-week period of speech practice among non-English major university students. The study participants were 46 students from public universities in Shiga Prefecture, all of whom agreed to participate in the study. Ethical approval for the study was granted by the appropriate institutional review board, and informed consent was obtained from all participants.

Procedures

1 Students read materials related to the theme in reading classes and collected. ideas for. the. content. of their speeches through group discussions.

2 Based on the group discussions, the students prepared a draft of their speeches.

3 Students practiced reciting their speeches based on the prepared speech drafts.

4 The speeches were recorded using smartphones and uploaded to Padlet. This recorded. data. was converted to text with a text-to-speech app, then further analyzed and fed back into ChatGPT.

5 students peer-rated their classmates' speeches using Google Forms. This evaluation. served as feedback to promote student self-improvement and collaborative learning.

6 The experimental period was concluded through a final speech presentation and a. summative. activity using a worksheet.

Figure 1

Class Flow

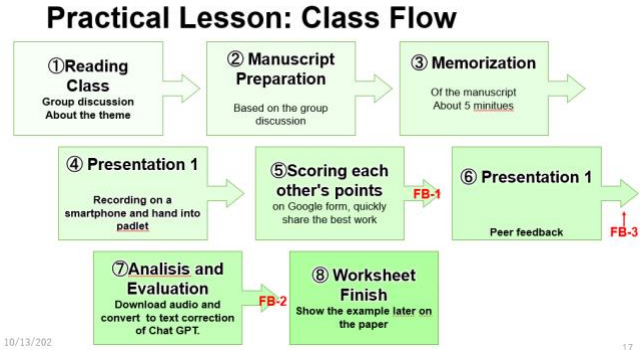
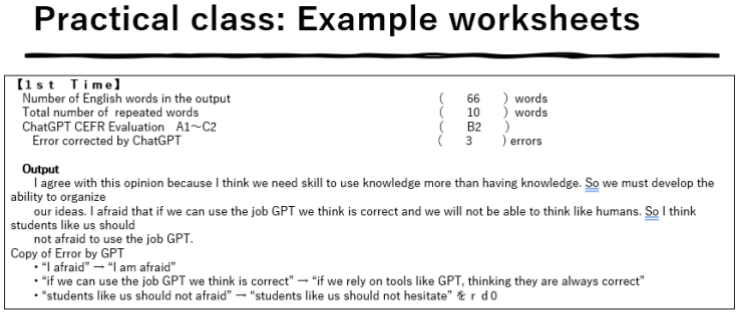


Figure 2

Example Worksheet



[Theme]1. Do you agree with Chat GPT? Why or why not?
2. Where would you like to go? Why?
3. What can be done to correct the wage gap between men and women?
4. What can be done to attract tourists to remote islands?
5. What can be done to combat global warming?

Data Collection and Analysis

Recordings were transcribed, and textual analysis was performed using AI tools. Data analysis was conducted using the following methods.

Quantitative analysis: Changes in motivation to learn before and after speech practice were measured using a student survey.

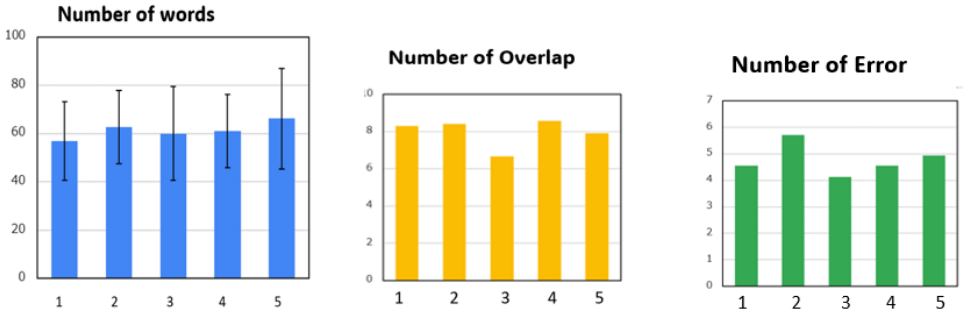
Qualitative analysis: Based on feedback from the AI tool (ChatGPT) and peer evaluations, areas for improvement in the quality and content of speeches were identified.

Results

Look at Figure 3 and Figure 4.

Figure 3

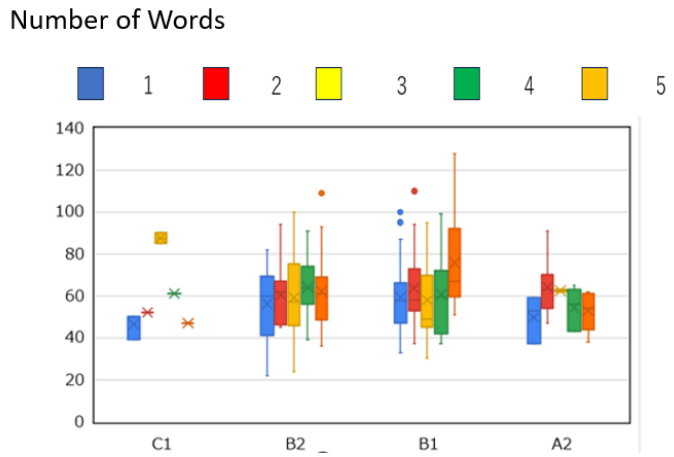
Changes in the quality of the English text of students' speeches as a result of the implementation of the class



The study analyzed students' English speech fluency through three metrics: total word counts, overlapping words, and error words. Transcripts generated via Notta and analyzed with a free application showed a steady increase in total word counts across sessions, indicating improved fluency with practice. However, overlapping words (indicating disfluencies) and error words flagged by ChatGPT exhibited no consistent trends, fluctuating unpredictably between sessions. While students relied on translation apps to minimize errors (averaging fewer than 5 per speech), the variability in disfluencies and errors suggests that these metrics are less reliable indicators of progress than overall word output.

Figure 4.

Number of words



Next, we discuss the data presented in Figure 4. This figure shows a whisker diagram of the total word counts calculated by ChatGPT for each level, showing that for students at the B2 and B1 levels, fluency increased with each session, and the total number of words increased steadily.

For C1 and A2 level students, on the other hand, it was difficult to find a clear trend due to the small sample size. This result may be attributed to the small number of students in the sample

Next. Look at Figure 5 and Figure 6.

Figure 5.

Questionnaire

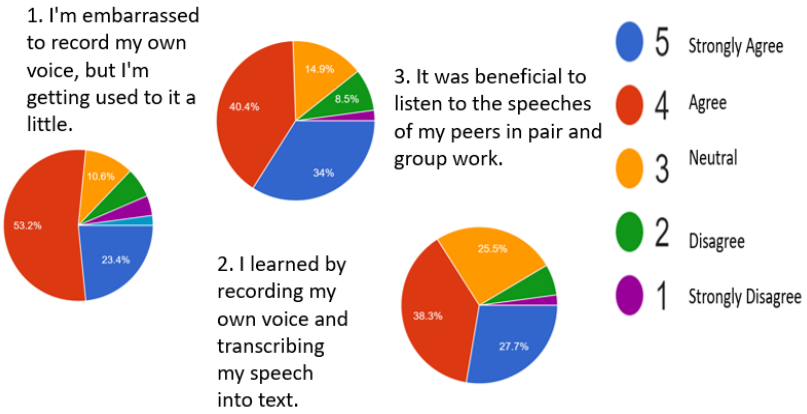


Figure 6

Questionnaire

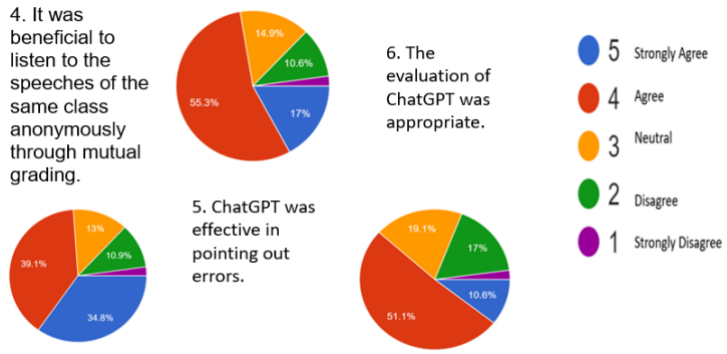
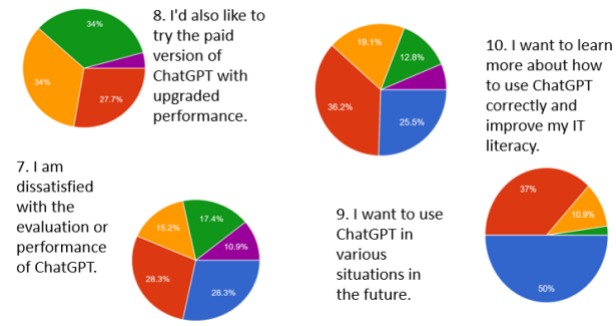


Figure 7.



The study investigated the impact of thematic complexity on students' English speech fluency and the role of voice recording, AI feedback, and peer evaluation in learning. Initially, most students were uncomfortable recording their voices, but 76.6% adapted and found it beneficial for speech practice. Clear pronunciation motivated 74.4% of learners, and 34% strongly valued recording for learning. Peer and group work were effective; 66% found listening to classmates' speeches meaningful, enhancing motivation through shared success.

Regarding AI feedback, 61.7% rated ChatGPT's speech evaluation as adequate, though 27% were dissatisfied. Over half (56.6%) expressed interest in a paid version, anticipating better academic support, and 87% wished to learn to use AI more effectively. Peer feedback was also beneficial, with 66% endorsing it and 61.7% wanting continued use of ChatGPT.

Fluency increased on familiar, simpler topics, aligning with Bygate (2001), though this sometimes led to more vocabulary errors due to overconfidence. Complex topics reduced fluency but improved accuracy, consistent with Skehan's (1998) task complexity framework. These results support prior findings that cognitive demands affect fluency and accuracy (Ellis, 2003; Foster & Skehan, 1996).

The study concludes that self-assessment via voice recording promotes fluency, supporting Zimmerman's (2002) self-regulated learning theory. AI tools show promise but require institutional investment for advanced features. Peer evaluation boosts motivation and confidence, suggesting a combined AI-peer framework enhances speaking skills. Future research should explore paid AI versions, compare AI and human feedback, and standardize peer assessments. Limitations include a small sample size, a brief intervention, and technical issues.

Regarding AI feedback, 61.7% rated ChatGPT's speech evaluation as adequate, though 27% were dissatisfied. Over half (56.6%) expressed interest in a paid version, anticipating better academic support, and 87% wished to learn more to use AI more effectively. Peer feedback was also beneficial, with 66% endorsing it and 61.7% wanting continued use of ChatGPT. _This tendency is consistent with recent findings showing that learners generally perceive ChatGPT-based support positively because of its immediacy, accessibility, and interactive feedback functions (Phuong, 2024).

Discussion

The study examined how thematic complexity influences student speech fluency. As shown in Figure 1, students' word counts increased, especially during the second session on the familiar topic, "Which country do you want to go to and why?" Familiarity with everyday vocabulary boosted engagement and fluency, making speech preparation easier. However, consistent with Bygate (2001), increased fluency on simple topics may also lead to more vocabulary errors due to overconfidence and a prioritization of speed over accuracy.

In contrast, the complex topic "How can we address the gender wage gap?" in the third session resulted in lower fluency. Students focused more on accuracy, using less vocabulary due to the topic's complexity. This supports Skehan's (1998) framework, which suggests that complex tasks encourage deliberate, accurate language use. These findings align with Ellis (2003) and Foster and Skehan (1996), who argue that cognitive task demands influence fluency and error rates.

The student survey found that smartphone-based voice recording and transcription effectively motivated English speech practice. AI-generated feedback from ChatGPT was valued for identifying weaknesses, though some noted limitations in the free version, suggesting a need for institutional investment in advanced tools. Assessment results showed many students at

CEFR B1/B2 levels, with a narrow score distribution possibly reflecting homogenization in AI assessments, unlike personalized human feedback (Kumar, 2023).

Peer evaluation enhanced motivation and participation. Listening to peers' speeches fostered shared learning and confidence, supporting collaborative language acquisition (Topping, 2009). Combined with AI assessment, peer feedback offers a balanced framework for promoting proficiency.

AI-generated feedback from ChatGPT was valued for identifying weaknesses, though some noted limitations in the free version, suggesting a need for institutional investment in advanced tools. Recent TESOL research has similarly suggested that AI tools such as ChatGPT are most effective when combined with teacher mediation and reflective learning activities rather than used independently (Pham & Cao, 2025)

Peer evaluation enhanced motivation and participation. Listening to peers' speeches fostered shared learning and confidence, supporting collaborative language acquisition (Topping, 2009). The present findings also support recent AI-assisted language learning studies that emphasize the motivational benefits of ChatGPT-supported speaking practice in EFL contexts (Nhan, 2025).

Conclusion

Self-assessment using voice recording and transcription improves English speech fluency, supporting Zimmerman's (2002) self-regulated learning. AI tools show promise, but advanced features may require institutional support. Peer feedback motivates learners and builds confidence. Institutions should integrate AI and peer evaluation to enhance speaking skills. Future research should explore paid AI versions, compare AI and human feedback, address technology gaps, and standardize peer assessments.

Limitations include a small sample size, a short intervention, and technological challenges, all of which warrant further investigation.

Acknowledgments

I would like to express my sincere gratitude to the anonymous reviewers for their valuable feedback and insightful comments, which have greatly improved the quality of this work.

Furthermore, I would like to extend my heartfelt thanks to the University of Shiga Prefecture for their cooperation in facilitating the practical experimental classes. Their support was indispensable in carrying out this research.

References

Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan, & M. Swain (Eds.), Researching pedagogic tasks: Second language learning, teaching and testing (pp. 23-48). Longman. https://doi.org/10.4324/9781315838267-3

Chun, D. M (2016). The role of technology in teaching and researching speaking. *Annual Review of Applied. Linguistics*, *36*, 128-144. https://doi.org/10.64152/10125/44463

Doughty, C. J., & Long, M. H. (2003). The handbook of second language acquisition. Blackwell. Publishing https://doi.org/10.1002/9780470756492

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language

performance. *Studies in Second Language Acquisition, 18*(3), 299–323. DOI: 10.1017/S0272263100015047

Godwin-Jones, R. (2015). The evolving roles of language teachers: Trained coders, local researchers, global. citizens. *Language Learning & Technology, 19*(1), 10–22. https://doi.org/10.64152/10125/44395

Heil, C. R., Wu, J. S., Lee, J. J., & Schmidt, T. (2016). A review of mobile language learning applications: Trends, challenges, and opportunities. *The EUROCALL Review, 24*(2), 32-50. https://doi.org/10.4995/eurocall.2016.6402

Kim, H.-S. (2013). Emerging mobile apps to improve English listening skills. *Multimedia-Assisted Language Learning, 16*(2), 11–30. https://doi.org/10.15702/mall.2013.16.2.11

Kumar, P. (2023). Faculty members' use of artificial intelligence to grade student papers: A case of. implications. *International Journal for Educational Integrity, 19*(1), 1-10.

https://doi.org/10.1007/s40979-023-00130-7

Laal, M., & Ghodsi, S. M. (2012). Benefits of collaborative learning. *Procedia - Social and Behavioral. Sciences, 31*, 486-490. https://doi.org/10.1016/j.sbspro.2011.12.091

Nhan, D. L. T. (2025). Examining the impact of AI-assisted preparation on Vietnamese EFL learners' willingness to communicate in English. *ICTE Conference Proceedings*. https://doi.org/10.54855/979-8-9870112-9-4_6

Padlet. https://padlet.com

Pham, T., & Cao, M. (2025). The practice of ChatGPT in English teaching and learning in Vietnam: A systematic review. *International Journal of TESOL & Education.* https://doi.org/10.54855/ijte.25513

Phuong, H. P. X. (2024). Using ChatGPT in English language learning: Students' perceptions and experiences. *International Journal of TESOL & Education. https://doi.org/10.54855/ijte.24414*

Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied Linguistics*, *17*(1), 38-62. https://doi.org/10.1093/applin/17.1.38

Stockwell, G. (2012). Computer-assisted language learning: Diversity in research and practice. Cambridge. University Press. https://www.cambridge.org/core/books/computerassisted-language-learning/A28AF8FF17764316B56F7DC06A2AC51E

Sung, Y. T., Chang, K. E., & Liu, T. C. (2015). The effects of integrating mobile devices with teaching and. learning on students' learning performance: *A meta-analysis and research synthesis. Computers & Education, 94*, 252-275. https://doi.org/10.1016/j.compedu.2015.11.008

Suwantarathip, O., & Wichadee, S. (2014). The effects of collaborative writing activity using. Google Docs on students' writing skills. *The Turkish Online Journal of Educational Technology, 13*(2), 148–156.

Topping, K. J. (2009). Peer assessment. *Theory into Practice, 48*(1), 20-27. https://doi.org/10.1080/00405840802577569

Exploring the Relationship Between Word Count, Proficiency, and CEFR Evaluation Fluency Development in AI-Assisted English Writing: A Study of Kosen Students Using Transable

Yuko Ito

Introduction

The rapid advancement of AI is expanding its beneficial applications in education (UNESCO, 2023). Kosen (National Institute of Technology) students have a high concentration in specialized fields such as mechanical engineering, electrical and electronic engineering, and chemical bioengineering (National Institute of Technology[KOSEN], 2022). However, Kosen students are expected to become future engineers who will lead Japan and thus must acquire an international sensibility during their studies. They need to enhance their English proficiency through various opportunities, including interactions with international students studying at Kosen, and to strengthen their English writing skills for future project presentations (Dasaradhi et al., 2016).

In this context, recent language education has increasingly emphasized the development of communicative competence and practical language use (Ellis, 2003).

This study compares English compositions written independently by students with those revised using Transable, an AI-powered tool equipped with automatic correction, automatic scoring, and machine translation functionalities (Tsuichibaru & Yamashita, 2023). We specifically focused on fluency among the three key indicators for evaluating language proficiency—accuracy, fluency, and complexity (Wolfe-Quintero et al., 1998)—to investigate the correlation between changes in word count and evaluation, as well as students' perceptions of the tool. The experiment aims to verify the potential of utilizing Transable in English writing instruction.

Literature Review

Writing Instruction Utilizing Transable

Tsuichibaru and Yamashita (2023) compared Transable (Tr), an English writing support tool with automatic correction, scoring, and machine translation, with Write and Improve with Cambridge (WIC), which lacks automatic scoring. They examined (1) writing quality improvement, (2) the validity of automatic scoring, and (3) student perceptions. For (1) and (2), they analyzed changes in word count, lexical diversity, and error frequency between pre- and post-revision essays, and their relation to CEFR scores. WIC showed a positive correlation between word count and evaluation, whereas Tr did not, suggesting that Tr emphasizes content features such as structure and logic over grammar. Thus, Tr was found to be more effective at enhancing overall writing quality. While their study compared two tools, the present study focuses solely on Tr and its effectiveness for Kosen students, using the same methodology. CEFR-based scoring in Tr will be used for evaluation. Unlike the previous study, this research also considers students' English proficiency levels and explores their influence on writing outcomes, supported by a questionnaire.

Reliability and Validity of AI-Assisted Language Data

Mizumoto and Yamaguchi (2023) highlighted the usefulness of tools that provide automatic correction and feedback to enhance content and organization in learners' writing, supporting autonomous development. They also addressed the reliability and validity of the e-rater system. Oki (2024) compared the reliability of AI-based English learning diary analysis with text mining, a method for objectively analyzing large language datasets. After finding similar results

from both methods, Oki concluded that language-generation AI offers sufficient reliability for language analysis.

Effectiveness of Machine-Translated English Compositions by AI

Niimi and Umeki (2024) reported that grammar and orthography improved significantly following revision activities using Grammarly. Furthermore, Ota and Sadoshima (2013) stated that fostering autonomous writers is crucial for writing instruction. Given that Transable includes a back-translation feature that converts independently written expressions back into Japanese, allowing writers to confirm their intended meaning, it is considered to contribute to "fostering autonomous writers."

English Writing Support Tool: Transable

Transable is an English writing support tool developed by Kohei Sugiyama at Ritsumeikan University(Yamashita et al., 2024). Designed to promote autonomous learning through multimodal approaches, it integrates machine translation, automatic correction, and scoring to move beyond traditional teacher-centered instruction. The tool encourages learners to make justified decisions among multiple possible answers, thereby fostering independent language use. While its core functions—translation, correction, and CEFR-based scoring—have been previously examined (Tsuichibaru & Yamashita, 2023), Transable also features a ChatGPT-linked “simple question” function that enables users to revise their writing through guided prompts. Additionally, it incorporates Grammarly-based feedback aligned with multiple proficiency scales (CEFR, GTEC, IELTS, TOEFL iBT), allowing learners to identify and address weaknesses autonomously. These features collectively position Transable as a pedagogically grounded tool that supports both linguistic accuracy and learner autonomy.

Fluency as a Component of Writing Quality

While accuracy and complexity have traditionally been emphasized in second language writing research, fluency has also been recognized as a key indicator of writing development. Wolfe-Quintero, Inagaki, and Kim (1998) identified total word count and lexical variety as reliable measures of fluency that correlate with writing proficiency. Their comprehensive review positioned fluency alongside accuracy and complexity as essential dimensions of writing assessment. More recently, Tavakoli (2023) emphasized the pedagogical importance of fluency, noting its strong association with learners’ overall communicative competence and motivation. These findings support the inclusion of fluency as a meaningful construct in evaluating the effectiveness of AI-assisted writing tools such as Transable.

Purpose of the Study

Building upon the preliminary research conducted by Tsuichibaru and Yamashita (2023), the present study aims to examine whether the quality of English compositions improves in terms of fluency when using Transable. In addition, the study examines how these improvements may vary across learners’ English proficiency levels. To gain further insight into the pedagogical utility of Transable, a questionnaire will be administered to explore students’ perceptions of its usability, effectiveness, and limitations in the context of English writing instruction at Kosen institutions.

Research Questions

1. *In writing activities where learners independently select, evaluate, and revise suggestions provided by the English writing support tool, Transable, do word counts increase, does fluency improve, and is there an overall enhancement in the quality of English compositions after revision?*

2. *Is there a correlation between changes in word count after revision and the scores generated by the automatic scoring system? Furthermore, do these correlations differ according to learners' English proficiency levels?*
3. *Does the use of the English writing support tool foster autonomous learning and enhance learners' motivation for future English study?*

Methodology

Participants

The participants were 40 fourth-year students from the Department of Chemical and Biological Engineering at a Kosen. All participants had completed at least three years of English instruction and were familiar with basic academic writing tasks.

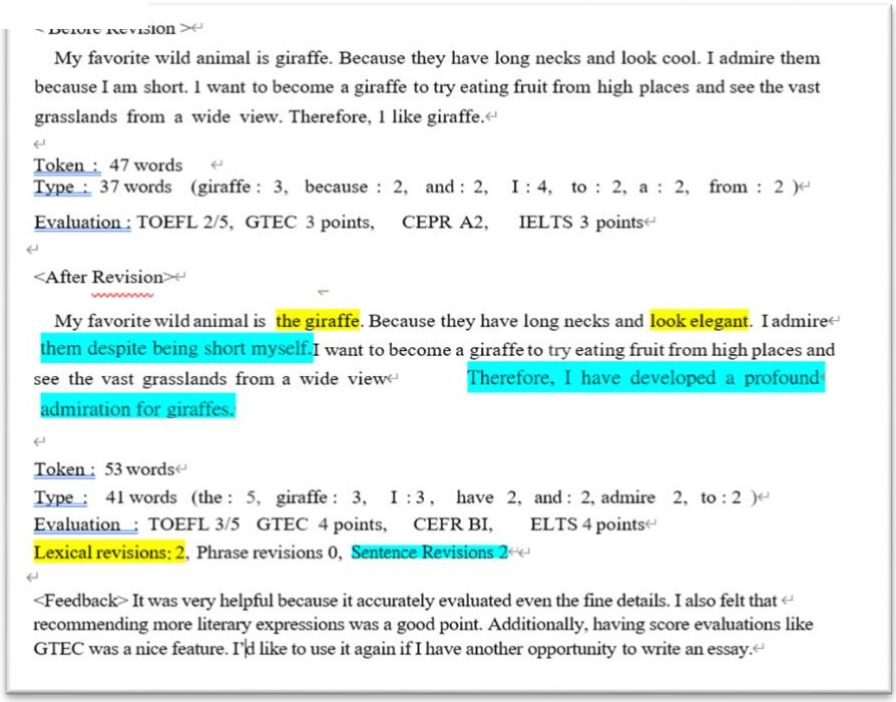
Materials and Procedure

Students completed two writing tasks based on Unit 4, "Express Your Ideas," from *Science Inspiration*:

1. Do you have a favorite wild animal?
2. Do you think we should keep wild animals in zoos?

The first prompt was descriptive, while the second required opinion and reasoning. Students initially wrote their essays independently, then revised them using Transable. Using Transable's back-translation and CEFR-based scoring, they reviewed feedback and revised their writing. The "Get Opinion" function provided additional suggestions to support revision. For submission, students compiled both pre- and post-revision essays in an A4 Word document. They included total and unique word counts, CEFR scores, and highlighted corrections using color codes. They also recorded the number of corrections and wrote a brief reflection on their experience with Transable (see Figure 1).

Figure 1



Analysis

This study examined whether changes in total word count (Tokens) and unique word count (Types) before and after revision, as measured with Transable, correlated with CEFR-based evaluations. For statistical analysis, CEFR levels were converted numerically as follows: A2 = 2 points, B1 = 3 points, and B2 = 4 points.

Overall Trend

Table 1 Changes in Each Item and Correlations Between Evaluation and Increases/ Decreases in Tokens and Types

Essay 1 Title: "Do you have a favorite wild animal or animals? Wy do you like them?"

N=38

Category	Item	Before	After	Change Rate	Correlation
Lexical Measures	Tokens	50.7	52.7	3.4%	-0.18
Lexical Measures	Types	36.7	39.4	6.5%	-0.31
Revisions	Word Corrections		2		
Revisions	Expression Corrections		2		
Revisions	Sentence Corrections		1		
Proficiency	CEFR Level	2.7	3.1	13.7%	

Changes in total and unique word counts and CEFR level before and after revision of Essay 1 (N = 38), showing a weak negative correlation between lexical diversity and evaluation.

For Essay 1, the average of total word count increased from 50.6 to 52.4 (+1.8), and the unique word count from 37.2 to 39.6 (+2.4), with an average CEFR score increase of +0.4. A weak negative correlation was observed between unique word count and evaluation, suggesting that increased lexical variety did not necessarily lead to higher scores (see Table 1).

For Essay 2, the total word count rose slightly from 54.5 to 55.1 (+0.6), and the unique word count from 40.6 to 41.5 (+0.9), again resulting in a +0.4 increase in the CEFR score. However, no significant correlations were found between either lexical measure and evaluation (see Table 2).

Table 2

Essay 2 Title: "'Do you think we should keep wild animals in zoo?"

N=36

Category	Item	Before	After	Change Rate	Correlation
Lexical Measures	Tokens	54.5	55.1	1.1%	-0.14
Lexical Measures	Types	39.8	41.1	3.2%	-0.10
Revisions	Word Corrections		1		
Revisions	Expression Corrections		2		
Revisions	Sentence Corrections		1		
Proficiency	CEFR Level	3.3	3.7	12.1%	

Summary of lexical and CEFR changes in Essay 2 (N = 36), with slight increases in word counts and no significant correlation with evaluation.

Proficiency-Level Analysis

To explore differences by proficiency, students were grouped into two levels (Upper, Lower) based on regular test scores and three levels (Upper, Middle, Lower) based on external test scores. A total of 12 correlation analyses were conducted across both essays.

In Essay 1, weak negative correlations were found between total word count and evaluation for both low-proficiency groups. Interestingly, a moderate negative correlation emerged between unique word count and evaluation for the high-proficiency group (external test), suggesting that more varied vocabulary may have introduced complexity that the scoring system did not favor.

In Essay 2, a strong positive correlation was observed between total word count and evaluation for the high-proficiency group (external test), indicating that fluency may have contributed to higher scores. Conversely, moderate negative correlations were observed in low-proficiency groups, suggesting that increased length did not translate into improved quality.

These findings suggest that the relationship between lexical measures and evaluation is nonlinear and may vary with learners' proficiency. High- and low-proficiency learners appear more sensitive to changes in word count than intermediate learners, highlighting the need for differentiated feedback strategies.

Table 3

Comparison of Token and Types word Counts Pre-and Post-Revison and Their Correlation with CEFR Score

		Pre			Post			Difference			Correlation	
Essay 1												
	N	1	2	3	1	2	3	1	2	3	1	2
All	38	50.6	37.2	2.7	5.2	39.6	3.1	1.8	2.4	0.4	-0.18	-0.31
R. Upper	24	51.0	37.9	2.8	52.5	40.5	3.1	1.5	2.6	0.3	-0.18	-0.30
E. Upper	3	51.3	39.3	2.7	53.3	41	2.3	2.3	1.7	-0.3	0.08	-0.50
E. Middle	26	50.9	37.2	2.7	53.3	40.2	3.1	2.1	3	0.4	-0.14	-0.32
R. Lower	14	49.9	36.0	2.5	52.2	38.1	3.1	2.4	2.1	0.6	-0.25	-0.32
E. Lower	9	49.8	39.2	2.4	50.2	40.8	3.2	0.4	1.6	0.8	-0.26	-0.35

Note. 1=Token. 2=Types. 3=CEFR. R=Regular Test. E=External Test.

Changes in tokens, types, and CEFR level before and after revision of Essay 1, showing a weak negative correlation between lexical diversity and evaluation

Table 4

		Pre			Post			Difference			Correlation	
Essay 2												
	N	1	2	3	1	2	3	1	2	3	1	2
All	36	54.5	39.8	3.3	55.1	41.1	3.7	0.58	1.28	0.39	-0.14	-0.10
R. Upper	24	56.5	40.6	3.4	56.5	41.5	3.7	0.1	0.9	0.3	0.08	0.07
E. Upper	3	58.5	39.7	3.3	58	41	3	-0.7	1.3	-0.3	0.84	0.19
E. Middle	24	55	40	3.3	55.3	40.9	3.8	0.3	0.9	0.5	0.00	0.06
R. Lower	12	50.5	38.2	3.2	52.2	40.2	3.7	1.6	2	0.5	-0.42	-0.20
E. Lower	9	51.8	39.3	3.3	53.7	41.6	6.6	1.9	2.2	0.2	-0.60	-0.40

Note. 1=Token. 2=Types. 3=CEFR. R=Regular Test. E=External Test.

Changes in tokens, types, and CEFR level for Essay 2, highlighting a strong positive correlation between total word count and evaluation for high-proficiency learners.

Results

RQ1: Changes in Fluency and Composition Quality

After revision using Transable, both essays showed slight increases in total and unique word counts. Essay 1 increased by +1.8 (tokens) and +2.4 (types), while Essay 2 increased by +0.6 and +0.9, respectively. Although the increases were modest, they suggest a general improvement in fluency. Notably, low-proficiency groups showed the highest gains in both total and unique word counts, indicating that the tool may be particularly effective for learners with limited output.

In terms of composition quality, CEFR-based evaluations improved by an average of +0.4 points for both essays. Essay 1 saw the greatest improvement among the low-proficiency group (external test: +0.8), while Essay 2 showed similar gains in both low- and middle-proficiency groups (+0.5). These results suggest that Transable contributed to measurable improvements in writing quality, especially among lower-proficiency learners.

RQ2: Correlation Between Word Count and Evaluation

Across both essays, most proficiency groups showed no positive correlation between word count and CEFR evaluation. In Essay 1, five of six groups showed no correlation with total word count, and all groups showed a negative correlation with unique word count. In Essay 2, similar patterns emerged: moderate negative correlations in low-proficiency groups and no significant correlations in middle-proficiency groups. The only strong positive correlation was observed in the high-proficiency group (external test) for Essay 2, where increased total word count was associated with higher evaluation scores ($r = 0.838$). This suggests that for advanced learners, fluency may contribute more directly to perceived writing quality.

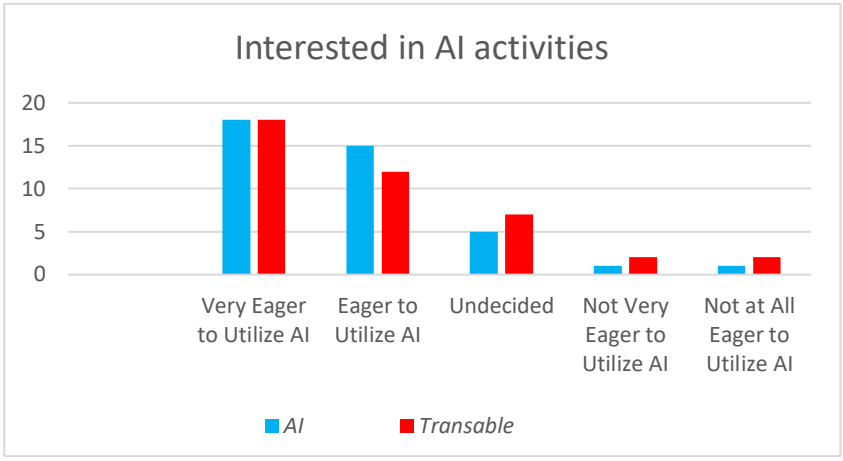
In contrast, low-proficiency learners showed consistent negative correlations across both essays. Although their word counts increased, their initial output was limited, and improvements in evaluation were modest. This implies that improvements in vocabulary and grammatical accuracy may be necessary before gains in fluency translate into higher scores. Middle-proficiency learners showed mixed results, with only one negative correlation across four measures. This suggests that for this group, further improvement may depend more on content development and logical structure than on lexical measures alone.

RQ3: Learner Perceptions and Motivation

Survey responses indicated that most students found Transable helpful and expressed a desire to continue using AI tools in the future. As shown in Figure 2, most students expressed a strong interest in continuing to use AI tools. This suggests that the tool contributed positively to learner motivation. As shown in Figure 3, students most frequently received AI-generated corrections for grammar and expression, accounting for 36% of the total corrections. In free-response comments, 31 of 40 students provided positive feedback, with “Promotion of Learning and Understanding” as the most common theme (19 responses). This indicates that students viewed Transable not only as a correction tool but also as a means of supporting autonomous learning.

Figure 2.

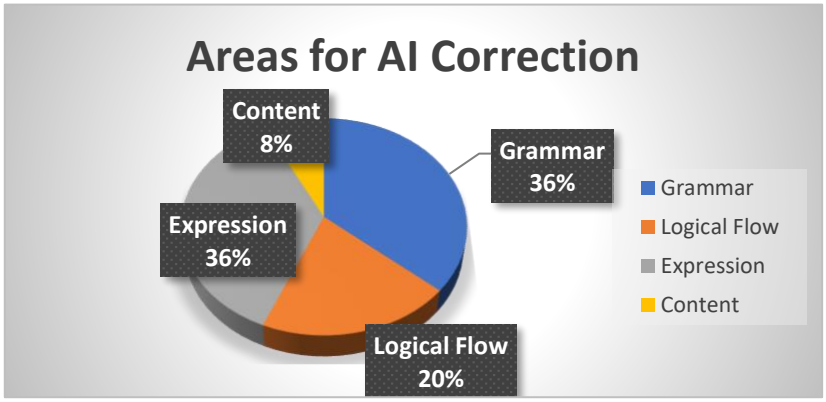
Students' interest in AI activities after using Transable



Note. The chart shows that more students expressed a strong interest in using AI tools

Figure 3.

Areas targeted by AI for correction in student writing

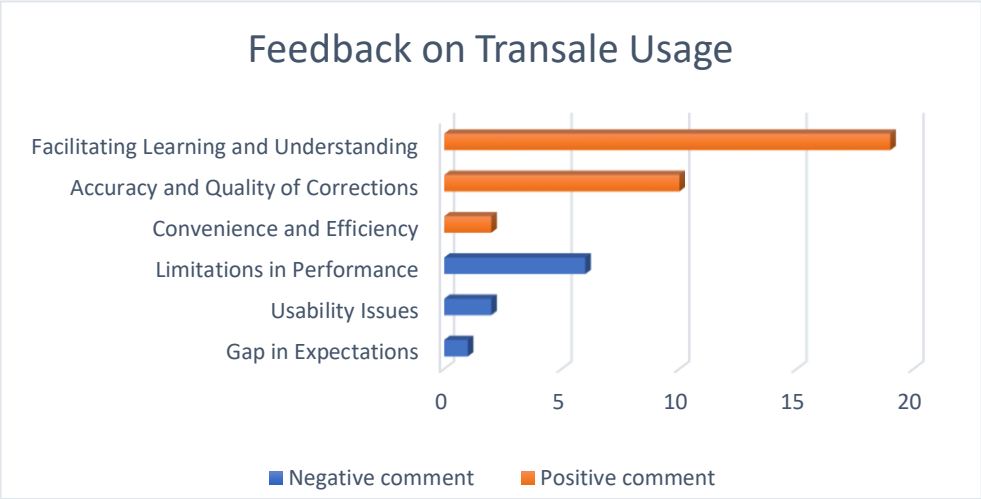


Note. Grammar and expression were the most frequently corrected areas, each accounting for 36% of the total.

As shown in Figure 4, students most frequently praised the AI tool, Transable for facilitating learning and understanding, while also noting limitations in its performance as a key area of concern.

Figure 4.

Student feedback on Transable usage: Positive and negative aspects



Note. The chart compares positive and negative student comments on various aspects of Transable

Discussion

This study aimed to examine whether using Transable, an AI-assisted English writing tool, improved fluency and composition quality among Kosen students. While slight increases in word count and CEFR evaluations were observed, strong statistical correlations were not consistently found. This suggests that lexical expansion alone may not directly influence automated evaluation outcomes.

One possible explanation is that Transable's integrated ChatGPT prioritizes logical coherence over surface-level fluency, as prior research supports. Additionally, the short length of the essays (approximately 50 words) may have limited the potential for meaningful revision. Future studies should consider longer writing tasks (e.g., 100–300 words) to better assess the tool's impact.

Another factor may be the level of instructional support. Previous studies have shown that the effectiveness of AI in language learning depends not only on learners' readiness but also on the presence of appropriate instructional guidance and support (Nguyen et al., 2026; Nguyen, 2026). Although students received basic guidance on using the tool, many required further scaffolding, such as examples on how to evaluate suggestions, set revision goals, and apply grammar rules. Insufficient pre-instruction may have limited the tool's effectiveness, particularly for lower-proficiency learners. For these learners, it is especially important to provide explicit strategies for using the AI tool's functions effectively, such as back-translation, scoring interpretation, and revision support.

Interestingly, Essay 2, which involved a more cognitively demanding topic, showed fewer negative correlations. This suggests that task complexity may influence how learners engage with AI feedback and how their revisions are evaluated. Further research is needed to explore how topic type and learner autonomy interact with AI-supported writing development.

Conclusion

This study explored the effects of using Transable on English writing among Kosen students. While modest gains in word count and CEFR evaluations were observed, strong correlations between lexical measures and evaluations were limited. These findings highlight that automated

scoring systems consider more than just word count, including content quality, logical structure, and task complexity.

The results also underscore the importance of teacher involvement in helping students to critically engage with AI-generated suggestions. Learners must develop the ability to evaluate and apply feedback effectively, a process that requires instructional support tailored to their proficiency levels.

Ultimately, while AI tools like Transable offer valuable opportunities to enhance writing instruction, they should be integrated thoughtfully into a pedagogical framework. This study contributes to a growing understanding of how AI can support autonomous learning and writing development in technical education contexts.

Acknowledgements

This research expresses gratitude to Professor Hisami Tsuichibaru, who introduced the practice of Transable at the University of Shiga Prefecture and allowed us to observe classes at Ryukoku University; Mr. Kohei Sugiyama of Ritsumeikan University, Ritsumeikan Global Innovation Research Organization, who developed Transable and provided guidance on its utilization; the various professors who provided valuable feedback at the JACET Kansai Writing Section Meeting and Professor Takayuki Obari, who provides guidance at the JACET Kanto AI Branch Meeting; and the teachers of our school who permitted the use of AI in classes, and the students of Fukushima Kosen who cooperated in the research.

References

Dasaradhi, K., Venkata, R A., & Sai, D P.V.(2016). Need of 'Proficiency in English' For Engineering Graduates. *International Journal of English Language, Literature and Humanities*, 4(1).

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.

Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics, 2*(2), 100050. https://doi.org/10.1016/j.rmal.2023.100050

National Institute of Technology (KOSEN). (n.d.). Retrieved April 22, 2026, from https://www.kosen-k.go.jp/en/

Nguyen, T. K.T., Nguyen, T. T. D., & Trinh, D. D. (2026). English majors' and non-English majors' perceptions and readiness for AI-assisted learning in English language courses: An exploratory study. *ICTE Conference Proceedings*, 9, 107–120. https://doi.org/10.54855/979-8-9870112-9-4_7

Nguyen, T. P. L. (2026). Vietnamese EFL lecturers' perspectives on leveraging AI tools to enhance students' autonomous learning. *International Journal of Language Instruction, 5*(1), 1–17. https://doi.org/10.54855/ijli.26511

Niimi, N., & Umeki, R. (2024). An exploratory practice of English writing instruction aided by automated writing evaluation system (Grammarly). *Remedial Education Research, 18*, 69–80. https://doi.org/10.18950/jade.2023.07.18.01

Oki, T. (2024). Reliability of analyzing English learning diaries using language generation AI: Comparison with text mining. *Hakuoh Journal of the Faculty of Education, 18*(2), 109–135.

Ota, Y., & Sadoshima, S. (2013). Tutor Training and PAC Analysis of Two Tutors' Awareness

towards Tutorial Sessions: Waseda University Writing Center's Case [in Japanese] .Waseda Global Forum, 9, 237-277

Tavakoli, P. (2023). Making fluency research accessible to second language teachers: The impact of a training intervention. *Language Teaching Research, 27*(2), 368–393. https://doi.org/10.1177/1362168820951213

Jidō tensaku oyobi jidō saiten to kikai hon'yaku o mochiita writing shidō no kanōsei: Eisakubun shien tsūru Transable o shiyō shite [The potential of writing instruction using automated correction, scoring, and machine translation: A case study with the English composition support tool "Transable"]. *Japan Journal of English Language and Literature, 33*, 195–209. https://jaell.org/wp-content/uploads/2024/04/Tsuichibaru_yamashita.pdf

UNESCO. (2023). Guidance for generative AI in education and research. https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research

Wolfe-Quintero, K., Inagaki, S., & Kim, H.-Y. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity.* University of Hawaii Press.

Yamashita, M., Yamanaka, T., & Sugiyama, K. (2024). AI tsūru o ikashita eigo writing jugyō: Eibun sakusei shien tsūru Transable o dōnyū shite [AI-enhanced English writing classes: Introducing the English writing support tool Transable] [in Japanese]. *Ritsumeikan Kōtō Kyōiku Kenkyū, 24*, 75–87.

AI-Mediated Assessment in Japanese University EFL Programs: A Mixed-Methods Evaluation of Scribo and Progos

Hiroyuki Obari

Abstract

The rapid development of artificial intelligence (AI) is transforming English education by enabling personalized learning trajectories and automated assessment. This mixed-methods study investigates the pedagogical impact of two AI-based assessment tools—**Scribo** for academic writing and **Progos** for oral proficiency—across three Japanese universities. Sixteen engineering and information-science majors (13 males, 3 females) engaged in a 14-week blended program that combined weekly AI-mediated tasks with instructor-led workshops emphasizing critical thinking and intercultural awareness. Quantitative analyses of pre-/post-writing portfolios and Progos Speaking Test scores revealed significant gains in grammatical accuracy (t (15) =4.32, p<.001) and oral fluency (t (15) =2.12, p=.025). Qualitative data from focus-group interviews indicated heightened learner motivation, appreciation for immediate feedback, and emerging metacognitive awareness of genre conventions. Nevertheless, students and instructors alike underscored the irreplaceable role of human evaluation for discourse-level coherence, pragmatic nuance, and equitable scoring. The findings support a **blended assessment model** in which AI expedites formative feedback while human judgment safeguards reliability and fosters higher-order skills. Recommendations include scaled deployments of AI tools, longitudinal monitoring of proficiency trajectories, and faculty development on AI-human orchestration to meet the complex demands of second-language acquisition in the 21st century.

Keywords: AI-enhanced assessment; Scribo; Progos; learner motivation; blended feedback; Japanese EFL

Introduction

Artificial intelligence (AI) has shifted from a peripheral novelty to a mainstream component of language education, with algorithmic feedback engines now embedded in many writing and speaking platforms (Chun, 2022). For English as a Foreign Language (EFL) programs in Japan—where class sizes are often large and instructional time is constrained—AI-assisted assessment promises scalable, individualized feedback (Kikuchi, 2023). Yet concerns remain about bias, hallucination, and the potential deskilling of teachers. The present study focuses on the classroom deployment of **Scribo**, an AI writing analyzer that provides sentence-level diagnostics, and **Progos**, an automated speaking test aligned to the Common European Framework of Reference (CEFR).

We ask whether these tools measurably improve language proficiency, how they influence student motivation, and where human evaluators continue to add value. By concentrating on engineering majors—who typically prioritize technical accuracy over rhetorical nuance—we probe the affordances and limitations of AI assessment for discipline-specific English learning.

Research Questions

1. *Proficiency effect: To what extent do Scribo and Progos improve writing accuracy and oral fluency over one academic term?*
2. *Affective effect: How do students perceive AI feedback in terms of motivation and self-efficacy?*
3. *Pedagogical balance: Which assessment dimensions require human judgment to ensure fairness and reliability?*

Literature Review

AI Feedback and Writing Development

Automated writing evaluation (AWE) systems trace their roots to e-rater® (Burstein, 1998) and have since diversified into tools such as Grammarly, Criterion, and Scribo. Meta-analysis shows small-to-medium effects on grammatical accuracy (Li,2021), though gains in discourse organization are modest unless coupled with teacher mediation ~~(Link et al.,2022)~~. (Ranalli & Yamashita, 2022)

Automated Oral Assessment

Progos leverages automatic speech recognition and neural scoring to generate CEFR-aligned ratings within minutes (Progos, 2024). While studies report high concurrent validity with human scoring (Nakamura & Yoshida, 2024), concerns persist regarding pronunciation bias for non-standard accents (Yoon, 2023).

Motivational affordances of AI

Immediate, personalized feedback is linked to higher intrinsic motivation and reduced foreign-language anxiety (Dewaele & Dewaele, 2022). Nevertheless, over-reliance on AI may throttle deeper strategic learning and critical thinking unless accompanied by reflective pedagogy.

Human–AI Complementarity

Sociocultural theory posits that learning emerges through mediated interaction (Vygotsky, 1978). In AI-rich contexts, the mediator may be both tool and teacher; optimal learning arises from their orchestration (Luckin et al., 2022). Therefore, evaluating AI tools in situ must examine not only accuracy gains but also the interplay of human mentorship and

machine analytics.

Method

Participants and Context

Sixteen postgraduate students enrolled in English for Engineering Communication courses at three Tokyo-area universities volunteered for the study (age M = 24.2 years). Entry CEFR speaking levels ranged from B1 to B2. All participants had prior exposure to Grammarly but no formal training with Scribo or Progos.

Design

A convergent mixed-methods design integrated repeated-measures proficiency tests with qualitative inquiry (Creswell & Plano Clark, 2018). Over 14 weeks, students alternated between weekly Scribo-mediated writing tasks (e.g., technical abstracts) and Progos practice interviews. Each AI task was followed by a 30-minute seminar where instructors highlighted critical-thinking strategies, genre conventions, and intercultural pragmatics.

Instruments

- **Writing accuracy**: Error density (errors/100 words) in a timed argumentative essay, scored by two blind raters (κ = .87).
- **Oral fluency**: Progos overall score plus words-per-minute (WPM) from recorded interviews.
- **Motivation**: A 10-item Likert questionnaire adapted from the L2 Motivational Self System (Dörnyei & Ushioda, 2009; α =.81).
- **Focus-group interviews**: Semi-structured sessions (n = 3) exploring perceptions of AI vs. human feedback.

Data Analysis

Paired-sample t tests examined pre-/post-changes. Interview transcripts were coded inductively, then mapped to motivation and metacognition themes using NVivo 14. Trustworthiness was ensured through member checking and peer debriefing.

Results

Writing Accuracy

Mean error density dropped from 15.4 to 9.7 errors/100 words ($\Delta = -5.7$, $t(15) = 4.32$, $p < .001$, $d = 1.08$), with the largest reductions in verb-tense (−41%) and article usage (−38%). Students attributed improvements to Scribo's color-coded alerts and inline explanations.

Oral Fluency

Progos overall scores increased from M = 5.25 (B1) to 6.12 (B1 High), while WPM increased by 18% ($t(15) = 2.12$, $p = .025$). Pronunciation sub-scores improved modestly but remained below CEFR B2 thresholds for five participants, consistent with previous accent-bias findings.

Motivational Shifts

Self-efficacy ratings rose significantly ($M_{pre} = 3.2$, $M_{post} = 4.1$ on a 5-point scale; $t(15) = 3.85$, $p < .01$). Qualitative data revealed three salient drivers:

1. **Immediacy**: "Seeing feedback pop up instantly made revising less frustrating."
2. **Agency**: "I could choose which Scribo suggestions to accept, so I felt in control."

3. **Benchmarking**: Progos CEFR levels offered "clear goals" and "game-like progression."

Yet several students cautioned that AI "cannot tell if my example is persuasive" or "if my tone is polite."

Discussion

Efficacy of AI Tools

The sizeable reduction in grammatical errors corroborates prior AWE studies (Li, 2021) and suggests that Scribo's engineering-specific vocabulary database aligns well with learners' academic needs. Progos gains, though smaller, mirror Yoshida's (2024) report of +0.8 CEFR bands after intensive AI-speaking practice.

Motivational Benefits and Caveats

Echoing AI feedback bolstered confidence and reduced revision anxiety. Nevertheless, participants' critique of discourse-level omissions underscores the indispensability of teacher guidance for argument quality and pragmatics.

Human Evaluation for Fairness and Reliability

Both students and instructors voiced concerns about opaque scoring algorithms and occasional misrecognition of Japanese-accented consonant clusters. Human moderation thus remains essential for equitable high-stakes assessment, aligning with testing-standard guidelines (American Educational Research Association, 2014).

Toward a Blended Assessment Model

Findings support a blended model whereby AI handles micro-level accuracy and rapid diagnostics, freeing instructors to cultivate critical thinking, intercultural pragmatics, and genre sophistication. Such orchestration resonates with the "intelligence infrastructure" framework (Luckin et al., 2022).

Implications for Practice

1. **Faculty development**: Workshops on interpreting AI analytics and designing follow-up tasks can maximize learning transfer.
2. **Curriculum integration**: Embedding Scribo/Progos checkpoints into project-based learning can scaffold iterative improvement.
3. **Ethical stewardship**: Transparent communication about data privacy and algorithmic limitations is vital for trust.

Limitations and Future Research

The small, discipline-specific sample limits generalizability. Future studies should examine humanities cohorts and compare alternative AWE systems. Longitudinal tracking beyond 14 weeks would clarify whether gains plateau or consolidate. Finally, computational discourse analysis could quantify improvements in coherence—an area relatively untouched by current AI tools.

Conclusion

This study demonstrates that Scribo and Progos, when embedded in a pedagogically guided framework, significantly enhance grammatical accuracy and oral fluency while energizing learner motivation. Yet AI alone cannot adjudicate rhetorical persuasiveness, cultural appropriateness, or equity in high-stakes contexts. A blended assessment approach, leveraging

AI's speed and human judgment's nuance, is therefore essential for meeting the multifaceted challenges of 21st-century second-language acquisition.

Acknowledgment

The authors employed ChatGPT and Grammarly to refine the English in this proceedings paper; however, they assume full responsibility for all remaining errors and interpretations.

References

American Educational Research Association, American Psychological Association, & National. Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. AERA.

Burstein, J. (1998). A description of the e-rater scoring engine. *ETS Research Report Series*, *1998*(2), 1–30. https://doi.org/10.1002/j.2333-8504.1998.tb01764.x

Chun, D. M. (2022). Artificial intelligence in language assessment: A critical overview. *Language Assessment Quarterly*, *19*(1), 1–20. https://doi.org/10.1080/15434303.2021.2003897

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE.

Dewaele, J-M., & Dewaele, L. (2022). The impact of CAPTCHA-style feedback on foreign-language anxiety. *System*, *108*, 102853. https://doi.org/10.1016/j.system.2022.102853

Dörnyei, Z., & Ushioda, E. (2009). *Motivation, language identity and the L2 self.* Multilingual Matters.

Kikuchi, K. (2023). AI-mediated feedback in large Japanese EFL classes. *JACET Journal*, *67*, 45–60.

Li, L. (2021). Automated writing evaluation and second language writing: A meta-analysis. *Journal of Second Language Writing*, *54*, 100813.

Ranalli, J., & Yamashita, T. (2022). Automated written corrective feedback: Error-correction performance and timing of delivery, *Language Learning & Technology, 26*(1), 1–25. : http://hdl.handle.net/10125/73465

Luckin, R., George, K., & Cukurova, M. (2022). *AI for school teachers.* CRC Press. https://doi.org/10.1201/9781003193173

Nakamura, T., & Yoshida, R. (2024). Validating an AI-based speaking test in Japan. *Language. Testing in Asia*, *14*(6), 1–24.

Progos. (2024). *Progos speaking test technical manual*. https://progos.ai/en

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Yoon, S. Y. (2023). Accent bias in automated speech scoring: A systematic review. *Language. Assessment Quarterly*, *20*(2), 202–225. https://doi.org/10.1080/15434303.2023.2183214

Conclusion

This paper synthesizes three empirical studies involving technical college and university students, examining the latest developments in AI-assisted English language education using

diverse AI tools—namely Transable, Scribo/Progos, and ChatGPT. The findings highlight three key insights that can shape the future of English language teaching and learning.

Building upon these insights, it is clear that the future of English language education must embrace a collaborative learning environment where AI and humans—teachers and learners alike—work together. The strengths of AI—rapid diagnosis and personalized adaptation—should be combined with the unique capabilities of human educators, who excel in fostering higher-order thinking, value judgment, and intrinsic motivation. This "hybrid instructional model" leverages the best of both worlds, creating a more effective and engaging learning experience for all students.

Looking ahead, there is a need for further empirical research involving a wider range of learner populations and educational settings. Continuous improvement of AI tools, enhanced AI literacy among teachers, and the development of ethical and institutional frameworks are all essential next steps. As AI technology evolves, it is our responsibility as educators to continually explore new approaches that empower every learner to engage in autonomous and creative lifelong learning. The mission for English language educators is to remain at the forefront of this transformation, ensuring that the integration of AI enriches rather than diminishes the human aspects of teaching and learning.

In conclusion, this paper advocates for a balanced, collaborative approach to English language education, where AI and human expertise are seamlessly integrated. By doing so, we can unlock the full potential of every learner and prepare them for the challenges and opportunities of a rapidly changing world.

Biodata

Hisami Tsuichibaru is a part-time lecturer at The University of Shiga Prefecture (USP), specializing in English language education. Her research focuses on English writing instruction and ICT-integrated language learning. She studies automated writing evaluation and machine translation for language education.

Yuko Ito is an experienced English educator with a Master's in English Language Education from the University of Tsukuba, Japan, and a TESOL Certificate from Anaheim University. Co-author of *Foundational Knowledge of English Language Education for Enhancing Teaching Skills* and *AI and University English Education: Toward the Next Stage* (2026) with members of the JACET (The Japan Association of College English Teachers) AI Special Interest Group. As a member of both the JACET AI and Writing Special Interest Groups, my research focuses on innovative AI-assisted teaching methods in EFL writing and speaking. Currently serving as Assistant Professor at National Institute of Technology (KOSEN), Fukushim College, committed to advancing student engagement and English proficiency.

Dr Hiroyuki Obari is a Professor at Globiz Professional University and Professor Emeritus at Aoyama Gakuin University. He specializes in English-medium instruction (EMI), CLIL, AI-mediated learning, and intercultural competence development. He also teaches at the Institute of Science, Tokyo, and serves as a visiting researcher at the National Institute of Advanced Industrial Science and Technology (AIST). He holds degrees from the University of Oklahoma (B.A.), International Christian University (M.A.), Columbia University (M.A. TESOL/Applied Linguistics), and the University of Tsukuba (PhD in Computer Science), and

has conducted research at the University of Oxford as a visiting scholar. Dr. Obari's research integrates AI, ICT, and multimodal learning environments to enhance language education and global competence. He has published extensively on CALL, EMI program design, flipped learning, and AI-supported assessment, with over 40 books, 100 papers, and 300 conference presentations. His current work focuses on AI-driven learning analytics, research ethics, and innovative curriculum design for higher education.