



Title	A Collaborative Journey : Practical Report on a Joint Research Project with Graduate Students
Author(s)	Lee, Shzh-chen Nancy
Citation	言語文化共同研究プロジェクト. 2025, 2024, p. 9-16
Version Type	VoR
URL	https://doi.org/10.18910/102276
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

A Collaborative Journey: Practical Report on a Joint Research Project with Graduate Students

Lee Shzh-chen Nancy

This paper presents a reflective report of a collaborative research project I conducted with graduate students examining the impact of AI-driven and machine translation-driven feedback on English as a second language (ESL) writing. The project emerged from my introduction to applied linguistics class at the Graduate School of Humanities, Osaka University, where we revised the syllabus to align with students' research interests, background, and abilities. Through extensive discussions after conducting literature review, we agreed on a shared topic of second language acquisition which led us to investigate the effects of AI and machine translation feedback on ESL writing. This paper consists of a literature review on corrective feedback and the original paper that we replicated (Sasaki et al., 2024). It provides a detailed reflection on the development of this project and its processes, including brainstorming and topic selection, literature review, research design, ethical considerations and applying for approval, research redesign, data collection, result analysis and interpretation. Our journey culminated in an academic conference presentation in Osaka in 2025, marking a significant milestone in our academic collaboration. This paper also discusses the importance of collaborative research projects and suggests implications for future classroom and research practices. It is hoped that this reflective account can be useful for teachers and students who might be interested in starting their own research projects. By sharing this practical report, readers could hopefully take away some of the learning experiences and avoid possible pitfalls.

1. Introduction

This joint research project originated from a graduate course I taught called Introduction to Applied Linguistics where only international students signed up and they were interested in topics different from my original syllabus. To accommodate their needs and interests, I decided to modify the syllabus for the above course. I invited the students to collaboratively develop a new syllabus and through reconstructing the syllabus, we found a gap in the existing literature. Eventually, the course evolved from reading and understanding existing literature to actually conducting a joint research project. Based on this research gap and students' strong interests, we discussed the feasibility of conducting a research project as a class project within the remaining semester time and our potential obstacles. Students also actively debated over appropriate topics for this possible class-based research project. Finally, students decided to replicate a writing study looking at machine translation as a form of feedback on L2 writing, comparing the impact of machine translation and teacher feedback on ESL writing (Sasaki et al., 2024).

This paper includes a brief literature review on the effects of corrective feedback and the original study on the comparison between the impact of machine translation (MT) and teacher feedback on writing followed by a reflective account of the collaborative research project. Collaborative research provides a unique opportunity for graduate students to engage in hands-on academic inquiry while developing critical thinking, analytical, and research skills. In this project, we adapted the original syllabus to accommodate student interests and foster an interactive learning experience. The present reflective summary describes the project starting from the birth of the project, brainstorming of ideas, topic selection, research design, applying to ethical committees for approvals, research redesign, data collection, data analysis, and finally presenting the research results at an academic conference. Our research focused on the comparative effectiveness of AI-driven writing assistance and machine translation tools in improving ESL writing quality. This practical report paper documents our research journey and hopes to offer some practical insights for future teachers and students starting their own joint research projects.

2. Literature Review

2.1 *Corrective Feedback*

Corrective feedback directs learners' attention to their linguistic errors, and learners exhibit more noticing when feedback is given (Mackey, 2006). According to Schmidt's (1990) Noticing Hypothesis, noticing is a necessary prerequisite for L2 development. Previous studies on the effects of corrective feedback have mostly produced robust evidence for its effectiveness (Lyster & Saito, 2010), while a lack of corrective feedback inhibits syntactic accuracy development (Goh & Burns, 2012). Feedback could encourage learners to notice discrepancies between their own language production and target language forms. Therefore, the more corrective feedback learners receive, the higher the probability they will improve on the target forms. However, there is a continuing controversy over the effectiveness of corrective feedback because there is evidence that implicit types of feedback are not always interpreted by learners as error correction. If learners do not realize a corrective form, they are unlikely to notice the gap between their production and the target language rule (Williams, 2005). In addition to the concerns with using implicit types of feedback, feedback can also be ineffective itself when extensive correction is given, in which learners receive feedback on too many types of errors or they experience fatigue in receiving feedback (Williams, 2005). In addition, even if learners notice errors in their own language output, it is not certain that they will successfully integrate what they have noticed into their interlanguage (Nicholas et al., 2008). Lyster (2004) further argued that only explicit forms of corrective feedback can push learners to precisely and appropriately reformulate their output and criticized the value of implicit feedback. Ineffective corrective feedback can result in students' inattention and negative attitudes toward receiving feedback (Paulus, 1999).

The role of feedback in second language writing has been widely studied, with teacher feedback traditionally considered to be the gold standard. Studies have shown that direct feedback from teachers helps learners improve accuracy and develop metacognitive skills. However, there are also some criticisms against teacher feedback because it can be vagueness and "rubber stamp" consisting of no constructive feedback. In addition, teacher comments can have an over reliance on grammar correction (Paulus, 1999). Finally, providing timely and individualized feedback is labor-intensive, leading to increasing interests in technological alternatives (Sasaki et al., 2024).

2.2 *Comparing Teacher and Machine Translation (MT) Feedback*

Sasaki et al. (2024)'s study examined machine translation as a form of feedback on L2 writing by comparing teacher corrective feedback (TCF) and machine translation (MT) feedback in an L2 writing classroom. A total of 23 Japanese university students whose first language is Japanese participated in this study where they were divided into two groups: TCF and MT. The TCF group described a picture in English where they were instructed to write one paragraph. They received teacher feedback for the paragraph they have written before they repeated the same English description task. The MT group described the same picture in English and were asked to compose a small paragraph in Japanese of what they composed in English. The Japanese paragraph was entered into DeepL and an English translation was produced. Differences between what they wrote in English and the English translation of their composed Japanese paragraph were highlighted and given to the MT group as feedback before they repeated the same English writing task. The TCF and MT groups reversed their order and received the same feedback before completing the English writing task. All participants were asked to complete a survey after completing their writing tasks. Sasaki et al. (2024) found that TCF improved complexity whereas MT improved accuracy and fluency. They suggested that differences in the engagement level of the teacher feedback and machine translation feedback might have influenced the different learning outcomes.

3. Reflective Account of the Project

3.1 *Origin of the Project*

The idea for this research project originated from my graduate class titled introduction to applied linguistics in October 2024, where I recognized a need to revise the syllabus to cater students' academic interests and backgrounds. My original syllabus concerned language learning and teaching from a scientific research perspective and how research findings can be applied into language classrooms. I designed my original course for students who want to be future language teachers and those interested in the mechanism of language development, especially the development of speaking proficiency. It was meant to serve as an introduction to students who want to experience postgraduate

studies in English as it included a strong component of reading and understanding literature written in English. Students were supposed to read research papers related to language proficiency development, participate in active class discussions, and to complete writing assessments in English on research paper critiques.

However, it became evident through discussions with the students that they had limited interest in my original selection of readings because of their backgrounds. Three students enrolled in this course and they are all native speakers of English and they are interested in JSL (Japanese as second language) research because they are international students who came to Japan to pursue their studies and research interests. As a result, we decided to replace the original list of class readings which were predominantly related to the research of the learning and teaching of English as a second language with readings related to the learning and teaching of JSL.

I uploaded the class readings list onto a Google doc document so that students could take ownership in developing the new reading list (see Appendix A). Students proactively searched and uploaded research articles. Students also took turns briefly introducing the papers they uploaded and reasons for uploading those articles. However, after searching online for several weeks, we found that there are limited numbers of research papers written in English on the learning and teaching of JSL. Through the process of looking for JSL literature, a gap surfaced that there is a need for expanding the research of the learning and teaching of Japanese as a second language.

Based on this research gap and students' high motivation, an idea appeared that we could change the course syllabus where students could jointly conduct a small research project to extend the line of JSL research instead of reading and critiquing JSL related research papers. Quickly, we discussed the feasibility of conducting a research project within the remaining semester time and our potential obstacles. Students also actively debated over appropriate topics for this potential class research project. Students enrolled in this class were interested in both speaking and writing proficiency development related topics. They debated the amount of time required for conducting writing proficiency development research and speaking proficiency development research. Finally, they decided to replicate a writing study looking at machine translation as a form of feedback on L2 writing, comparing the impact of machine translation and teacher feedback on ESL writing (Sasaki et al., 2024).

3.2 Replicating Previous Research

After students decided to replicate the study by Sasaki et al. (2024) comparing the impact of machine translation and teacher feedback on ESL writing, they discussed if they could do a complete replication of the study but change the research target from ESL to JSL. The original study by Sasaki et al. was a longitudinal study involving a complicated within-participant crossover design (Lui, 2016). We decided to shorten and simplify our study by changing our research design to once off intervention treatment and also by allocating participants into three groups (including two interventional groups and a control group) instead of crossover two interventional groups without a control group. We also changed the research target from ESL writing to JSL writing because all three students were interested in Japanese related studies. Finally, given students' interests in AI, we decided to conduct our study comparing the impact of AI driven feedback and machine translation driven feedback on JSL writing.

With the rise of artificial intelligence (AI), researchers have explored its efficacy as supplementary or alternative educational sources. However, while AI is rapidly changing the existing language learning and teaching practices, it is sometimes being misused or misinterpreted. AI and MT often overlap as AI also includes MT tools and vice versa. AI can offer context-aware suggestions whereas MT tools such as Google Translate or DeepL offer linguistic transformations that can assist learners in understanding complex structures. Some research suggests that AI tools like Grammarly enhance grammatical accuracy and coherence, helping learners refine their writing iteratively (Li & Hegelheimer, 2013). Conversely, MT feedback is often criticized for producing translations that lack nuance and contextual appropriateness. We wanted to replicate Sasaki et al. (2024) study but to further investigate on how learners interact with different feedback modalities in various learning contexts.

3.3 Ethical Committee Approval

A significant component of this research project involved navigating approval from research ethics committees. It was also the longest and most difficult stage of the project, which proved to be a valuable and albeit demanding learning experience for the students and myself. After we designed our research, we looked for suitable research participants through our affiliates. Since our original research

topic focused on JSL (Japanese as a second language), we consulted with a teaching fellow at a Japanese language program at our university with the possibility of asking her JSL students to participate in our study. We also invited the Japanese teaching fellow to join our project as she is also interested in the JSL research. After she consulted with her supervisor, we were asked to prepare an application to submit to their research ethics committee. The application documentation was very thorough and consisted of questions to carefully consider and protect participant privacy and to avoid misuse of research data. We also prepared a written consent for the participants in the program (See Appendix B). The following was our original research purpose that we included in the research ethics application: The study investigates how AI-powered translation and machine translation (MT) feedback impact Japanese as a second language (JSL) learners. While AI and MT translation tools have been widely used, their differences have often been overlooked as well as their potential different effects on learners. Participants will write one paragraph in Japanese and then they will receive either AI or MT generated feedback before they rewrite their paragraph. Participants' revised paragraph will be examined and differences in their Japanese composition will be examined.

Our initial submission was positive and was met with requests for resubmission due to concerns about data privacy and participant consent. We revised our application and resubmitted our application for but was unfortunately rejected. The ethics committee pointed out our mistakes in completing the research ethics application forms as we made a number of careless mistakes in completing the research ethics forms. It was a meaningful yet rather frustrating process as while we learned that we need to spend more time and not rush with the research ethics application applications, the students were concerned with the time running out in the semester and the fear of not being able to collect data for the project. The ethics committee finally granted us approval in our first submission attempt but requested for the research to be conducted outside of class time in order to not disadvantage participants from receiving normal class contact time. Due to the nature of our project looking at the effect of AI and MT generated feedback on writing and on the participants, we decided it would be infeasible to conduct the research outside the classroom. The students feared that they might need to terminate their project as we were unable to recruit enough number of JSL participants for our project.

3.4 Redesigning Study and Changing Research Focus

We eventually revised and changed the focus of our research from JSL to ESL writing and submitted an application to a different ethical committee board, which ultimately granted our approval in Feb 2025. Instead of collecting data from JSL learners, we decided to collect data from ESL learners from my own general English classes. While students were disappointed that they were unable to continue the line of JSL research, they quickly became eager to continue their project with a new focus on ESL. Our new research focused on comparing AI-generated feedback and machine translation (MT) feedback by evaluating their impact on ESL learners' writing improvement. We had two questions for our study: 1) How does AI feedback and MT feedback differ in terms of their usefulness? 2) What are the effects of each type of feedback on writing development? We collected data from three general English classes that I taught myself and randomly allocated the participants into two experimental groups (AI and MT) and one control group. The new project was designed where AI group received AI generated feedback and the MT group received MT generated feedback. The control group received no feedback intervention. The research was conducted during class time. We hypothesized that one interventional group would improve more in writing than the other interventional group and the control group.

3.5 Data Collection

The data for this study were collected from three undergraduate general English classes I taught at our university. A total of 49 participants were divided into three groups: AI-generated feedback, MT-generated feedback, and a control group. In order to make sure we have equal number of participants completing tasks in both experimental groups, we randomly assigned participants into one of the two experimental groups. A total of 19 students were assigned to the AI group and they received feedback from ChatGPT. A total of 19 students were assigned to the MT group and they received feedback from DeepL. Only eleven students were in the control group. Data collection lasted for two weeks. In Week 1, all participants were asked to describe a four-panel comic in English. They were asked to write a paragraph describing the comic and to aim to write for 150 words in 15 minutes. They were also asked to write a Japanese paragraph of equivalent to what they wrote in English. The data was collected

using Google form during class time. Research purpose was explained to the participants and a consent form was distributed before data collection.

We used ChatGPT to generate feedback for the AI group by copying and pasting participants' English paragraphs into ChatGPT. We instructed ChatGPT to produce feedback for improving participants' English writing. We used DeepL to generate feedback for MT group by copying and pasting the participants' Japanese paragraphs into DeepL. We highlighted differences between participants' original English text and the DeepL generated English translation by following Sasaki et al. (2024)'s research methodology. We printed out the feedback onto A4 sheets stapled the feedback with original English text. In Week 2, we distributed the feedback to individual participants and gave them 10 minutes to review the feedback before we collected the feedback from participants. Participants were asked to write the same task. Participants in the control group were not given any feedback and they simply repeated the same English writing task. We also collected the second writing samples using Google form during class time. Finally, we asked participants to complete a survey concerning their experiences with using AI and MT and their perceptions toward their repeated writing tasks.

3.6 Data Analysis

The project progressed very quickly and we moved onto the next stage of data analysis and interpretation. We evaluated changes in participants writing from the impact of feedback by looking at four types of measurements: holistic human rater scoring (overall evaluation by the graduate students), syntactic accuracy (percentage of accurate verb tense and global accuracy, fluency (total number of words), and complexity (MLT (mean length of T-units) and clause/T-unit ratio). We evaluated changes (improvements) in participants' writing from the effect of feedback by looking at scores of their first writing and second writing. We compared the differences between and across the three groups.

Results of our project found different results using different measurement tools. Based on holistic human rater scoring, both AI and MT groups outperformed the control group (who received no feedback). In addition, the AI group had a larger improvement than the MT group. However, based on the accuracy, fluency and complexity measurements, control group surprisingly outperformed the experimental groups. The AI group also decreased in performance in writing 2. These results indicate that analytical results can be very different from human holistic ratings. The survey found that most participants use AI and MT once a week or more. It also found out that some participants do not perceive AI and MT feedback to be effective. Some participants indicated that they wanted teacher feedback.

3.7 Presenting Research Results

The highlight of this research project was our presentation at the ACIE Conference in Osaka in March 2025, marking a major academic milestone for both myself and my students. The process of preparing for this conference involved many meetings with the students inside and outside of class time, in person and online, on weekdays and over the weekends. At the presentation, we delivered our well-practiced presentation where we all took active parts and enjoyed our parts. We engaged in discussions with the audience and further enriched our experiences as researchers of our own project and also researchers in the field of applied linguistics.

4. Conclusion

This collaborative research project provides a unique opportunity for graduate students to engage in hands-on academic inquiry while developing critical thinking, analytical, and research skills. It fosters a sense of academic ownership, enhances motivation, and equips students with practical experience in navigating research. The research process was a dynamic one, involving multiple stages such as brainstorming and topic selection, literature review, research design, ethical considerations and research ethics applications, data collection, analysis, and presentation. This paper aims to document our journey, providing an in-depth reflection on the processes we undertook, the challenges we encountered, and the key takeaways from our collaborative efforts.

References

- Goh, C. C. M., & Burns, A. (2012). *Teaching speaking: A holistic approach*. Cambridge University Press.
- Li, Z., & Hegelheimer, V. (2013). Mobile-assisted grammar exercises: Effects on self-editing in L2 writing. *Language Learning & Technology*, 17(3), 135-156.
- Lui, K. J. (2016). *Crossover design: Testing, estimation, and sample size*. Wiley.
- Lyster, R. (2004). Research on form-focused instruction in immersion classrooms: Implications for theory and practice. *Journal of French Language Studies*, 14(3), 321-341. doi:10.1017/S0959269504001826
- Lyster, R., & Saito, K. (2010). Oral feedback in classroom SLA: A meta-analysis. *Studies in Second Language Acquisition*, 32(2), 265-302.
- Mackey, A. (2006). Feedback, noticing and instructed second language learning. *Applied Linguistics*, 27(3), 405-530.
- Nicholas, H., Lightbown, P., & Spada, N. (2008). Recasts as feedback to language learners. *Language Learning*, 51(4), 719-758.
- Paulus, T. M. (1999). The effect of peer and teacher feedback on student writing. *Journal of Second Language Writing*, 8(3), 265-289.
- Sasaki, M., Mizumoto, A., & Matsuda, P. K. (2024). Machine translation as a form of feedback on L2 writing. *International Review of Applied Linguistics in Language Teaching*, 1-26.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129-158.
- Williams, J. (2005). Form-focused instruction. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 671-691). Routledge.

Appendix A

Class reading list collaboratively developed by students

- Ishizaka et al. (2013). Development of word fluency and expressive vocabulary in Japanese kindergarten children. *Kitasato Med Journal*, 44, 26-30.
- Quackenbush, H. C. (1977). English Loanwords in Japanese: Why Are They Difficult for English-Speaking Students? *The Journal of the Association of Teachers of Japanese* 12, no. 2/3: 149–73. <https://doi.org/10.2307/489164>.
- Chang, Y. H. (2023). Flipped classroom for developing Japanese language proficiency: A study of non-Japanese major learners in University of Taiwan, *台灣日語教育學報*, 41, 31-60.
- Kubota, M. (2017). Post study abroad investigation of kanji knowledge in Japanese as a second language learners. *System (Linköping)*, 69, 143–152. <https://doi.org/10.1016/j.system.2017.07.006>
- Junjie, S., Nishihara, Y., Yamanishi, R., & Fukumoto, J. (2017). Analysis of Dialogues Difficulty in Anime Comparing with JLPT Listening Tests. *Procedia Computer Science*, 112, 1345–1352. <https://doi.org/10.1016/j.procs.2017.08.022>
- Fukunaga, N. (2006). Those Anime Students’: Foreign Language Literacy Development through Japanese Popular Culture. *Journal of Adolescent & Adult Literacy* 50, no. 3: 206–22. <http://www.jstor.org/stable/40013700>
- Siegal, M., & Okamoto, S. (2003). Toward Reconceptualizing the Teaching and Learning of Gendered Speech Styles in Japanese as a Foreign Language. *Japanese Language and Literature* 37, no. 1: 49–66. <https://doi.org/10.2307/3594875>.
- Goss, S. A (2008). Critical Pedagogy of Lexical Accent in L2 Japanese: Insights into Research and Practice. *Japanese Language and Literature* 52, no. 1: 1–24. <https://www.jstor.org/stable/10.2307/26739436>
- Masuda, K. (2018). The Second Language Acquisition of Challenging Japanese Locative Particles, ni and de: A Usage-Based and Discourse Approach. In *Cognitive Linguistics and Japanese Pedagogy: A Usage-Based Approach to Language Learning and Instruction* edited by Kyoko Masuda. De Gruyter Mouton.
- Bengtsson, A. (2023). The Effects of Extramural Language: Relationships Between Engagement in Japanese Language Activities and General Japanese Language Proficiency. PhD diss., Stockholm University. ISBN: 978-91-8014-555-8 (electronic). <https://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-222652>
- Hotta, T. & Horie, K. (2018). L2 Acquisition of the Japanese Verbal Hedge omou: A Prototype Approach. In *Cognitive Linguistics and Japanese Pedagogy: A Usage-Based Approach to Language Learning and Instruction* edited by Kyoko Masuda. De Gruyter Mouton.
- Fukunaga, T. (2023). L2 Writing Development Through Two Types of Writing Task Repetition. *International Review of Applied Linguistics in Language Teaching* 61, no. 3: 1109–38. <https://doi.org/10.1515/iral-2021-0144>
- Sasaki, M., Mizumoto, A., & Matsuda, P. K. (2024). Machine Translation as a Form of Feedback on L2 Writing. *International Review of Applied Linguistics in Language Teaching*. <https://doi.org/10.1515/iral-2023-0223>

Appendix B
Research Participation Consent Form
Date to be decided

You are being asked to participate in a Japanese proficiency development research. The purpose of this research is to improve the writing ability of JSL (Japanese as a second language) learners. Your help is important in order to improve the current Japanese education. Please understand the following points before participating in this research.

- Participation is voluntary (you may withdraw your consent at any point).
- You are free to ask questions about this research anytime.
- Your personal information will not be disclosed in any paper or presentation related to this study.
- Your personal information will only be accessed for the purpose of this research and for this duration of this study.
- The results of your participation will not affect your course grade.
- Upon completion of this study you may request the results at lee.nancy.hmt@osaka-u.ac.jp

I agree to take part in the study on Japanese proficiency development by four graduate students. This paper presents a reflective report of a collaborative research project I conducted with some graduate students at the Graduate School of Humanities, Osaka University. The project examined the impact of AI-driven and machine translation-driven feedback on ESL (English as a second language) writing. This joint research project originated from a graduate course I taught called Introduction to Applied Linguistics where only international students signed up and they were interested in topics different from my original syllabus. To cater for their needs and interests, I decided to modify the syllabus for the above course. I invited the students to collaboratively develop a new syllabus and through reconstructing the syllabus, we found a gap in the existing literature. Eventually, the course evolved from reading and understanding existing literature to actually conducting a joint research project. Based on this research gap and students' high interests, we discussed the feasibility of conducting a research project as a class project within the remaining semester time and our potential obstacles. Students also actively debated over appropriate topics for this possible class based research project. Finally, students decided to replicate a writing study looking at machine translation as a form of feedback on L2 writing, comparing the impact of machine translation and one professor.

Shzh-chen Nancy Lee
Graduate student names omitted

Signature _____ Date..... (E.g. 2025/1/25)