



# Investigating the effect of computer-mediated feedback via an LMS integration in a large-scale Japanese speaking class

Yuka Kataoka<sup>1</sup> · Achmad Husni Thamrin<sup>1</sup> · Rodney Van Meter<sup>2</sup> · Jun Murai<sup>3</sup> · Kotaro Kataoka<sup>4</sup>

Received: 6 June 2021 / Accepted: 24 July 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

## Abstract

In second or foreign language (SFL) education, oral corrective feedback (OCF) is widely used to individually correct students' erroneous utterances during classroom hours. However, students cannot have sufficient opportunities for oral production and personalized feedback during classroom hours if a class is large-scale with many students. This paper addresses the lack of OCF opportunities in a large-scale class, assuming the causes to be the severe time constraints and the teachers' labor intensiveness in examining students' utterances and generating OCF. This research proposes using computer-mediated feedback (CMF) outside classroom hours to complement OCF in an online, semiautomated, and scalable fashion. This paper implements Oral Repetition Practice (ORP) Gym to provide students with sufficient opportunities for speaking practice through two types of CMFs; Hybrid Recast to enhance the recognition of errors and Explicit Error Correction to make errors detectable and correctable. Online External Assistant (OEA) is a mechanism used to increase the amount and quality of feedback by distributing the workload for scoring and generating CMF. The evaluation was conducted as a classroom observational study by introducing ORP Gym to a spoken Japanese SFL basics course with 55 students at an Indian university. Compared with the students who did not utilize ORP Gym, those who utilized ORP Gym performed more ORP and exhibited significant score improvement in the posttest. This research contributes to enabling CMF in large-scale SFL classes and empirically and statistically proving the improvement of the learning effect, including *uptake* and *repair*, by CMF using ORP Gym and an OEA.

**Keywords** Second or foreign language · Oral corrective feedback · Computer-mediated feedback · Oral repetition practice · Learning management system

---

✉ Yuka Kataoka  
shori@sfc.wide.ad.jp

Extended author information available on the last page of the article

## 1 Introduction

Many studies have investigated the effect of oral corrective feedback (OCF) on second or foreign language (SFL) development and acquisition during the past two decades (Ellis et al., 2001; Loewen & Philp, 2006; Lyster & Ranta, 1997; Panova & Lyster, 2002; Sheen, 2004) as Fu & Nassaji, (2016) summarized. OCF has contributed to the language acquisition process, and played an important role as a scaffolding for involving students in the conversation (Lyster et al., 2013). The previous studies have been traditionally conducted in laboratory studies that ‘involve interaction between two individuals, usually a researcher and a learner’ and recently expanded to classroom studies that ‘involve interaction between a teacher and an intact class of students’ (Lyster et al. (2013), p.2). The classroom studies can be categorized to quasi-experiments (Li, 2010), a type of observational study, based on conversation analysis by audio recordings of generally small to medium-sized classes. There is a practical problem that teachers’ OCF cannot fully detect and correct individual students’ errors in an actual classroom owing to time constraints and poor listening quality due to the distance between the teacher and the student. Li (2010) also emphasized that OCF is unlikely directed towards individual students in a classroom context because there is more distraction than in a laboratory context (p.316). Therefore, classroom studies are important to understand the actual learning effect of OCF.

In the context of the classroom studies, many previous studies revealed that OCF is ineffective in terms of student *uptake* (students’ response to OCF) and *repair* (students’ error correction) in a classroom. The supporting theory behind OCF studies dates back to the noticing hypothesis. Schmidt (1990) proposed that noticing, which means the learner’s conscious process, is an essential cognitive process to promote language acquisition. However, Lyster’s studies revealed that *recasts* were frequently provided by class teachers, accounting for 55% of all six OCF types, resulting in lower effectiveness in terms of student uptake and repair than other forms of OCF in French immersion classes (Lyster & Ranta, 1997) and adult English as foreign language (EFL) classes (Panova & Lyster, 2002). Recasts were implicitly (Long, 1996; Long & Robinson, 1998) favoured in the discourse because it is a non-threatening way to correct students’ errors and helpful for students to participate in conversations without interrupting the flow of conversations during classroom hours (Ellis et al., 2001; Seedhouse, 1997; Sheen, 2004). To compensate for the ineffectiveness of recasts in grammatical accuracy (Kim & Han, 2007; Lyster, 1998; Nabei, 2005), existing approaches such as choral repetition and read-aloud in isolation (Lyster & Mori, 2006) and “focus on form (FonF)” were effective at making students notice linguistic forms in the classrooms (Ellis et al., 2001; Long & Robinson, 1998). Loewen and Philp (2006) also claimed that teachers’ phrasal, prosodic, and discursal cues could reduce the ambiguity of recasts (p.541). In addition, students were more likely to benefit from FonF when it occurred in small group and one-on-one interactions than in whole-class interactions (Ellis, 2016; Nassaji, 2013). Especially if the proficiency level of students is low, they benefit more from interactions with a teacher (Ammar & Spada, 2006).

While the previous studies mainly focused on the importance of students' oral production and one-on-one interactions between a teacher and a student, the approaches based on the existing OCF have limitation in a large-scale class, defined as having 50 or more students in a classroom per teacher (Holliday, 1994). In fact, the previous OCF studies were rarely conducted in a large-scale class. Due to the high demand for international industry and commerce, the number of Japanese learners abroad is increasing significantly over the past 40 years (Japan Foundation, 2018). There have been cases where more than 100 students attended a Japanese basics class at an Indian university since 2016 due to the increasing demand for Japanese language education in higher education (Kataoka et al., 2018). The number of Japanese language teachers and native speakers who work abroad is very small compared to the number of students abroad. Therefore, such a large-scale SFL class suffers from limited human resources and time constraints to provide learning opportunities with detailed individual support.

Although students need sufficient oral practice and individual support to improve their erroneous utterances, especially in basics to intermediate language classes (Fukada, 2013; Ikeda & Fukada, 2012), there is a lack of opportunities for students' individual oral repetition practice to detect and correct errors in a large-scale class. Teachers' OCF was supposed to meet individual students' needs, but such personalized feedback was hard to provide to all students during classroom hours due to severe time constraints (Kataoka et al., 2018). Therefore, this study aimed to propose computer-mediated feedback (CMF) outside classroom hours using an online oral repetition practice support system, which is called ORP Gym, to complement the ineffectiveness and limitation of the existing OCF in a large-scale class. Since the existing studies had not established approaches to complement OCF in a large-scale SFL class, this research plays an important role in establishing an approach using CMF to tackle the problem. In addition, many existing studies had claimed the importance of quasi-experimental (classroom observational) studies because experiments in laboratory studies were not enough to prove the effectiveness of the approach in an actual SFL class. To evaluate the effect of CMF in a classroom, we conducted a classroom observational study using ORP Gym as a part of a large-scale course at an Indian university. Therefore, this research provides invaluable findings and insights into developing the approach using ORP Gym and evaluating the effect of CMF in a large-scale class.

## 2 Literature review

With the progress of technology development, CMF, which is teachers' manual feedback via computer, has drawn attention as promising support for language learners' performance improvement such as speaking or writing skills in a computer-assisted language learning (CALL) environment (Tuffley & Antonio, 2015). However, the effect of CMF in a large-scale class is still unclear. According to the latest systematic survey of CMF studies in speaking classes for the past ten years (Zhang, 2021), the number of existing studies of CMF in speaking classes is very small compared with writing classes. A study developed a mobile application with audio recordings

and computer-generated feedback (CGF) (Ahn & Lee, 2016) as one of the existing approaches in a large-scale class. However, their evaluation focused on surveying over 300 students regarding students' perceptions of the mobile application, not investigating their actual speaking performance. As another approach, Fang et al. (2021) developed a mobile application with information gap activities without audio recordings to complement the lack of task-based classroom activities in a large-scale class. Since few existing studies installed audio recording speaking practice with CMF for speaking classes, a lack of understanding of the effect of CMF in a large-scale class persists.

While speaking classes suffer from the limitation of opportunities for oral practice during classroom hours (Buckingham & Alpaslan, 2017), CMF and web-based language learning (WBLL) have been studied to complement the opportunities for oral practice outside classroom hours in the past decade. The following overview discusses the necessity of CMF and WBLL for a large-scale speaking class using a learning management system (LMS).

## 2.1 The effect of computer-mediated feedback (CMF) in a speaking class

Rassaei (2019) claimed that investigating the effect of CMF needs to expand studies in language classrooms. In the past decade, the previous studies investigated the effect of CMF outside classroom hours in two fashions using 1) asynchronous audio-visual speaking activities (A/Vs) such as speaking quizzes with teachers' voice recording, which require students' monologue audio recordings, and 2) synchronous computer-mediated communication (CMC) such as oral conversational activities over audio-video conferencing tools such as Skype.

Buckingham and Alpaslan (2017) showed the effectiveness of A/Vs with CMF for improving speaking scores compared with the paper-based worksheet activities. The previous studies using voice-blogs or podcasts outside classroom hours (Ducate & Lomicka, 2009; Hsu, 2016; Sun, 2012) did not monitor students' oral performance periodically nor provide linguistic feedback. Therefore, Buckingham and Alpaslan (2017) examined the effect of CMF outside classroom hours in two Turkish language classrooms (Grade 3) by a controlled experiment. The experimental group practiced asynchronous A/Vs using audio recordings and got CMF (written feedback by a teacher). On the other hand, the control group practiced using a paper-based worksheet in the same form as the A/Vs and got written feedback from the teacher. As the result of the 4-month experiment, the experimental group significantly improved the speaking test scores between pretest and posttest compared to the control group. Furthermore, the posttest scores in the experimental group were significantly higher than the posttest scores in the control group.

Some studies demonstrated the effectiveness of different forms of CMF for second language (L2) speaking development. Rassaei (2019) investigated the effect of computer-mediated text-based and audio-based CMF in Iranian EFL classes. The result indicated that audio-based CMF was more effective than text-based CMF in speaking improvement. Tseng and Yeh (2019) also examined the two types of CMF in EFL classes: text-based and audio-based CMF. It was proven that 1) text-based

CMF was effective in linguistic accuracy, and 2) audio-visual (A/V) CMF was effective in students' pronunciation. The study concluded that further research on combining two types of CMF is needed to understand the potential benefits for students in both roles.

CMC is also a practical approach based on one-on-one online video communication to enhance L2 learners' speaking proficiency. Kato et al. (2016) investigated the effect of CMC outside classroom hours. The study showed significant improvement between pretest and posttest in the speaking abilities of Japanese and American participants. However, the previous CMC studies did not focus on corrective feedback (CF) to reduce erroneous utterances. The following two studies investigated the effect of CMF on CMC. Rassaei (2017) investigated the effect of two different conditions of recasts: 1) face-to-face recast and CMF recast through Skype video call in an Iranian EFL class. As the result of a 10-day experiment, CMF recast was as effective as face-to-face recast in facilitating L2 speaking development. Akiyama (2017) also investigated the effect of three types of OCF (recast, explicit correction, and clarification request) on CMC, which is named 'eTandem'. As a result, the study clarified that the uptake rates were higher when the students got their preferred type of CF than when they got the unpreferred type of CF.

As automatic immediate feedback, some studies demonstrated the potential of CGF. Neri et al. (2008) and Cucchiaroni et al. (2009) presented the effect of CGF employing automatic speech recognition (ASR) in Dutch L2 speaking classes. Although the accuracy of CGF was not 100%, the studies exhibited the positive effect of CGF in improving students' mispronunciation. Ahn and Lee (2016) also proved the positive responses from EFL students to their mobile-based application, 'Speaking English 60 Junior' by enabling ASR. Vries et al. (2016) examined the effect of CF on an ASR-based CALL system with CGF, which was called 'GREET system'. The system improved Dutch speaking and grammar between pretest and posttest in both groups with-CF and without-CF. However, there was no significant difference in speaking and grammar tests between the two groups at the posttest. This result indicates that the additional learning effect of CGF in error correction has not been proven yet.

Compared with CGF, CMF is still labour-intensive (Ashwell & Elam, 2017). However, CGF cannot wholly replace CMF in students' effective error correction (Sherafati et al., 2020; Thomson, 2011). In L2 writing, CMF significantly improved students' writing ability on a delayed posttest compared with CGF (Sherafati et al., 2020). While CMF requires teachers' manual evaluation, CMF can generate and provide meaningful linguistic feedback covering all categories of morphology, syntax, and pronunciation errors.

## 2.2 The effective use of web-based language learning (WBLL) in a speaking class

WBLL can provide multimedia-supported interactive web-based learning activities with CMF, especially in computer-assisted pronunciation training in speaking (Thomson, 2011; Veselovska, 2016). According to Cong-Lem (2018), Web 2.0, including a web-based LMS, has played an important role in enhancing L2 speaking

performance. An LMS is a great platform for students to engage in language learning without time and space restrictions. It is also helpful for teachers to monitor students' learning progress and provide better feedback, even in a large-scale class (Kataoka et al., 2018). Ideally, students need high-quality feedback on their assessment in any field. Although generating high-quality feedback is labour-intensive (Ashwell & Elam, 2017), Tuffley and Antonio (2015) insist that computer-mediated assessment and CMF via an LMS integration are cost-effective and can be applied to large-scale courses.

Since LMSs such as Moodle (Moodle Project, 2022) have become increasingly available and popular since around the year 2010, many studies have investigated the effective use of LMSs for language classrooms. With audio capabilities available on an LMS, teachers can provide sufficient opportunities for their students' oral practice and assessments outside classroom hours by taking advantage of students' portable devices and Internet access. Compared with other self-paced learning applications (Bajorek, 2017), the learning contents on an LMS can be managed by teachers flexibly related to the classroom activities. 'Speak Everywhere' (Fukada, 2013) and 'Voice Shadow' (Kumai & Paul, 2013) enabled multiple computer-mediated oral assessments and CMF outside classroom hours in small to medium-sized classes using an LMS. Such teachers' CMF, which is given to students electronically and asynchronously, is categorized as delayed feedback compared with OCF during classroom hours (Ikeda & Fukada, 2012). However, oral repetition practice (ORP) with such CMF is effective for students to improve word accentuation (Yoshida & Fukada, 2014).

These studies, covering the past decade, proved the effect of CMF and the effective use of WBLL outside classroom hours to complement the opportunities for ORP in an SFL class. However, Elola and Oskoz (2016) claimed that the affordances and limitations of providing feedback in diverse, technology-driven, and multimodal ways had not been widely explored in SFL education. Rassaei (2019) also asserted that the effects of various forms of CMF on L2 development have still not been established adequately, especially in speaking classes. In addition, the previous CMF studies assumed small and medium-sized classes of around 20 to 30 students, not large-scale classes. Buckingham and Alpaslan (2017) and Rassaei (2019) enabled audio-based CMF using Microsoft Office or Adobe Acrobat audio annotation functions. However, none of the previous studies focused on teachers' workload reduction/mitigation/distribution in managing audio files and examining students' audio answers. Therefore, implementation issues have not been discussed well when CMF studies are applied to large-scale language classes.

Since our research focuses on a large-scale class, providing A/V speaking quiz activities with CMF via an LMS integration is a relatively feasible and reasonable option for deployment in a large-scale class (Tuffley & Antonio, 2015). CMC is also effective in enhancing L2 learners' speaking proficiency (Akiyama, 2014, 2017). However, CMC is practically difficult to implement in our educational setting because it requires the same number of native language instructors as the number of students. The previous WBLL approaches demonstrated the benefit of multiple oral assessments by integrating an LMS (Fukada, 2013; Ikeda & Fukada, 2012; Kumai & Paul, 2013). Although Yoshida and Fukada (2014) proved that ORP with CMF

effectively improved students' word accentuation on the LMS, the uptake and repair rates have not been detailed in the various categories.

Therefore, this study aims to 1) propose CMF using an online oral repetition practice support system via an LMS integration and 2) investigate the effect of different types of CMF by measuring uptake and repair rates in a large-scale SFL class. Our system provides two types of CMF: 1) synchronous hybrid (audio and text) feedback and 2) teachers' asynchronous personalized linguistic text-based feedback. However, consistent error correction is arduous during classroom hours from the students' perspective (Lasagabaster & Sierra, 2005). In addition, Internet or web technologies are not always welcomed by all language learners readily (Wang & Sun, 2001). This study introduces the proposed approach with new technology into an actual language class by practically examining the learning effect and students' satisfaction throughout the course. This paper provides valuable practical knowledge and insight into the effective use of CMF to enhance students' erroneous utterances.

The research questions (RQs) in this paper, which aim to provide sufficient and individual ORP with CMF on the online support system, are as follows.

1. What type of CMF will be utilized actively by students?
2. What type of CMF will contribute to the students' retries and repaired answers?
3. Does the increased frequency of ORP improve uptake and repair rates?
4. To what extent does the increased frequency of ORP improve the exam score?
5. What is the amount of the teacher's workload and time outside classroom hours when we provide CMF?
6. Are the students satisfied with the course introducing the proposed system?

### **3 Approach: ORP Gym**

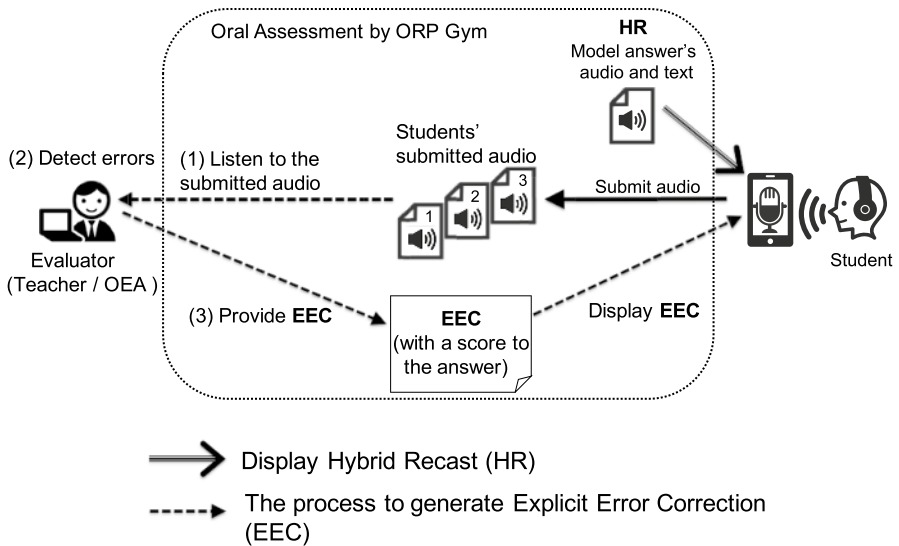
This paper proposes ORP Gym, which is a web-based oral repetition practice support system for a large-scale Japanese SFL education. The main features of ORP Gym are summarized as follows.

- A web-based LMS with incentive management that encourages students to perform ORP as often as possible outside classroom hours.
- Hybrid Recast (HR) and Explicit Error Correction (EEC) for making errors detectable and correctable for each student with audio and text.
- Online External Assistant (OEA) for ensuring the quality of EEC and mitigating the workload of teachers to produce EEC.

#### **3.1 System overview and design**

ORP Gym is used by students 1) to attempt an online oral assessment by submitting an audio answer file, 2) to acquire HR and EEC to the answer, 3) to perform ORP for the improvement of Japanese utterance, and 4) to retry for the score improvement of the assessment. Figure 1 illustrates the system design of ORP Gym, which





**Fig. 1** The proposed online oral assessment by ORP Gym

extends an LMS to interactively conduct an online course activity between teachers and students. Teachers can observe the uptake and repair of each student with the audio answer files submitted by students, the log data of the student activity on ORP Gym, and the assessment score available on a single platform.

The remainder of this section details HR, EEC, and how they are integrated as part of ORP Gym.

### 3.2 Hybrid Recast (HR)

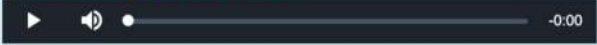
HR is a combination of 1) audio of a correct answer to a question and 2) its correct transcription (in *romaji*, which is a writing method of Japanese using the Roman alphabet, with pitch symbol), as shown in Fig. 2, and can be categorized as *recast*. HR is available immediately after any audio answer is submitted by a student regardless of its correctness. However, the use of HR is not mandated, and a student needs to voluntarily read the text and listen to the audio answer by clicking the play button on ORP Gym. In most cases, a student would utilize HR without knowing their score and any mistakes they made. Because HR itself does not contain specific instructions to perform an ORP, a student would voluntarily perform it by imitating the model answer provided by a native Japanese speaker.

The concept of HR can be applied to any question that has a correct answer, including time expression, verb forms, and read-aloud. Currently, the workload to produce HR increases in proportion to the number of answers. Such workload can be reduced by optimizing the HR preparation process by partial recording and audio concatenation to produce HRs for a question with many answer patterns, such as time expression.




Look at the picture below and listen to the question.  
Answer the question using the audio submission.

↓ **audio question**




↓ **student's submitted audio**



**HR**

(Model answer - もはんかいとう)

ima, yo\~ji ni-ju\~ppun desu. ← **model answer's text  
(Romaji and Pitch symbols)**



↑ **model answer's audio**

Fig. 2 An example of HR

### 3.3 Explicit Error Correction (EEC)

EEC is an explicit correction that 1) points out an error in the answer and 2) gives specific instructions to correct the error. EEC is provided in the form of the notation for both the erroneous part in the original answer and its correction. Figure 3 shows an example of EEC when a student has a pronunciation error in a Japanese word, *tegami* (which means a letter in English). The score of the original answer is accompanied together with EEC. Therefore, a student is informed of the gap between the original answer and the ideal utterance through the EEC and the score.

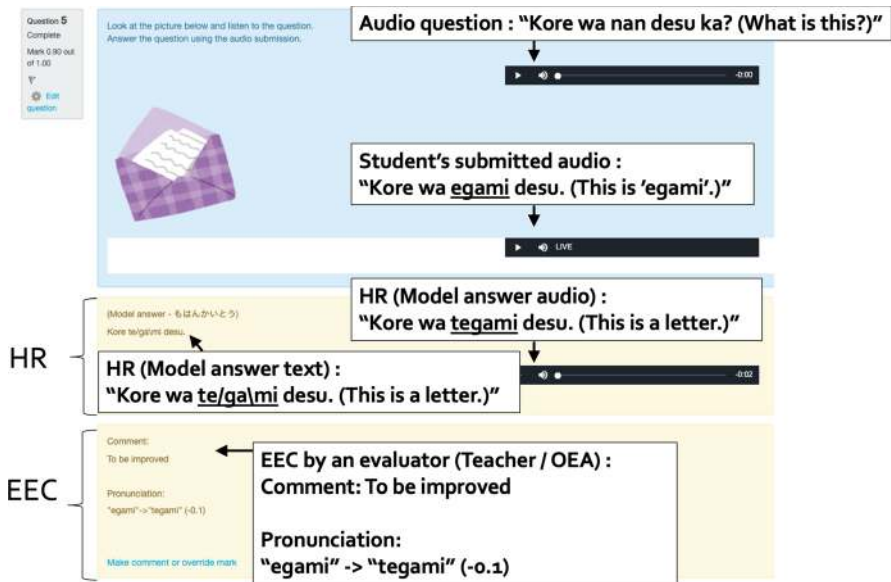


Fig. 3 An example of EEC

In this system, all errors are manually identified by a teacher. Producing EEC involves the following three steps as shown in Fig. 1: 1) listening to the submitted audio, 2) detecting errors, and 3) providing EEC for each answer on the system. Therefore, these steps take a significant amount of time, and the workload of the teacher increases in proportion to the number of EEC. In our system, the evaluation task is coordinated so that students can receive EEC within 48 hours after submission in the Japanese SFL course with approximately 50 students.

### 3.4 The procedure of enhancing uptake and repair by HR and EEC

ORP Gym takes advantage of the incentive to improve the assessment score by 1) allowing students to retry the quiz multiple times and 2) recording the latest score as the final score. Figure 4 shows a flow diagram of student uptake and repair enhancement by HR and EEC.

While HR depends on students themselves to recognize, detect, and correct an error, EEC allows the students to have assistance from evaluators to do so. The ideal flow to maximize the benefit of HR and EEC is to use both. Once a student generates uptake using HR or EEC, the student can retry the oral assessment and exhibit repair which includes improved utterance or error correction. On the other hand, if a student does not attempt a retry, it is presumed *no uptake* by ORP Gym.

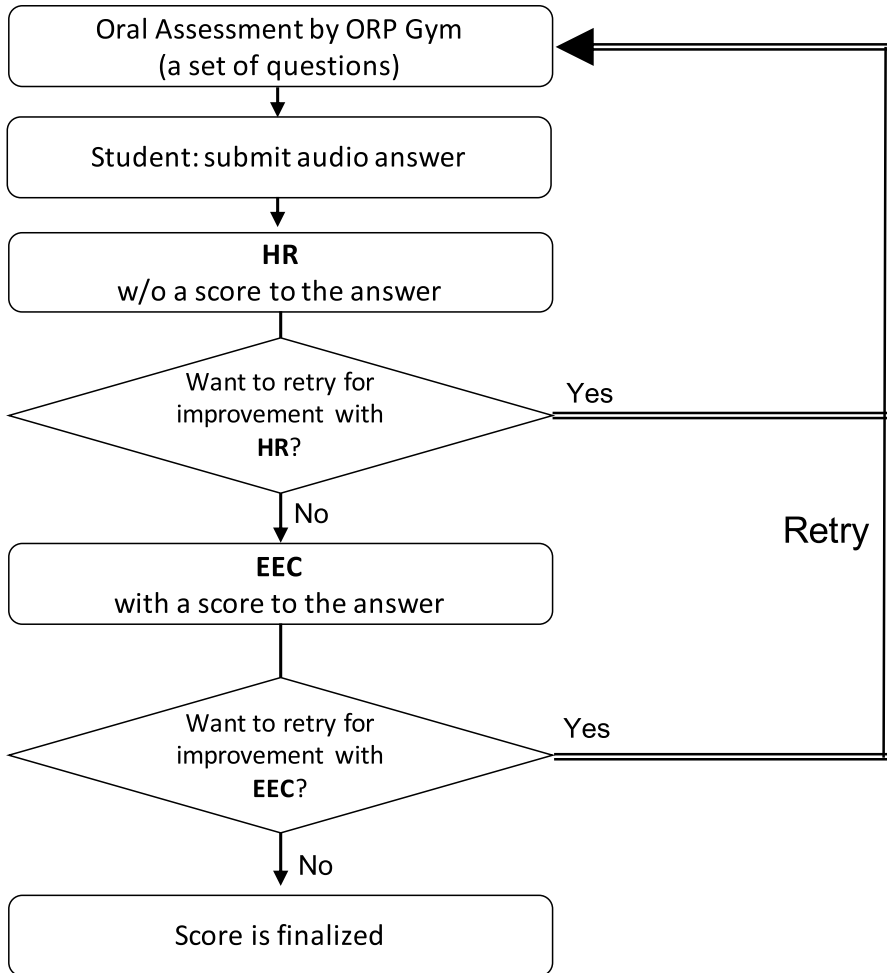


Fig. 4 Flow diagram of student uptake and repair enhancement through ORP Gym

### 3.5 Online External Assistant (OEA)

While producing HR is a one-time task that happens only when an assessment question is created, producing EEC is labour-intensive work for a teacher (Ashwell & Elam, 2017). However, we believe that manual feedback could be flexibly customized to all students' levels including absolute beginners' level and encourage students' motivation and interest in their language learning. Especially, characteristic accents and intonations due to students' mother tongue tend to cause various erroneous utterances. Manual evaluation allows teachers to understand such utterances and provide appropriate personalized feedback and support.

ORP Gym introduces OEA to take advantage of the online nature of the system. OEA enables the distribution of evaluation workload by allowing reasonably

qualified assistants to evaluate the answers in addition to the teacher. Even when the number of students or the frequency of oral assessments increases during the course, OEA ensures the amount and quality of EEC will be as high as that given in the class, and the benefit of OEA will be significant for a short-term, large-scale class.

To maintain the quality of EEC by OEA, the clarity and consistency of the scoring criteria of the audio answers are important. We introduced an instruction handbook written in Japanese (Supplementary file 1) for OEA to instruct how to evaluate the submitted audio answers and generate EEC without evaluation discrepancy. The handbook contains a frequently asked questions (FAQ) section, and the know-how of assessment evaluation can also be shared between teachers and OEAs. Online training was conducted for OEA with simulated scoring modules on ORP Gym before the course started.

### 3.6 System implementation

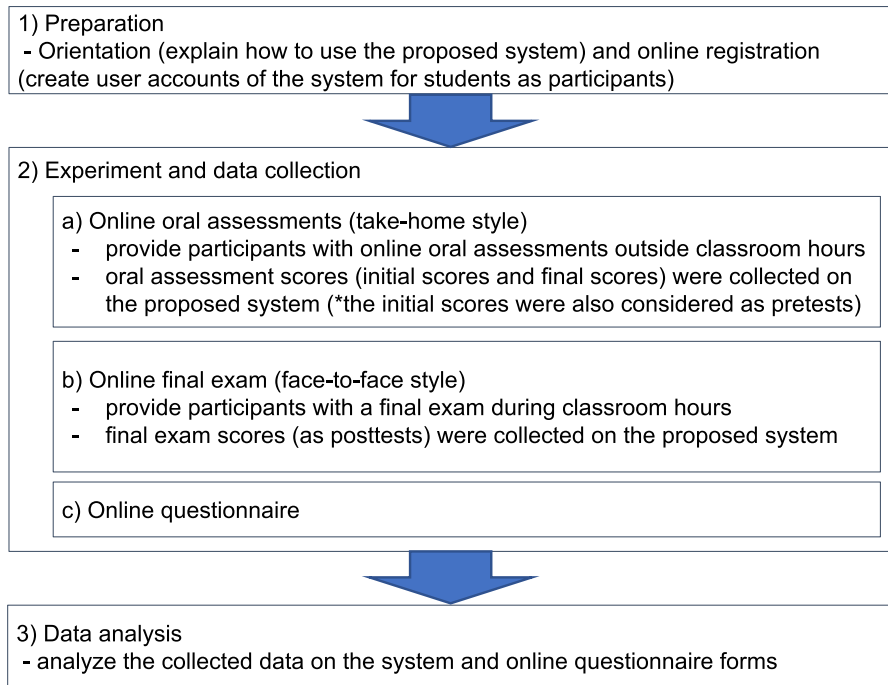
ORP Gym runs on top of Moodle (Moodle Project, 2022), a widely used LMS, and uses RecordRTC (Real-Time-Communication) for Atto Plugin (Federico, 2018) for the audio and video recording and submission. A total of 108 different questions with HR were prepared on the question bank including 18 yes-no questions, 18 when/what/how questions, and 72 time expression (how to tell the time) questions. In addition, 1 read-aloud quiz of self-introduction was also included. Students were allowed to record and submit up to a few sentences (within 30 seconds) to answer a quiz.

## 4 Methods

This research attempted to incorporate ORP Gym in a large-scale SFL class of more than 50 students per teacher to improve students' utterances outside classroom hours. To investigate the effect, we deployed ORP Gym as a part of an elective Japanese SFL speaking course, 'Spoken Japanese Basics (SJB)' at an Indian university from January 2nd to February 8th (around five weeks), 2019. The course was a total of 15 face-to-face classroom hours.

### 4.1 Experimental procedure

Figure 5 shows the experimental procedure to investigate our six research questions in this research. There were 3 phases: 1) preparation, 2) experiment and data collection, and 3) data analysis. The initial two phases were conducted during the 5-week course. The data analysis phase followed afterward. We conducted orientation and online registration during the preparation phase to explain how to use ORP Gym and create user accounts before the experiment started. During the experiment and data collection phase, we conducted a) online oral assessments outside classroom hours, b) an online final exam during classroom hours, and c) an online questionnaire.



**Fig. 5** The experimental procedure in this research

#### 4.1.1 Oral assessments

ORP Gym was used for conducting two online oral assessments: the 1st assessment from January 22nd to 30th (9 days), and the 2nd assessment from February 1st to 8th (8 days). The 1st assessment consists of 4 quizzes (maximum score = 10). The quizzes were randomly selected from a question bank of 20 yes-no/what questions about nouns, and 72 time expression questions. One common read-aloud quiz of their self-introduction was given to all the students along with the 1st assessment to investigate the uptake and repair rates. The 2nd assessment consists of 5 quizzes (maximum score = 10) that were randomly provided from 16 when/how questions about adjectives and verbs. Although the oral assessments were not mandatory to complete the course, they were provided for students' voluntary practice. The students were allowed to attempt the quizzes multiple times during each assessment period.

#### 4.1.2 Pretest and posttest

We used the first trial of the oral assessments as a pretest. As a posttest, the individual online final speaking exam was conducted during classroom hours on Feb 8th after the oral assessment period. The exam contained 1 common read-aloud quiz of their self-introduction and Part 1 (a set of 4 quizzes) and Part 2 (a set of 5 quizzes)

in the same style as the 1st and 2nd assessments. Although the quizzes on the exam were different from the quizzes on oral assessments to eliminate the effect of repeating exactly the same quizzes, the degree of difficulty of the quizzes was the same level. For example, sentence patterns of assumed answers were the same, but words such as nouns, adjectives, and verbs in the sentence were different. However, words were randomly selected from a given unit of instruction during the course to ensure the homogeneity between pretest and posttest. Forty-five students took the exam on ORP Gym using their mobile devices (laptop computer or smartphone) and earphones, with a time limit of 20 minutes (Fig. 6). The functions of HR and retry were deactivated on the system, and EEC was not provided on the exam.

## 4.2 Online questionnaire

An online questionnaire was created by Google Forms (Google, [n.d.](#)) and provided to students who completed the course and the experiment. The questionnaire consists of two parts: 1) three questions with a 5-point Likert scale and 2) an open-ended question (free descriptive style) regarding the course introducing the proposed system.

## 4.3 Participant details

A total of 45 students out of 55 registered students completed the course by meeting the grading criteria (passing the final exam). Ten students were audit students who only participated in the class activities during classroom hours and did not take oral



**Fig. 6** The individual online final speaking exam using students' mobile devices and earphones

assessments and examinations on our LMS. Their nationality was Indian, and they were undergraduate students ranging from 18 to 21. The level of all the participants was absolute beginner, and all had no prior experience learning Japanese. However, they were capable of commanding multilingual languages such as English, Hindi, or other different state languages in India.

#### 4.4 Data analysis

Regarding RQ1 and RQ2, we conducted a statistical test to compare the learning gain between the initial and the final scores during their oral assessments period and between the pretest and the posttest. The Wilcoxon signed-rank test was used as a nonparametric test for the two related samples because the data did not follow a normal distribution based on the result of the Shapiro-Wilk normality test ( $p < 0.01$ ). IBM SPSS Statistics (version 27) was used for the statistical analysis.

### 5 Result

To investigate the six research questions in this paper, the following six indicators are extracted from the data collected through the deployment of ORP Gym and a questionnaire.

1. Relationship between utilization of ORP Gym and score improvement
2. The number of retries and repaired answers
3. Linguistic uptake and repair rates by ORP Gym
4. Comparison of score improvement in the final examination
5. The workload distribution by OEA and the average time to produce EEC
6. Students' satisfaction with the course and CMF

#### 5.1 Relationship between utilization of ORP Gym and score improvement

The students exhibited different usage of ORP Gym and were classified into two categories: *no retry using ORP Gym* and *retry using ORP Gym* according to the access log and the audio data on the system. These categories were classified into four groups based on the action patterns observed through ORP Gym: Group 1 (No uptake and no repair), Group 2 (Uptake and repair by HR only), Group 3 (Uptake and repair by HR and EEC), and Group 4 (Uptake and repair by EEC only) as summarized in Table 1.

The maximum score is 10 for both assessments. Out of 45 students who completed the course, 41 students attempted the 1st assessment, and 40 students attempted the 2nd assessment. Tables 2 and 3 summarize the median and interquartile range (IQR) of initial and final scores achieved by the students on ORP Gym in the 1st and 2nd assessments. The detailed result of the Wilcoxon signed-rank test is shown in Table 9 in the Appendix section.



**Table 1** Groups classification by students' voluntary retries (Y=Yes, N=No)

	Group	Description of action pattern	Retry by HR (Y/N)	Retry by EEC (Y/N)
No retry group by ORP Gym	1	No uptake and no repair	N	N
Retry group by ORP Gym	2	Uptake and repair by HR only	Y	N
	3	Uptake and repair by HR and EEC	Y	Y
	4	Uptake and repair by EEC only	N	Y

**Table 2** The median and IQR of the 1st oral assessment and the result of the Wilcoxon signed-rank test

	Group	N	Initial score	Final score	Z
			Median (IQR)	Median (IQR)	
No retry group by ORP Gym	1	15	9.50 (8.25 - 10.00)	9.50 (8.25 - 10.00)	N/A <sup>a</sup>
Retry group by ORP Gym	2	25	8.00 (5.00 - 9.25)	9.75 (8.80 - 10.00)	-4.02***
	3	1	2.25 (- <sup>b</sup> )	9.50 (- <sup>b</sup> )	N/A <sup>a</sup>
	4	0	- <sup>b</sup>	- <sup>b</sup>	- <sup>b</sup>

\*\*\*  $p < .001$

<sup>a</sup>The statistical test was not applicable (N/A) because Group 1 did not retry and Group 3 was only 1 student

<sup>b</sup>No data because the number of the students in the groups was 0 or 1

**Table 3** The median and IQR of the 2nd oral assessment and the result of the Wilcoxon signed-rank test

	Group	N	Initial score	Final score	Z
			Median (IQR)	Median (IQR)	
No retry group by ORP Gym	1	16	9.80 (9.35 - 9.80)	9.80 (9.35 - 9.80)	N/A <sup>a</sup>
Retry group by ORP Gym	2	23	9.40 (7.40 - 9.80)	9.80 (9.60 - 9.80)	-2.66**
	3	0	- <sup>b</sup>	- <sup>b</sup>	- <sup>b</sup>
	4	1	9.40 (- <sup>b</sup> )	10.00 (- <sup>b</sup> )	N/A <sup>a</sup>

\*\*  $p < .01$

<sup>a</sup>The statistical test was not applicable (N/A) because Group 1 did not retry and Group 4 was only 1 student

<sup>b</sup>No data because the number of the students in the groups was 0 or 1

In both assessments, the no retry group recorded the highest initial median scores, and they did not retry for score improvement. We suspect that no retry group did not feel a strong need to retry due to reasonably high scores. The final median scores of the no retry group were the same or lower than the retry group. On the other hand, the majority of the students in the retry group belong to Group 2. The students in Group 2 showed significant score improvement with a large effect size in the 1st assessment ( $Z = -4.02, p < .001, r = -.80$ ) and the 2nd assessment ( $Z = -2.66, p < .01, r = -.56$ ). This result means students' retries, especially with HR by ORP Gym contributed to improving students' utterances. However, the number of other students who used EEC for voluntary ORP during the assessment period (Groups 3 and 4) was only zero or one, although the student achieved a score improvement through retries. Although EEC directly contributes to enhancing student repair, EEC was not used actively for their retries by students. We believe that one of the reasons was the delay in providing EEC and a score, which was 48 hours after submission.

Based on these results, we observe significant score improvement by HR. On the contrary, EEC was not actively used by the students. EEC helps to further improve the utterances by allowing the student to notice the mistakes which cannot be recognized nor repaired using HR. However, the total number of EEC users was quite small in this experiment and the latter statement cannot be generalized until EEC is evaluated with a larger sample size.

## 5.2 The number of retries and repaired answers

Table 4 summarizes the number and the ratio of retries and repaired answers in each assessment. Erroneous answer contained a mistake and scored less than one (Full Mark). Repaired answer is a correct answer that satisfies all the criteria and its score is one (Full Mark) as a result of retry.

Both assessments exhibited reasonably high rates of retry and repair. Retry was attempted against 85.9% and 78.0% of erroneous answers, and 28.9% and 31.0% of the erroneous answers were repaired. ORP Gym was actively utilized based on the average number of retries per student, 1.90 and 2.02 in the 1st and 2nd assessments. Even though approximately 70% of the answers of retry did not receive the full mark, the score improvement is still significant in the retry group with HR (Group

**Table 4** Improvement of utterances by retries on ORP Gym

	Total no. of answers	Erroneous answers	HR	EEC <sup>a</sup>	Retries	Repaired answers
1st assessment	390	128 (100.0%)	128 (100.0%)	128 (100%)	110 (85.9%)	37 (28.9%)
2nd assessment	405	74 (100.0%)	74 (100.0%)	74 (100.0%)	58 (78.0%)	23 (31.0%)

<sup>a</sup>In this experiment, EEC was provided to all erroneous answers. However, because of the small number of students who actually used EEC, the effect of EEC is excluded from the causality of utterance improvement as well as uptake and repair rates

2), as observed in Tables 2 and 3. Since HR was used by the majority of the students, HR contributed to the students' retries and repaired answers. However, EEC should be excluded from the causality of utterance improvement as well as uptake and repair rates due to the small number of students who used EEC for retries.

### 5.3 Linguistic uptake and repair rates by ORP Gym

In this experiment, uptake rate and repair rate were measured using a total of 78 audio answers (including initial answers and retried ones) to the common read-aloud quiz submitted by 41 students in the 1st assessment. The quiz was a student's self-introduction with the same and fixed sentences where the student's name must be included. Table 5 shows the breakdown of the number of uptake and repair as well as uptake rate and repair rate based on the error category: pronunciation, grammar, and vocabulary. 'Errors' were counted by the part of speech (called *hinshi* in Japanese).

Generally, as a benchmark among the existing OCF studies for the past few decades, 59.2% uptake rate and 45.3% repair rate in total are considered to be high (Fu & Nassaji, 2016). Although we analyzed the students' audio answers collected by ORP Gym and our analysis is not based on audio recording data during classroom hours, our overall 84.8% uptake rate and 45.5% repair rate by ORP Gym are reasonably high considering the benchmark. The majority of the errors occurred in pronunciation and grammar categories. The uptake rates were high in both categories, but the repair rate in grammar was much lower than in pronunciation.

### 5.4 Comparison of score improvement in the final examination

To measure their speaking improvement by ORP Gym, this paper compares the scores of Parts 1 and 2 in the final exam as posttests with the initial scores of the 1st and 2nd assessments as pretests by the Wilcoxon signed-rank test as shown in Tables 6 and 7 respectively. The detailed result of the Wilcoxon signed-rank test is shown in Table 9 in the Appendix section.

The median scores achieved by the *no retry group by ORP Gym*, who did not attempt a retry on ORP Gym during the oral assessments period, were high in both the pretest and posttest in Part 1. For Part 2, the degradation of the median score was

**Table 5** Uptake and repair rates by ORP Gym

Error Category	Error	HR	EEC <sup>a</sup>	Uptake	Uptake rate	Repair	Repair rate
Pronunciation	17	17	17	14	82.4%	10	58.8%
Grammar	13	13	13	11	84.6%	2	15.4%
Vocabulary	3	3	3	3	100%	3	100%
Total	33	33	33	28	84.8%	15	45.5%

<sup>a</sup> In this experiment, EEC was provided to all erroneous answers. However, because of the small number of students who actually used EEC, the effect of EEC is excluded from the causality of utterance improvement as well as uptake and repair rates

**Table 6** The comparison between pretest and posttest in Part 1 by the result of the Wilcoxon signed-rank test

	Group	N	Initial score	Final score	Z
			Median (IQR)	Median (IQR)	
No retry group by ORP Gym	1	15	9.50 (8.25 - 10.00)	10.00 (9.66 - 10.00)	-2.83**
Retry group by ORP Gym	2	25	8.00 (5.00 - 9.25)	10.00 (9.66 - 10.00)	-3.68***
	3	1	2.25 (- <sup>b</sup> )	9.66 (- <sup>b</sup> )	N/A <sup>a</sup>
	4	0	- <sup>b</sup>	- <sup>b</sup>	- <sup>b</sup>

\*\*  $p < .01$ , \*\*\*  $p < .001$

<sup>a</sup>The statistical test was not applicable (N/A) because Group 3 was only 1 student

<sup>b</sup>No data because the number of the students in the group was 0

**Table 7** The comparison between pretest and posttest in Part 2 by the result of the Wilcoxon signed-rank test

	Group	N	Initial score	Final score	Z
			Median (IQR)	Median (IQR)	
No retry group by ORP Gym	1	6	9.80 (9.35 - 9.80)	9.74 (9.10 - 10.00)	-.11
Retry group by ORP Gym	2	23	9.40 (7.40 - 9.80)	10.00 (9.49 - 10.00)	-2.80***
	3	0	- <sup>b</sup>	- <sup>b</sup>	- <sup>b</sup>
	4	1	9.40 (- <sup>b</sup> )	10.00 (- <sup>b</sup> )	N/A <sup>a</sup>

\*\*  $p < .01$

<sup>a</sup>The statistical test was not applicable (N/A) because Group 4 was only 1 student

<sup>b</sup>No data because the number of the students in the group was 0

observed between the pretest and the posttest. Significant score improvement with a large effect size was observed in the no retry group for Part 1 ( $Z = -2.83$ ,  $p < .01$ ,  $r = -.73$ ) but not for Part 2 ( $Z = -.11$ ,  $p > .05$ ,  $r = .03$ ). The no retry group improved their scores through the course contents, classroom activities including OCF, and voluntary ORP that cannot be assessed using the data collected on ORP Gym. However, the no retry group (Group 1) exhibited degradation of the score, which is reasonable because they did not use ORP Gym.

The majority of the students (Group 2), who attempted a retry on ORP Gym during the assessments, used HR for their retries. Their median scores were low in the pretest but high in the posttest. Significant score improvement with a large effect size has been observed for both Part 1 ( $Z = -3.68$ ,  $p < .001$ ,  $r = -.74$ ) and Part 2 ( $Z = -2.80$ ,  $p < .01$ ,  $r = .58$ ) in Group 2. In addition, IQR in the posttest is smaller than in the pretest. This indicates that ORP Gym, especially with HR, helped the students' utterances to improve significantly and effectively between pretest and posttest (exam score). Regarding the students who used EEC for their retries (Groups 3 and 4) in the *retry group by ORP Gym*, we could not conduct statistical tests due to the small sample size, and we cannot generalize the relationship between EEC and the score improvement.

## 5.5 The workload distribution by OEA and the average time to produce EEC

The instructor was a certified Japanese language teacher, who is a native Japanese speaker with four years of teaching experience in the course. One OEA (age 28, bachelor's degree, a native Japanese speaker, without Japanese language teacher certification) was trained using the handbook which was developed by the instructor (Supplementary file 1) and the evaluation trial using 172 sample audio data sets. In the handbook, the evaluation criteria and scoring instruction of borderline cases were mentioned, and the instructor gave the results and feedback of the scoring simulation to the OEA. The duration of the training period was 1 month and the training was completed before the course began.

During the experiment, the OEA evaluated 790 audio answers in two oral assessments over a total of 11.05 hours, and the instructor evaluated 426 audio answers on the final exam over a total of 6.6 hours. The average time to produce EEC was 50.4 seconds per erroneous answer by OEA, and 55.8 seconds per erroneous answer by the instructor.

## 5.6 Students' satisfaction with the course and CMF

Table 8 shows the result of a questionnaire with a 5-point Likert scale taken by 41 students who completed the course successfully and responded to the questionnaire. The summary of the free descriptive comments from 27 students (68.85%), which were extracted from the answers to the open-ended question, also follows afterward. Since the data was nonparametric based on the Shapiro-Wilk normality test ( $p < 0.01$ ), the median and IQR are used in Table 8. The median of each item was above 4 with a small IQR, which means that most of the students showed high satisfaction and comprehension with the course where ORP Gym was incorporated as oral assessments. In addition, most of the students gave a positive reaction to the availability of CMF (HR and EEC) as personalized feedback in oral assignments as well as the chances of score improvement through multiple attempts outside classroom hours by ORP Gym.

Regarding the free descriptive comments, all the comments were roughly categorized into positive comments including positive words such as fun, interesting,

**Table 8** The questionnaire with a 5-point Likert scale (41 students responded on Feb 8th - Feb 12th, 2019)

Items	Median (IQR)
1. Did you enjoy this course?	5.0 (4.0 - 5.0)
2. Was this course easy to understand for you?	4.0 (4.0 - 5.0)
3. Were the several times of oral assessments (personalized feedback to your speaking errors) helpful for your study?	5.0 (4.0-5.0)

helpful, useful, and happy. Specifically, 5 students gave positive detailed comments<sup>1</sup> as follows. These comments indicate that the course with ORP Gym increased the amount of student participation in speaking and contributed to high students' confidence and satisfaction. It is also noted that a student also mentioned that speaking tests were graded fairly and provided good insights.

1. Learning a completely new language is definitely a herculean task and you've made it very simple. Managing the academics of the course in moodle is a very good idea. The minitests, listening and speaking tests are very helpful.
2. The various listening tests and spoken tests were graded well, and they provided good insight.
3. It covered speaking and listening part very well, it also had good amount of student participation which really helped learning the language better.
4. It had a very good approach of teaching, especially the online assessment and learning.
5. The online platform for all the related coursework was a nice way to learn.

On the other hand, a student pointed out a limitation of the ORP Gym. This comment<sup>2</sup> indicates that the student feels that interactive communication with instructors or between peers during classroom hours is helpful for their language development. ORP Gym supported individual students' oral repetition practice with CMF outside classroom hours throughout Japanese SFL basics speaking course. However, ORP Gym does not support interactive communication during classroom hours.

- When I was learning other languages, I found it very helpful to my development when we were asked to speak (try to speak) only in that specific language to the other students and the professor during the class hours. Maybe this will be difficult for an introductory course.

## 6 Discussion

Previous OCF studies during classroom hours have revealed the ineffectiveness of recasts in terms of student uptake and repair (Lyster & Ranta, 1997; Panova & Lyster, 2002), especially in grammatical error correction (Kim & Han, 2007; Lyster, 1998; Lyster & Mori, 2006; Nabei, 2005). Choral repetition and read-aloud (Lyster & Mori, 2006) and one-on-one interactions between a teacher and a student by FonF (Ammar & Spada, 2006; Nassaji, 2013) are well-known effective approaches to improve students' utterances in small to medium-sized classes. However, we focused on the lack of opportunities for students' oral production and individual support during classroom hours in a large-scale class. Using technology-enhanced learning approaches, computer-mediated feedback (CMF) and web-based language learning (WBLL) have been studied to complement the opportunities for oral practice

---

<sup>1</sup> The comments were extracted from the questionnaire as it is.

<sup>2</sup> The comment was extracted from the questionnaire as it is.

during the past decade. Audio-visual speaking activities (A/Vs) with CMF outside classroom hours were proven to be effective for enhancing students' speaking skills (Buckingham & Alpaslan, 2017; Rassaei, 2019; Tseng & Yeh, 2019). Students' ORP through A/Vs and CMF on a learning management system (LMS) effectively improved word accentuation (Yoshida & Fukada, 2014). CMF is essential for students to get personalized feedback to improve their erroneous utterances, and an LMS is also a cost-effective approach to implement CMF for large-scale courses (Tuffley & Antonio, 2015). Although some previous studies demonstrated the effective use of an LMS in a speaking class (Fukada, 2013; Ikeda & Fukada, 2012; Kataoka et al., 2018; Yoshida & Fukada, 2014), the effect of CMF integrated into an LMS in a large-scale SFL class has not been investigated and proven by measuring uptake and repair rates.

As the solution, we developed ORP Gym to provide students with sufficient ORP and CMF, and to improve their utterances outside classroom hours. This research contributed to enabling CMF within a large-scale SFL class, and empirically and statistically proving the learning effect, including *uptake* and *repair*, by CMF using ORP Gym and OEA for the first time to the best of our knowledge. Since ORP Gym is developed based on an open-source LMS (Moodle Project, 2022), the proposed approach is deployable and reproducible by implementing ORP Gym and OEA as a module on the LMS proven to be working in a large-scale class. This paper shares the practical experience and know-how to effectively utilize the proposed approach through the actual deployment including the online documentation for OEA to produce CMF on ORP Gym. In addition, this research also contributed to establishing an approach to investigate the effect of CMF to fill in the paucity of existing CMF studies in a large-scale speaking class. ORP Gym is reproducible on an LMS, and OEA can be introduced with reasonable manpower cost. Further investigation of the effect of CMF is affordable in other large-scale SFL classes.

While teachers' OCF was provided electronically and asynchronously as CMF, ORP Gym establishes personalized feedback to promote students' oral production and enhances their utterances by 45.5% as the result shows. The benefit of teachers' HR and EEC on ORP Gym could support individual students' ORP visually and audibly without time constraints. Regarding **RQ1**, this research demonstrated that recasts, which were previously known to be ineffective during classroom hours, were actively used by students and effective in improving students' utterances on ORP Gym outside classroom hours. Recasts are proven to be effective in computer-mediated communication (CMC) based on one-on-one communication (Rassaei, 2017), and this study also exhibited the effectiveness of recast-like CMF on the LMS outside classroom hours. According to Tseng and Yeh (2019), the combination of A/V-based CMF and text-based CMF is beneficial. This result also shows the possibility that students can recognize and detect their erroneous utterances especially pronunciation by themselves using audio-based and text-based CMF without teachers' explicit instruction as Thomson (2011) also claimed.



The result obtained in this study demonstrated that ORP Gym promoted students' voluntary retries and different action patterns based on the access log and audio data. Concerning **RQ2**, the majority of the students significantly improved their scores through retries during the two oral assessments period by HR (model answer's text and audio) on ORP Gym. HR was proven to enhance students' self-generated uptake and repair by showing significant score improvement in the oral assessments. This paper measured student uptake and repair rates using a common read-aloud quiz on ORP Gym for the first time (**RQ3**). We found that ORP Gym encouraged their uptake through retries at high rates in pronunciation and vocabulary categories, but the repair rate in the grammar category was low. This result is probably attributed to the fact that the majority of the students used HR more than EEC because implicit recasts, which HR is categorized into, is known to be ineffective against grammatical errors (Kim & Han, 2007; Lyster, 1998; Nabei, 2005). Although HR is provided to students visually and audibly on ORP Gym, it might be insufficient for students to notice the grammatical error that causes their erroneous utterances.

In addition, the active users, who retried with HR on ORP Gym, showed more score improvement between the pretest and the posttest than those who did not retry on ORP Gym (**RQ4**). This indicates that students' oral production through retries especially with HR on ORP Gym during the two oral assessments period improved their speaking accuracy and proficiency in the final exam. However, EEC (pointing out the problematic part) was not actively used by students compared with HR. We believe the reason may be the length of delay to provide EEC, or students' lowered motivation for repairing or improvement after they got the score. To improve the utilization of EEC by students, we have to consider improving EEC's latency or giving higher incentives for repair or further improvement. Visualizing or notifying the repair and improvement including perfection to the individual student can be explored to motivate them to use EEC.

With respect to **RQ5**, the OEA, who works in parallel with teachers and remotely, distributes the workload and time of examining the audio answers and producing EEC. OEA is a relatively reasonable option to ensure the quality of feedback and reduce the workload related to the teacher's CMF. However, the current method of producing EEC relies on manual error detection and scoring by the teacher and OEA. While increasing the number of OEAs may reduce the time to process a certain number of answers, such an approach may not be sustainable considering the availability and cost of reasonably skilled OEAs. OEAs can be non-certified native Japanese speakers, but still require guidance and training before the course starts. The training includes how to use ORP Gym, evaluate submitted students' audio answers, and generate EEC. The handbook (Supplementary file 1) documents how to detect each part of speech errors and generate EEC using transcription (in romaji with pitch symbol) to avoid evaluation discrepancy. Maintaining the quality of EEC by OEAs requires human labour and human resource management costs at present. A crowdsourcing platform to manage OEAs can be explored to improve the scalability of ORP Gym as the number

of students and the frequency of oral assessments increase (Takahashi et al., 2015). Taking the advantage of its active deployment and availability, automatic speech recognition (ASR) is helpful for teachers to examine students' utterances and produce EEC quickly. Computer-generated feedback (CGF) employing ASR gives supplemental information in addition to CMF (Sherafati et al., 2020). For example, Fu et al. (2020) have developed an ASR-based system for estimating the similarity of English pronunciation between Japanese speakers and native English speakers using Deep Learning. ASR can also be extended for EEC by converting submitted audio answers to raw text, marking an erroneous pronunciation so that the error can be recognized by the students (Kataoka et al., 2019), because ASR is promising to enhance students' pronunciation improvement (Bajorek, 2017; Cucchiarini et al., 2009).

Regarding **RQ6**, based on the questionnaire result, we found that ORP Gym was utilized actively outside classroom hours by students and is promising for students to improve their utterances and maintain their interest and motivation. Lasagabaster and Sierra (2005) pointed out that consistent error correction is arduous for students during classroom hours. However, ORP Gym enabled consistent error detection and correction outside classroom hours and brought students' high satisfaction and comprehension. Furthermore, students also positively expressed that ORP Gym benefited their language learning through the response in the questionnaire.

## 7 Implications and limitations

In this experiment, we prioritized introducing of ORP Gym and providing a variety of quizzes randomly from a question bank in a classroom activity instead of conducting a dedicated investigation of uptake and repair rates using more common quizzes in a controlled environment. In addition, the effectiveness of EEC could not be validated due to the small sample size in this experiment. Further data collection and investigation of uptake and repair rates as well as the effect of EEC should be conducted for statistical analysis and learning analytics in the future.

Since this research is a classroom observational study in a Japanese basics SFL course at an Indian university, the age group, language level, and nationalities of the participants were limited in this experiment. The participants were undergraduate Indian students ranging from 18 to 21. The level of participants was beginners without any prior learning experience in Japanese language. Therefore, further research needs to be conducted for different age groups, different levels of learners, and different countries to investigate the effectiveness of the approach.

The use of ORP Gym is limited to enabling language instructors to provide asynchronous computer-mediated practice and feedback outside classroom hours. Therefore, it does not support real-time OCF using CMF during classroom hours. On the other hand, we believe that ORP Gym can support fully online language classes, which are offered in a completely online environment,

due to pandemic issues such as COVID-19. However, our experiment did not cover such a use case, and therefore, the effectiveness of ORP Gym has not been proven for large-scale fully online Japanese SFL speaking classes.

## 8 Conclusion

This paper presented ORP Gym, an online oral repetition practice support system to provide sufficient opportunities and support for students' ORP aiming to improve the rates of uptake and repair in a large-scale Japanese SFL education. Given the fact that the existing OCF of erroneous utterances during classroom hours suffers from enhancing uptake and repair due to severe time constraints, especially in a large-scale SFL class, ORP Gym provides a platform to perform ORP enhanced with two types of CMF outside classroom hours: HR and EEC. As the evaluation shows, the erroneous answers were corrected through retries encouraged by ORP Gym in each assessment. The evaluation of the errors in read-aloud of all 41 students' self-introduction exhibited an 84.8% uptake rate and a 45.5% repair rate during the assessment period. The majority of the students retried using ORP Gym and improved their speaking scores significantly in the posttest than in the pretest compared with those who did not utilize the system. OEA contributed to mitigating the instructor's workload and shortening the time duration to produce EEC by distributing the correction tasks.

As a future work, a separate evaluation of EEC should be conducted to measure its effectiveness. The course management plan can be improved so that EEC will be more actively utilized. In addition, the number of OEAs involved in the proposed approach was one. We can explore the direction to improve the efficiency of OEA management even though the number of OEA can also be more than one or even larger. The automatic scoring of audio answers by ASR is promising to mitigate the workload of the instructor and OEA as well as to make the production of EEC faster. Utilizing ORP Gym, further data collection and analysis should be conducted by collaborating with multiple educational institutes in India as well as other countries. We need to analyze error tendencies among various students whose age group, level, and native language are different so that types of CMF could be adjusted flexibly according to the error tendencies. We are also planning to introduce ORP Gym for Japanese SFL courses in Southeast Asian countries in collaboration with SOI (School of Internet) (Okawa et al., 1999) and SOI Asia (SOI Asia Project, 2022) as a part of the broader deployment, collection of best practices of ORP Gym, and further improvement of the system. Since this experiment proved the effect of HR outside classroom hours, we will extend ORP Gym to support the real-time oral repetition practice and investigate the effect of HR during classroom hours. The use of Artificial Intelligence will be considered to enable the real-time EEC. In addition, due to COVID-19 pandemic issues, the form of many SFL classes shifted to fully online. We will investigate the effectiveness of ORP Gym in fully online speaking classes.

## Appendix

**Table 9** The result of the Wilcoxon signed-rank test

Groups	Variables	Neg/Pos/Ties	N	Mean Rank	Sum of Ranks	Z	p
1	Posttest - Pretest (Part 1)	Negative Ranks	1	3.00	3.00	-2.83**	0.005
		Positive Ranks	11	6.82	75.00		
		Ties	3				
	Posttest - Pretest (Part 2)	Negative Ranks	6	10.33	62.00	-.11	.91
		Positive Ranks	9	6.44	58.00		
		Ties	1				
2	final score - initial score (OA1 <sup>a</sup> )	Negative Ranks	0	.00	.00	-4.02**	< 0.001
		Positive Ranks	21	11.00	231.00		
		Ties	4				
	final score - initial score (OA2 <sup>b</sup> )	Negative Ranks	5	4.90	24.50	-2.66**	0.008
		Positive Ranks	13	11.27	146.50		
		Ties	5				
	Posttest - Pretest (Part 1)	Negative Ranks	2	8.50	17.00	-3.68**	< 0.001
		Positive Ranks	21	12.33	259.00		
		Ties	2				
	Posttest - Pretest (Part 2)	Negative Ranks	6	5.00	30.00	-2.80**	0.005
		Positive Ranks	14	12.86	180.00		
		Ties	3				

\*\*  $p < .01$

<sup>a</sup>OA1 = 1st oral assessment

<sup>b</sup>OA2 = 2nd oral assessment

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10639-022-11262-7>.

**Acknowledgements** The authors are grateful to Dr. Osamu Nakamura, Dr. Keiko Okawa, Dr. Yuko Nakahama at Keio University, and Prof. Tim Ashwell at Komazawa University for their advice.

**Author Contributions** Ms. Yuka Kataoka designed and implemented the research and carried out the course and experiment to collect and analyze the data. Dr. Achmad Husni Thamrin and Dr. Kotaro Kataoka contributed to shaping the design, implementation, and analysis of the research. Dr. Rodney Van Meter and Dr. Jun Murai supervised the overall research. All the authors read and approved the final manuscript.

**Funding** This work was supported by Grant-in-Aid for JSPS Fellows.

**Data Availability** Experimental data without participants' personal information can be provided upon request.

## Declarations

**Conflicts of interest** The authors declare that they have no competing interests in this research.

**Ethical approval** The personal information of students is hidden to protect them. This experiment was conducted based on students' voluntary participation, fair grade policy, and equal opportunity in education during an actual language course.

**Consent to participate** The experiment was conducted based on students' voluntary participation.

**Consent for publication** The authors consented to the publication. The university where we conducted the experiment consented to the publication as long as the participants' personal information is protected.

## References

- Ahn, T. Y., & Lee, S. M. (2016). User experience of a mobile speaking application with automatic speech recognition for EFL learning. *British Journal of Educational Technology*, 47(4), 778–786. <https://doi.org/10.1111/bjet.12354>
- Akiyama, Y. (2014). Using Skype to focus on form in Japanese telecollaboration: Lexical categories as a new task variable. In Li, S., Swanson, P. (Eds.), *Engaging Language Learners through Technology Integration: Theory, Applications, and Outcomes* (pp. 181–209). IGI Global. <https://doi.org/10.4018/978-1-4666-6174-5.ch009>
- Akiyama, Y. (2017). Learner beliefs and corrective feedback in telecollaboration: A longitudinal investigation. *System*, 64, 58–73. <https://doi.org/10.1016/j.system.2016.12.007>
- Ammar, A., & Spada, N. (2006). One size fits all?: Recasts, prompts, and L2 learning. *Studies in Second Language Acquisition*, 28(4), 543–574. <https://doi.org/10.1017/S0272263106060268>
- Ashwell, T., & Elam, J. R. (2017). How accurately can the Google Web Speech API recognize and transcribe Japanese L2 English learners' oral production? *The JALT CALL Journal*, 13(1), 59–76. <https://eric.ed.gov/?id=EJ1141025>
- Bajorek, J. P. (2017). L2 Pronunciation in CALL: The unrealized potential of Rosetta Stone, Duolingo, Babbel, and Mango Languages. *Issues and Trends in Educational Technology*, 5(1), 24–42. [https://doi.org/10.2458/azu\\_itet\\_v5i1\\_bajorek](https://doi.org/10.2458/azu_itet_v5i1_bajorek)
- Buckingham, L., & Alpaslan, R. S. (2017). Promoting speaking proficiency and willingness to communicate in Turkish young learners of English through asynchronous computer-mediated practice. *System*, 65, 25–37. <https://doi.org/10.1016/j.system.2016.12.016>
- Cong-Lem, N. (2018). Web-based language learning (WBLL) for enhancing L2 speaking performance: A review. *Advances in Language and Literary Studies*, 9(4), 143–152. <https://doi.org/10.7575/aialc.all.v9n.4p.143>
- Cucchiari, C., Neri, A., & Strik, H. (2009). Oral proficiency training in Dutch L2: The contribution of ASR-based corrective feedback. *Speech Communication*, 51(10), 853–863. <https://doi.org/10.1016/j.specom.2009.03.003>
- de Vries, B. P., Cucchiari, C., Bodnar, S., Strik, H., & van Hout, R. (2016). Spoken grammar practice and feedback in an ASR-based CALL system. *Computer Assisted Language Learning*, 28(6), 550–576. <https://doi.org/10.1080/09588221.2014.889713>
- Ducate, L., & Lomicka, L. (2009). Podcasting: An effective tool for honing language students' pronunciation? *Language Learning & Technology*, 13(3), 66–86. <https://doi.org/10.125/44192>
- Ellis, R. (2016). Focus on form: A critical review. *Language Teaching Research*, 20(3), 405–428. <https://doi.org/10.1177/1362168816628627>
- Ellis, R., Basturkmen, H., & Loewen, S. (2001). Learner uptake in communicative ESL lessons. *Language Learning*, 51(2), 281–318. <https://doi.org/10.1111/1467-9922.00156>
- Elola, I., & Oskoz, A. (2016). Supporting second language writing using multimodal feedback. *Foreign Language Annals*, 49(1), 58–74. <https://doi.org/10.1111/flan.12183>
- Fang, W. C., Yeh, H. C., Luo, B. R., & Chen, N. S. (2021). Effects of mobile-supported task-based language teaching on EFL students' linguistic achievement and conversational interaction. *ReCALL*, 33(1), 71–87. <https://doi.org/10.1017/S0958344020000208>
- Federico, J. (2018). RecordRTC for Atto. Retrieved from [https://moodle.org/plugins/atto\\_recordrtc](https://moodle.org/plugins/atto_recordrtc)
- Fu, T., & Nassaji, H. (2016). Corrective feedback, learner uptake, and feedback perception in a Chinese as a foreign language classroom. *Studies in Second Language Learning and Teaching*, 6(1), 159–181. <https://doi.org/10.14746/ssl.2016.6.1.8>

- Fu, J., Chiba, Y., Nose, T., & Ito, A. (2020). Automatic assessment of English proficiency for Japanese learners without reference sentences based on deep neural network acoustic models. *Speech Communication, 116*, 86–97. <https://doi.org/10.1016/j.specom.2019.12.002>
- Fukada, A. (2013). An online oral practice/assessment platform: Speak Everywhere. *IALLT Journal of Language Learning Technologies, 43*(1), 64–77. <https://doi.org/10.17161/iallt.v43i1.8518>
- Google (n.d.). Google Forms. Retrieved from <https://www.google.com/forms/about>
- Holliday, A. (1994). *Appropriate methodology and social context*. Cambridge: Cambridge University Press
- Hsu, H. C. (2016). Voice blogging and L2 speaking performance. *Computer Assisted Language Learning, 29*(5), 968–983. <https://doi.org/10.1080/09588221.2015.1113185>
- Ikeda, J., & Fukada, A. (2012). Designing a speaking-oriented course integrated with Speak Everywhere and its classroom implementation practice. *Journal of Japanese Language Teaching, 152*, 46–60. [https://doi.org/10.20721/nihongokyoiku.152.0\\_46](https://doi.org/10.20721/nihongokyoiku.152.0_46)
- Japan Foundation (2018). Survey report on Japanese-language education abroad 2018. Retrieved from <https://www.jpff.go.jp/j/project/japanese/survey/result/dl/survey2018/all.pdf>
- Kataoka, Y., Thamrin, A.H., Murai, J., Kataoka, K. (2018). Effective use of learning management system for large-scale Japanese language education. In *Proceedings of the 10th International Conference on Education Technology and Computers (ICETC'18)* (pp. 49–56). <https://doi.org/10.1145/3290511.3290564>
- Kataoka, Y., Thamrin, A.H., Murai, J., Kataoka, K. (2019). Employing automatic speech recognition for quantitative oral corrective feedback in Japanese second or foreign language education. In *Proceedings of the 11th International Conference on Education Technology and Computers (ICETC'19)* (pp. 52–58). <https://doi.org/10.1145/3369255.3369285>
- Kato, F., Spring, R., & Mori, C. (2016). Mutually beneficial foreign language learning: Creating meaningful interactions through video-synchronous computer-mediated communication. *Foreign Language Annals, 49*(2), 355–366. <https://doi.org/10.1111/flan.12195>
- Kim, J., & Han, Z. (2007). Recasts in communicative EFL classes: Do teacher intent and learner interpretation overlap? In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A series of empirical studies* (pp. 269–297). Oxford: Oxford University Press.
- Kumai, N., Paul, D. (2013). Moodle module and app development for shadowing practice on mobile devices. *Language, Culture and Society, 11*, 115–130. <https://cir.nii.ac.jp/crid/1050001202934792576>
- Lasagabaster, D., & Sierra, J. M. (2005). Error correction: Students' versus teachers' perceptions. *Language Awareness, 14*(2–3), 112–127. <https://doi.org/10.1080/09658410508668828>
- Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning, 60*(2), 309–365. <https://doi.org/10.1111/j.1467-9922.2010.00561.x>
- Loewen, S., & Philp, J. (2006). Recasts in the adult English L2 classroom: Characteristics, explicitness, and effectiveness. *The Modern Language Journal, 90*(4), 536–556. <https://doi.org/10.1111/j.1540-4781.2006.00465.x>
- Long, M.H. (1996). The role of linguistic environment in second language acquisition. In Ritchie, W.C., Bhatia, T.K. (Eds.) *Handbook of Second Language Acquisition* (pp. 413–468). San Diego: Academic Press.
- Long, M.H., Robinson, P. (1998). Focus on form: Theory, research and practice. In Doughty, C., Williams, J. (Eds.), *Focus on Form in Classroom Second Language Acquisition* (pp. 15–41). Cambridge: Cambridge University Press.
- Lyster, R. (1998). Negotiation of form, recasts, and explicit correction in relation to error types and learner repair in immersion classrooms. *Language Learning, 48*(2), 183–218. <https://doi.org/10.1111/1467-9922.00039>
- Lyster, R., & Mori, H. (2006). Interactional feedback and instructional counterbalance. *Studies in Second Language Acquisition, 28*(2), 269–300. <https://doi.org/10.1017/S0272263106060128>
- Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition, 19*(1), 37–66. <https://doi.org/10.1017/S0272263197001034>
- Lyster, R., Saito, K., & Sato, M. (2013). Oral corrective feedback in second language classrooms. *Language Teaching, 46*(1), 1–40. <https://doi.org/10.1017/S0261444812000365>
- Moodle Project. (2022). moodle. Retrieved from <https://moodle.org>
- Nabei, T. (2005). Recasts in a Japanese EFL classroom. Osaka: Kansai University Press.

- Nassaji, H. (2013). Participation structure and incidental focus on form in adult ESL classrooms. *Language Learning*, 63(4), 835–869. <https://doi.org/10.1111/lang.12020>
- Neri, A., Cucchiari, C., Strik, H. (2008). The effectiveness of computer-based speech corrective feedback for improving segmental quality in L2 Dutch. *ReCALL*, 20(2), 225–243. <https://doi.org/10.1017/S0958344008000724>
- Okawa, K., Ijuin, Y., Murai, J. (1999). School of Internet - building a university on the Internet. *Transactions of Information Processing Society of Japan*, 40(10), 3801–3810. <http://id.nii.ac.jp/1001/00012527/>
- Panova, I., & Lyster, R. (2002). Patterns of corrective feedback and uptake in an adult ESL classroom. *TESOL Quarterly*, 36(4), 573–595. <https://doi.org/10.2307/3588241>
- Rassaei, E. (2017). Video chat vs. face-to-face recasts, learners' interpretations and L2 development: A case of Persian EFL learners. *Computer Assisted Language Learning*, 30(1–2), 133–148. <https://doi.org/10.1080/09588221.2016.1275702>
- Rassaei, E. (2019). Computer-mediated text-based and audio-based corrective feedback, perceptual style and L2 development. *System*, 82, 97–110. <https://doi.org/10.1016/j.system.2019.03.004>
- Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–158. <https://doi.org/10.1093/applin/11.2.129>
- Seedhouse, P. (1997). The case of the missing "No": The relationship between pedagogy and interaction. *Language Learning*, 47(3), 547–583. <https://doi.org/10.1111/0023-8333.00019>
- Sheen, Y. (2004). Corrective feedback and learner uptake in communicative classrooms across instructional settings. *Language Teaching Research*, 8(3), 263–300. <https://doi.org/10.1191/1362168804lr1460a>
- Sherafati, N., Largani, F. M., & Amini, S. (2020). Exploring the effect of computer-mediated teacher feedback on the writing achievement of Iranian EFL learners: Does motivation count? *Education and Information Technologies*, 25, 4591–4613. <https://doi.org/10.1007/s10639-020-10177-5>
- SOI Asia Project. (2022). SOI Asia. Retrieved from <https://www.soi.asia>
- Sun, Y. C. (2012). Examining the effectiveness of extensive speaking practice via voice blogs in a foreign language learning context. *CALICO Journal*, 29(3), 494–506. <https://www.jstor.org/stable/10.2307/calicojournal.29.3.494>
- Takahashi, E., Hatasa, Y., Yamamoto, H., Maekawa, S., Hatasa, K. (2015). The development of an automatic evaluation system of L2 pronunciation in Japanese. In *Proceedings of IPSJ SIG Computers and the Humanities Symposium* (pp. 59–64). <http://id.nii.ac.jp/1001/00146524/>
- Thomson, R.I. (2011). Computer assisted pronunciation training: Targeting second language vowel perception improves pronunciation. *CALICO Journal*, 28(3), 744–765. <https://www.jstor.org/stable/calicojournal.28.3.744>
- Tseng, S S., Yeh, H C. (2019). The impact of video and written feedback on student preferences of English speaking practice. *Language Learning & Technology*, 23(2), 145–158. <https://doi.org/10.125/44687>
- Tuffley, D., & Antonio, A. (2015). Enhancing educational opportunities with computer-mediated assessment feedback. *Future Internet*, 7, 294–306. <https://doi.org/10.3390/fi7030294>
- Veselovska, G. (2016). Teaching elements of English RP connected speech and CALL: Phonemic assimilation. *Education and Information Technologies*, 21, 1387–1400. <https://doi.org/10.1007/s10639-015-9389-1>
- Wang, Y., & Sun, C. (2001). Internet-based real time language education: Towards a fourth generation distance education. *CALICO Journal*, 18(3), 58–73. <https://doi.org/10.1558/cj.v18i3.539-561>
- Yoshida, K., Fukada, A. (2014). Effects of oral repetition on learners' Japanese word accentuation. *IALLT Journal of Language Learning Technologies*, 44(1), 17–37. <https://doi.org/10.17161/iallt.v44i1.8533>
- Zhang, W. (2021). The efficacy of computer-mediated feedback in improving L2 speaking: A systematic review. *Theory and Practice in Language Studies*, 11(12), 1591–1601. <https://doi.org/10.17507/tpls.1112.11>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.



## Authors and Affiliations

Yuka Kataoka<sup>1</sup>  · Achmad Husni Thamrin<sup>1</sup> · Rodney Van Meter<sup>2</sup> · Jun Murai<sup>3</sup> · Kotaro Kataoka<sup>4</sup>

Achmad Husni Thamrin  
husni@sfc.wide.ad.jp

Rodney Van Meter  
rdv@sfc.wide.ad.jp

Jun Murai  
jun@keio.jp

Kotaro Kataoka  
kotaro@cse.iith.ac.in

<sup>1</sup> Graduate School of Media and Governance, Keio University, 5322 Endo, Fujisawa-shi, Kanagawa 252-0882, Japan

<sup>2</sup> Faculty of Environment and Information, Keio University, 5322 Endo, Fujisawa-shi, Kanagawa 252-0882, Japan

<sup>3</sup> Keio University, 2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan

<sup>4</sup> Department of Computer Science and Engineering, Indian Institute of Technology Hyderabad, Kandi, Sangareddy, Telangana 502285, India