Doctoral Dissertation

Improving EFL Communicative and Peer Feedback Skills through a Learning Analytics-enhanced Microlearning App

Thomas GORHAM
Supervisor: Professor Hiroaki OGATA

December 19, 2024

Department of Social Informatics
Graduate School of Informatics
Kyoto University

Abstract

This dissertation investigates the design and implementation of the asynchronous microlearning app Pebasco to enhance peer feedback skills and communicative competence in English as a Foreign Language (EFL) contexts. The research addresses the persistent challenge Japanese learners face in developing oral proficiency. While peer feedback has been recognized as an effective strategy to improve communicative competence in face-to-face settings, limited research exists on its application in peer feedback training in asynchronous online environments. This dissertation bridges that gap by integrating peer feedback training with learning analytics, enabling more effective and engaging peer feedback practices on spoken content in remote learning contexts.

Pebasco was designed to facilitate peer feedback training and enhance students' internal feedback skills. It encourages reflective learning by allowing users to compare their feedback with that of peers and instructors. Through the integration of learning analytics, the app provides insights into student engagement and learning progress, offering a data-driven approach to personalized education in asynchronous environments.

The research follows an Educational Design Research (EDR) approach, which uses an iterative approach to developing the Pebasco system over the course of three studies to evaluate Pebasco's effectiveness. The first study occurred during the abrupt shift to Emergency Remote Teaching (ERT) during the COVID-19 pandemic. It found that structured peer feedback activities with a prototype version of Pebasco, alongside reflective discussions and feedback literacy instruction, contributed to significant improvements in students' communicative performance.

The second study examined the standalone mobile microlearning app version of Pebasco and its role in improving the quality of peer feedback in an asynchronous setting. The findings indicate that students who actively engaged with the app significantly enhanced their ability to provide detailed and constructive feedback on spoken content. This study highlights Pebasco's capacity to bridge the gap between remote learning environments and the hands-on nature of communicative language teaching (CLT)-based instruction. Additionally, the data collected through the app's learning analytics system provided valuable insights into student engagement and learning behaviors.

The third study explored the transferability of peer feedback skills to broader communicative tasks. Results showed that students who consistently engaged with Pebasco not

only improved their peer feedback abilities but also developed greater confidence in using English in communicative settings. Notably, students who initially struggled with communicative tasks demonstrated marked improvement by the end of the study, achieving performance levels comparable to their initially higher-performing peers. This suggests that consistent engagement with mobile microlearning apps like Pebasco can support deeper learning and improve language skills even for students facing initial difficulties.

The implications of this research are far-reaching for EFL educators and instructional designers, particularly those working in asynchronous or remote contexts. The findings demonstrate that mobile microlearning platforms like Pebasco can effectively support the development of both peer feedback and internal feedback skills, helping students become more reflective learners. By incorporating learning analytics, educators gain valuable data on student progress, allowing them to provide more personalized and effective support. Pebasco's iterative design also underscores the potential of no-code development tools, showing how educators with limited technical skills can create powerful educational platforms tailored to their pedagogical goals.

This research's broader contribution lies in its demonstration of how technology can be integrated with peer feedback strategies to improve communicative competence in EFL learners, even in ERT contexts. The findings offer practical insights for educators looking to adopt similar approaches, and the integration of learning analytics opens new possibilities for understanding and supporting student learning in diverse settings.

Future research could explore Pebasco's applicability across a broader range of educational and cultural contexts to validate its effectiveness in different environments. Additionally, refining the system's learning analytics capabilities could offer more granular insights into student engagement, enhancing its impact on learning outcomes. Nonetheless, this dissertation provides a solid foundation for educators and researchers interested in using mobile microlearning platforms to improve peer feedback training and communicative competence in language education.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Research background

The global rise of English as a lingua franca has driven substantial efforts to enhance English language education in Japan, particularly in developing students' communication skills. Despite these efforts, Japanese learners continue to struggle with oral proficiency, a skill that is crucial in communicative language teaching (CLT) contexts. CLT emphasizes the practical use of language through authentic, task-based learning experiences that encourage students to engage in real-world communication (Richards, 2006; Sauvignon, 2005). Unfortunately, Japanese students' performance on international English proficiency tests, such as the IELTS, remains below global averages, and Japan is consistently ranked low in English proficiency indices (Education First, n.d.; MEXT, n.d.). These trends highlight a persistent gap between educational policy and student outcomes, particularly in oral communication skills.

A promising pedagogical approach for addressing this gap is peer feedback, a strategy where students evaluate and provide feedback on each other's performance. Peer feedback encourages reflective learning, fosters critical thinking, and enhances students' understanding of the subject matter (Hattie, 2012)). In the context of English as a Foreign Language (EFL) learning, it has been shown to improve communicative competence, as well as students' ability to identify and correct language errors (Fujii et al., 2016; Sato & Lyster, 2012). However, while the benefits of peer feedback are well-documented, the processes by which it can be effectively integrated into classroom practice—especially using technology—remain underexplored.

The COVID-19 pandemic added another layer of complexity to these challenges. The rapid shift to Emergency Remote Teaching (ERT) due to lockdowns disrupted traditional language learning environments. In the context of CLT, which prioritizes interaction and real-world communication, these challenges were further amplified due to the sudden loss of in-person interaction (Le Cor & Coutherut, 2020; Luporini, 2020). These circumstances created an urgent need for innovative, technology-driven solutions to maintain communicative skill development and support peer feedback in asynchronous and remote learning environments.

1.2 Problem to be addressed

Given the persistent challenges faced by Japanese EFL learners in improving their second/foreign language (L2) oral communication skills, it is clear that new strategies are required. Tools that can maintain engagement and effectiveness in asynchronous and remote settings are needed. Although peer feedback holds significant potential for enhancing communicative competence, more research is needed on how to train students to provide high-quality feedback in various learning contexts.

Moreover, calls have been made for further research into several areas. Researchers have emphasized the need for studies on task-based learning and task-based language teaching (TBL/TBLT) in technology-mediated contexts (Chong & Reinders, 2020; Richards, 2006). There are also calls for more experimental and quasi-experimental studies examining peer feedback, especially those that explore the social aspects of learning through multimodal learning analytics (Ouhaichi et al., 2023). Additionally, scholars have underscored the need for research into how feedback training can improve both the provision and reception of peer feedback (Sato, 2017), as well as its impact on students' ability to transfer learning from one task to another. This dissertation aims to address these gaps by integrating learning analytics and peer feedback in EFL contexts, particularly in the challenging landscape of ERT.

1.3 Our solution

In response to these challenges, we developed a prototype system called Pebasco, designed to enhance peer feedback skills in asynchronous and remote learning environments. Pebasco stands for "Peer Feedback on Spoken Content," and the system has undergone two iterations during its development. Version 1 of Pebasco was introduced as a prototype during the 2020 Fall term. It allowed students to provide peer feedback on recorded spoken content asynchronously, using a combination of readily available technologies such as Microsoft Teams, Flipgrid, and Google Data Studio to facilitate the process. This initial prototype demonstrated potential but also revealed areas for improvement.

Building on the lessons learned from Version 1, Version 2 of Pebasco was developed as a standalone mobile microlearning app, designed using a no-code platform. This second version integrated learning analytics, providing real-time feedback on students' performance in peer assessment tasks. The mobile app not only allowed for asynchronous learning but also offered personalized feedback to help students refine their peer feedback skills in a more flexible and scalable manner. This iterative design process reflects the adaptive nature of our approach, focusing on continuous refinement based on both user feedback and empirical data collected during its implementation.

The uniqueness of this research lies in its integration of learning analytics with peer feedback training in an EFL context, particularly within asynchronous and remote learning environments. Pebasco provided students with the opportunity to engage in reflective learning through a mobile platform that was accessible and adaptable to different learning contexts. Additionally, Pebasco helped identify student behavior patterns by supporting analyzing data collected from peer feedback activities and categorizing students into different profiles. This novel use of learning analytics allowed for better understanding of student engagement and progress throughout the peer feedback training process. This research also contributes to the growing field of mobile microlearning, which emphasizes short, flexible learning activities that fit into students' daily routines.

This dissertation addresses the following three research questions (RQs):

- **RQ1**: How can task-based approaches like Communicative Language Teaching (CLT) be positively impacted through peer feedback to improve language learning outcomes in asynchronous or remote learning environments?

- **RQ2**: How does peer feedback training impact communicative competence?

- **RQ3**: In what ways can learning analytics and mobile microlearning platforms be leveraged to enhance peer feedback processes and support language skill development in CLT for EFL education?

1.4 Published research

In order to answer the preceding research questions, the following works were submitted for peer review in academic publications:

Journals (peer-reviewed)

1. Tom Gorham, Rwitajit Majumdar, & Hiroaki Ogata. (2023). Analyzing learner profiles in a microlearning app for training language learning peer feedback skills. Journal of Computers in Education, 10(3), 549-574. https://doi.org/10.1007/s40692-023-00264-0. Impact Factor 4.3 (2023); Emerging Sources Citation Index (ESCI)

2. Tom Gorham, Rwitajit Majumdar, & Hiroaki Ogata. (2024). Learning analytics of peer feedback on communicative skills in an EFL course across different learning modalities. Studies in Educational Evaluation, 81, 101352. Impact Factor 2.6 (2023); Social Sciences Citation Index (SSCI)

3. Tom Gorham, Rwitajit Majumdar, & Hiroaki Ogata. (Under Editorial Review). A microlearning app for peer feedback training and its effect on learning performance and self-confidence during an EFL speaking task. Smart Learning Environments.

International conferences (peer-reviewed)

1. Tom Gorham & Hiroaki Ogata. (2020a). Improving Skills for Peer Feedback on Spoken Content Using an Asynchronous Learning Analytics App. 28th International Conference on Computers in Education Conference Proceedings (2), 782-785. https://ci.nii.ac.jp/naid/120006940277/en/

2. Tom Gorham & Hiroaki Ogata. (2020b). Professional Learning Community's Views on Accessibility during Emergency Remote Teaching. 28th International Conference on Computers in Education Conference Proceedings (1), 570-572. https://ci.nii.ac.jp/naid/120006940280/en/

3. Tom Gorham, Rwitajit Majumdar, & Hiroaki Ogata. (2022). Pebasco: An asynchronous learning analytics app for communicative language teaching built using no-code technology. Proceedings of the 1st APSCE International Conference on Future Language Learning (ICFULL) 2022, New Territories, Hong Kong SAR, 60.

1.5 Dissertation Structure

Chapter 2 presents a detailed literature review, focusing on Communicative Language Teaching (CLT), peer feedback training, and learning analytics-driven technologies for designing microlearning applications.

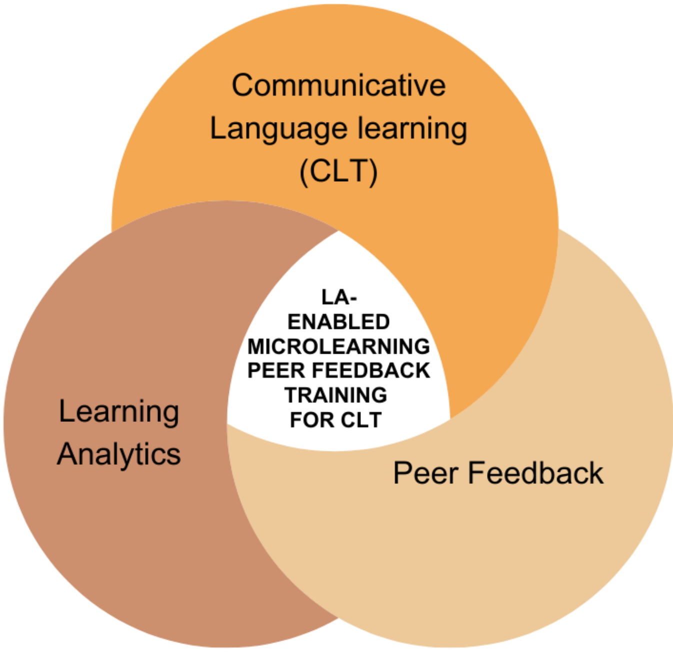
Chapter 3 introduces the development of the Pebasco system, detailing its iterative design process from Version 1 to Version 2. This chapter also describes how the system was implemented to enhance peer feedback skills in an EFL context using an educational design research approach.

Chapters 4, 5, and 6 present the results from three studies conducted to evaluate the effectiveness of the Pebasco system. Chapter 4 explores the impact of the prototype system on peer feedback and communicative skill development by comparing three cohorts of students during a shift to ERT. Chapter 5 investigates how the standalone version of Pebasco improved peer feedback provision skills in an asynchronous learning environment. Chapter 6 assesses the transfer of peer feedback training to broader communicative language learning tasks, focusing on the system's role in fostering learners' internal feedback skills.

Chapter 7 discusses the results of the three studies, comparing findings and analyzing the broader implications of using learning analytics and mobile microlearning for peer feedback in language learning. Finally, Chapter 8 concludes the dissertation by outlining the contributions, limitations, and directions for future research. An overview diagram of the dissertation is presented in Figure 1.1 and a visualization of the positioning of this dissertation's research is presented in Figure 1.2.

Main issues	- Japanese students need to improve their oral communication skills in the context of English as Foreign Language (EFL) course. - Peer feedback can help improve academic outcomes, but more research has been called for in the context of EFL.		
Overall Solution	Develop an asynchronous mobile microlearning app (Pebasco) for training peer feedback skills, with learning analytics for identifying student behavior patterns in the EFL context.		
Gaps	**A.** Calls for more research in CLT/TBL/TBLT in technology-mediated contexts **B.** Calls for more consideration of the use of technology and learning analytics in feedback processes	**A.** Calls for research examining how training can enhance the quality of peer feedback **B.** Calls for more experimental and quasi-experimental research that separates the provision and reception of peer feedback	**A.** Call for more research into the factors that modulate the effects of peer assessment on learning **B.** Relatively small number of studies that look at the effect of peer feedback on the transfer of learning to other tasks
Research Questions	**Design of CLT activity with LA support** **RQ1**: How and to what extent can different activities that support the development of peer feedback skills improve the students' communicative skills in a mandatory basic EFL course based on CLT(TBL/TBLT), as measured by communicative speaking tests?	**Evaluation of CLT activity with LA support** **RQ2**: How and to what extent can the asynchronous microlearning app Pebasco improve students' skills for providing peer feedback on spoken content in CLT-based English as a foreign language (EFL) education?	**Longer-term effect of CLT activity with LA support** **RQ3**: How and to what extent does the use of a microlearning app for training peer feedback skills in the context of university CLT EFL classes affect the transfer of learning to a CLT speaking activity? **RQ4**: How and to what extent does the use of a microlearning app for training peer feedback skills impact students' self-reported perceptions of improvement of peer- and internal feedback skills, as well as general L2 confidence?
Solutions	Developed a prototype learning analytics system for training peer feedback skills	Built an asynchronous mobile microlearning app for training peer feedback provision skills and conducted quasi-experimental research	Examined the use of the asynchronous mobile microlearning app on student performance in an external speaking activity
Results	**(n=320)** ERT cohorts, especially those using the prototype system, outperformed the face-to-face cohort in communicative speaking tests, highlighting the effectiveness of peer feedback activities in boosting communicative skills	**(n=87)** Learning analytics data were used to identify five behavioral profiles. The pattern of profile migration over the course of using Pebasco indicates that many participants improved or maintained desirable patterns of behavior and outcomes, suggesting a positive impact on the quality of peer feedback skills and L2 skills, and on the ability to detect L2 errors	**(n=89)** Students who engaged with Pebasco reported significant self-perceived improvements in both peer and internal feedback skills. Notably, students with initially lower performance on the constructive CLT speaking activity showed substantial progress, achieving parity with their initially higher-performing peers, suggesting the app's effectiveness in bridging performance gaps
Academic Contributions	Study 1 (SEE 2024, ICCE[1] 2020, ICCE[2] 2020) Chapter 4	Study 2 (JLCE 2023, ICFULL 2022) Chapter 5	Study 3 (SMLE) [in editorial review] Chapter 6

Figure 1.1: Research approach



Previous Work	**Study 1**	**Study 2**	**Study 3**
Calls for more research in CLT/TBL/TBLT, particularly in technology-mediated contexts (**Chong & Reinders, 2020**; **Richards, 2006**)	0	X	0
Calls for more consideration of the use of technology and learning analytics in feedback processes (**Carless & Boud, 2018**; **Lui & Andrade, 2022**)	0	0	0
Call for research examining how training can enhance the quality of peer feedback (**Dao & Iwashita, 2021**)	X	0	X
Call for more use of experimental and quasi-experimental methodologies in peer feedback research (**Topping, 2013**)	0	0	0
Call for future experimental research that separates the provision and reception of peer feedback (**Sato, 2017**)	X	0	0
Relatively small number of studies examining the effect of peer feedback on the transfer of learning (**Yu & Schunn, 2023**)	X	X	0
Call for more research into the factors that modulate the effects of peer assessment on learning (**Double et al., 2020**)	0	X	0
Version of Pebasco used:	Ver. 1	Ver. 2	Ver. 2

Figure 1.2: Positioning of current research

Chapter 2

Literature review

2.1 Literature review overview

This chapter reviews key literature relevant to the research reported in this dissertation. Section 2.2 covers Communicative Language Teaching (CLT). Section 2.3 focuses on the efficacy of peer feedback in education, the key theoretical frameworks that inform feedback processes, peer feedback and the transfer of learning, and peer feedback training. Finally, Section 2.4 discusses the use of learning analytics-driven technologies for designing microlearning applications.

2.2 Communicative language teaching

CLT is an approach to second and foreign language learning that prioritizes the development of learners' functional competency through the performance of authentic communication-focused tasks (Sauvignon, 2005). It started in the 1970s as a reaction to the grammar-translation method and the audio-lingual method (Dörnyei, 2009). One of the most common subsets of CLT is called task-based learning and teaching (TBLT) or sometimes simply task-based learning (TBL). Students are asked to use the target language to complete authentic tasks rather than learning decontextualized grammatical patterns or memorizing pre-made dialogues (Richards, 2006). It has become so popular that Dörnyei (2009) reports that the terms TBLT/TBL have started to replace CLT. Furthermore, the need for communicative skills training has been identified in other areas, such as social work (Reith-Hall & Montgomery, 2023).

Timed Paired Practice (TPP) is a TBL approach that "incorporates random impromptu interactions and Corrective Feedback to help students identify and repair their errors" (Elam, 2014, p. 17). The creator of TPP (Moe, 2005) describes its most basic form, beginning with the random selection of two students who are asked to start a conversation on a given topic while the teacher times the exchange. Once one partner makes an uncorrected error, the dialogue is ended, the teacher stops the timer, and the conversation's duration is recorded as a component of the score for the activity. The teacher explains the uncorrected error before selecting another random pair of students who continue the process. The entire class goes through multiple rounds of TPP testing during one session and multiple sessions per term, which means that TPP is a type of low-stakes formative assessment. Students are encouraged to listen to their classmates' conversations and pay attention to the mistakes they make. TPP has been shown to improve both the length and quality of student conversations (Pipe, 2015).

In the context of Study 1, TPP testing was used as a means of low-stakes formative assessment in an introductory CLT-focused EFL course. TPP was employed before and after the switch to ERT, so it was used as a benchmark to measure students' learning outcomes to evaluate this study's research question. There have been calls for more research in TBL/TBLT, particularly in technology-mediated contexts (Chong & Reinders, 2020; Richards, 2006). Study 1 can add to the literature.

2.3 Peer feedback

2.3.1 Efficacy of peer feedback

Peer feedback is sometimes considered a variant of peer evaluation that includes providing more detailed qualitative comments (see Alqassab et al., 2018; Liu & Carless, 2006). However, for this literature review, we will follow Panadero & Alqassab's (2019) lead and treat peer assessment and peer feedback as synonymous. We consider peer feedback as the act of one learner evaluating the performance of another learner. We also include peer corrective feedback (PCF) as a synonym, which is differentiated from corrective feedback provided by a teacher.

The practice of peer feedback can positively impact student learning and achievement (Hattie, 2009, 2012; Hattie & Clarke, 2019; Kerr, 2020), even in online learning environments (Jongsma et al., 2023), including MOOCs (Kasch et al., 2021); it has been shown to reduce classroom anxiety (Motallebzadeh et al., 2020). It has also been found to improve students' self-reflection/internal feedback skills (To & Panadero, 2019).

In the context of language learning, peer feedback can improve learners' L2 language skills and public speaking presentation skills (El Mortaji, 2022; Patri, 2002). Rodríguez-González and Castañeda (2016) found that peer feedback can improve L2 speaking skills and feelings of self-confidence and self-efficacy. Peer feedback practice can improve learners' ability to detect language errors (Fujii et al., 2016), and PCF has been found to be a "feasible option for helping learners to attend to form in effective ways during peer interaction" (Sato & Lyster, 2012, p. 618). Conversely, Adams et al. (2011) found that peer feedback had a negative effect on L2 learning. This may point to the importance of improving the quality of peer feedback through training.

While there are many potential reasons for the efficacy of peer feedback, Nicol et al. (2014) demonstrated that the act of providing peer feedback can encourage students to think more deeply and critically about their own work. Nicol (2021) theorizes that performing peer feedback harnesses the brain's natural inclination to engage in comparison, a process that underlies much of cognition (Gentner et al., 2001; Goldstone et al., 2010; Hofstadter & Sander, 2013). Nicol (2021) theorizes that engaging in peer feedback contributes to the development of "internal feedback" skills, in which learners reflect on and regulate their own learning processes. In the context of second language learning, Levelt's perceptual theory of monitoring has been used to explain how a learner can improve their internal feedback skills by providing peer feedback (Sato, 2017; Sato & Lyster, 2012). Furthermore, there has been a call for further research in training students' peer feedback skills (Kasch et al., 2021).

The Dual Model Theory of peer corrective feedback (PCF) posits that there may be differing impacts on learning when comparing the provision and reception of PCF (Sato, 2017). There is a growing body of research that suggests that providing PCF has a larger positive impact on learning (Lu & Law, 2012; Nicol et al., 2014; Y. Wu & Schunn, 2023; Yu & Schunn, 2023; Zong et al., 2021). In the context of the ICAP framework, Wu & Schunn (2023) argue that one explanation for this is that the provision of PCF is a more constructive, overt activity and, therefore, more cognitively engaging. There has been a call for future research that can investigate these two sides independently (Sato, 2017).

2.3.2 Theoretical basis of feedback frameworks

The ICAP framework (Chi, 2009) is a popular heuristic that can be used as a starting point when considering peer feedback. It suggests that learners' overt behaviors can be identified as belonging to one of the following four categories, which make up the acronym in descending order of cognitive engagement: “interactive,” “constructive,” “active,” and “passive.” Chi and Wylie (2014) assert that transfer of learning can be facilitated when students are engaged in constructive activities because they require learners to create new inferences and link schemas. Wu and Schunn (2023) generally identified the provision of peer feedback as a constructive activity and that it positively impacted student learning. However, it has been noted that one of the drawbacks of the ICAP framework is that it does not attend to internal processes (Chi, 2009; Thurn et al., 2023). For this reason, Studies 1, 2, and 3 use the ICAP framework as a jumping-off point supplemented by additional theoretical perspectives (e.g., see Table 3.2).

The frameworks presented by Panadero & Lipnevich (2022), Yang (2021), Lui & Andrade (2022), Carless & Boud (2018), and Sutton (2012) offer different perspectives on feedback and feedback literacy, but they can be viewed holistically as they complement each other in understanding the complex nature of feedback in educational settings. The MISCA model by Panadero & Lipnevich (2022) provides a comprehensive framework that encapsulates the various elements of feedback, including the message, implementation, student, context, and agents. This model serves as a foundation, emphasizing the importance of the feedback message, its delivery, the role of the student, the learning environment, and the sources of feedback.

Yang’s (2021) concept of feedback orientation builds upon the MISCA model by focusing on the student’s perspective. It highlights the importance of students’ attitudes and beliefs toward feedback, which can significantly influence how they receive and utilize feedback. This model complements the student element of the MISCA model, providing a deeper understanding of the individual characteristics that influence feedback reception and use. In the context of language learning, there has been a growing interest in research regarding the effects of Individual Differences (ID) among students from the perspective of second language acquisition and computer-assisted language learning (Pawlak, 2022).

Lui and Andrade’s (2022) work on the internal mechanisms of feedback processing further expands on the student element by exploring how students process and respond to feedback. It emphasizes the influence of contextual and external factors, aligning with the context element of the MISCA model.

Carless & Boud’s (2018) strategies for developing student feedback literacy in higher education can be seen as practical applications of the MISCA model and Yang’s feedback orientation concept. They provide concrete methods to enhance students' understanding and use of feedback, aligning with the implementation element of the MISCA model.

Finally, Sutton's (2012) concept of feedback literacy provides a broader perspective, encompassing epistemological, ontological, and practical dimensions. This concept integrates the elements of the MISCA model, Yang's feedback orientation, and Lui and Andrade's internal mechanisms of feedback processing, emphasizing the multifaceted nature of feedback literacy.

In summary, these frameworks can be viewed holistically as they each contribute to a comprehensive understanding of feedback and feedback literacy. They highlight the importance of the feedback message, the delivery method, the student's characteristics and attitudes, the learning environment, and the sources of feedback. They also emphasize the need for practical strategies to enhance feedback literacy and the importance of further research to understand the complex nature of feedback processing.

In Study 1, these interlocking theoretical frameworks provide a means to understand not only what elements comprised the different interventions during the three phases of the study, but also how those interventions may have positively impacted learning outcomes. This is important because feedback is a complex and multifaceted endeavor. For example, when trying to understand Study 1 through the five components of the MISCA (Message-Implementation-Student-Context-Agent) model, it becomes clear that although the feedback messages (M) largely remain the same, the change in context (C) due to the switch to ERT necessitated a change in implementation (I). By the third phase of Study 1, these changes included peer feedback activities, which expanded the range of feedback agents (A) and were supported by feedback literacy training, reflective discussions, and meta-dialogues about some of the internal mechanisms of feedback processing. Part of the novelty of Study 1 is that it shines a light on the use and applicability of these theoretical feedback frameworks in the challenging context of ERT.

2.3.3 Peer feedback and transfer of learning

Transfer has been called one of "the most powerful principles of learning" and "the core of problem solving, creative thinking, and all other higher mental processes, inventions, and artistic products... [and] one of the ultimate goals of teaching and learning" (Sousa, 2022, p. 122-123). Transfer of learning occurs when a person is able to apply prior learning to a new situation. When the new situation is similar to the context of the original prior learning, it is often called “near transfer,” and when there is more difference between the two, it is referred to as “far transfer” (Barnett & Ceci, 2002; Double et al., 2020).

Transfer is not easy to accomplish successfully (Barnett & Ceci, 2002); even educators sometimes have trouble transferring knowledge that they have learned in professional development training sessions to their teaching practice (Chi et al., 2018). To counter this challenge, researchers have suggested that Transfer Appropriate Processing (TAP)—the attempt to make the cognitive processes that occur during learning as similar as possible to the ones that will occur in future situations—can be an effective way to promote transfer (Lightbown, 2008; Morris et al., 1977). This assertion has been echoed in the context of peer feedback, too. A recent meta-analysis of research on the effects of peer assessment found that only cases in which the peer assessment activities were the same as the activities that were measured as performance indicators were the ones that demonstrated statistically significant correlations (Double et al., 2020). This meta-analysis led to a call for more research into the factors that modulate the effects of peer assessment on learning (Double et al., 2020). Study 3 adds to the literature that addresses this call by examining the effect of training peer feedback provisioning skills through a microlearning app on near transfer. Similarly, Yu and Schunn (2023) note the relatively small number of studies that look at the effect of peer feedback on the transfer of learning to other tasks; Study 3 aims to add to that body of research.

2.3.4 Peer feedback training

Training is essential to maximizing the benefits of peer feedback. In the context of a teacher training course, Sluijsmans et al. (2004) concluded that “peer assessment is a skill that can be trained” (p. 74). In a survey of 27 studies conducted at the university level, van Zundert et al. (2010) found evidence that training can improve peer assessment skills and students’ attitudes toward peer assessment.

Similarly, in the context of language learning, Sato (2013) found that peer feedback training could improve the attitude that learners have toward peer feedback while also improving the learners’ ability to notice grammar (i.e., form) mistakes in both their peers’ speaking and in their own (i.e., internal feedback). Training was also found to improve the amount of peer feedback given and its quality in studies with Japanese adult English language learners and children in Canadian French immersion programs (Sato & Ballinger, 2016).

One challenge in implementing peer feedback in language learning contexts is that "learners need a threshold level of the target language" (Lyster et al., 2013, p. 29) for them to offer accurate feedback. Toth (2008) identified this same problem of students' peer feedback accuracy and suggested the need for training. Foreign language students have reported a reluctance to give peer feedback because of a lack of L2 proficiency (Philp et al., 2010). It was hypothesized in Studies 2 and 3 that students will improve their peer feedback skills and L2 skills by using the Pebasco mobile microlearning app.

2.4 LA-driven technologies for designing microlearning applications

2.4.1 Learning analytics

Learning analytics (LA) involves gathering and analyzing data about learners and their environments to improve both learning outcomes and the conditions in which learning takes place (Misiejuk & Wasson, 2023; Ouhaichi et al., 2023; Scherer et al., 2012; Siemens & Baker, 2012). LA has been gaining popularity for over a decade, expanding on the promise of educational data mining to power functions such as goal-oriented visualizations, learning dashboards, learning recommenders, and data-driven pedagogical research (Duval, 2011; Siemens & Baker, 2012).

In language learning, the Learning and Evidence Analytics Framework (LEAF) has been used to integrate daily educational practices such as active reading and self-directed learning in language classes (Ogata et al., 2022, 2023). LA has also been applied to explainable artificial intelligence for language learning (Ogata et al., 2024) and to enhance EFL vocabulary learning through recommender systems (Takii et al., 2021). LA has improved self-regulated learning in EFL courses (Chen, 2024) and modeled patterns in vocabulary retention and forgetting (Ma et al., 2023). It has been used to assess student oral performance in flipped language classrooms (Lin & Hwang, 2018) and optimize group formation in EFL settings (Liang et al., 2022, 2024).

When combined with peer feedback, LA helps analyze how students interpret and respond to feedback (Misiejuk et al., 2021) and supports dialogic peer feedback (Er et al., 2021). Darvishi et al. (2022) explored an AI-supported peer assessment system that improved students' ability to provide more effective peer feedback on written assignments in two first-language science courses. While the study demonstrated improvements in feedback quality, the application of such systems in language learning remains underexplored. A recent review further highlights the lack of research on training peer feedback skills for spoken content, particularly in communicative language teaching (CLT) contexts (Misiejuk & Wasson, 2023).

2.4.2 Microlearning and peer feedback apps

The concept of microlearning has been around for nearly 20 years, and the practice has been evolving along with the technologies that enable it (Hug et al., 2006). To best understand microlearning, it is worthwhile to consider some of its inherent design principles. Jahnke et al. (2020) identified the design principles of mobile microlearning (i.e., microlearning that is accessed on mobile devices), including the following key ideas: (a) the learning activities should be practical and interactive; (b) the activities should be "snackable," meaning students should be able to complete activities at their convenience in just a few minutes; (c) microlearning activities should include multimedia content; (d) the activities should provide immediate formative feedback; and (e) finally, microlearning apps should be available across multiple types of devices and operating systems. Lee et al. (2021) suggest that mobile microlearning is better suited for training lower-order thinking skills.

Microlearning is inherently flexible and can be connected with "multiple learning theories and approaches" (Garshasbi et al., 2021, p. 241). For example, Khong and Kabilan (2022) developed a theoretical model of microlearning for second-language instruction. Their theory fuses multiple existing theories related to cognitive science, motivation, and multimedia technology. One interesting element of their theory is the incorporation of Cognitive Load Theory (Sweller, 1994, 2020). This suggests that one of the advantages of the brief duration of microlearning activities is that the effort required does not exceed the cognitive capacity that a learner has available at the time to allocate to a particular activity.

While microlearning is often used in business and industry contexts, Drakidou and Panagiotidis (2018) cite examples of microlearning in foreign language learning. A significant amount of research has focused on using microlearning to improve L2 vocabulary. For example, Dingler et al. (2017) created a mobile microlearning app called "QuickLearn," which offered L2 vocabulary practice with flashcards and multiple-choice questions that could be pushed by mobile device notifications. They found that their study's participants used the app more while in transit between two locations.

Schneegass et al. (2022) investigated the integration of L2 vocabulary practice into common mobile device interactions, such as unlocking the device through an authentication action and responding to notifications. Inie and Lungu (2021) developed an internet browser extension that requires the user to spend a specified amount of time engaged in language learning activities such as (but not limited to) L2 vocabulary study before allowing them to access a designated "time-wasting" website. They reported a slight improvement in the post-tests following the use of the system, but not all the participants reported enjoying using it in its current implementation.

L2 vocabulary microlearning has also been investigated with an app that leverages user location data (Edge et al., 2011); with passive exposure to vocabulary and collocated phrases via an automatically updated wallpaper screen on the user's mobile device (Dearman & Truong, 2012); with an Internet browser extension that overlays L2 words on first language website content (Trusty & Truong, 2011); with a system that inserts interactive vocabulary quizzes into the user's social media feed (Kovacs, 2015); and with machine learning (natural language processing) to offer vocabulary practice that is embedded into the learner's daily life (Arakawa et al., 2022). L2 vocabulary microlearning has been inserted in the time wasted while people are waiting for an elevator, an instant message reply, or a wi-fi connection (Cai et al., 2015, 2017). Zhao et al. (2018) even suggested using smartwatches for L2 vocabulary learning.

Despite this considerable focus on L2 vocabulary learning, there is a lack of research on using microlearning to train peer feedback skills. However, there is a relatively long history of using technology to support the provision of peer feedback. van den Bogaard and Saunders-Smits (2007) describe the PeEv system from the Delft University of Technology and the SPARK (Self and Peer Assessment Resource Kit) from the University of Technology Sydney, which date back to the turn of the 21st century (Sridharan et al., 2019). Two more recent tools that are used for providing peer feedback include the University of Technology Sydney's updated SPARKPLUS system (Knight et al., 2019) and the Comprehensive Assessment of Team Member Effectiveness (CATME) system (Layton et al., 2012; Loughry et al., 2014; Ohland et al., 2012).

Both systems have some functionality for training students to provide peer feedback. CATME's calibration function allows students to practice on fictional student content using the rating scale for a particular activity. Similarly, SPARKPLUS has a "benchmarking" feature where students can check to see if their peer feedback matches the instructor's. While these systems are examples of technology-assisted peer feedback training, a search of the literature did not return any research investigating the use of the SPARKPLUS benchmarking function or the CATME rater calibration feature for training peer feedback skills in the context of spoken content in foreign language learning.

Çelik et al. (2018) describe a digital tool used in teacher training that allows users to tag videos of their peers' teaching practice with feedback comments. In the context of foreign language learning, PeerEval is a mobile app for students to provide real-time peer feedback during classmates' L2 presentations (Wu & Miller, 2020). While both tools provide a platform for providing peer feedback on spoken content, neither one offers asynchronous microlearning training of peer feedback skills. In fact, Wu and Miller (2020) cite the need for training students how to confidently provide feedback as the main challenge with using the PeerEval app.

There is a gap in the literature of research that investigates the use of asynchronous mobile microlearning peer feedback training in the context of spoken content in foreign language learning. The development of the mobile microlearning Pebasco app aims to address this gap.

Chapter 3

Research methodology and Pebasco development

3.1 Background

3.1.1 Emergency remote teaching

A significant change caused by the COVID-19 global pandemic was the call from health officials to implement social distancing protocols, which impacted educational institutions. To protect the health of students, faculty, and staff, many opted to switch to distance education (most commonly fully online learning), often without adequate notice, training, preparation, or infrastructure in place. This abrupt shift has been called Emergency Remote Teaching (ERT). Educational technology researchers differentiate ERT from properly designed and delivered distance education programs, which can take significant time and effort to create effectively. ERT is designed "not to re-create a robust educational ecosystem but rather to provide temporary access to instruction and instructional supports in a manner that is quick to set up and is reliably available during an emergency or crisis" (Hodges et al., 2020, para. 1).

At first glance, one might assume that this shift to ERT would not hurt EFL learning outcomes. Research has found that online learning modalities can improve foreign/second language learning and intercultural communication competencies via computer-mediated communication (e.g., Alshahrani, 2016; Angelova & Zhao, 2016; Austin et al., 2017; Avgousti, 2018; Shahrokhi Mehr et al., 2013; Terhune, 2016; Yeh & Lai, 2019).

However, studies examining ERT during the COVID-19 pandemic (e.g., Aucejo et al., 2020; Bao, 2020; Gorham & Ogata, 2020b; Khan et al., 2021; McDaniel et al., 2020; Means & Neisler, 2020) noted some common themes of difficulties, including problems with student motivation and attitude, issues with a lack of peer social interaction (feelings of social isolation), accessibility problems, and problems locating suitable learning environments within the students' homes. Furthermore, studies that examined the specific context of foreign/second language learning during ERT in France, Italy, and Indonesia (Le Cor & Coutherut, 2020; Luporini, 2020; Nartiningrum & Nugroho, 2020) all identified a similar pattern of difficulties.

3.1.2 Research context

The context for the studies described in this dissertation is an introductory CLT-focused EFL course called "Kiso Eigo" that was mandatory for all first-year students in the Faculty of Letters at a university in Tokyo, Japan. Study 1 began in 2019, prior to the COVID-19 pandemic, with a cohort of students who were enrolled in the face-to-face version of the class that had no peer feedback training. That cohort's performance was a baseline to compare the performance of the spring 2020 ERT cohort and the fall 2020 ERT cohort, which received additional peer feedback training with the prototype Pebasco system.

Studies 2 and 3 took place during the Spring 2021 term. Although it was the same CLT-focused EFL course, the ongoing COVID-19 pandemic resulted in the declaration of multiple states of emergency in Tokyo, necessitating the switch to an online asynchronous modality. These studies explored the development, deployment, and effects of the asynchronous mobile microlearning app, Pebasco, which was built based on findings from the use of the prototype Pebasco system from Study 1. See Table 3.1 for a comparison of the two versions of Pebasco and issues that were addressed in designing the second version.

Table 3.1: A comparison of Pebasco versions 1 and 2

	Pebasco Version 1	**Pebasco Version 2**	**Issue Addressed**
Focus	Training peer feedback skills through the provision and reception of peer feedback	Specific focus on the training of peer feedback provision	Call for research that separates the provision and reception of peer feedback (Sato, 2017)

Continued on next page...

Table 3.1: A comparison of Pebasco versions 1 and 2 (continued)

	Pebasco Version 1	**Pebasco Version 2**	**Issue Addressed**
Design	Bricolage approach: orchestrating multiple available technologies into one prototype system	Single standalone mobile microlearning app built with the no-code platform, Bubble	Lack of customization due to conflicts of seamless navigation and functionalities in different tools assembled for building the prototype system
Learning Analytics Affordances	The system lacked common functionalities of other learning analytics systems (e.g., learning logs and user profiling)	The app allowed for the collection of user learning logs and user profiling	Calls for more research on the use of technology and learning analytics in feedback processes (Carless & Boud, 2018; Lui & Andrade, 2022)
Instructor Workload	**Higher**: The instructor provides feedback on each video and manually enters all teacher and peer feedback into the Google Data Studio student-facing dashboard	**Lower**: All of the instructor comments are pre-loaded in the app, and the system evaluates the students' accuracy in peer feedback provision	The instructor was burdened by the additional time required to manually evaluate student submissions and set up the student-facing dashboard each time

Continued on next page...

Table 3.1: A comparison of Pebasco versions 1 and 2 (continued)

	Pebasco Version 1	**Pebasco Version 2**	**Issue Addressed**
Timing of formative assessment	Approximately 3-5 days elapse between student submissions of peer feedback and the dashboard data update	The system automatically provides immediate formative assessment of students' performance	The lag in response time between student submissions and the availability of the visualizations in the student-facing dashboard did not align with the Microlearning Design Principles (Jahnke et al., 2020), the MISCA Model (Panadero & Lipnevich, 2022) and the approach of formative assessment (Thurn et al., 2023)

3.2 Educational design research approach

Educational Design Research (Brown, 1992; Collins, 1992; McKenney & Reeves, 2018) is an iterative meta-methodology that is frequently used by educators who are researcher practitioners. It encourages multiple modes of data collection and analysis; it takes into account the complex dynamic systems that are inherent in real-world teaching and learning; and it pushes for a reciprocal cycle of theory-inspired research and research-inspired theory which is always aimed at improving educational processes and outcomes. EDR is a framework that has provided flexibility to react to the unpredictable changes caused by COVID-19 while maintaining research rigor. In other words, McKenney and Reeves (2018) define EDR as:

> a genre of research in which the iterative development of solutions to practical and complex educational problems also provides the context for empirical investigation, which yields theoretical understanding that can inform the work of others. Its goals and methods are rooted in, and not cleansed of, the complex variation of the real world. (p.6)

One result of this rootedness leads EDR researchers to hold a unique perspective on variation: "where experimentally oriented researchers mostly try to control or plan variation, design researchers welcome unexpected variation to see how robust their ideas and designs are" (Bakker, 2018, p. 118). This perspective is manifested in the EDR approach's flexibility, even during experiments. Within EDR, "educational ideas for student or teacher learning are formulated in the design, but can be adjusted during the empirical testing of these ideas – for example if a design idea does not quite work as anticipated" (Bakker, 2018, p. 5). Bakker (2018) explains the rationale: "In the process of designing and improving educational materials, for example, it does not make sense to wait until the end of the teaching experiment before changes can be made. This would be inefficient" (p. 48).

Another important aspect of EDR is that it encourages multiple modes of data collection and analysis. Furthermore, it borrows Lévi-Strauss's (1962) concept of bricolage and it encourages researchers to use "all the material that is at hand, including theoretical insights and practical experience with teaching and designing" (Bakker, 2018, p. 60).

Taken as a whole, Educational Design Research provides an extremely flexible and grounded perspective from which to conduct research during a period of a global pandemic. It gives structure for reacting to a changing research context and, through the idea of bricolage, it provides guidance on how to evaluate the affordances of readily available technology that could be used in leu of the creation of a bespoke app. Easterday et al. (2018) describe the iterative phases of EDR as: Focus, Understand, Define, Conceive, Build, Test, and Present (see Figure 3.1).

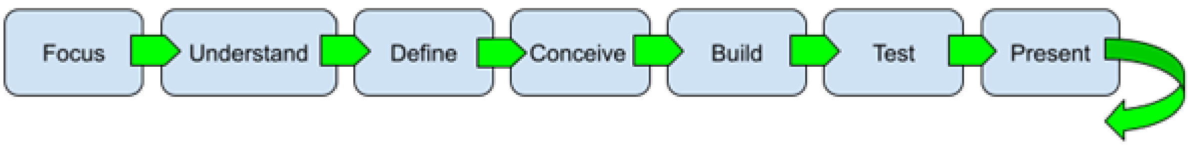


Figure 3.1: Iterative cycle of educational design research. Adapted from Easterday et al. (2018)

3.3 Pebasco ver. 1: Prototype

In study 1, a prototype learning analytics system was designed to improve students' peer feedback skills on spoken content. The system is called "Pebasco." The naming comes from the underlined letters in the following phrase: "peer feedback on spoken content."

The Pebasco system takes inspiration from the TPP testing (Gorham & Ogata, 2020a). During TPP sessions, the student observers have multiple opportunities to see how the instructor identifies the mistakes of the performing pairs. The Pebasco system allows the students to mimic the role of the instructor. They watch a brief asynchronous video of a classmate's recording of spoken content. If the student identifies a mistake, they can annotate the recording with a description of the type of mistake and the timestamp when it occurs. In this manner, the students can practice providing peer feedback on spoken content modeled on the expert practice that they observe when they watch the instructor feedback during the TPP tests. One difference is that during a TPP session, the instructor will only identify one mistake before stopping the timer and ending the conversation. However, while using Pebasco, students can identify multiple errors in one recording. After the student adds their timestamped feedback comments to a recording, the Pebasco system allows them to compare their comments to the ones the instructor added to the same recording.

When designing the prototype Pebasco system, the authors chose to use the Educational Design Research approach of bricolage (Bakker, 2018, p. 60) to create the system by orchestrating multiple available discrete technologies (see Figure 3.2). The instructor used many of the tools for the first time during the spring 2020 ERT term.

After assessing the affordances of the available tools, the authors identified a stack of technologies that, if orchestrated together, could provide the desired functionality of the Pebasco system. The first is Microsoft Teams, which is a learning management system (LMS) that could handle student personal information, communication via chat, and the assignment of topic prompts. The second is Microsoft Flipgrid, which easily allows students to record, edit, and share video and audio content with their classmates in the "Kiso Eigo" course (see Figure 3.3). Figure 3.3 showcases some of the creative ways that students used Flipgrid as viewed on smartphones, including the addition of digital stamps, frames, photographs, and even realia, such as manga comic books. It is worth noting that Microsoft shortened the name of Flipgrid to Flip in 2022. The third is Microsoft Forms—a survey creator that can be used by students giving peer feedback to provide comments tied to particular timestamps in their classmates' Flipgrid videos (see Figure 3.4). The final one is Google Data Studio, which is a data visualization platform. It can be used to create a student-facing dashboard to facilitate the comparison of peer feedback that is

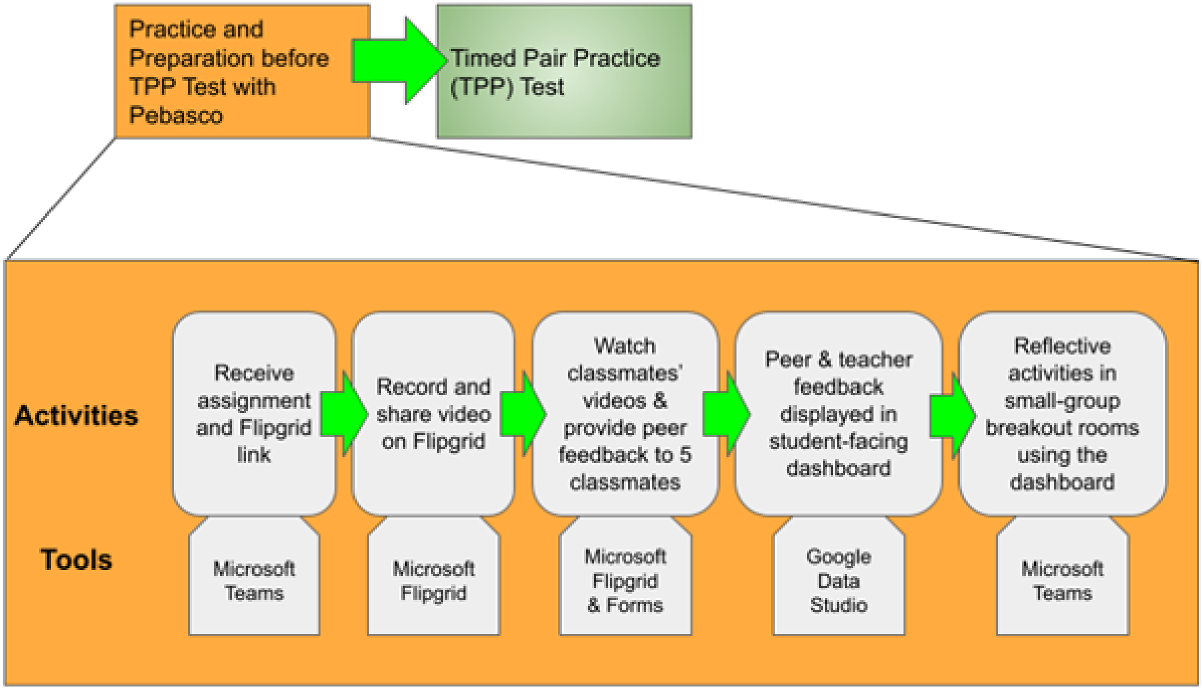


Figure 3.2: Student workflow using Pebasco for practice and preparation before a TPP test

given and received to feedback provided by the teacher.

After the student and instructor feedback comment data was collected via Microsoft Forms for a unit of Pebasco, the data were added to an interactive student-facing dashboard created using Google Data Studio. In one view of the dashboard (see Figure 3.5), students could see the feedback that other students gave on their recording (marked in blue) compared with the feedback given by the instructor (marked in pink). In another dashboard view (see Figure 3.6), a student can check their individual performance as a feedback provider compared to the instructor's feedback on each of the recordings that the student submitted feedback on. Please note that the student names have been pseudonymized with animal names in the dashboard screenshots shown in Figures 3.5 and 3.6.

The final step in using the Pebasco system is a set of collaborative, reflective activities, similar to the meta-dialogues described by Banister (2023) and Carless and Boud (2018). During the synchronous class session, usually held the week before a TPP test, small groups of students were assigned to breakout rooms. They used a shared collaborative document in the Microsoft Teams Notebook and the Pebasco student-facing dashboard. Together, they engaged in various reflective activities (e.g., comparing their performances;

Figure 3.3: Examples of the student-created Flipgrid videos for the "Kiso Eigo" basic English course

jointly writing the corrected versions of the mistakes that the instructor had identified; voicing disagreement with any of the data displayed in the dashboard).

3.4 Pebasco ver. 2: Asynchronous mobile microlearning app

The second iteration of the Pebasco system is a mobile microlearning app designed to train peer feedback skills on spoken content within a CLT EFL context (see Table 3.2 for a description of the motivating theories/frameworks behind features of the Pebasco system). Students enhance their peer feedback abilities by interacting with audio clips recorded by their peers from a mandatory CLT EFL course. Each audio clip is annotated by the course instructor with timestamped comments identifying errors in grammar, vo-

Figure 3.4: Collecting Comments and Timestamps with Microsoft Forms

Figure 3.5: Pebasco student-facing dashboard: Video creator view

cabulary, pronunciation, or pacing. Students listen to these clips and attempt to predict the nature and timing of each instructor comment. The system immediately confirms correct predictions, allowing students to make unlimited attempts to refine their accuracy.

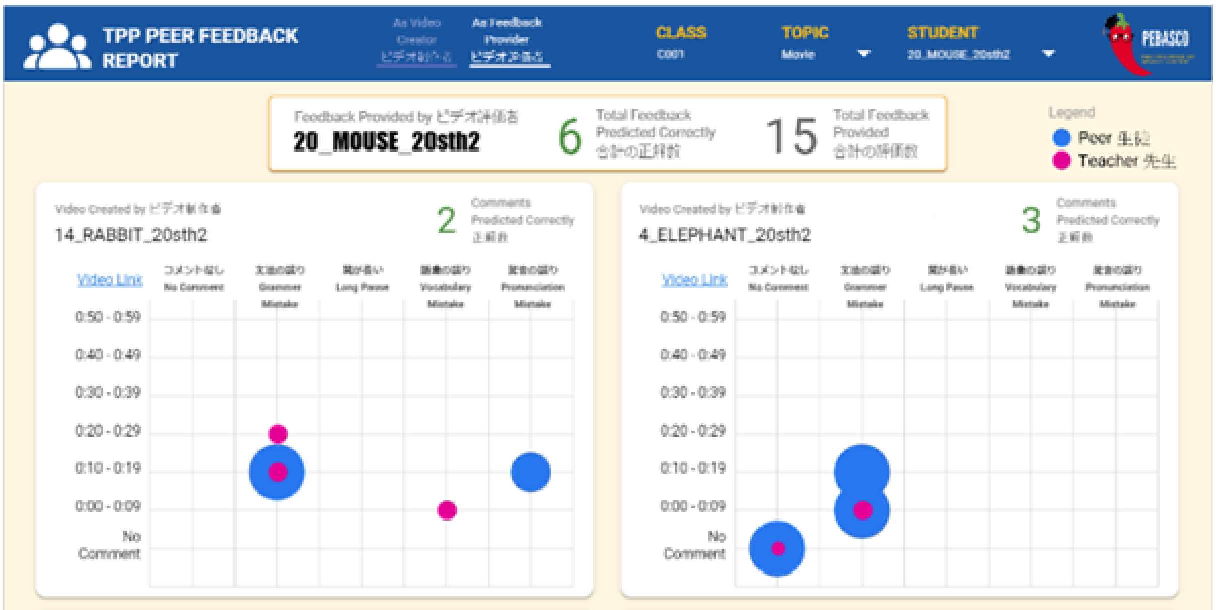


Figure 3.6: Pebasco Student-Facing Dashboard: Feedback Provider View

The following subsections will provide a more detailed description of the design of the second iteration of the Pebasco system.

Table 3.2: Motivating theories/frameworks for learning activity and application design

Theory/ Framework	**Key Concepts**	**Learning Activity Design Components**	**Learning Application Design Components**
ICAP Framework (Chi, 2009)	Interactive, Constructive, Active, Passive engagement	Used as a heuristic for designing peer feedback activities based on visible signs of engagement.	**Timestamped Feedback Creation**: Students must create timestamped feedback on their peers' audio clips.
Formative Assessment (Thurn et al., 2023)	Ongoing assessment to provide continuous feedback	Integrated into the app's design to provide immediate feedback and support internal feedback mechanisms, it acts as a supplement to the ICAP Framework.	**Immediate Feedback**: Confirms correct predictions instantly, allowing students to make unlimited attempts to refine their accuracy.

Table 3.2: Motivating theories/frameworks for learning activity and application design (continued)

Theory/ Framework	Key Concepts	Learning Activity Design Components	Learning Application Design Components
Dual Model Theory of PCF (Sato, 2017)	Differing impacts of providing vs. receiving peer corrective feedback	Basis for emphasizing training in the provision of feedback, which has a greater positive impact on learning.	**Peer Feedback Provision**: The training is focused on the provision of peer feedback.
Transfer Appropriate Processing (TAP) (Lightbown, 2008; Morris et al., 1977)	Aligning cognitive processes during learning with future situations	Applied in the study to ensure that peer feedback training activities are relevant to the final CLT speaking task.	**Content Selection**: The selected audios reflect common problems that students in the course experience.
Cognitive Load Theory (Khong & Kabilan, 2022; Sweller, 1994, 2020)	Managing cognitive load through brief activities	Influenced the design of microlearning activities in the Pebasco app to enhance learning efficiency.	**Hint System**: Provides three levels of assistance to support students in predicting teacher comments.
			Brevity of Activities: Brief activities designed to manage cognitive load and enhance engagement.
MISCA Model (Panadero & Lipnevich, 2022)	Message, Implementation, Student, Context, Agents	Used to inform the comprehensive design of feedback messages and their delivery within the app.	**Immediate Feedback**: Confirms correct predictions instantly, allowing students to make unlimited attempts to refine their accuracy.

Table 3.2: Motivating theories/frameworks for learning activity and application design (continued)

Theory/ Framework	Key Concepts	Learning Activity Design Components	Learning Application Design Components
Model of the Internal Mechanisms of Feedback Processing (Lui & Andrades, 2022)	Contextual and external factors can influence how individuals process and respond to feedback	Emphasized the role of contextual external factors in feedback processing, supplementing the ICAP framework's focus on overt behaviors.	**Adaptable to the Individual**: Learners can make unlimited attempts and can adjust the amount of scaffolding they receive to meet their needs.
Microlearning Design Principles (Jahnke et al., 2020)	Mobile Microlearning follows a set of common design principles	The design of the Pebasco app aligns with all of these design principles.	**Multi-Device Accessibility**: Allows students to access the app and their progress across different devices, ensuring flexibility and ease of use. **Brevity of Activities**: Brief activities designed to manage cognitive load and enhance engagement.

3.4.1 No-code design of Pebasco ver. 2

The second version of Pebasco was built as a standalone asynchronous microlearning app using the no-code development platform Bubble (Bubble, 2022). No-code technology empowers people who are less proficient in computer programming to develop websites and interactive apps using visual programming languages and more intuitive human-readable syntax. No-code technology is rapidly gaining implementation in business and industry; the former CEO of the massive software development platform GitHub, Chris Wanstrath, has said that in the future, all coding will be done by some form of no-code (Johannessen & Davenport, 2021). This is an important element of this paper's novelty because a literature search reveals a lack of research on the use of no-code technology to develop educational apps, learning analytics apps, or apps for computer-assisted language learning (CALL).

Bubble is a multi-functional no-code platform. It can serve as an editing tool for creating the user interface and the underlying workflows of an app. It can also host the app on the web and maintain the dynamic database of all of the app's data. Figure 3.7 displays some of the collapsed workflows that underlie one of the pages in the Pebasco app; each of the rectangles can be clicked to expand and display the step-by-step breakdown of the actions that comprise that particular workflow.

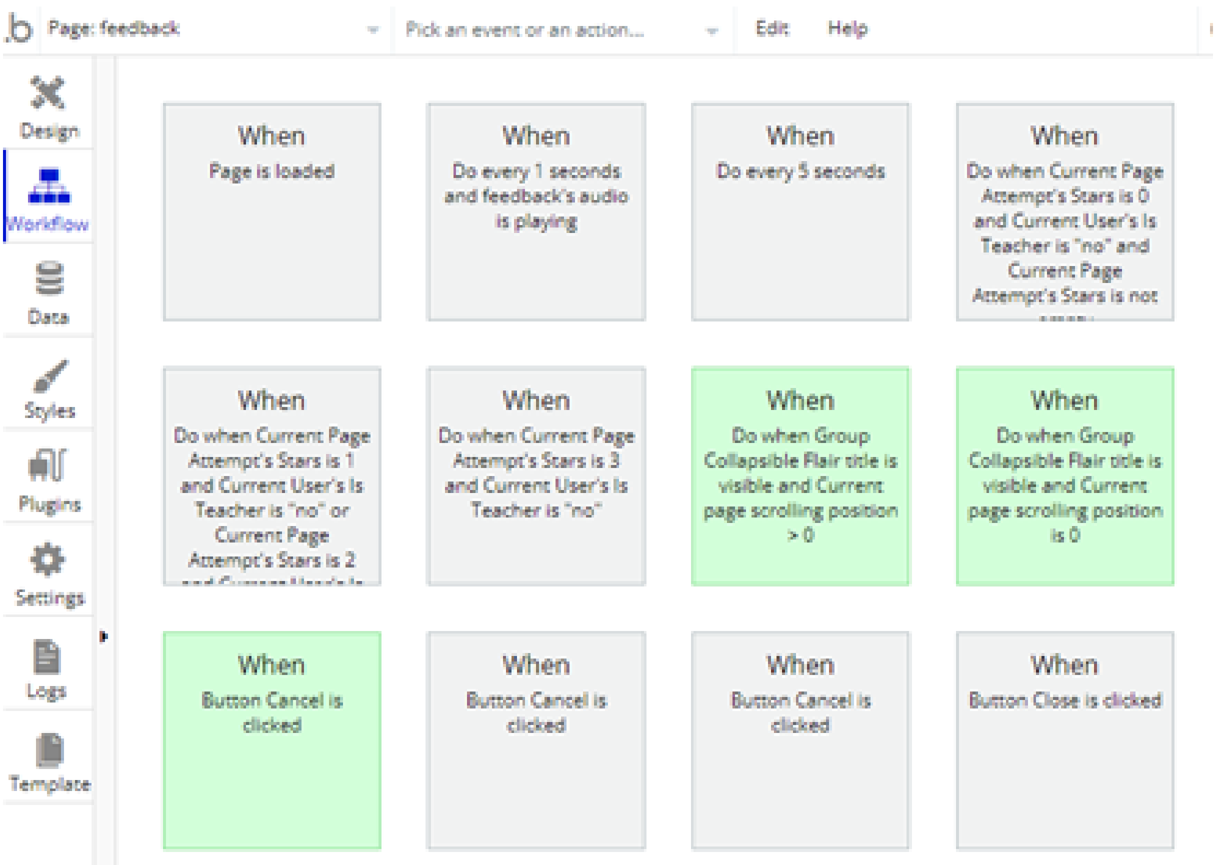


Figure 3.7: A screenshot of some Bubble workflows underlying a page in Pebasco

Just because Bubble does not require a user to know how to program in a particular computer coding language does not mean that it is necessarily easy to learn. Bubble offers more flexibility in the range of websites and apps that can be built than many other no-code tools do. However, that results in more complexity and a steeper learning curve. As none of the authors of Study 2 had used a no-code tool to build an app before, it was essential to reach out to the no-code community for support. This support included an official Bubble Bootcamp offered from the Bubble homepage (Bubble, 2022) and no-code design and educational groups such as ProNoCoders (2022) and Buildcamp (2022).

3.4.2 The Pebasco student interface

The purpose of Pebasco is to help students improve their peer feedback skills and internal feedback skills by comparing the feedback that they give to the expert feedback from a teacher. The overview of Pebasco is that students listen to audio clips that were recorded by students enrolled in the same course. For each audio, the teacher has made timestamped comments that indicate some mistake within the audio. Each audio has between one and seven teacher comments. The teacher comments either indicate a grammar/vocabulary mistake, a pronunciation mistake, or an unnecessarily long pause in the audio. The goal is for the student to listen to an audio and to try to predict the type and timestamp of each of the teacher's comments. If a student can successfully predict all the teacher's comments in an audio, they earn three stars for that audio. Students can make an unlimited number of attempts per audio. After a student logs in to Pebasco, they first see the unit list page (see Fig. 3.8). There are currently four units with 15 audios per unit. Students were assigned one unit per week over four weeks. The student then selects an audio by clicking the "New Attempt" button. This takes the student to the attempt page (see Fig. 3.9).

On the attempt page, students have a simple audio player that allows them to play, pause, and skip forward or backward by 5 seconds. If a student identifies a section of the audio where they think the teacher may have left a comment, they will pause the audio, which then displays the current timestamp. The student can then annotate the audio at that timestamp by clicking on the red "Feedback" button. Next, they select the type of teacher comment (e.g., a pronunciation mistake) from a dropdown list. Then they indicate how confident they feel that their feedback annotation matches with the teacher comment. The system then announces if the student annotation matches with a teacher comment near the same timestamp.

3.4.3 Scaffolding learning with hints

One of the challenges of designing asynchronous learning experiences is that teachers cannot directly give their students scaffolding support as they might in a traditional face-to-face class. For this reason, research on primarily asynchronous massive open online

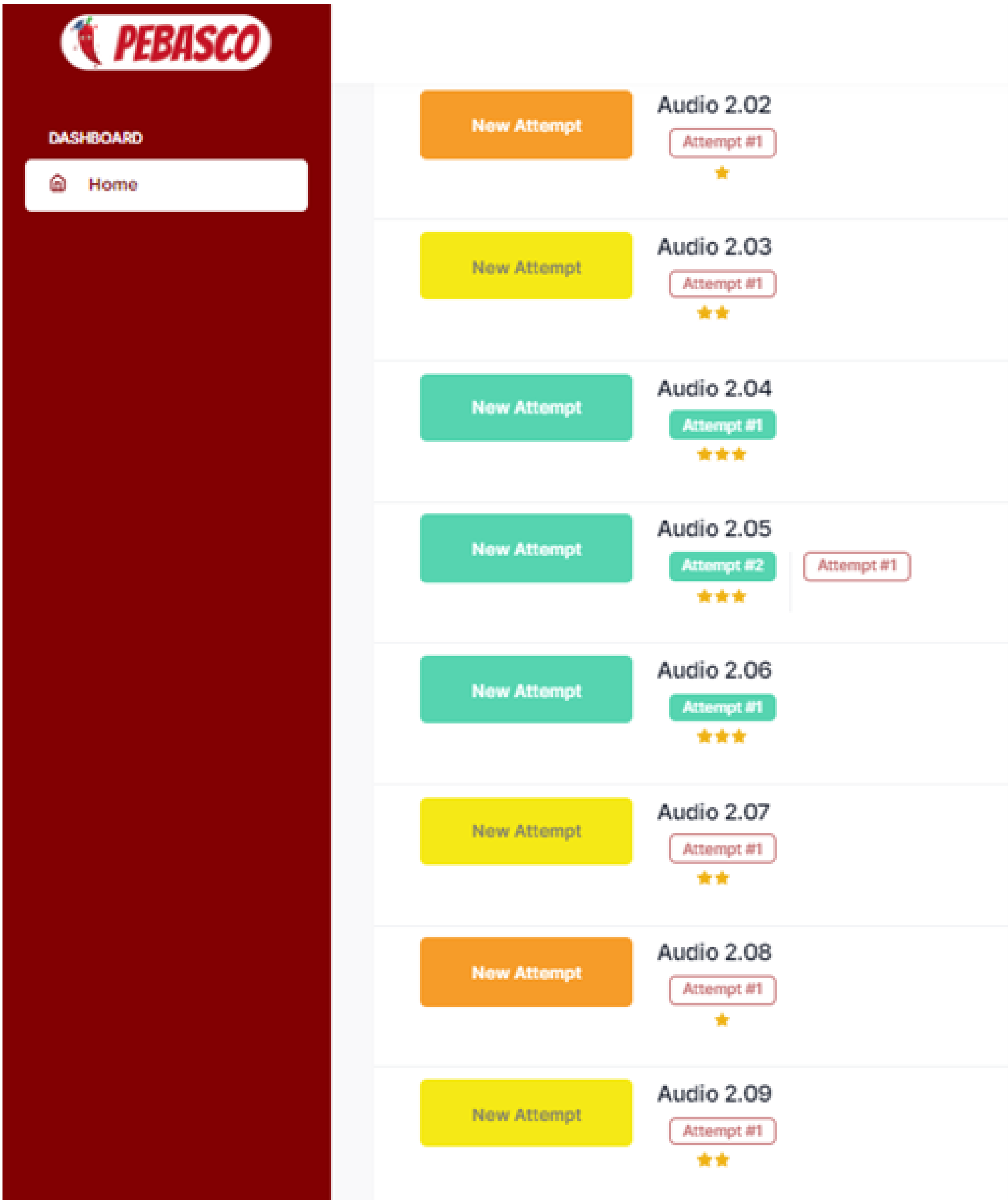


Figure 3.8: A screenshot of Pebasco's unit list page

courses (MOOCs) can guide possible approaches to take because much of it has investigated ways to provide students with feedback and support (Heffernan & Heffernan, 2014; Kasch et al., 2021; Meek et al., 2017). Heffernan and Heffernan (2014) describe

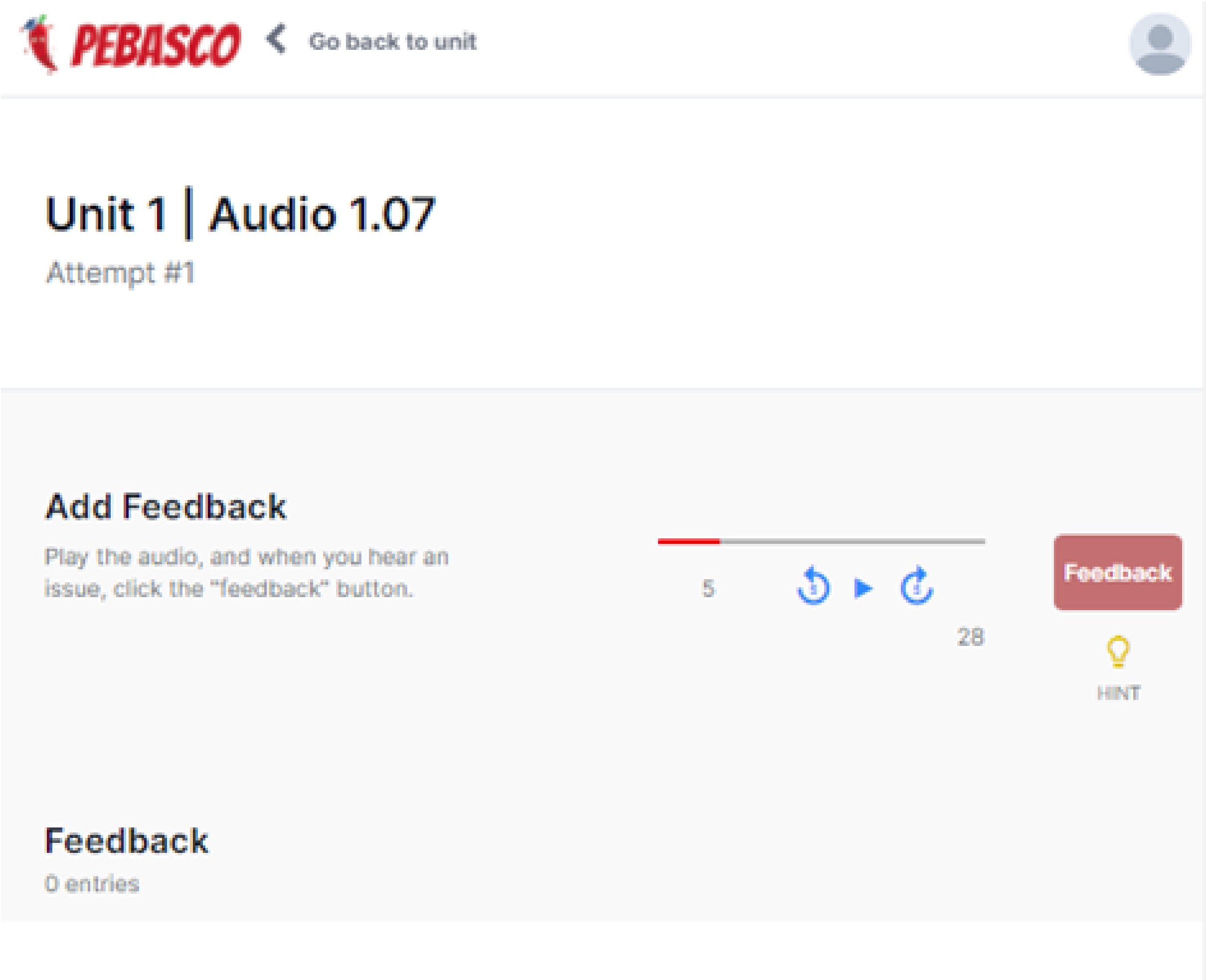


Figure 3.9: A Screenshot of Pebasco's attempt page

the ASSISTments platform, which allows researchers to investigate the effects of different types of hints and scaffolding in online educational contexts such as MOOCs. Zhou et al. (2021) used the ASSISTments platform in concert with the edX MOOC Big Data and Education. They found that adult learners with lower prior knowledge of the material performed worse when they were required to complete an entire sequence of scaffolding support, regardless of how relevant it was to them at the time. They suggested that adults who prefer to have more learner autonomy may prefer to select only the hints that they need. Another option provided to the participants in that study was that participants could request for the system to reveal the correct answers to items. This option was also described by Moreno-Marcos et al. (2019). In the mobile microlearning version of Pebasco app, the students can choose what level of support they receive from the system, including requesting the reveal of the correct answers for each item. This design choice aligns with the theoretical perspective of Cognitive Load Theory as described earlier.

3.4.4 Hints and stars

If a student successfully matches their peer feedback annotations to all of the teacher comments in an audio, the system makes a congratulatory announcement and awards them three stars for that audio. In the current version of Pebasco, users are not penalized for making "false positive" annotations (i.e., an annotation that does not match a teacher comment). This design decision was made to not discourage students from attempting annotations even if they did not have high confidence that they were correct.

The Pebasco app also has a hint engine to scaffold students in this activity. Below the red "Feedback button" on each audio's attempt page is a "Hint" icon (see Fig. 3.9). Students can click on that icon up to three times to access increasing levels of support from the system. For hint number one, the system tells the student how many teacher comments are in the current audio. For hint number two, the system provides the timestamp for each of the teacher comments and links that jump the audio player to a few seconds prior to the timestamps and begin playing the audio. The links can be clicked multiple times. The third hint is a request to reveal the answers, similar to the functionality used in prior studies (Moreno-Marcos et al., 2019; Zhou et al., 2021). When hint three is used, the following information is shared with the user: the number of teacher comments; the timestamps and links to each timestamp in the audio for each teacher comment; the type of each teacher comment; the corrected text of the mistake (in the case of a grammar/vocabulary mistake) or a link to audio examples of British or North American pronunciation of words (in the case of a pronunciation mistake); and a full transcript of the audio clip (see Fig. 3.10).

The number of hints used in an attempt determines the number of stars awarded when a student matches with all the teacher comments. If a student uses zero hints, they are awarded three stars. If they use one hint, they get two stars. If they use two hints, they get one star. Finally, if they use all three hints (i.e., request to reveal the answers), they get zero stars. Whenever a student makes a new attempt at an audio, the number of hints used for that attempt is reset to zero. The number of attempts made and stars earned is dynamically reflected on the unit list page. The "New Attempt" button changes color from red (zero stars), to orange (one star), to yellow (two stars), to green (three stars), reflecting the maximum number of stars earned across all attempts made on an audio.

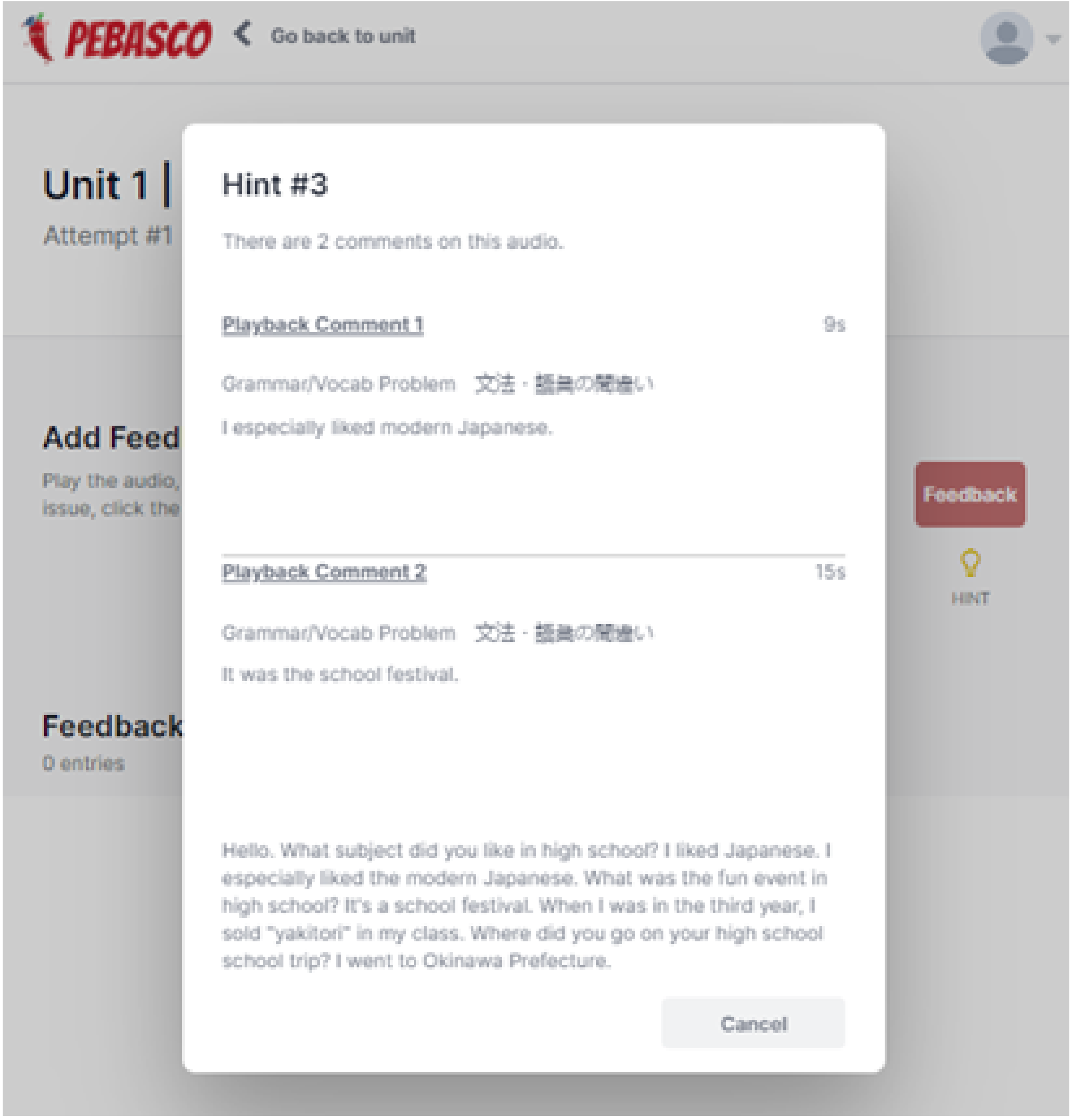


Figure 3.10: A screenshot of Pebasco's hint number three

For example, in Fig. 3.8, in the second audio of unit 2 (i.e., Audio 2.02), the student had made one attempt, using two hints while matching with all of the teacher comments, to earn one star. In the fifth audio of unit two (i.e., Audio 2.05), the student did not earn any stars on their first attempt (either because they did not match with all of the teacher comments or they used all three hints and requested the reveal of the correct answers), but they were able to earn three stars on their second attempt. Please note that the screenshot shown in Fig. 3.8 is early in the student's engagement with Pebasco's unit 2;

the student could still make unlimited future attempts to earn three stars on each audio.

3.5 Presenting the studies

The preceding chapter laid out the methods for developing the Pebasco system using an Educational Design Research approach. This flexible and pragmatic approach allowed us to adapt to the changing situation as the course was disrupted by a shift to ERT. In the next three chapters, the studies that comprise the core of this dissertation will be presented in detail. The first two studies have been published in peer-reviewed international academic journals. The third is currently under editorial review by a peer-reviewed international academic journal.

3.6 Ethical considerations

Rissho University's research ethics review board approved this dissertation's research methods, data collection, and data handling.

In the interest of transparency, the following declaration of the use of generative AI (artificial intelligence)/LLMs (large language models) is provided. During the preparation of this work, the author(s) used ChatGPT and Grammarly in order to improve the readability and language of some sections. After using these tools/services, the author(s) reviewed and edited the content as needed and take full responsibility for the content of the publication.

Chapter 4

Research study 1: "Learning analytics of peer feedback on communicative skills in an EFL course across different learning modalities"

4.1 Research background

The start of the COVID-19 pandemic necessitated a swift transition to Emergency Remote Teaching (ERT) (Hodges et al., 2020), posing unique challenges for educators, particularly in the realm of Communicative Language Teaching (CLT) (Gorham & Ogata, 2020b; Le Cor & Coutherut, 2020; Luporini, 2020; Nartiningrum & Nugroho, 2020). This study delves into these challenges within a mandatory English-as-a-Foreign-Language (EFL) course at a Japanese university, with a specific focus on the role of feedback and peer feedback in technology-enhanced learning.

Task-Based Learning and Teaching (TBLT), a popular subset of CLT, has been extensively researched in traditional, face-to-face contexts (Sauvignon, 2005). However, the sudden shift to ERT has underscored the previous calls for further research in technology-mediated TBLT (Chong & Reinders, 2020). This study responds to this call, investigating the use of Timed Paired Practice (TPP), a task-based learning activity known to improve students' performance in CLT contexts (Pipe, 2015), in an ERT setting.

Feedback, and peer feedback in particular, is a central theme of this study. Drawing on the comprehensive feedback frameworks of Panadero and Lipnevich (2022), Yang (2021), Lui and Andrade (2022), Carless and Boud (2018), and Sutton (2012), this study explores how feedback can be effectively implemented in an ERT context. Peer feedback has been shown to positively impact student learning and achievement (Hattie, 2009, 2012; Hattie & Clarke, 2019; Kerr, 2020), and this study investigates the potential of technology to enhance these effects.

The research involved 320 students across three cohorts, with the latter two experiencing the ERT modality. The final cohort additionally utilized a prototype learning analytics system to enhance peer feedback skills, responding to calls for more consideration of the use of technology and learning analytics in feedback processes (Carless & Boud, 2018; Lui & Andrade, 2022). This also echoes a recent call by Ouhaichi et al. (2023) for more research into the social aspects of learning by using multimodal learning analytics (MMLA).

RQ 1: How and to what extent can different activities that support the development of peer feedback skills improve the students' communicative skills in a mandatory basic EFL course based on CLT, as measured by communicative speaking tests?

By providing empirical evidence supporting modern feedback frameworks and exploring the effectiveness of peer feedback activities in an ERT context, this study makes a significant contribution to the literature on feedback in language learning.

4.2 Methods

4.2.1 Course goals

The context for this study is an introductory ("Kiso Eigo") CLT-focused EFL course that was mandatory for all first-year students in the Faculty of Letters at a university in Tokyo, Japan. The primary goal of the course is to improve students' basic English communicative skills.

4.2.2 Participants and study design

The participants in this study were first-year students enrolled in the course. Only students who completed both the first and last (fifth) TPP communicative speaking tests were included in this study. There are three cohorts in this study: 1) the face-to-face spring/fall 2019 "F2F" cohort (n=117); 2) the spring 2020 ERT online cohort (n=109); and 3) the fall 2020 ERT online + Pebasco cohort (n=94). The naming of the third cohort comes from the prototype learning analytics system that was used. Furthermore, the cohorts were independent samples; no students were in multiple cohorts.

The current design is a quasi-experimental study using convenience sampling. It loosely considers the Phase 1 cohort as a control group against which the feedback interventions in the second and third phases can be compared. Limitations with this experimental design will be raised in the discussion section.

4.2.3 Common flow of instructional design and adaptation to each phase

Regardless of the phase of this study that each cohort participated in, there is a common flow of instructional design that remained consistent throughout. First, the students would be informed of the topic of the following week's TPP test, and they would be given time to practice. Then, in the following week's class, the students would participate in the TPP test. In the TPP test, they would do multiple rounds of TPP testing with randomly assigned partners. Each student's average time and average score across all the rounds of TPP that they completed would be calculated and recorded. This pattern of practice in one week's class, followed by TPP testing in the following week's class, continued until students completed five sessions of TPP testing. Building off this common flow, each phase of this study featured different learning strategies and technology affordances to accommodate the change to ERT (see Table 4.1). **Phase 1: 2019 Spring/Fall Cohort**: Face-to-Face. Some of the activities assigned in the face-to-face classes included textbook work, pair work, and group work. One of the main activities and the primary focus of this study was Timed Pair Practice (TPP), which was used as a form of low-stakes formative CLT assessment. There were 5 TPP testing sessions; each session was based on a broad, general-interest topic related to a unit in the textbook. In each TPP session, students participated in multiple rounds.

The version of TPP that was used in this study had two success metrics: 1) an average score and 2) an average time. In the basic form of TPP, a randomly assigned pair of students begins a conversation, and they talk until one of them makes an uncorrected error, and the teacher stops them. If either student corrects the error before the teacher stops them, the pair can continue talking. However, once the conversation has been stopped, the teacher records the length time of the duration of the conversation. This duration is not simply a measure of time on task but rather an indicator of English fluency and accuracy. The second success metric is a score. The score is calculated based on a target length of time (e.g., 30 seconds) and the quality of the conversation. If the pair cannot

Table 4.1: Adaptation of pedagogic strategy and technology integration to emergency remote teaching conditions

Mode	Phase	Course elements: Learning strategy (practice for TPP)	Course elements: Technology affordances
F2F	Phase 1 2019 Spring/Fall (n=117)	Preparation-focused small-group practice (F2F)	No specific student-facing technology was used. Students themselves organized group practice sessions.
Online	Phase 2 2020 Spring (n=109)	Preparation-focused small-group practice (Online synchronous)	The video chat function of Microsoft Teams was used to host small-group practice breakout rooms for TPP practice.
Online	Phase 3 2020 Fall (n=94)	Reflection-focused self-practice with Pebasco (Online asynchronous) + Preparation-focused small-group practice (Online synchronous)	The video chat function of Microsoft Teams was used to host small-group practice breakout rooms. During the weeks between TPP assessments, students used the Pebasco system to practice their peer feedback skills.

reach the target time, they get 1 point. If the pair reaches or exceeds the target length of time, they get 2 points. If the pair reaches or exceeds the target length of time and has a high-quality conversation (i.e., demonstrates good logical flow, incorporates follow-up questions, provides supporting examples, uses unique questions/responses/phrasing, etc.), they get 3 points. Each student's average score and average time are calculated across all the rounds of TPP that they completed during that session of TPP testing, which reflect their basic English communicative skills at that point in time.

Here is a typical example of a conversation in the first TPP test (topic: campus life) that lasted 9 seconds and earned 1 point. Note that the teacher stops the students because neither of them corrected the error:

Student A: *What was your favorite subject in high school?*

Student B: *My like subject was science.*

Teacher: *Stop. My favorite subject was science. Or you could say, "I liked science."*

9 seconds. One point.

Here is a typical example of a conversation in the last TPP test (topic: Japan) that lasted 50 seconds and earned 3 points. Note that at one point in the conversation, Student C notices and corrects their mistake, indicating internal feedback skills:

Student C: *Where do you think is the most beautiful place in Japan?*

Student D: *That's a good question. I think Okinawa is. We went there on our school trip in high school. The color of the water there is such a pretty color of blue. I think Okinawa has the best beaches in Japan. Have you ever visited there?*

Student C: *No, I haven't. But I really want to go someday. We went to Tokyo Disneyland for our high school trip. By the way, did you bought. . .Oh, sorry. . .Did you buy any souvenirs?*

Student D: *Yes, I did. Okinawa is famous for salt. I bought some salt for my mother. The shop can buy many different kinds of salt.*

Teacher: *Stop. The shop sells many different kinds of salt. 50 seconds and three points.*

Phase 2: 2020 Spring Term: Online. Due to COVID-19, the start of the school year was delayed until mid-May 2020, and all classes were moved online in a sudden shift to ERT. For most first-year students, it was their first experience taking online classes. Similarly, it was the first time the course instructor had used Microsoft Teams, Microsoft Forms, and Microsoft Flipgrid—the main technologies used.

Like in the F2F context, there were five sessions of TPP testing, held every other week and interspersed with a session for practice and preparation (unless the university schedule required a change in the weekly pacing). However, students did TPP via the video conferencing function of Microsoft Teams rather than at the front of a classroom.

Preparing for the Following Term. Near the end of the second phase of this study, the students were asked to supplement their last two practice/preparation Flipgrid video activities with peer feedback surveys using Microsoft Forms. The purpose was to test one of the elements that would be used to make a prototype learning analytics system for improving peer feedback skills on spoken content that would be used in the third phase of this study. The creation and use of this system are described in more detail in the second chapter.

Phase 3: 2020 Fall Term: Pebasco. In the 2020 fall term, the syllabus, schedule, and activities were nearly identical to the 2020 spring term. Beyond that, a prototype learning analytics system, Pebasco, was introduced to improve students' ability to provide peer feedback on spoken content (see Chapter 2 for a more detailed description of the system). The system allowed students to watch asynchronous recordings of classmates' spoken performances and annotate them with feedback, similar to how the instructor provides feedback in TPP testing (Gorham & Ogata, 2020a). Unlike TPP, where only one mistake is identified, students using Pebasco could mark multiple errors in a single recording. They could then compare their feedback with the instructor's comments, promoting reflective learning and feedback skill development.

Pebasco was built by integrating various available technologies to create a seamless system. Microsoft Teams served as the LMS for handling communications and assignments, while Microsoft Flipgrid was used for video recording and sharing. Students provided timestamped feedback using Microsoft Forms, and the data were visualized through Google Data Studio, which generated dashboards that allowed students to compare peer and instructor feedback on their recordings. Collaborative reflection activities followed, where students worked in small groups to discuss and reflect on their feedback before upcoming TPP tests.

4.2.4 Data collection

The students' TPP test performance data were collected during normal class activities as a means to evaluate the research question. The university's research ethics board approved this study's methods and data collection. The university's research ethics board determined that informed consent was not required for the quantitative data collection used to address the research question because the data is anonymized and aggregated.

4.2.5 Statistical methods

Two sets of quantitative data were collected for all three cohorts: (1) the students' scores after the TPP tests and (2) the students' time on the TPP test (i.e., duration of time a pair could speak without making an uncorrected error as a measure of English fluency and accuracy). JASP software (Love et al., 2019) was used to analyze the data.

The raw data was cleaned and checked for outliers. Descriptive statistics were calculated. The dependent variables were checked for normal distribution and homogeneity of variance before analysis. Pair samples t-tests were used to describe the difference in average times from TPP1 to TPP5 for each cohort. Wilcoxon signed-rank t-test was used to describe the difference in average scores from TPP1 to TPP5 for each cohort. Finally, analysis of variance (ANOVA) and post hoc tests (Tukey) were used to compare the change in average time and the change in average score among the three cohorts.

4.3 Results

This study's research question is addressed by analyses of the two success metrics of the TPP tests. The following two sections will report the results from the analyses of the students' time on the TPP tests and their scores on the TPP tests, respectively. Each section will begin by sharing the relevant descriptive statistics. Then, the results of t-tests will demonstrate the intra-group changes of the cohorts from TPP1 to TPP5. Finally, the results of analysis of variance (ANOVA) and post hoc tests (Tukey) will be used to compare the changes among the three cohorts.

4.3.1 Average time gain

Descriptive statistics for all cohorts' average times from the TPP1 (baseline) and the TPP5 (final session) are displayed in Table 4.2. Paired sample t-tests for each cohort's average time comparing TPP1 to TPP5 are shown in Table 4.3. This demonstrates intra-group improvement on average time that reached a significant level for each cohort: the F2F cohort with $t=17.16$ ($p<.001$) with $d=1.59$; showing a large effect size; the Online cohort with $t=20.06$ ($p<.001$) with $d=1.92$ showing a large effect size; and the Pebasco Cohort with $t=31.1$ ($p<.001$) with $d=3.21$ showing a large effect size (Cohen, 1988). The first metric that was examined was the average time gain (in seconds) from TPP1 to TPP5: F2F ($M=25.53$, $SD=16.1$, $n=117$), Online ($M= 39.19$, $SD=20.4$, $n=109$), and Pebasco ($M=79.24$, $SD=24.67$, $n=94$). Before comparing the average time gain data for the three groups, the average times for the first TPP test (TPP1) were analyzed to determine if the three groups were starting from approximately the same point. It was found that the average time data for F2F ($M=25.34$, $SD=6.47$) vs Online ($M=26.73$, $SD=11.1$) vs Pebasco ($M=25.63$, $SD=7.75$) were not significantly different as indicated by Analysis of Variance (ANOVA), $F(2, 196.45)=0.65$, $p= .525$. Appropriate Welch homogeneity correction was conducted as Levene's test found that equality of variances was not satisfied. Once it indicated no significant differences among the three

Table 4.2: Descriptive statistics average time TPP1 and TPP5

	TPP1 AVE Time			TPP5 AVE Time		
	F2F	Online	Pebasco	F2F	Online	Pebasco
Count of learners	117	109	94	117	109	94
Mean	25.34	26.73	25.63	50.87	65.92	104.84
Std. Deviation	6.47	11.1	7.75	16.26	18.93	25.09
Skewness	0.484	0.583	0.244	0.417	0.685	0.784
Kurtosis	-0.069	-0.608	-0.434	-0.008	0.023	0.549
Minimum	11.2	8.5	11	19.6	34	64
Maximum	44.3	52	46.5	98.7	116.5	179

groups when comparing their starting times in TPP1, subsequent average time gain from TPP1 to TPP5 for the three cohorts was measured.

The average time gain data for F2F vs Online vs Pebasco was then compared using ANOVA (see Table 4.4), $F(2/193.74)=165.46$, $p= <.001$, with $\eta^2=0.546$, showing a large effect size. Levene's test for equality of variances was conducted and found to be statistically significant, therefore, the Welch homogeneity correction was used. Tukey's HSD Test for multiple comparisons found that the mean value of Average Time Gained was significantly different between F2F and Online ($p = < .001$); between F2F and Pebasco ($p = < .001$); and between Online and Pebasco ($p = < .001$).

4.3.2 Average score gain

Descriptive statistics for all cohorts' average scores from TPP1 and TPP5 are displayed in Table 4.5. Student's Paired sample t-tests for each cohort's average score comparing TPP1 to TPP5 are shown in Table 4.6. Wilcoxon signed-rank test was used to adjust for non-normality. This demonstrates intra-group improvement on average score that reached a significant level for each cohort: the F2F cohort with $z=8.55$ ($p<.001$) with $r_B=0.93$; showing a large effect size; the Online cohort with $z= 8.14$ ($p<.001$) with $r_B=0.96$ showing a large effect size; and the Pebasco cohort with $z=8.42$ ($p<.001$) with $r_B=1$ showing a

Table 4.3: Intra-group average time paired samples T-Tests

	t	df	p	Cohen's d
F2F Cohort TPP5 - TPP1 AVE Time (N=117)	17.16	116	$< .001$	1.59
Online Cohort TPP5 - TPP1 AVE Time (N=109)	20.06	108	$< .001$	1.92
Pebasco Cohort TPP5 - TPP1 AVE Time (N=94)	31.1	93	$< .001$	3.21

Note: Student's t-test.

Table 4.4: ANOVA of average time gained

Group	N	Mean	Std. Dev	F (2,193.74)	p	η^2
F2F	117	25.53	16.1	165.46	$<.001$	0.546
Online	109	39.19	20.4			
Pebasco	94	79.24	24.67			

Note. $p < 0.05$*

large effect size (Cohen, 1988).

The second metric that was examined was the average score gain from TPP1 (baseline) to TPP5 (final session) for the three cohorts: F2F (M=0.58, SD=0.49, n=117), Online (M=0.79, SD=0.57, n=109), and Pebasco (M=1.48, SD=0.34, n=94). Before comparing the average score gain data for the three groups, the average scores for the first TPP test (TPP1) were analyzed to determine if the three groups were starting from approximately the same point. It was found that the average score data for F2F (M=1.33, SD=0.22) vs Online (M=1.39, SD=0.39) vs Pebasco (M=1.28, SD=0.28) were not significantly different as indicated by Analysis of Variance (ANOVA), $F(2, 193.85)=2.77$, $p= .065$). Appropriate Welch homogeneity correction was conducted as Levene's test found that equality of variances was not satisfied. Once it indicated no significant differences among the three groups when comparing their starting scores in TPP1, subsequent average score

Table 4.5: Descriptive statistics average score TPP1 and TPP5

	TPP1 AVE score			TPP5 AVE score		
	F2F	Online	Pebasco	F2F	Online	Pebasco
Count of Learners	117	109	94	117	109	94
Mean	1.33	1.39	1.28	1.9	2.19	2.78
Std. Deviation	0.22	0.39	0.28	0.5	0.43	0.28
Skewness	0.138	0.392	0.799	0.59	-0.098	-1.17
Kurtosis	-0.752	-1.261	-0.105	-0.508	-0.287	0.548
Minimum	1	1	1	1	1	2
Maximum	1.8	2	2	3	3	3

gain from TPP1 to TPP5 for the three cohorts was measured.

Next, the average score gain data for F2F vs Online vs Pebasco were then compared using ANOVA (see Table 4.7), $F(2, 207.66)=141.53$, $p= <.001$, with $\eta^2=0.38$, showing a large effect size (Cohen, 1988). Levene's test for equality of variances was conducted and found to be statistically significant; therefore, the Welch homogeneity correction was used. Tukey's HSD Test for multiple comparisons found that the mean value of Average Score Gained was significantly different between F2F and Online ($p = 0.003$); between F2F and Pebasco ($p = < .001$); and between Online and Pebasco ($p = < .001$).

4.4 Discussion

This study, conducted within a mandatory basic English-as-a-Foreign-Language (EFL) course at a Tokyo university, aimed to determine the extent to which peer feedback activities could enhance students' communicative skills. As a result of a shift to online emergency remote teaching (ERT) caused by the COVID-19 pandemic, this study was able to trace the effects of different instantiations of feedback activities across three cohorts of students taking the same course under different conditions. Part of the novelty of this study is that this unique situation provided an opportunity to address previous calls for further research in technology-mediated Task-Based Learning and Teaching (TBLT)

Table 4.6: Intra-group average score paired samples T-Tests

	W	z	p	Rank-Biserial Correlation
F2F Cohort TPP5 -TPP1 AVE Score (N=117)	236.5	8.55	$< .001$	0.93
Online Cohort TPP5 -TPP1 AVE Score (N=109)	87.5	8.14	$< .001$	0.96
Pebasco Cohort TPP5 -TPP1 AVE Score (N=94)	0	8.42	$< .001$	1

Note. Wilcoxon signed-rank test.

Table 4.7: ANOVA average score gained

Group	N	Mean	Std. Dev	F (2, 207.66)	p	η^2
F2F	117	0.58	0.49	141.53	$<.001$	0.38
Online	109	0.79	0.57			
Pebasco	94	1.48	0.34			

$p<0.05$*

(Chong & Reinders, 2020), specifically through the use of Timed Paired Practice (TPP), a task-based learning activity known to improve students' performance in Communicative Language Teaching (CLT) contexts (Pipe, 2015).

In Phase 1, the feedback mechanism was traditional and teacher-centric, conducted in a face-to-face classroom setting. The focus was on immediate verbal feedback from the teacher following each round of the TPP formative assessments, primarily addressing language errors in grammar, vocabulary, and pronunciation. The performance of this cohort, as measured by the TPP tests' average time change and average score change, served as a baseline and control for comparison with subsequent phases.

Phase 2 integrated online elements into the feedback process, as a result of the shift to Emergency Remote Teaching (ERT). Despite retaining the teacher-led feedback approach, this phase saw an improvement in student performance compared to Phase 1, as evidenced by the statistically significant findings in both TPP test metrics. This suggests that even modest changes in the feedback process and implementation can positively impact student learning outcomes.

Phase 3 introduced a multifaceted feedback approach with an emphasis on peer feedback which significantly outperformed both Phase 1 and Phase 2 cohorts in the TPP success metrics. This phase was characterized by a tightly coupled loop of feedback activities which included (a) teacher feedback during TPP tests, focusing on immediate and practical language usage errors; (b) meta-dialogues about the rationale and benefits of peer feedback and its impact on internal feedback skills; (c) active participation in the Pebasco system, where students provided and received peer feedback on Flipgrid videos; (d) comparison of peer feedback with teacher feedback through the Pebasco student-facing dashboard, fostering self-reflection; and (e) small group reflection sessions based on the feedback received, using the student-facing dashboard to facilitate discussions and insights. This comprehensive approach in Phase 3 not only enhanced the students' engagement with feedback but also promoted a deeper understanding of feedback processes and their relevance to language learning. To better understand how feedback activities impacted the learners, it is important to view them through the context of the theoretical feedback frameworks that were discussed earlier.

4.4.1 Contextualizing results within theoretical feedback frameworks

This study's phased approach allows for a nuanced exploration of the MISCA model's application (Panadero & Lipnevich, 2022), particularly how each phase's distinct feedback mechanism aligns with the model's components. In Phase 1, the traditional face-to-face feedback predominantly focused on the 'Message' aspect of the MISCA model, where the teacher's feedback was immediate and error-focused. This phase established a baseline for feedback effectiveness, primarily addressing the content and clarity of feedback, but with limited exploration of student individuality and context. The transition to the online modality in Phase 2 marked a change in the 'Implementation' aspect of the MISCA model. This shift demonstrated that modifying the feedback delivery method (while retaining its core message) could lead to improved student performance. Phase 3's tightly-coupled feedback loop significantly outperformed the previous phases, as it closely aligned with multiple elements of the MISCA model. The integration of teacher feedback with peer feedback in the Pebasco system enhanced both the message's relevance and the variety of the implementation. The teacher's meta-dialogues explaining the purpose and benefit of peer feedback activities emphasized the importance of students developing internal feedback skills, thereby addressing the 'Student' component in the MISCA model. This was furthered by actively involving students in the feedback process, thereby acknowledging and fostering their individual feedback skills. The shift to ERT provided a new 'Context,' demonstrating the applicability of the MISCA model across diverse settings. The addition of peer feedback alongside teacher feedback diversified the 'Agents,' illustrating how different sources of feedback can collectively enhance the learning experience.

While the MISCA model is a broad framework for consideration, Yang's (2021) Feedback Orientation framework provides a lens to examine how the phased approach in this study influenced students' engagement with and perception of feedback. In Phase 1, feedback was direct and teacher-centric, primarily focusing on language accuracy. This phase aligns with Yang's 'Feedback Utility' dimension, as students received clear and immediate feedback on their language use. However, the approach was limited in fostering 'Feedback Self-Efficacy' as students primarily received feedback without engaging in self-reflection or peer interaction. With the introduction of online tools in Phase 2, the 'Feedback Utility' aspect was maintained, and the feedback delivery's digital nature offered a new dynamic to students' feedback experience. The improved performance metrics from this phase suggest that students may have found value in this modified feedback approach, possibly indicating a positive shift in their perception of feedback's utility in learning.

Phase 3 offered a more holistic application of Yang's framework; its integration of peer feedback through the Pebasco system, alongside teacher feedback, directly engaged students in the feedback process. This engagement may have enhanced their belief in the feedback's utility and fostered feedback self-efficacy, as students were not only recipients but also providers of feedback. The teacher's meta-dialogues and group reflection sessions in Phase 3 promoted social awareness around feedback. These discussions may have improved students' understanding of the social dynamics involved in giving and receiving feedback, fostering a more collaborative learning environment. The tightly coupled feedback loop in Phase 3, involving teacher feedback, peer feedback, and reflective practices, nurtured a sense of accountability among students. They were responsible not only for internalizing feedback but also for contributing constructively to their peers' learning, aligning with Yang's notion of feedback accountability. The significant improvements in the TPP test metrics in Phase 3 corroborate the effectiveness of aligning feedback practices with Yang's Feedback Orientation dimensions. The study's phased approach demonstrates that a comprehensive and active engagement with feedback – encompassing utility, self-efficacy, social awareness, and accountability – can support improved language learning outcomes.

Beyond Yang's (2021) Feedback Orientation framework, the findings of this study also support other theoretical feedback frameworks. For example, Carless & Boud's (2018) feedback literacy model identifies several implications that are influential in the Phase 3 feedback activities. For example, they suggest the importance of meta-dialogues which illuminate feedback processes, peer feedback as a core curricular element, and the use of reflective group discussions to make sense of feedback activities. All of these were clearly enacted in Phase 3 of this study, and the findings of this study suggest that they may have played a part in the improved performance in the third phase compared to earlier phases.

This also echoes Sutton's (2012) three dimensions of feedback literacy. The meta-dialogues in Phase 3 bolster the epistemological dimension by helping the students to have a deeper understanding of feedback processes. Then the ontological dimension reflects the students' growing sense of self-confidence, which may be linked to the practical dimension of Sutton's model, as indicated by the students' improved TPP performances.

Taken together, the findings of this study add evidence suggesting how the use of these interconnected theoretical feedback models may be used to improve the learning process. This adds evidence to the literature, demonstrating how these feedback models can be successfully used in CLT classes during ERT.

4.4.2 Implications for instructors

There are two sets of implications for instructors that are raised by this study. The first is that the Educational Design Research approach of bricolage (Bakker, 2018, p. 60) was used to create the prototype learning analytics system by orchestrating multiple available discrete technologies. The findings of this study suggest that such a prototype system can have a positive influence on learning outcomes in the context of peer feedback in a CLT EFL course during ERT. This is potentially important because Carless & Boud (2018) assert that "the possibilities of learning analytics to inform feedback research and the related implications for student feedback literacy are also likely to be an expanding area of research focus" (p. 9). Not only can a prototype learning analytics system yield empirical data for peer feedback research, but it can also provide guidance for the development of future learning analytics systems, such as the updated version of Pebasco, which was created as a standalone app based on the lessons learned in the current study (see Gorham et al., 2023). Instructors might find that the lower barrier to entry of this bricolage approach allows them to begin to conduct research at the intersection of learning analytics and frameworks of feedback and feedback literacy.

The second set of implications is that this study adds to the evidence that overarching frameworks, such as the MISCA model (Panadero & Lipnevich, 2022), are correct in emphasizing that to successfully implement peer feedback training, an instructor needs to approach the process from a holistic point of view which incorporates elements such as the learning environment, the sources of feedback and also the internal states and prior experiences of students (Lui & Andrade, 2022; Yang, 2021). If this can work in a challenging environment such as the ERT context that was described in the present study, then it may also be effective in a more traditional classroom environment. Instructors may want to start utilizing these feedback frameworks not only because they can be effective in their regular classrooms but also because they can positively guide curriculum design in the case of an ERT disruption caused by an incident such as a natural disaster or pandemic.

4.4.3 Limitations

There are a number of limitations to this study. As mentioned in the Methods section, this study loosely treated the Phase 1 cohort as a control group. However, Pawlak (2022) raises the issue of individual differences, particularly in the broader context of Complex Dynamic Systems Theory (CDST), and the idea that there may be some confounding variables that are not controlled for in this experimental design. For example, students may have different inherent motivations to learn. Alternatively, in this study, almost all the students in the spring 2020 term were experiencing their first time in an online synchronous class. The students in the fall 2020 term already had experience with several asynchronous classes during the spring term. This difference in prior experience could have impacted the outcomes of this study. Hiver and Al-Hoorie (2019) recommend design-based research (DBR) as one methodological approach to building and testing theories while acknowledging the complex dynamic nature of the systems that are involved in educational research.

Another limitation is that the Pebasco prototype system lacked some of the most important functionalities of other learning analytics systems, such as user learning logs and user profiling (for example, see Majumdar et al., 2021; Ogata et al., 2018). With better data collection and analysis tools, the Pebasco system could potentially have a more positive effect on learning outcomes.

Finally, the design of the prototype Pebasco system followed an Educational Design Research approach. It was being designed while the course was being taught, and the design was adapted to the changing conditions (most notably the shift to ERT). For this reason, the Microsoft Forms data collection part of the prototype Pebasco system was first trialed with students at the end of the second phase. Although they did fill out the Microsoft forms, they did not receive any feedback from their classmates, they did not see any of the instructor feedback, they did not see the student-facing dashboard, and they did not participate in any small-group reflective activities using the Pebasco system. Nevertheless, this small interaction with a subsection of the Pebasco system could have potentially impacted the results of this study.

4.5 Conclusion

The novelty of this study is that it is the first study of its kind to explore the use of a prototype learning analytics system that promotes peer feedback skills on spoken content during a shift ERT within the context of modern feedback frameworks. Future possibilities for research and development include the creation of a standalone Pebasco application that has more functionalities of learning analytics systems (e.g., the collection of learning logs). An example of this can be found in Gorham et al., (2023). In a recent systematic literature review of the use of films in EFL teaching, (Sánchez-Auñón et al., 2023), the authors suggested that future research could be expanded to include different types of video content, it is possible that the present study could be used as part of a systematic literature review study that examines the use of student-created videos in the context of peer feedback in EFL courses. Another avenue for future research could be to better identify individual differences (Pawlak, 2022) among the students in areas such as motivation to reduce potentially confounding variables. Finally, future studies could replicate this study in different academic institutions and cultural contexts.

Chapter 5

Research study 2: “Analyzing learner profiles in a microlearning app for training language learning peer feedback skills”

5.1 Research background

Peer feedback can be described as the act of one learner evaluating the performance of another learner. It has been shown to positively impact student learning and achievement (Hattie, 2009, 2012; Hattie & Clarke, 2019; Kerr, 2020). It also has been found to improve students’ self-reflection/internal feedback skills (To & Panadero, 2019). In the context of language learning, the practice of peer feedback (both provision and reception) can improve learners’ second/foreign language (L2) skills (El Mortaji, 2022; Patri, 2002; Rodríguez-González & Castañeda, 2016); and peer feedback practice can improve learners’ ability to detect L2 errors (Fujii et al., 2016). Peer feedback is a skill that can be trained (Sluijsmans et al., 2004; van Zundert et al., 2010).

There have been calls for research examining how training can enhance the quality of peer feedback (Dao & Iwashita, 2021) and for more use of experimental and quasi-experimental methodologies in peer feedback research (Topping, 2013). In the Dual Model Theory of peer corrective feedback (PCF), it is theorized that there are educational benefits to both the provision and reception of PCF, and there is a call for research to “experimentally tease apart the two sides” (Sato, 2017, p. 30). The current study addresses this call for research by isolating just the provisioning side of peer feedback.

Mobile microlearning is a type of technology-enhanced learning that is notable for its short duration and flexibility in the time and place of learning (Hug et al., 2006). Although there are several studies that research novel uses of microlearning to improve L2 vocabulary skills (e.g., Arakawa et al., 2022; Dingler et al., 2017; Inie & Lungu, 2021; Kovacs, 2015), there is a gap in the literature of research that investigates the use of asynchronous mobile microlearning peer feedback training in the context of spoken content in foreign language learning. The current study aims to fill this gap.

The opportunity to address these gaps in the literature arose due to an unexpected change in the teaching context. In response to the ongoing COVID-19 pandemic, the city of Tokyo was placed under a declared state of emergency multiple times in 2021 (The Japan Times, 2021). As a result, the university where the current study was conducted decided that some classes would be held in person while others would be taught using an online asynchronous modality. The present study examines the continued development and use of the Pebasco app to support the adaptation of a mandatory Communicative Language Teaching (CLT)-based English as a foreign language (EFL) class to this asynchronous modality.

The first version of Pebasco was a prototype learning analytics system used in synchronous online classes held in a Japanese university in the first year of the COVID-19 pandemic (Gorham & Ogata, 2020a). It was designed to improve students' peer feedback skills on spoken content in a CLT-based EFL class context. Following an iterative design process, the second version of Pebasco was built as a standalone asynchronous microlearning app using the no-code development platform Bubble (Bubble, 2022).

The second version of Pebasco allows for the collection of learning analytics data. Based on the system affordances and learner interactions with the system, this paper first defines the learner behavior profiles of the participants (n=87) for this language microlearning task. Using these learner profiles, this paper attempts to answer the following research question by tracking the migration of learner profiles over the course of using the Pebasco system:

RQ 1: *How and to what extent can the asynchronous microlearning app Pebasco improve students' skills for providing peer feedback on spoken content in CLT-based English as a foreign language (EFL) education?*

5.2 Methods

5.2.1 Research design

This study used convenience sampling and a quasi-experimental single-group pre-/post-research design. Participants used the Pebasco system over the course of four units. The first unit was treated as a tutorial for the students to get familiarized with the system's functionality. The participants' usage patterns with the system and their end-of-activity performance in the second (pre) and fourth (post) units were compared. The usage patterns and performance data were used to create profiles. Based on work by Majumdar and Iyer (2016), each profile is considered a stratum, and the transition pattern is measured across the two phases (i.e., unit 2 and unit 4). The migration in profile types from the second to fourth units was described and further compared by appropriate statistics to evaluate if there was an improvement in the group of participants' peer feedback skills over the course of using the Pebasco system.

5.2.2 Course goals

The context for this study is an asynchronous introductory CLT-focused EFL course that was mandatory for all first-year students in the Faculty of Letters at a university in Tokyo, Japan. The primary goal of the course is to improve students' basic English communicative skills. One of the sub-goals of the course is to improve the students' peer feedback and internal feedback skills for spoken content.

5.2.3 Participants

The participants in this study were first-year university students enrolled in the course described in the previous section. They were selected using convenience sampling. A total of 131 students were enrolled in the course. However, only students who provided informed consent and who had at least tried the second and last (fourth) units in Pebasco were included in this study (n=87).

5.3 Data collected and analysis method

The Pebasco app collects learning analytics data from user interactions with the system. Table 5.1 shows the number of attempts that all the students made in each unit of Pebasco.

For each record of a student's attempt, the system collected additional data, including the number of stars earned on the attempt, the number of hints used, the number of false positives (annotations that did not match with a teacher comment), the number of seconds that the audio was played during the attempt, the level of confidence a student reported for each annotation, and other clickstream data created when navigating the audio player. The highest level of analysis provided by the Pebasco system is at the

Table 5.1: Pebasco attempt data

Analysis flow	Data
Total number of students	131
Subset of students who gave informed consent and completed unit 2 and unit 4	87
Total audio attempts (units 1-4)	9,979 from 131 users
Audio attempts (unit 1)	3,327 from 131 users
Audio attempts (unit 2)	2,321 from 131 users
Audio attempts (unit 3)	2,180 from 131 users
Audio attempts (unit 4)	2,151 from 131 users

student outcome level. It shows how many stars an individual earned for each unit. Each unit has 15 audios with an opportunity to earn three stars per audio, resulting in the potential to earn a maximum of 45 stars per unit. The students are told when they are assigned a Pebasco unit that their score for the activity is based on the number of stars they earn in the unit, regardless of the number of attempts they make. The score breakdown is 10 points (45 stars), 9 points (40-44 stars), 8 points (35-39 stars), 7 points (30-34 stars), 6 points (25-29 stars), 5 points (20-24 stars), 4 points (15-19 stars), 3 points (10-14 stars), 2 points (5-9 stars), 1 point (1-4 stars), and 0 points (0 stars).

For the present study, unit 1 of Pebaso was used for onboarding and training purposes. The students' outcomes and behaviors in unit 2 and unit 4 were compared for this study. Figure 5.1 shows the distribution of points (which reflect the sum of the maximum number of stars students earned for each audio in a unit) earned by students in unit 2 compared to unit 4. Students who earned seven points or more (i.e., 30 stars or more) were placed in the "High Stars" category, while the remaining students were placed in the "Low Stars"

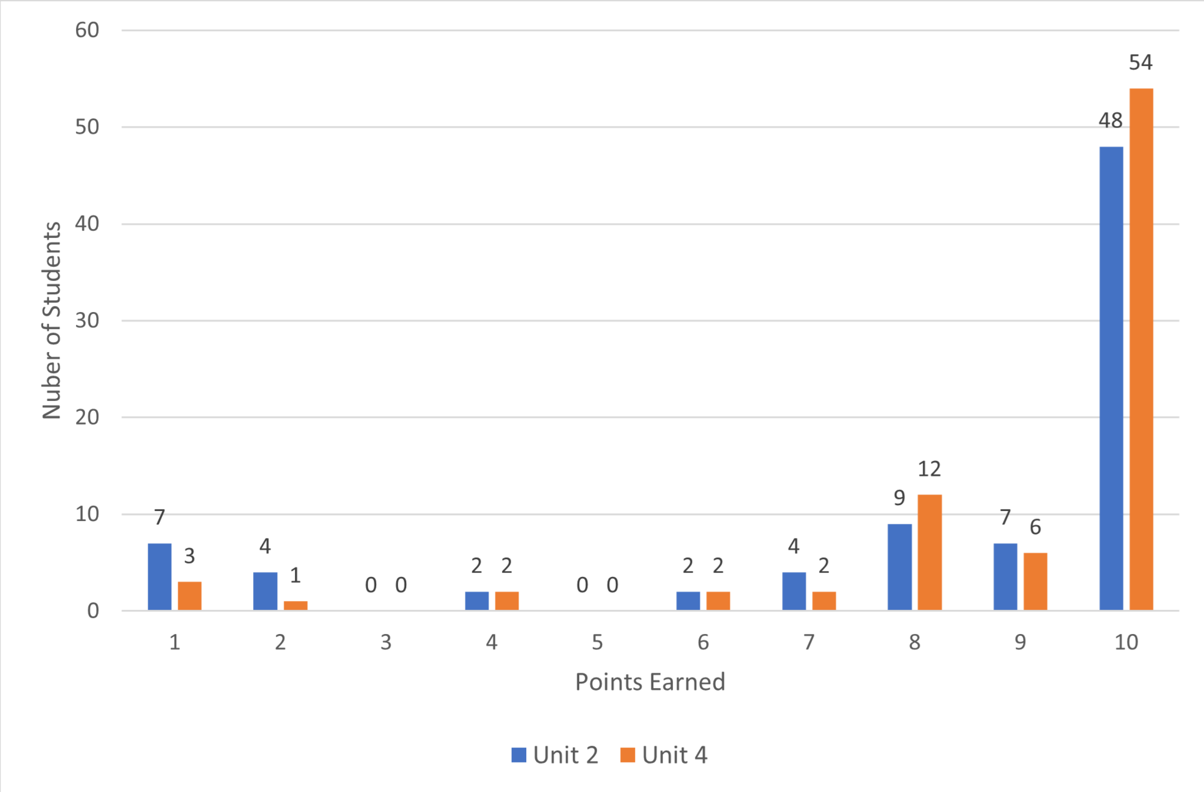


Figure 5.1: Distribution of points earned in unit 2 vs. unit 4 (n=87 students)

category. Figure 5.2 shows the migration of students between the two categories from unit 2 to unit 4 (Flourish, 2022).

5.3.1 Data analysis method

To answer the research question posed by this study, it is necessary to examine the data collected from Pebasco at a finer level of detail than shown by the student outcomes in Figure 5.2. The additional behavioral data can be mined and combined with the outcome data to identify student profiles in the context of Pebasco use.

A collaborative no-code visual interactive data analytics platform called Einblick (Zgraggen et al., 2017) was used to identify student usage patterns within the learning analytics data generated by the Pebasco app. Einblick was used to explore and manipulate the data and identify behavioral patterns of interest within the data. Einblick (2022) offers a video of one example session using this platform to explore the Pebasco data used in this study. Through this data exploration and manipulation, two behavioral patterns of interest emerged.

Figure 5.2: Migration of students between the High and Low Stars categories from unit 2 to unit 4

The first behavioral pattern is when students request the correct answer by using three hints during an attempt. The second behavioral pattern is when students' behavior while completing an audio conforms to multiple success criteria that the authors identified and combined into an aggregate behavioral pattern called "Top audio." Students displaying this pattern of behavior demonstrate higher levels of peer feedback skills. Table 5.2 shows the elements and thresholds of the collected learning analytics data used to create the "Top audio" category. For example, a student in the "Top audio" category would have only made one attempt in which they matched with all of the teacher comments and earned three stars without using any hints. The decision was made to allow up to two false positives (student feedback annotations that did not match with a teacher comment) because, in the usage data, many students used the fact that the system did not penalize false positives rather than using one hint to check the number of teacher comments in an audio. Finally, the system recorded the average number of seconds a student had played an audio per annotation they made. If it was too low, it might indicate a "spamming" technique in which the student quickly makes annotations without adequately listening to the audio, thereby abusing the lack of penalization of false positives. On the other end of the spectrum, if the time was too high, it might indicate that the student may not understand the audio well and may be hesitating or indecisive.

Table 5.2: Elements comprising the “Top Audio" behavioral pattern

Elements	**Threshold**	**Notes**
Max Number of Attempts	=1	Identifies audios in which the student only tried one attempt
Max Stars	=3	Identifies audios in which the student earned three stars
Max Hints	=0	Identifies audios in which the student did not rely on any hints
False Positives	≤2	Identifies audios in which the student made two or fewer annotations that did not match with the teacher comments
Seconds per Annotation	≥2 seconds	Eliminates audios in which the student was “spamming” annotations without listening
Seconds per Annotation	≤200 seconds	Eliminates audios in which the student was highly indecisive while listening

The behavioral patterns of “correct answer request” and “Top audio” were combined with the outcome data of the number of stars earned to identify learner profiles for each unit (see 5.3). Note that in Figure 5.3, the two rectangles on the far left and right sides only have one output because there were no participants in the other possible output. The “1_Desired” profile is a student who performed well across all of the success criteria: they earned more than 30 stars in a unit, more than 50% of their audios in the unit meet the “Top audio” criteria, and they requested the correct answers for fewer than 50% of the audios in the unit. The “2_Successful with Support” profile is a student who earned more than 30 stars in a unit, but they had under 50% of audios that met the “Top audio” criteria and 50% of audios that had correct answer requests. These students may have used other supports in Pebasco, such as the ability to make multiple attempts and to seek hints that do not reveal the correct answers. The “3_Successful with Answers” profile is a student who frequently relied on the system’s ability to reveal the correct answers to earn more than 30 stars in a unit. The “4_Unsuccesful with Answers” profile is a student who requested correct answers for more than 50% of the audios in the unit, but they still did not manage to get 30 stars or more. Only one student in one unit fell into this unusual category. Finally, the “5_Disengaged” profile is a student who made at least one attempt at a unit, but they disengaged from the system without earning 30 stars or more or displaying either the “correct answer request” or the “Top audio” behavioral pattern in more than 50% of the audios in a unit.

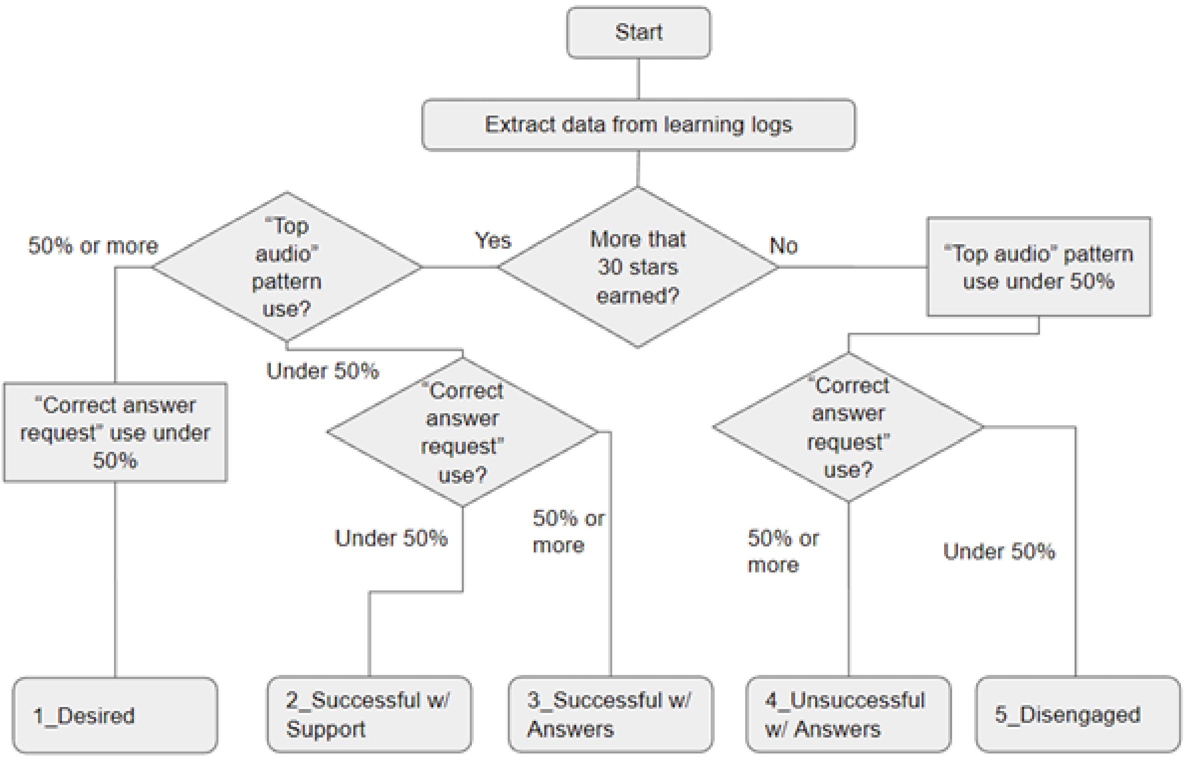


Figure 5.3: Flowchart showing the decision chain used to identify the learner profiles

Figure 5.4 shows the migration of the student profiles (n=87) within Pebasco from unit 2 to unit 4 (Flourish, 2022). Based on work by Majumdar and Iyer (2016), each profile is considered as a stratum, and the transition pattern is measured across the two phases (i.e., unit 2 and unit 4). Of the 36 students who started in the “1_Desired” profile in unit 2, 32 remained in that profile, while 2 migrated to the “2_Successsful with Support” profile, and two migrated to the “3_Successful with Answers” profile. Of the 18 students who started in the “2_Successsful with Support” profile in unit 2, only two remained in that profile, while 15 migrated to the “1_Desired” profile and one migrated to the “4_Unsuccessful with Answers" profile. Of the 14 students who started in the "3_Successful with Answers" profile, 13 remained, while one migrated to the "2_Successful with Support" profile. Of the 19 students who started in the "5_Disengaged" profile, 12 remained, while three migrated to the "2_Successsful with Support" profile and four migrated to the "1_Desired" profile.

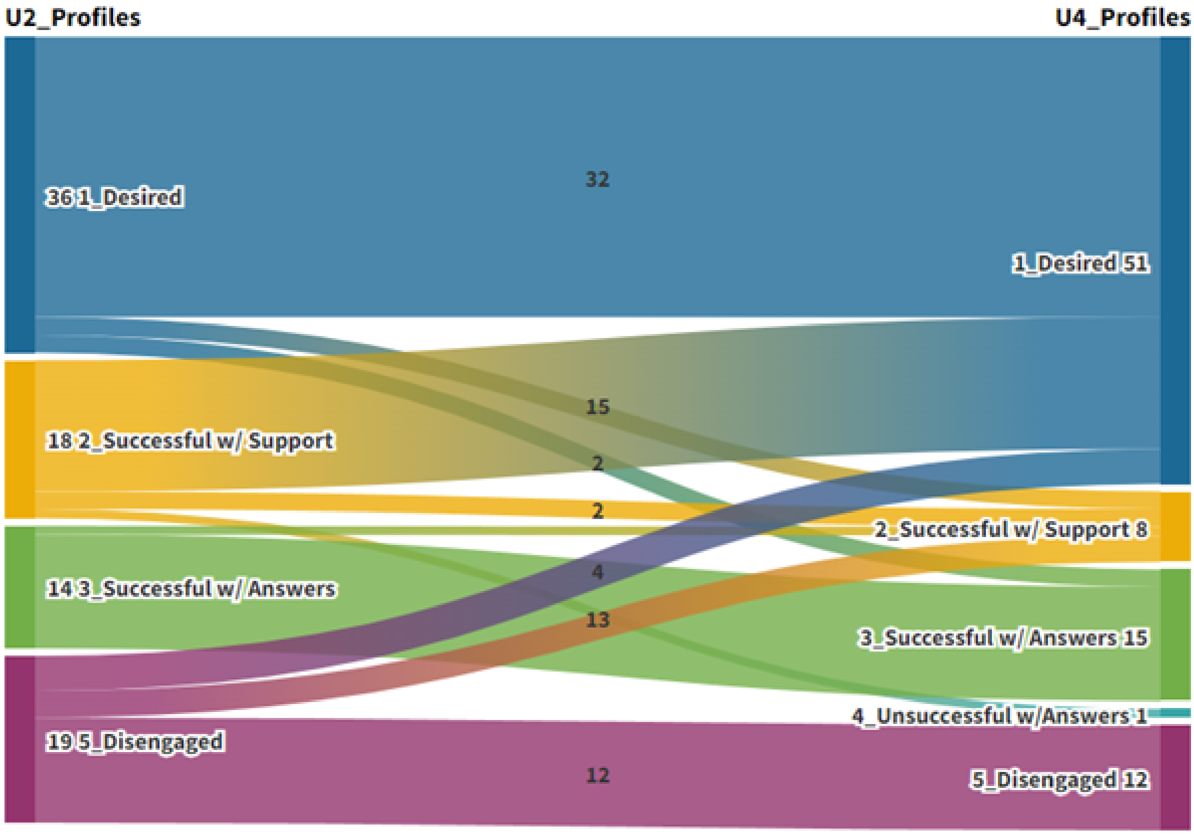


Figure 5.4: Migration of student profiles from unit 2 to unit 4

5.3.2 Addressing the research question

This study sought to answer the following research question: *How and to what extent can the asynchronous microlearning app Pebasco improve students' skills for providing peer feedback on spoken content in English as a foreign language (EFL) education?* By mining behavioral data from the Pebasco learning analytics system and combining it with the outcome data (i.e., the number of stars earned per unit), it became possible to identify five profiles of the students using the app. The profile migration that the students made over the course of using Pebasco indicates changes in the students' level of peer feedback skills.

The “1_Desired” profile is for students with the highest level of peer feedback skills that can be demonstrated within the app. The profile migration pattern shown in 5.4 shows that in unit 2, only 41% (n=36) of the students fit the “1_Desired” profile. However, by unit 4, 59% (n=51) of students fit the “1_Desired” profile. This includes 15 students who migrated up from the “2_Successsful with Support” profile and four students who migrated up from the “5_Disengaged” profile. This migration shows that using the Pebasco microlearning app can positively affect students’ peer feedback skills in the context of English as a foreign language (EFL) education.

It is also worth noting two groups of students who remained static in their profiles from unit 2 to unit 4. The first group is 13 out of 14 students who started in the “3_Successful with Answers” profile and remained there. This may be because they felt they could get a good grade on the Pebasco assignments without needing to engage deeply with the content in the app. The second group is the 12 out of 19 students who started in the “5_Disengaged” profile and remained there. This might indicate a problem with student motivation. Alternatively, perhaps they had difficulty understanding how to use the app.

To conduct a statistical analysis of the profile changes from unit 2 to unit 4, the profile types were considered as ordinal variables running from 1 (i.e., “1_Desired”) to 5 (i.e., “5_disengaged”), with lower numbers indicating a better-performing profile. Unit 2 (M=2.402, SD=1.551, n=87) was compared with unit 4 (M=2.023, SD=1.438, n=87). The Shapiro-Wilk Test of Normality was conducted and found to be statistically significant, suggesting a deviation from normality, so the Wilcoxon signed-rank one-tailed, paired samples t-test was used (see Table 5.3). It was hypothesized that the unit 4 profile "rank" would be a lower number (because lower numbers indicate higher performance/better behavior) than that of unit 2. The results showed a statistically significant difference (P=0.002), indicating an improvement from unit 2 to unit 4 as measured by profile migration.

5.4 Discussion

5.4.1 Contrast with previous work

The present research first interacts with previous work by addressing multiple calls for research. There have been calls for research examining how training can enhance the quality of peer feedback (Dao & Iwashita, 2021) and for more use of experimental and

Table 5.3: Paired samples T-test for profile migration from unit 2 to unit 4

Measure 1	Measure 2	W	Z	P	Effect-size
U2_Profiles	U4_Profiles	327	2.824	0.002	0.611

Note. For all tests, the alternative hypothesis specifies that U2_Profiles is greater than U4_Profiles.
Note. Wilcoxon signed-rank test to compare the difference and Rank-Biserial Correlation for effect size.

quasi-experimental methodologies in peer feedback research (Topping, 2013). In the Dual Model Theory of peer corrective feedback, it is theorized that there are educational benefits to both the provision and reception of PCF, and there is a call for research to "experimentally tease apart the two sides" (Sato, 2017, p. 30). The present study addresses these calls for research by using a quasi-experimental research methodology to investigate the effect of practicing only the provision of peer feedback on the quality of the feedback provided.

Jahnke et al. (2020) proposed a set of mobile microlearning design principles described earlier in this paper. The design of the second version of the Pebasco app, which was used in the present study, adheres to the mobile microlearning design principles described by Jahnke et al. (2020): (a) the L2 listening and peer feedback practice are practical and interactive; (b) the Pebasco activities can be completed in only a few minutes; (c) multimedia audio clips are integral to the user experience; (d) the system provides immediate formative feedback in the form of text notifications, animations, and icon color changes; (e) and finally, Pebasco is accessible across multiple devices and operating systems. Although there are several studies that research novel uses of microlearning to improve L2 vocabulary skills (e.g., Arakawa et al., 2022; Dingler et al., 2017; Inie & Lungu, 2021; Kovacs, 2015), there is a gap in the literature of research that investigates the use of asynchronous mobile microlearning peer feedback training in the context of spoken content in foreign language learning. The current study aimed to fill this gap. The results of this study show that the use of an asynchronous mobile microlearning system resulted in a statistically significant shift in user profiles over time. This expands on the previous literature by demonstrating the use of mobile microlearning to improve peer feedback skills in the context of spoken content in foreign language learning.

The Pebasco app builds on previous research showing that peer assessment skills can be trained (Sluijsmans et al., 2004; van Zundert et al., 2010); that the practice of peer feedback (both provision and reception) can improve learners' L2 language skills (El Mortaji, 2022; Patri, 2002; Rodríguez-González & Castañeda, 2016); and that peer feedback practice can improve learners' ability to detect L2 errors (Fujii et al., 2016). The results of this study show a pattern of statistically significant profile migration over the course of using Pebasco that indicates that many participants improved or maintained desirable patterns of behavior and outcomes. This suggests a positive impact on the quality of peer feedback skills and L2 language skills, as well as the ability to detect L2 errors. This improvement in L2 language skills is important because a lack of L2 proficiency can hamper the effectiveness of peer feedback in language learning (Lyster et al., 2013; Philp et al., 2010; Toth, 2008). The current study adds to this body of literature by demonstrating that mobile microlearning can be an effective means of training peer feedback skills and L2 language skills.

One possible explanation of this improvement is based on the multiple repetitions of practice that the participants had to practice their L2 skills and their peer feedback skills as described in Skill Acquisition Theory (Sato & Lyster, 2012). However, deeper consideration must be given to the performance of the students who remained in the "3_Successful with Answers" profile. It may suggest a difficulty in balancing two theoretical approaches. On the one hand, the design decision to include scaffolding hints and the ability to request the correct answers was informed by Cognitive Load Theory (Khong & Kabilan, 2022; Sweller, 1994, 2020) and the desire not to overwhelm the users, regardless of their cognitive capacity at the time. However, the option to view the correct answers seems to have created a situation for this particular group of students in which they could demonstrate lower cognitive engagement as described by the ICAP framework (Chi et al., 2018) while still achieving the desired outcome (i.e., receiving a high number of stars).

5.4.2 Limitations and future work

In this study, the context was limited to a Japanese university. In future studies, it might be worthwhile to expand to different contexts. Likewise, this study used the no-code tools Bubble and Einblick. Perhaps researchers might be interested in trying to recreate an app similar to Pebasco using different tools. Finally, the purpose of Pebasco is to provide microlearning opportunities for students to improve their peer feedback skills on spoken content in a CLT context while collecting learning analytics data. Researchers may want to try to conduct similar studies in other domains.

One limitation of this paper is that the pre-/post- data used was limited to the internal performance behavior data generated by the students using Pebasco. From the evaluation of the learning perspective, this study focused on learning behaviors while using the application. The profiles included the evaluation provided based on the task performance. While such rank indicators helped to understand the dynamics of improvement while using the application repeatedly over multiple similar tasks, one limitation of this remains that there is no control group to find the causal relationship of the intervention. Hence, further studies with independent measures of appropriate pre-/post-test performance would also help to establish the learning impact of using the Pebasco app.

5.4.3 Implications for instructors

The findings of this study demonstrate that a teacher or researcher with little or no knowledge of computer programming can use no-code tools to build an asynchronous microlearning app that serves as a learning analytics platform while supporting peer feedback skills on spoken content in a CLT-based EFL context. Hopefully, this will inspire others to try to build learning analytics apps and microlearning apps that fit their contexts and teaching/learning needs.

As described earlier in this paper, no-code tools sometimes have a learning curve and may require training and support, particularly when one tries to build a complex app. No-code novices should start with a simple plan for their app and then consider all the elements, such as data and workflows, that will be needed in their finished app. It is highly recommended that novices seek out support from the no-code community. Many freely available resources, such as online videos, training boot camps, and educational communities, can help improve a novice's skill set quickly (see Buildcamp, 2022; ProNoCoders, 2022).

5.5 Conclusion

This study looked at the use of Pebasco. This asynchronous mobile microlearning app supports the development of students' peer feedback skills on spoken content while simultaneously collecting usage data as a learning analytics platform. The system has a hint engine that offers a range of scaffolded support for students. The findings in this study indicate that the Pebasco app can help students improve their peer feedback skills on spoken content in the context of CLT-based EFL. The findings also suggest some areas of improvement for the system that can be made in future design iterations. This research is novel because of a current lack of research on the use of no-code technology to develop educational apps, learning analytics apps, or apps for computer-assisted language learning (CALL), particularly in the context of microlearning for the improvement of peer feedback skills in CLT-based EFL. Furthermore, this research addresses the call for further research in training students' peer feedback skills (Kasch et al., 2021) and the provision of peer feedback in particular (Sato, 2017).

Chapter 6

Research study 3: “A microlearning app for peer feedback training and its effect on learning performance and self-confidence during an EFL speaking task”

6.1 Research background

Peer feedback, often synonymous with peer assessment, involves one learner evaluating another's performance and includes peer corrective feedback (PCF) as a form of qualitative feedback (Alqassab et al., 2018; Liu & Carless, 2006; Panadero & Alqassab, 2019). It has been shown to enhance student learning and achievement (Hattie, 2009, 2012; Hattie & Clarke, 2019; Kerr, 2020) and improve self-reflection and internal feedback skills (To & Panadero, 2019). In language learning, peer feedback can bolster second language (L2) skills, (El Mortaji, 2022; Patri, 2002; Rodríguez-González & Castañeda, 2016). It also helps learners identify L2 errors (Fujii et al., 2016).

However, the impact of peer feedback on L2 learning can vary. While some studies highlight its benefits for language skills and self-confidence (Rodríguez-González & Castañeda, 2016; Sato & Lyster, 2012), Adams et al. (2011) reported negative effects, suggesting that the quality of feedback is crucial. Therefore, effective training in peer feedback is necessary to ensure its positive impact on language learning (Sluijsmans et al., 2004; van Zundert et al., 2010).

The Dual Model Theory of PCF suggests differing impacts on learning between providing and receiving feedback, with growing evidence indicating that providing PCF has a greater positive effect (Lu & Law, 2012; Nicol et al., 2014; Wu & Schunn, 2023; Yu & Schunn, 2023; Zong et al., 2021). According to Wu & Schunn (2023), providing PCF within the ICAP framework is more constructive and cognitively engaging. Nicol (2021) argues that giving peer feedback enhances learners' "internal feedback" skills—the ability to reflect, evaluate, and regulate their learning—by leveraging the brain's natural tendency to engage in comparisons, a fundamental cognitive process (Gentner et al., 2001; Goldstone et al., 2010; Hofstadter & Sander, 2013). This study aims to contribute to ongoing research by using a microlearning app to train peer feedback provision, addressing the need to distinguish between the effects of providing and receiving feedback (Sato, 2017).

Despite extensive research on microlearning for L2 vocabulary (Arakawa et al., 2022; Cai et al., 2015, 2017; Dearman & Truong, 2012; Dingler et al., 2017; Edge et al., 2011; Inie & Lungu, 2021; Kovacs, 2015; Schneegass et al., 2022; Trusty & Truong, 2011; Zhao et al., 2018), its use for peer feedback training is underexplored. Peer feedback training significantly enhances learning outcomes (Lyster et al., 2013; Philp et al., 2010; Sato, 2013; Sato & Ballinger, 2016; Sluijsmans et al., 2004; Toth, 2008; van Zundert et al., 2010), yet asynchronous mobile microlearning in this context is scarcely studied. Previous research on the Pebasco app (Gorham et al., 2023) addressed this gap by showing its efficacy in improving peer feedback skills. The current study extends this work by assessing Pebasco's impact on an external communicative EFL activity.

Transfer of learning, the ability to apply prior knowledge to new situations, is crucial for problem-solving, creative thinking, and other higher cognitive processes (Sousa, 2022). It can be categorized into "near transfer" when the new situation is similar to the original learning context and "far transfer" when the new situation is markedly different (Barnett & Ceci, 2002; Double et al., 2020; Yu & Schunn, 2023). Achieving successful transfer is challenging (Barnett & Ceci, 2002), and even educators often struggle to apply professional development learning to their teaching practice (Chi et al., 2018). Transfer Appropriate Processing (TAP), which aligns cognitive processes during learning with those needed in future situations, can promote transfer (Lightbown, 2008; Morris et al., 1977). This principle is supported by a meta-analysis indicating that peer assessment activities need to match performance indicators to be effective (Double et al., 2020). The current study addresses the need for more research on the factors modulating peer assessment's impact on learning by examining the effect of a microlearning app on near transfer of peer feedback skills. It also contributes to the limited research on how peer feedback affects the transfer of learning to other tasks (Yu & Schunn, 2023).

This study is part of a larger, multi-year Ph.D. project investigating the effects of training students' peer feedback skills on spoken content within a mandatory university English as a Foreign Language (EFL) course that uses a communicative language teaching (CLT) approach. Gorham et al. (2024) reported on the development of a prototype learning analytics system for training peer feedback skills on spoken content. Subsequent research by Gorham et al. (2023) reported on the development of a standalone mobile microlearning app called Pebasco based on these findings. That study indicated that the app effectively improved students' peer feedback skills on spoken content as measured by their performance within the app. The current study goes further to explore how using the Pebasco app impacts students' ability to transfer their learning from peer feedback training to a constructive CLT speaking activity. The following two research questions are answered as part of this study:

RQ1: *How and to what extent does the use of a microlearning app for training peer feedback skills in the context of university CLT EFL classes affect the transfer of learning to a CLT speaking activity?*

RQ2: *How and to what extent does the use of a microlearning app for training peer feedback skills impact students' self-reported perceptions of improvement of peer- and internal-feedback skills, as well as general L2 confidence?*

6.2 Methods

6.2.1 Pebasco performance indicators

Based on previous work with Pebasco (Gorham et al., 2023), learning analytics data from the Pebasco practice/training platform were used to categorize students into distinct user profiles based on their interaction patterns and performance outcomes. Two key behavioral patterns were identified: the "correct answer request" pattern, where students used all three hints to reveal correct answers, and the "Top audio" pattern, characterized by students making one attempt for an audio, matching all teacher comments, earning three stars without hints, and having no more than two false positives. The "Top audio" pattern indicates when a student has successfully completed an audio without needing to rely on hints from the system.

Students were then classified into five user profiles (see Figure 6.1). The “1_Desired” profile included students who excelled across all criteria, earning more than 30 stars per unit, with 50% or more of the audios meeting the “Top audio” criteria and less than 50% of audios involving correct answer requests. The “2_Successful with Support” profile consisted of students who also earned over 30 stars but had less than 50% of audios in the “Top audio” category and less than 50% of the audios that had correct answer requests. Students frequently relying on correct answer requests (i.e., 50% or more of the audios) to achieve over 30 stars were classified as “3_Successful with Answers.” The “4_Unsuccessful with Answers” profile was rare and included students who requested correct answers for more than 50% of audios but did not earn 30 stars. Lastly, the “5_Disengaged” profile identified students who made attempts but did not earn 30 stars or exhibit either primary behavioral pattern in over 50% of audios.

6.2.2 Research design and context

This research employed a quasi-experimental single-group pre-/post- design, utilizing convenience sampling. For the pre-/post- measurements, the participants used Microsoft’s Flipgrid as the performance platform and the mobile microlearning app Pebasco as the practice and training platform. A more detailed description of the flow of activities with these two platforms will be provided in a subsequent section.

This study was conducted in a mandatory, asynchronous EFL course focused on Communicative Language Teaching (CLT) for first-year students in the Faculty of Letters at a Tokyo-based university. The course’s main objective was to enhance students’ fundamental English communication abilities, with a secondary aim of developing their peer and internal feedback skills for spoken content. The first author’s university’s research ethics board approved this study’s methods and data collection.

In the preceding study, Gorham et al. (2023) demonstrated that most students using the Pebasco app enhanced their peer feedback skills, as evidenced by their in-app performance. The current study extends this research by analyzing the impact of training on peer feedback skills within Pebasco on subsequent communicative speaking activities recorded on Flipgrid. Despite a largely consistent participant base, some adjustments to the cohort were necessary for the current focus on Flipgrid performance.

The initial cohort from Gorham et al. (2023) included 87 students. For the current

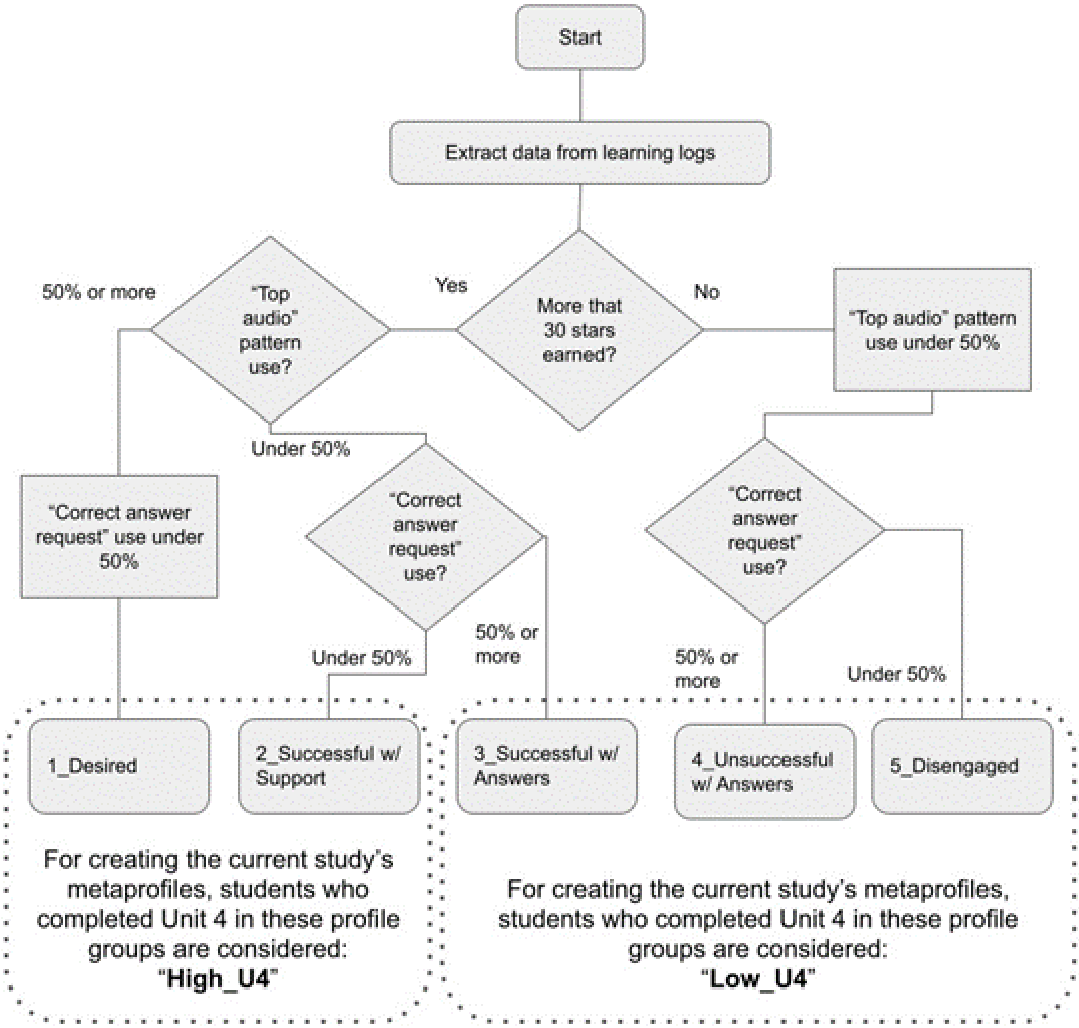


Figure 6.1: Pebasco User Profiles

study, five students who did not complete both the pre- and post- Flipgrid videos were excluded. Additionally, seven students who were excluded in the previous study for not completing both Units 2 and 4 in Pebasco, but who did complete the Flipgrid assessments in this study, were added to strengthen the analysis. These students are now classified under the profile type "5_Disengaged," akin to the least engaged users from the prior study, bringing the total number of participants to 89. Further explanation of the user profile types will be given in a subsequent section.

6.2.3 Description of the flow of activities

This study relies on two primary platforms: a performance platform and a practice/training platform. The performance platform is Microsoft's Flipgrid. Please note that Microsoft shortened the application's name from Flipgrid to Flip in 2022. Flipgrid is an educational application that allows students to create and edit short video clips they share with their classmates. In the context of the ICAP engagement model, this can be considered a constructive activity.

To demonstrate their pre-treatment performance on their constructive EFL speaking activity, students first uploaded a 30 to 90-second video to Flipgrid, responding in English to a prompt on a general daily communication topic. They then completed four units of peer feedback training on spoken content using the Pebasco app. Finally, to show their post-treatment performance, each student uploaded another video to Flipgrid, similar in length and topic type to the initial one.

6.2.4 Data collected

Data were collected from three sources: the performance platform Flipgrid, the practice/training platform Pebasco, and an end-of-class survey conducted with Microsoft Forms. From Flipgrid, the length in seconds of each submitted video was recorded. The course instructor then evaluated each video, noting the total number of mistakes and the timestamps of each mistake. The Pebasco app collects detailed learning analytics from users' interactions, tracking various data for each attempt on audio tasks, including stars earned, hints used, false positives (incorrect annotations), time spent playing the audio, and other navigation data. In this study, Unit 1 was used for onboarding and training, while the outcomes and behaviors of students in Unit 4 were analyzed. Based on the data, students were categorized into user profile groups as outlined by Gorham et al. (2023), ranging from most to least preferred: 1_Desired, 2_Successful with Support, 3_Successful with Answers, 4_Unsuccessful with Answers, and 5_Disengaged (see Figure 6.1). Finally, near the end of the school term, students completed an end-of-class (EOC) survey with Microsoft Forms. Using a 5-point Likert scale, they reported their perceptions on the improvement of their overall English confidence, as well as the effectiveness of Pebasco in enhancing their peer and internal feedback skills.

6.2.5 Data analysis method

First, a method for evaluating students' work on the Flipgrid performance platform was developed. To normalize student performance, a "construction quality index" was introduced. The naming of this index reflects the categorization of Flipgrid videos as overt constructions according to the ICAP model. It is calculated by dividing the number of mistakes made in a video by the video's length in seconds.

Next, students were sorted into meta-profiles that combined (a) students' performance on the pre-Flipgrid submissions as measured by the construction quality index and (b) the ending profile type that students had either maintained or migrated to by the end of the 4th unit of Pebasco. Students' starting performance on the "Pre- Construction Quality Index" was used to identify students who were starting at the same level. Those who had the median or better were classified as "High Pre-," while those who had worse than the median were classified as "Low Pre-." Additionally, students' Pebasco profile migration endpoints in unit 4 (U4) were categorized. Those who were static or migrated to "1_Desired" or "2_Success_with_support" were classified as "High_U4," whereas those who were static or migrated to "3_Success_with_answers," "4_unsuccessful_with_answers," or "5_Disengaged" were classified as "Low_U4."

A matrix of these two factors resulted in the creation of four metaprofiles: 1_High_Pre/High_U4 (n=29); 2_High_Pre/Low_U4 (n=16); 3_Low_Pre/High_U4 (n=26); and 4_Low_Pre/Low_U4 (n=18) (see Figure 6.2).

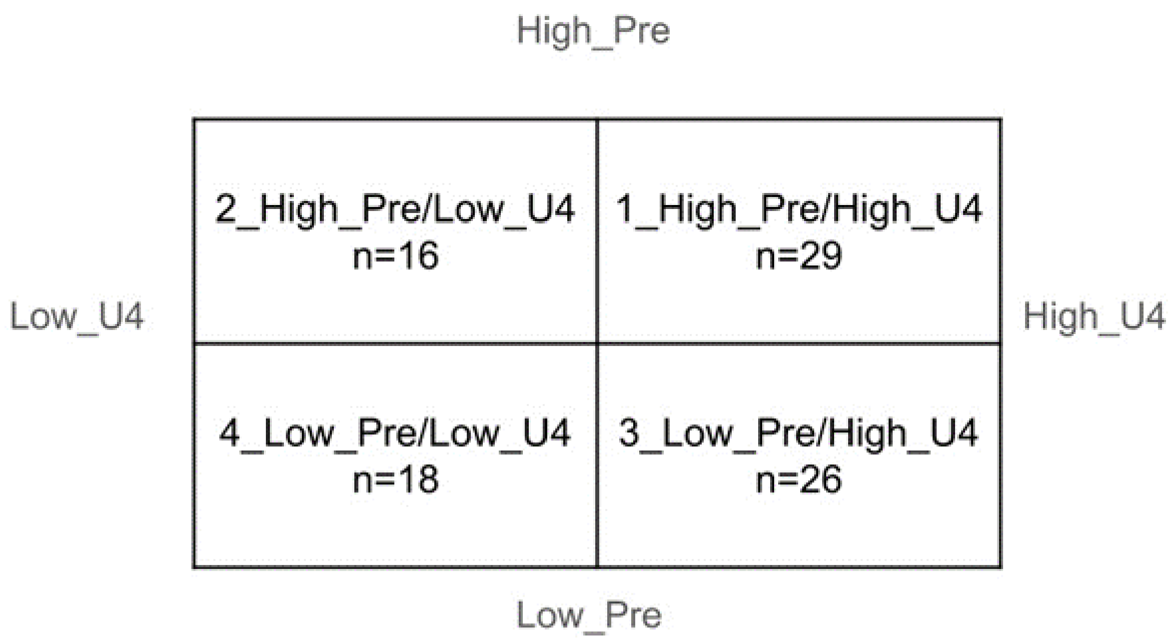


Figure 6.2: Metaprofile matrix

Based on these meta-profiles, ANOVA and post-hoc tests were conducted to compare student performance on the Post- Construction Quality Index and student perceptions on the EOC survey.

6.3 Results and interpretation

The two research questions posed by this study will be addressed by analyses of data collected by (a) the practice/training platform, Pebasco, and (b) the performance platform, Flipgrid. These two sets of analyses will be reported in the following two sub-sections. Each sub-section will share the relevant descriptive statistics, followed by the results of analysis of variance (ANOVA) and post hoc tests (Tukey) to compare the students assigned to the four metaprofile groups. Please note that when the metaprofile group names are mentioned in the tables, the terms "high" and "low" will be abbreviated as "hi" and "lo" (e.g., the metaprofile "2_High_Pre/LowU4" will be shortened to "2_HiPre_Lo_U4"). JASP software (Love et al., 2019) was used to analyze the data.

6.3.1 Post- Construction Quality Index analyses

The descriptive statistics for the pre- and post- Construction Quality Index scores are provided in Table 6.1. It is worth reiterating that the mean pre- Construction Quality Index scores for the 1_High_Pre/HighU4 group (n=29) and the 2_High_Pre/LowU4 (n=16) are similar to each other; and the mean pre- Construction Quality Index scores for the 3_Low_Pre/HighU4 group (n=26) and the 4_Low_Pre/LowU4 group (n=18) are also similar to each other. This is unsurprising because this score was one of the factors used in determining the metaprofiles.

The post- Construction Quality Index scores were compared using ANOVA (see Table 6.2), $F(3/85) = 6.567$, $p = < .001$, with $\eta^2 = 0.188$, indicating a large effect size (Cohen, 1988). Levene's test for equality of variances was conducted and found to be not statistically significant; therefore, no homogeneity correction was used. Tukey's HSD Test for multiple comparisons found that the mean value of the post- Construction Quality Index scores was significantly different between 1_High_Pre/HighU4 and 4_Low_Pre/LowU4 ($p = < .001$); between 2_High_Pre/LowU4 and 4_Low_Pre/LowU4 ($p = < .009$); and between 3_Low_Pre/HighU4 and 4_Low_Pre/LowU4 ($p = < .022$).

Table 6.1: Descriptive statistics pre- & post- Construction Quality Index

	Pre- Construction Quality Index				Post- Construction Quality Index			
	1_ HiPre _HiU4	2_ HiPre _LoU4	3_ LoPre _HiU4	4_ LoPre _LoU4	1_ HiPre _HiU4	2_ HiPre _LoU4	3_ LoPre _HiU4	4_ LoPre _LoU4
Count of Learners	29	16	26	18	29	16	26	18
Mean	0.051	0.052	0.1	0.119	0.05	0.056	0.064	0.097
Std. Deviation	0.018	0.019	0.017	0.033	0.033	0.032	0.032	0.05
Skewness	-0.983	-1.384	0.638	0.477	0.215	0.615	-0.26	1.153
Kurtosis	0.699	2.408	-0.601	-1.154	-0.83	0.851	-0.394	1.029
Minimum	0	0	0.078	0.079	0	0	0	0.031
Maximum	0.073	0.073	0.139	0.176	0.111	0.129	0.125	0.226

6.3.2 Analyses of EOC survey results

The first EOC survey question that was analyzed was one that asked the students to self-report if they felt as though their confidence in using English had improved as a result of taking this mandatory EFL class using a 5-point Likert scale. The descriptive statistics for this question are provided in Table 6.3. Please note that some of the learners were excluded because they did not complete the EOC survey.

The answers that students in different metaprofile groups provided to this first question were compared using ANOVA (see Table 6.4), F(3/74)= 4.132, p=0.009, with η^2=0.143, indicating a large effect size (Cohen, 1988). Levene's test for equality of variances was conducted and found to be not statistically significant; therefore, no homogeneity correction was used. Tukey's HSD Test for multiple comparisons found that the mean value of answers for this survey question was significantly different between 3_Low_Pre/HighU4 and 4_Low_Pre/LowU4 (p= < .005).

Table 6.2: ANOVA of post- Construction Quality Index

Cases	Sum of Squares	Mean Square	F (3,85)	p	η^2
Metaprofiles	0.027	0.009	6.567	< .001	0.188
Residuals	0.115	0.001			

Note. Type III Sum of Squares.

Table 6.3: Descriptive statistics EOC English confidence self-report

	1_ HiPre _HiU4	2_ HiPre _LoU4	3_ LoPre _HiU4	4_ LoPre _LoU4
Count of Learners	28	15	23	12
Excluded learners	1	1	3	6
Mean	3.929	3.8	4.174	3.333
Std. Deviation	0.604	0.676	0.576	0.985
Skewness	0.022	-1.344	0.018	0.559
Kurtosis	0.014	3.281	0.123	-0.309
Minimum	3	2	3	2
Maximum	5	5	5	5

The second EOC survey question that was analyzed was one that asked the students to self-report if they felt as though their use of the Pebasco app had improved their peer feedback skills using a 5-point Likert scale. The descriptive statistics for this question are provided in Table 6.5.

The answers that students in different metaprofile groups provided to this second question were compared using ANOVA (see Table 6.6), F(3/74)= 5.058, p= 0.003, with η^2=0.17, indicating a large effect size (Cohen, 1988). Levene's test for equality of variances was conducted and found to be not statistically significant; therefore, no homogeneity correction was used. Tukey's HSD Test for multiple comparisons found that the mean value of answers for this survey question was significantly different between 1_High_Pre/HighU4 and 4_Low_Pre/LowU4 (p= 0.001).

The third EOC survey question that was analyzed was one that asked the students to

Table 6.4: ANOVA of EOC English confidence self-report

Cases	Sum of Squares	Mean Square	F (3,74)	p	η^2
Metaprofiles	5.733	1.911	4.132	0.009	0.143
Residuals	34.228	0.463			

Note. Type III Sum of Squares

Table 6.5: Descriptive statistics EOC peer feedback skill improvement self-report

	1_HiPre _HiU4	2_HiPre _LoU4	3_LoPre _HiU4	4_LoPre _LoU4
Count of Learners	28	15	23	12
Excluded learners	1	1	3	6
Mean	4.357	3.8	3.913	3
Std. Deviation	0.78	0.775	1.276	1.206
Skewness	-1.25	-0.681	-1.261	-0.373
Kurtosis	1.662	1.081	0.689	-0.16
Minimum	2	2	1	1
Maximum	5	5	5	5

self-report if they felt as though their use of the Pebasco app had improved their internal feedback skills using a 5-point Likert scale. The descriptive statistics for this question are provided in Table 6.7.

The answers that students in different metaprofile groups provided to this second question were compared using ANOVA (see Table 6.8), $F(3/74)= 4.62$, $p= 0.005$, with η^2=0.158, indicating a large effect size (Cohen, 1988). Levene's test for equality of variances was conducted and found to be not statistically significant; therefore, no homogeneity correction was used. Tukey's HSD Test for multiple comparisons found that the mean value of answers for this survey question was significantly different between 1_High_Pre/HighU4 and 4_Low_Pre/LowU4 ($p= 0.003$); and between 3_Low_Pre/HighU4 and 4_Low_Pre/LowU4 ($p= 0.029$).

Table 6.6: ANOVA of EOC peer feedback skill improvement self-report

Cases	Sum of Squares	Mean Square	F (3,74)	p	η^2
Metaprofiles	15.717	5.239	5.058	0.003	0.17
Residuals	76.655	1.036			

Note. Type III Sum of Squares

Table 6.7: Descriptive statistics EOC internal feedback skill improvement self-report

	1_HiPre_HiU4	2_HiPre_LoU4	3_LoPre_HiU4	4_LoPre_LoU4
Count of Learners	28	15	23	12
Excluded learners	1	1	3	6
Mean	4.179	3.667	3.957	3
Std. Deviation	0.863	0.724	1.065	1.128
Skewness	-1.113	-0.676	-1.143	-0.912
Kurtosis	1.135	0.948	1.366	-0.337
Minimum	2	2	1	1
Maximum	5	5	5	4

6.3.3 Addressing the research questions

The current study aimed to address two research questions regarding the use of a mobile microlearning app, Pebasco, for training peer feedback skills in a mandatory university CLT EFL class. The first research question (RQ1) focused on how and to what extent the app affected the transfer of learning to a constructive CLT speaking activity. The second research question (RQ2) investigated the impact of the app on students' self-reported perceptions of improvement in peer- and internal feedback skills, as well as general confidence in using English.

Table 6.8: ANOVA of EOC internal feedback skill improvement self-report

Cases	Sum of Squares	Mean Square	F (3,74)	p	η^2
Metaprofiles	12.436	4.145	4.62	0.005	0.158
Residuals	66.397	0.897			

Note. Type III Sum of Squares

6.3.4 Addressing RQ1: Transfer of learning to a constructive CLT speaking activity

The transfer of learning was assessed using the post- Construction Quality Index scores, which measured the quality of student output in a constructive CLT speaking activity. The ANOVA results revealed a significant difference among the metaprofile groups, indicating that the use of Pebasco was correlated with the quality of students' constructed responses.

The 1_High_Pre/HighU4 and 2_High_Pre/LowU4 groups both started with high initial performance, as indicated by their pre- Construction Quality Index scores. However, it is noteworthy that the 2_High_Pre/LowU4 group did not fully engage with the Pebasco app, which suggests that factors other than the app might have contributed to their high post-Construction Quality Index scores, such as their baseline L2 skill level at the start of this study. In contrast, the 3_Low_Pre/HighU4 group, which started with lower initial performance but demonstrated active engagement with the Pebasco app, also demonstrated significantly higher post- Construction Quality Index scores compared to the 4_Low_Pre/LowU4 group and was statistically indistinguishable from the groups that started higher in the pre-performance.

These findings suggest that while high initial performance is a strong indicator of high post-performance, active engagement with the Pebasco app is correlated with significant improvements even for students who started with lower initial performance. This underscores the potential effectiveness of peer feedback training using a mobile microlearning app in transferring learning to enhanced performance in related CLT tasks in some cases, addressing the first research question.

6.3.5 Addressing RQ2: Self-reported perceptions of improvement

The impact of Pebasco on students' perceptions of their peer- and internal-feedback skills was evaluated through an end-of-class (EOC) survey, which included questions about confidence in using English, improvement in peer feedback skills, and improvement in internal feedback skills.

The ANOVA results for the EOC survey questions indicated significant differences in self-reported perceptions among the metaprofile groups. For confidence in using English, students in the 3_Low_Pre/HighU4 group reported significantly higher confidence levels compared to those in the 4_Low_Pre/LowU4 group. This suggests that even students who started with lower initial performance but actively engaged with the Pebasco app experienced substantial gains in their confidence.

Regarding peer feedback skills, the 1_High_Pre/HighU4 group perceived a greater improvement than the 4_Low_Pre/LowU4 group. Similarly, for internal feedback skills, both the 1_High_Pre/HighU4 group and the 3_Low_Pre/HighU4 group reported significantly higher improvements compared to the 4_Low_Pre/LowU4 group.

These results suggest that students who successfully engaged with the Pebasco app perceived greater improvements in their feedback skills. Of those students, the ones who started with lower performance scores reported significantly more confidence and improved internal feedback skills, while the ones who started with higher performance scores reported significantly improved peer and internal feedback skills, thereby addressing the second research question.

6.4 Discussion

6.4.1 Contrasting with previous research

The findings of the current study align with, yet extend beyond, the existing body of research on peer feedback and microlearning. Previous studies have established the benefits of peer feedback on student learning and achievement, internal feedback skills, and L2 abilities (Hattie, 2009, 2012; Hattie & Clarke, 2019; To & Panadero, 2019; El Mortaji, 2022; Patri, 2002; Rodríguez-González & Castañeda, 2016; Fujii et al., 2016; Kasch et al., 2021). The current study aligns with these findings, showing that students who effectively engaged with the Pebasco app perceived significant improvements in both peer and internal feedback skills. Notably, students who began with lower pre- Construction Quality Index scores demonstrated improvements in their post- Construction Quality Index scores, matching the performance of peers who initially started with higher scores.

However, the current study also addresses a notable gap in the literature by focusing on the use of a mobile microlearning app for peer feedback training in a language learning context. While extensive research has examined the application of microlearning for L2 vocabulary learning (Arakawa et al., 2022; Cai et al., 2015, 2017; Dearman & Truong, 2012; Dingler et al., 2017; Edge et al., 2011; Inie & Lungu, 2021; Kovacs, 2015; Schneegass et al., 2022; Trusty & Truong, 2011; Zhao et al., 2018), its use for training peer feedback skills has not been widely researched. This study contributes to filling this gap by providing empirical evidence that the Pebasco app can effectively enhance students' peer feedback skills and facilitate the transfer of these skills to a constructive CLT speaking activity in some cases.

The Dual Model Theory of peer corrective feedback (PCF) (Sato, 2017) suggests that providing PCF has a greater positive impact on learning compared to receiving it. This may be due to higher cognitive engagement as described in the ICAP framework or the development of internal feedback skills (Lu & Law, 2012; Nicol et al., 2014; Wu & Schunn, 2023; Yu & Schunn, 2023; Zong et al., 2021). The current study aligns with these findings, demonstrating that students who actively trained to provide feedback through the Pebasco app reported significant improvements in their internal feedback skills. This finding supports the argument that the provision of feedback, being a more constructive and cognitively engaging activity within the ICAP framework (Chi, 2009; Wu & Schunn, 2023), can lead to enhanced learning outcomes.

Furthermore, the study addresses calls for more research into the factors that modulate the effects of peer assessment on learning (Double et al., 2020). By examining the impact of a microlearning app on the near transfer of peer feedback skills, the study adds to the limited research on how peer feedback affects the transfer of learning to other tasks (Yu & Schunn, 2023). The significant improvements observed in the post- Construction Quality Index scores and the positive self-reported perceptions of feedback skills highlight the potential of microlearning apps to facilitate effective transfer of learning, at least in some cases.

The results also underscore the importance of active engagement with the microlearning app. While high initial performance was a strong indicator of high post-performance, the study found that students who started with lower initial performance but actively engaged with the Pebasco app (i.e., the 3_Low_Pre/HighU4 group) achieved post-performance levels statistically indistinguishable from those who started with higher initial performance (i.e., the 1_High_Pre/HighU4 and 2_High_Pre/LowU4 groups). This finding is significant as it suggests that targeted training using a microlearning app may be able to help bridge the gap between lower and higher initial performers.

6.4.2 Implications for instructors

The findings of this study suggest some practical implications for instructors using peer feedback training tools like the Pebasco app in EFL contexts. One key takeaway is the potential benefit of mobile microlearning apps in facilitating the transfer of learning from peer feedback training to related communicative tasks. The significant improvement observed in initially low-performing students who actively engaged with the app highlights the importance of consistent and meaningful interaction with such tools.

In the context of instructional design, the ICAP framework (Chi, 2009) has been a valuable heuristic for categorizing and encouraging constructive student activities. However, one of the drawbacks of this framework is its limited attention to external, visible processes (Chi, 2009; Thurn et al., 2023). Thurn et al. (2023) suggest that regular formative assessments are a proven approach that can effectively address this limitation by providing ongoing feedback.

The results of this study, along with insights from Nicol (2021), indicate that fostering internal feedback skills is a crucial internal process that the ICAP framework may overlook. Internal feedback skills, which involve students' ability to reflect on, evaluate, and regulate their own learning processes, are important for the learning process. Therefore, instructors should not only use the ICAP framework as a guide for designing interactive and constructive activities but also integrate strategies to promote internal feedback mechanisms.

Jahnke et al. (2020) described one of the key design principles of microlearning apps is that they provide immediate feedback; this formative assessment can give the opportunity to practice both peer and internal feedback processes. Encouraging students to engage in self-assessment and reflective practices while using peer feedback training apps and the provided scaffolding can enhance their internal feedback skills.

One possible interpretation of the implications of this process is that these improved internal feedback skills may provide a wider set of internal mental models from which learners can engage in Transfer Appropriate Processing (Lightbown, 2008; Morris et al., 1977) through which they can bridge their prior learning to future applications.

6.4.3 Limitations and future work

While the findings of this study offer valuable insights into the effects of using a mobile microlearning app for peer feedback training in CLT EFL contexts, some limitations need to be acknowledged. First, this study was conducted within the culturally homogeneous context of a Japanese university. The cultural factors influencing peer feedback dynamics and student engagement may differ significantly in more diverse settings. Therefore, future studies could explore the applicability and effectiveness of similar interventions in various cultural contexts to better understand the generalizability of the findings.

Second, the Flipgrid performance platform used in this study allows students to rerecord their videos before submission. While this feature can help reduce anxiety and encourage participation, it does not accurately reflect spontaneous spoken language production. To address this limitation, future research could employ synchronous performance tests, such as the Timed Paired Practice (TPP) tests (Moe, 2005), to measure students' real-time language use in more realistic and spontaneous settings. This approach would provide a more accurate assessment of the transfer of learning from peer feedback training to actual communicative performance.

6.5 Conclusion

In summary, this study has demonstrated the potential of using a mobile microlearning app, Pebasco, to enhance peer feedback skills and facilitate the transfer of these skills to a constructive CLT speaking activity in a Japanese university EFL context. The findings indicate that students who actively engaged with the app, particularly those who started with lower initial performance, showed significant improvements in their post-performance levels, reaching parity with their higher-performing peers. This underscores the importance of active and consistent engagement with microlearning tools for effective learning outcomes.

The study also highlights the complementary nature of peer feedback training and internal feedback skill development. By integrating both the ICAP framework and formative assessment strategies, educators can create a more holistic approach to peer feedback training that addresses both visible and internal cognitive processes. These insights contribute to the growing body of literature on the effectiveness of microlearning apps in language education and offer practical implications for instructors seeking to implement peer feedback training in diverse educational settings.

Chapter 7

Discussion

7.1 Findings Summary

This section addresses the research questions posed in this dissertation by synthesizing the findings across the three studies to show how the integration of peer feedback, learning analytics, and mobile microlearning can enhance EFL learners' communicative competence and feedback processes. The three studies in this dissertation contribute to understanding how LA-driven solutions can improve peer feedback in CLT-focused EFL contexts and help learners transfer feedback skills to communicative tasks (See Table 7.1).

Table 7.1: Summary of findings

Research Question	Study 1	Study 2	Study 3
RQ1: How can CLT be improved through peer feedback?	Pebasco cohort showed greater improvements in communicative skills (mean time gain of 79.24 seconds, F(2/193.74)=165.46, $p<0.001$, $\eta^2=0.546$).	Asynchronous peer feedback training led to significant profile migration, improving L2 error detection (W=327, Z=2.824, p=0.002, r=0.611).	Peer feedback skills transferred to CLT tasks. 3_Low_Pre/HighU4 group outperformed 4_Low_Pre/LowU4 in CQI scores (p=0.022).

Table 7.1: Summary of findings (continued)

Research Question	Study 1	Study 2	Study 3
RQ2: How does peer feedback training impact communicative competence?	Pebasco cohort saw a significant score gain of M=1.48 points in TPP tests, significantly higher than other cohorts (F(2, 207.66)=141.53, p<0.001, η^2=0.38).	Feedback training improved L2 error detection, enhancing internal feedback and communicative competence.	3_Low_Pre/HighU4 group reported higher confidence (M=4.174) than 4_Low_Pre/LowU4 (M=3.333, p=0.005, η^2=0.143), indicating enhanced confidence through peer feedback.
RQ3: How can learning analytics and microlearning platforms enhance peer feedback and language skills?	Pebasco Ver. 1 improved TPP test results, informing the development of Pebasco Ver. 2 with advanced learning analytics.	Learning analytics tracked engagement and profile migration, indicating improved peer feedback skills.	EOC survey showed greater peer feedback improvements for engaged groups (p=0.001, η^2=0.17) and internal feedback skills (p=0.005, η^2=0.158) through learning analytics.

7.1.1 RQ1: How can task-based approaches like Communicative Language Teaching (CLT) be positively impacted through peer feedback to improve language learning outcomes in asynchronous or remote learning environments?

Across the three studies, peer feedback activities consistently improved task-based CLT outcomes, particularly in asynchronous and remote settings.

In Study 1, the Pebasco cohort using the prototype Pebasco system saw significantly greater improvements in communicative skills, with a mean time gain of 79.24 seconds from TPP1 to TPP5, compared to 25.53 seconds for the F2F cohort and 39.19 seconds for the Online cohort. ANOVA results confirmed this difference was statistically significant with a large effect size, F(2/193.74)=165.46, p<0.001, η^2=0.546. These findings suggest that peer feedback, when integrated into CLT, can significantly enhance communicative language outcomes in remote environments.

Study 2 further demonstrated that asynchronous peer feedback training through the Pebasco app helped students improve feedback proficiency. By Unit 4, 59% of students reached the "1_Desired" profile, reflecting higher feedback proficiency. The Wilcoxon signed-rank test showed significant improvement from unit 2 (M=2.402, SD=1.551) to unit 4 (M=2.023, SD=1.438), W=327, Z=2.824, p=0.002, with a large effect size (r_B=0.611). This suggests that peer feedback training enhanced learners' L2 error detection even in asynchronous settings.

In Study 3, peer feedback skills transferred to broader CLT tasks. The 3_Low_Pre/HighU4 group, which started with lower performance but engaged deeply with Pebasco, showed significant gains in their Construction Quality Index (CQI) scores, outperforming the 4_Low_Pre/LowU4 group (p=0.022), and bringing their scores closer to those of initially higher-performing peers. This indicates that peer feedback training can improve CLT outcomes, even for initially lower-performing students.

Across all three studies, peer feedback activities significantly enhanced CLT learning outcomes, evidenced by improved TPP times and scores (Study 1), better L2 error detection (Study 2), and improved performance in CLT tasks (Study 3).

7.1.2 RQ2: How does peer feedback training impact communicative competence?

Peer feedback training significantly improved communicative competence across the three studies.

In Study 1, the introduction of peer feedback in the third phase led to a significant increase in students' communicative competence, measured by TPP test scores, which assess the quality of students' conversations. The Pebasco cohort experienced a mean score gain of M=1.48 points, significantly higher than the F2F cohort (M=0.58) and the Online cohort (M=0.79). ANOVA confirmed these differences were statistically significant with a large effect size (F(2, 207.66)=141.53, p<.001, η^2=0.38), illustrating how peer feedback activities can improve the quality of students' language output.

As noted earlier, Study 2 showed that using the Pebasco app for feedback training improved students' L2 error detection skills, as indicated by profile migration data. As students became more skilled at identifying and correcting errors in others' work, their ability to self-correct also improved through internal feedback, which likely contributed to their overall communicative competence.

In Study 3, peer feedback training not only transferred to broader CLT tasks but also boosted students' confidence in using English. Students in the 3_Low_Pre/HighU4 group, who started with lower performance but engaged deeply with Pebasco, reported significantly higher confidence (M=4.174) compared to the 4_Low_Pre/LowU4 group (M=3.333), who did not engage with Pebasco (p=0.005). ANOVA results showed a large effect size (η^2=0.143), highlighting that peer feedback training positively influenced learners' communicative confidence, particularly for those initially lower-performing students.

Across all three studies, peer feedback training had a positive impact on communicative competence. The results demonstrated improvements in communicative output (Study 1), better error detection and feedback skills (Study 2), and enhanced confidence and competence in broader CLT tasks.

7.1.3 RQ3: In what ways can learning analytics and mobile microlearning platforms be leveraged to enhance peer feedback processes and support language skill development in CLT for EFL education?

The evolution of the Pebasco system, from a prototype in Study 1 to a fully developed app in Studies 2 and 3, illustrates how learning analytics and mobile microlearning platforms can improve peer feedback and language skill development in CLT-based EFL education. In Study 1, the prototype Pebasco system (Ver. 1) was developed using a bricolage approach within an Educational Design Research framework. Although learning analytics were not fully implemented, the system contributed to significant improvements in the Pebasco cohort's TPP test results. These findings informed the development of Pebasco Ver. 2, which addressed the prototype's limitations by integrating advanced learning analytics, including learning logs and user profiling.

Study 2 showed how a proposed set of mobile microlearning design principles (Jahnke et al., 2020) was used to design Pebasco Ver. 2. The microlearning app's learning analytics capabilities were used to track student engagement and identify patterns of usage and improvement in peer feedback skills. As mentioned earlier, there was a statistically significant migration in user profiles, indicating not only an improvement in L2 error identification but also an improvement in peer feedback provisioning skills.

In Study 3, while CQI scores showed how peer feedback skills transferred to broader CLT tasks (as discussed in the RQ1 section above), the EOC survey results highlighted significant improvements in feedback skills. Students who actively engaged with the Pebasco app, such as those in the 1_High_Pre/HighU4 group, reported notably higher improvements in peer feedback skills (M=4.357) compared to the less-engaged 4_Low_Pre/LowU4 group (M=3.0). This difference was statistically significant with a large effect size (p=0.001, η^2=0.17). Similarly, both the 1_High_Pre/HighU4 (M=4.179) and 3_Low_Pre/HighU4 (M=3.957) actively engaged groups reported greater gains in internal feedback skills compared to the 4_Low_Pre/LowU4 group (M=3.0), with another statistically significant difference and large effect size (p=0.005, η^2=0.158). These student metaprofile groups, identified through Pebasco's learning analytics, demonstrate the importance of active engagement with the microlearning app in improving peer and internal feedback skills.

Across the studies, learning analytics and a mobile microlearning platform enhanced peer feedback processes and language skill development. Study 1's prototype informed the development of Pebasco Ver. 2, and Studies 2 and 3 showed how learning analytics-driven insights improved reports of peer and internal feedback skills.

7.2 Contributions

This dissertation contributes to the fields of language education, peer feedback, and learning analytics. It offers insights into the use of LA and technology to improve feedback skills and communicative competence in CLT-focused EFL contexts. These contributions are discussed in relation to each of the RQs.

7.2.1 Contributions to feedback theoretical frameworks (Study 1)

This research contributes to the application of theoretical feedback frameworks, particularly the MISCA model (Panadero & Lipnevich, 2022) and Yang's Feedback Orientation framework (2021). The findings from Study 1 demonstrate how the integration of peer feedback into CLT can improve learning outcomes in remote learning environments. For example, the tightly coupled feedback loop observed in Phase 3 of Study 1 suggests that the demonstrated improvements likely reflect the interplay of all four elements of Yang's (2021) Feedback Orientation Framework: feedback utility, feedback self-efficacy, feedback accountability, and feedback social-awareness.

7.2.2 Contributions to peer feedback and microlearning literature (Study 2)

Study 2 addressed gaps in the peer feedback literature by exploring how a mobile microlearning platform can improve peer feedback provisioning skills in EFL contexts. The use of the Pebasco app demonstrated how structured, LA-supported feedback training can lead to measurable improvements in students' feedback abilities. This research addresses a call for further research in training students' peer feedback skills (Kasch et al., 2021) and a call for research that can investigate the provisioning side of peer feedback independently from the receiving side of it (Sato, 2017).

7.2.3 Contributions to transfer of learning and microlearning engagement (Study 3)

Study 3 contributes to the understanding of how learning analytics and microlearning platforms can facilitate the transfer of peer feedback skills to communicative language tasks. The research supports the notion that mobile microlearning systems can successfully lead to learning transfer, especially for students with initially lower performance who actively engage with the systems. This research adds to the relatively small number of studies that look at the effect of peer feedback on the transfer of learning to other tasks (Yu & Schunn, 2023).

7.3 Implications for instructors

The findings of this dissertation have several implications for instructors, particularly regarding the integration of peer feedback frameworks and mobile microlearning technologies in EFL education. The results highlight the potential for using learning analytics and microlearning platforms to enhance both peer feedback and communicative language development in CLT-based courses.

7.3.1 Implications for feedback practices and theoretical frameworks

The use of comprehensive feedback frameworks such as the MISCA model (Panadero & Lipnevich, 2022) and Yang's Feedback Orientation framework (2021) in this research demonstrates the importance of approaching feedback holistically. By engaging students in structured peer feedback activities, instructors can foster both external feedback skills and internal feedback mechanisms, improving students' communicative competence. This research suggests that applying these frameworks to remote learning environments can lead to meaningful improvements in language learning outcomes that may also be applicable in in-person contexts, too.

7.3.2 Implications for the use of technology in peer feedback

The findings underscore the effectiveness of using no-code platforms and mobile microlearning apps to support peer feedback in EFL contexts. Instructors can build tools like Pebasco to provide asynchronous peer feedback opportunities while using learning analytics to monitor student engagement and progress. The ease of implementing such technologies, even without advanced technical expertise, allows the potential for scalable solutions that can be adapted to various teaching environments. This has significant implications for instructors seeking to integrate technology into their peer feedback processes, as it demonstrates that mobile microlearning platforms can improve peer feedback skills.

7.3.3 Implications for instructional design and learning transfer

The research also provides insights into how instructional design can facilitate the transfer of learning from peer feedback training to related communicative tasks. By incorporating mobile microlearning systems like Pebasco into the curriculum, instructors can help students transfer their feedback skills to broader communicative language tasks. The findings suggest that encouraging students to engage in reflective self-assessment alongside peer feedback activities can enhance their ability to transfer skills across different contexts, thereby supporting long-term language development. This is particularly relevant for students who initially perform at lower levels, as technology-driven feedback tools can help bridge performance gaps.

7.3.4 Implications for pedagogical flexibility and ERT preparedness

This research underscores the importance of pedagogical flexibility, particularly in preparing for future disruptions, such as those caused by the COVID-19 pandemic. The successful implementation of feedback frameworks and mobile learning technologies during Emergency Remote Teaching (ERT) suggests that instructors who incorporate these tools into their regular pedagogy can maintain high levels of student engagement even in disrupted learning environments. By using technology to facilitate peer feedback and integrate feedback processes into both synchronous and asynchronous learning, instructors can better adapt to changing educational contexts while ensuring that students continue to develop essential language skills.

Chapter 8

Conclusion: Limitations and future work

8.1 Limitations of study design and context

A key limitation across the studies is the quasi-experimental design, which did not fully control for individual differences such as student motivation and prior experience with online learning. These factors may have influenced the outcomes. Additionally, the studies were conducted in a culturally homogeneous context (a Japanese university), limiting the generalizability of the findings to more diverse educational settings. Future research should explore similar interventions in a wider range of cultural and linguistic contexts to better understand how peer feedback and mobile microlearning tools function across diverse learner populations.

8.2 Limitations of the technology used

The use of Flipgrid, which allowed students to re-record their videos before submission, may not accurately reflect spontaneous language production. While this feature may reduce student anxiety, it does not provide an authentic assessment of real-time language use. Future research should consider using synchronous CLT performance tasks, such as Timed Paired Practice (TPP) tests, to evaluate the transfer of peer feedback skills to more spontaneous, real-world communicative settings.

8.3 Future research directions

Future research could focus on replicating this work in more culturally diverse settings to test the generalizability of the findings and further explore the impact of peer feedback training in different language learning environments. Expanding the research beyond the Japanese university context could provide valuable insights into how peer feedback dynamics vary across educational settings and learner populations.

Additionally, future studies could investigate the role of individual differences, such as motivation and prior learning experience, in shaping the effectiveness of peer feedback training. As suggested by Complex Dynamic Systems Theory (CDST), individual variability can significantly influence learning outcomes. Future research should explore how different learner profiles respond to technology-supported feedback interventions.

Finally, the development and use of different no-code platforms for building mobile microlearning systems, as well as the potential integration of artificial intelligence (AI), offers promising directions for future exploration. By applying the Pebasco model to other educational fields, such as STEM or the social sciences, researchers could evaluate its broader applicability and examine how technology can support feedback processes across various learning domains.

Acknowledgement

First and foremost, I want to thank my supervisor, Dr. Hiroaki Ogata, for his kind, steady, and unflappable leadership throughout this process, even during the tumultuous disruptions caused by the COVID-19 pandemic. I am also grateful to my advisors, Dr. Adam Jatowt and Dr. Donghui Lin, and to my thesis committee members, Dr. Takayuki Kanda and Dr. Takayuki Ito. Their invaluable feedback and guidance not only shaped my research but also significantly improved this dissertation.

I want to express my sincere appreciation to the people who I met at Kyoto University's Graduate School of Informatics. I have been inspired by many of my colleagues in the Learning and Educational Technologies Research Unit over the years, including Dr. Gokhan Akcapinar, Dr. MeiRong Alice Chen, Dr. Hiroyuki Kuromiya, Dr. Huiyong Li, Dr. Yuko Toyokawa, Dr. Patrick Ocheja, Dr. Changhao Liang, Dr. Brendan Flanagan, Ryosuke Nakamoto, Kohei Nakamura, and Kensuke Takii. Special thanks go to Noriko Nakajima and Keiko Yamakawa for their patience and support. I am also especially grateful to my mentor, co-author, and friend, Dr. Rwitajit Majumdar, whose expertise and good humor bolstered my confidence and resilience throughout this process.

My heartfelt thanks extend to the many teachers, collaborators, and friends who have shaped my thinking and supported me over the years: Curtis Kelly, Robert Murphy, Ai Murphy, Tim Murphey, Marc Helgesen, Joseph Shaules, Louise Ohashi, John Duplice, Caroline Handley, Graham Jones, Douglas Jarrell, Gary Ross, Robert Chartrand, Edo Forsythe, Jonathan Bollick, Mitsuru Ito, Marc Hudson, Karen Brennan, Bertrand Schneider, Bruno della Chiesa, David Rose, Liz Hartmann, Jose Blackorby, Gigi Luk, Timothy McCarthy, Terry Tivnan, Colin Winter, Steve Joy, and Tonya Aiossa. I also wish to thank Tracey Tokuhama-Espinosa, Sharon Tettegah, and Howard Gardner for their insightful advice and guidance.

Finally, at the center of it all, I am supremely grateful to my family for their love, strength, and unwavering support. Thank you for making this possible.

References

Adams, R., Nuevo, A. M., & Egi, T. (2011). Explicit and implicit feedback, modified output, and sla: Does explicit and implicit feedback promote learning and learner–learner interactions? *The Modern Language Journal*, *95*(s1), 42–63. https://doi.org/10.1111/j.1540-4781.2011.01242.x

Alqassab, M., Strijbos, J.-W., & Ufer, S. (2018). Training peer-feedback skills on geometric construction tasks: Role of domain knowledge and peer-feedback levels. *European Journal of Psychology of Education*, *33*(1), 11–30. https://doi.org/10.1007/s10212-017-0342-0

Alshahrani, A. (2016). Communicating authentically: Enhancing efl students' spoken english via videoconferencing. *Computer Assisted Language Learning Electronic Journal*, *17*, 1–17.

Angelova, M., & Zhao, Y. (2016). Using an online collaborative project between american and chinese students to develop esl teaching skills, cross-cultural awareness and language skills. *Computer Assisted Language Learning*, *29*(1), 167–185. https://doi.org/10.1080/09588221.2014.907320

Arakawa, R., Yakura, H., & Kobayashi, S. (2022). Vocabencounter: Nmt-powered vocabulary learning by presenting computer-generated usages of foreign words into users' daily lives. https://doi.org/10.1145/3491102.3501839

Aucejo, E. M., French, J. F., Araya, M. P. U., & Zafar, B. (2020). The impact of covid-19 on student experiences and expectations: Evidence from a survey. *National Bureau of Economic Research Working Paper Series*, *No. 27392*. https://doi.org/10.3386/w27392

Austin, N., Hampel, R., & Kukulska-Hulme, A. (2017). Video conferencing and multimodal expression of voice: Children's conversations using skype for second language development in a telecollaborative setting. *System*, *64*, 87–103. https://doi.org/10.1016/j.system.2016.12.003

Avgousti, M. I. (2018). Intercultural communicative competence and online exchanges: A systematic review. *Computer Assisted Language Learning*, *31*(8), 819–853. https://doi.org/10.1080/09588221.2018.1455713

Bakker, A. (2018). *Design research in education. a practical guide for early career researchers*. Routledge.

Banister, C. (2023). Exploring peer feedback processes and peer feedback meta-dialogues with learners of academic and business english. *Language Teaching Research*, *27*(3), 746–764. https://doi.org/10.1177/1362168820952222

Bao, W. (2020). Covid-19 and online teaching in higher education: A case study of peking university. *Human Behavior and Emerging Technologies*, *2*(2), 113–115. https://doi.org/10.1002/hbe2.191

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin*, *128*(4), 612–637. https://doi.org/10.1037/0033-2909.128.4.612

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences*, *2*(2), 141–178. https://doi.org/10.1207/s15327809jls0202_2

Bubble. (2022). Bubble. https://bubble.io

Buildcamp. (2022). Buildcamp. https://buildcamp.io/

Cai, C. J., Guo, P. J., Glass, J. R., & Miller, R. C. (2015). Wait-learning: Leveraging wait time for second language education. https://doi.org/10.1145/2702123.2702267

Cai, C. J., Ren, A., & Miller, R. C. (2017). Waitsuite: Productive use of diverse waiting moments. *ACM Trans. Comput.-Hum. Interact.*, *24*(1), Article 7. https://doi.org/10.1145/3044534

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, *43*(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354

Çelik, S., Baran, E., & Sert, O. (2018). The affordances of mobile-app supported teacher observations for peer feedback. *International Journal of Mobile and Blended Learning (IJMBL)*, *10*(2), 36–49. https://doi.org/10.4018/IJMBL.2018040104

Chen, J. (2024). Effects of learning analytics-based feedback on students' self-regulated learning and academic achievement in a blended efl course. *System*, *124*, 103388. https://doi.org/10.1016/j.system.2024.103388

Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, *1*(1), 73–105. https://doi.org/10.1111/j.1756-8765.2008.01005.x

Chi, M. T. H., Adams, J., Bogusch, E. B., Bruchok, C., Kang, S., Lancaster, M., Levy, R., Li, N., McEldoon, K. L., Stump, G. S., Wylie, R., Xu, D., & Yaghmourian, D. L. (2018). Translating the icap theory of cognitive engagement into practice. *Cognitive Science*, *42*(6), 1777–1832. https://doi.org/10.1111/cogs.12626

Chi, M. T. H., & Wylie, R. (2014). The icap framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Chong, S. W., & Reinders, H. (2020). Technology-mediated task-based language teaching: A qualitative research synthesis. *Language Learning & Technology*, *24*(3), 70–86. http://hdl.handle.net/10125/44739

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)* L. Erlbaum Associates.

Collins, A. (1992). Toward a design science of education. In E. Scanlon & T. O'Shea (Eds.), *New directions in educational technology* (pp. 15–22). Springer Berlin Heidelberg.

Dao, P., & Iwashita, N. (2021). Peer feedback in l2 oral interaction. In H. Nassaji & E. Kartchava (Eds.), *The cambridge handbook of corrective feedback in second language learning and teaching* (pp. 275–300). Cambridge University Press.

Darvishi, A., Khosravi, H., Sadiq, S., & Gašević, D. (2022). Incorporating ai and learning analytics to build trustworthy peer assessment systems. *British Journal of Educational Technology*, *53*(4), 844–875. https://doi.org/10.1111/bjet.13233

Dearman, D., & Truong, K. (2012). Evaluating the implicit acquisition of second language vocabulary using a live wallpaper. https://doi.org/10.1145/2207676.2208598

Dingler, T., Weber, D., Pielot, M., Cooper, J., Chang, C.-C., & Henze, N. (2017). Language learning on-the-go: Opportune moments and design of mobile microlearning sessions. https://doi.org/10.1145/3098279.3098565

Dörnyei, Z. (2009). *The psychology of second language acquisition*. Oxford University Press.

Double, K. S., McGrane, J. A., & Hopfenbeck, T. N. (2020). The impact of peer assessment on academic performance: A meta-analysis of control group studies. *Educational Psychology Review*, *32*(2), 481–509. https://doi.org/10.1007/s10648-019-09510-3

Drakidou, C., & Panagiotidis, P. (2018). *The use of micro-learning for the acquisition of language skills in mlll context*.

Duval, E. (2011). Attention please! learning analytics for visualization and recommendation. https://doi.org/10.1145/2090116.2090118

Easterday, M. W., Rees Lewis, D. G., & Gerber, E. M. (2018). The logic of design research. *Learning: Research and Practice*, *4*(2), 131–160. https://doi.org/10.1080/23735082.2017.1286367

Edge, D., Searle, E., Chiu, K., Zhao, J., & Landay, J. A. (2011). Micromandarin: Mobile language learning in context. https://doi.org/10.1145/1978942.1979413

Education First. (n.d.). Ef english proficiency index: Japan 2020. https://www.ef.com/wwen/epi/regions/asia/japan/

Einblick. (2022). Byod: Bring your own data with thomas gorham, assistant professor at rissho university, tokyo, japan.

El Mortaji, L. (2022). Public speaking and online peer feedback in a blended learning efl course environment: Students' perceptions. *English Language Teaching*, *15*(2). https://doi.org/10.5539/elt.v15n2p31

Elam, J. R. (2014). Corrective feedback, uptake and repair in tpp (time-paired practice). *Journal of Global Media Studies*, *14*, 17–30. http://gmsweb.komazawa-u.ac.jp/wp-content/uploads/2014/04/h14-2.pdf

Er, E., Dimitriadis, Y., & Gašević, D. (2021). Collaborative peer feedback and learning analytics: Theory-oriented design for supporting class-wide interventions. *Assessment & Evaluation in Higher Education*, *46*(2), 169–190. https://doi.org/10.1080/02602938.2020.1764490

Flourish. (2022). https://flourish.studio/

Fujii, A., Ziegler, N., & Mackey, A. (2016). Peer interaction and metacognitive instruction in the efl classroom. In M. Sato & S. Ballinger (Eds.), *Peer interaction and second language learning: Pedagogical potential and research agenda* (pp. 63–90, Vol. 45). John Benjamins.

Garshasbi, S., Yecies, B., & Shen, J. (2021). Microlearning and computer-supported collaborative learning: An agenda towards a comprehensive online learning system. *STEM Education*, *1*(4), 225–255. https://doi.org/10.3934/steme.2021016

Gentner, D., Holyoak, K. J., & Kokinov, B. N. (2001). *The analogical mind: Perspectives from cognitive science*. The MIT Press.

Goldstone, R. L., Day, S., & Son, J. Y. (2010). Comparison. In B. Glatzeder, V. Goel, & A. Müller (Eds.), *Towards a theory of thinking: Building blocks for a conceptual framework* (pp. 103–121). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-03129-8_7

Gorham, T., Majumdar, R., & Ogata, H. (2023). Analyzing learner profiles in a microlearning app for training language learning peer feedback skills. *Journal of Computers in Education*, *10*(3), 549–574. https://doi.org/10.1007/s40692-023-00264-0

Gorham, T., Majumdar, R., & Ogata, H. (2024). Learning analytics of peer feedback on communicative skills in an efl course across different learning modalities. *Studies in Educational Evaluation*, *81*, 101352. https://doi.org/10.1016/j.stueduc.2024.101352

Gorham, T., & Ogata, H. (2020a). Improving skills for peer feedback on spoken content using an asynchronous learning analytics app. *28th International Conference on Computers in Education Conference Proceedings*, (2), 782–785. https://ci.nii.ac.jp/naid/120006940277/en/

Gorham, T., & Ogata, H. (2020b). Professional learning community's views on accessibility during emergency remote teaching. *28th International Conference on Computers in Education Conference Proceedings*, (1), 570–572. https://ci.nii.ac.jp/naid/120006940280/en/

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Hattie, J. (2012). Feedback in schools. In R. Sutton, M. J. Hornsey, & K. M. Douglas (Eds.), *Feedback: The communication of praise, criticism, and advice* (pp. 265–278). Peter Lang Publishing.

Hattie, J., & Clarke, S. (2019). *Visible learning: Feedback*. Routledge.

Heffernan, N. T., & Heffernan, C. L. (2014). The assistments ecosystem: Building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. *International Journal of Artificial Intelligence in Education*, *24*(4), 470–497. https://doi.org/10.1007/s40593-014-0024-x

Hiver, P., & Al-Hoorie, A. H. (2019). *Research methods for complexity theory in applied linguistics* (Vol. 137). Multilingual Matters.

Hodges, C., Moore, S., Lockee, B., Trust, T., & Bond, A. (2020). The difference between emergency remote teaching and online learning. https://er.educause.edu/articles/2020/3/the-difference-between-emergency-remote-teaching-and-online-learning

Hofstadter, D. R., & Sander, E. (2013). *Surfaces and essences: Analogy as the fuel and fire of thinking*. Basic Books.

Hug, T., Lindner, M., & Bruck, P. (2006). *Microlearning: Emerging concepts, practices and technologies after e-learning*. Innsbruck University Press. https://www.researchgate.net/publication/246822097_Microlearning_Emerging_Concepts_Practices_and_Technologies_after_e-Learning

Inie, N., & Lungu, M. F. (2021). Aiki - turning online procrastination into microlearning. https://doi.org/10.1145/3411764.3445202

Jahnke, I., Lee, Y.-M., Pham, M., He, H., & Austin, L. (2020). Unpacking the inherent design principles of mobile microlearning. *Technology, Knowledge and Learning*, *25*(3), 585–619. https://doi.org/10.1007/s10758-019-09413-w

Johannessen, C., & Davenport, T. (2021). When low-code/no-code development works — and when it doesn't. https://hbr.org/2021/06/when-low-code-no-code-development-works-and-when-it-doesnt

Jongsma, M. V., Scholten, D. J., van Muijlwijk-Koezen, J. E., & Meeter, M. (2023). Online versus offline peer feedback in higher education: A meta-analysis. *Journal of Educational Computing Research*, *61*(2), 329–354. https://doi.org/10.1177/07356331221114181

Kasch, J., van Rosmalen, P., Löhr, A., Klemke, R., Antonaci, A., & Kalz, M. (2021). Students' perceptions of the peer-feedback experience in moocs. *Distance Education*, *42*(1), 145–163. https://doi.org/10.1080/01587919.2020.1869522

Kerr, P. (2020). Giving feedback to language learners: Part of the cambridge papers in elt series. https://www.cambridge.org/gb/files/4415/8594/0876/Giving_Feedback_minipaper_ONLINE.pdf

Khan, R., Jahan, A., Sultana, S., NaushaadKabir, M. M., Haider, M. Z., & Roshid, M. M. (2021). Accessing online instruction amidst covid-19 in bangladesh: Barriers and coping strategies. *Language Teaching Research Quarterly*, *22*, 33–48.

Khong, H. K., & Kabilan, M. K. (2022). A theoretical model of micro-learning for second language instruction. *Computer Assisted Language Learning*, *35*(7), 1483–1506. https://doi.org/10.1080/09588221.2020.1818786

Knight, S., Leigh, A., Davila, Y. C., Martin, L. J., & Krix, D. W. (2019). Calibrating assessment literacy through benchmarking tasks. *Assessment & Evaluation in Higher Education*, *44*(8), 1121–1132. https://doi.org/10.1080/02602938.2019.1570483

Kovacs, G. (2015). Feedlearn: Using facebook feeds for microlearning. https://doi.org/10.1145/2702613.2732775

Layton, R. A., Loughry, M. L., Ohland, M. W., & Pomeranz, H. (2012). Workshop: Training students to become better raters: Raising the quality of self- and peer-evaluations using a new feature of the catme system. *2012 Frontiers in Education Conference Proceedings*, 1–2. https://doi.org/10.1109/FIE.2012.6462267

Le Cor, G., & Coutherut, M. (2020). Online courses in times of pandemic: Esp and applied english classes at université paris 8. *ASp*, *78*, 7–17. https://doi.org/10.4000/asp.6481

Lee, Y.-M., Jahnke, I., & Austin, L. (2021). Mobile microlearning design and effects on learning efficacy and learner experience. *Educational Technology Research and Development*, *69*(2), 885–915. https://doi.org/10.1007/s11423-020-09931-w

Lévi-Strauss, C. (1962). *The savage mind*. University of Chicago Press.

Liang, C., Gorham, T., Horikoshi, I., Majumdar, R., & Ogata, H. (2022). Estimating peer evaluation potential by utilizing learner model during group work. In L.-H. Wong, Y. Hayashi, C. A. Collazos, C. Alvarez, G. Zurita, & N. Baloian (Eds.), *Collaboration technologies and social computing* (pp. 287–294). Springer International Publishing.

Liang, C., Toyokawa, Y., Majumdar, R., Horikoshi, I., & Ogata, H. (2024). Group formation based on reading annotation data: System innovation and classroom practice.

Journal of Computers in Education, *11*(3), 667–695. https://doi.org/10.1007/s40692-023-00274-y

Lightbown, P. M. (2008). Transfer appropriate processing as a model for classroom second language acquisition. In Z. Han (Ed.), *Understanding second language process* (pp. 38–80). Multilingual Matters. https://doi.org/10.21832/9781847690159-005

Lin, C.-J., & Hwang, G.-J. (2018). A learning analytics approach to investigating factors affecting efl students' oral performance in a flipped classroom. *Journal of Educational Technology & Society*, *21*(2), 205–219. http://www.jstor.org/stable/26388398

Liu, N.-F., & Carless, D. (2006). Peer feedback: The learning element of peer assessment. *Teaching in Higher Education*, *11*(3), 279–290. https://doi.org/10.1080/13562510600680582

Loughry, M. L., Ohland, M. W., & Woehr, D. J. (2014). Assessing teamwork skills for assurance of learning using catme team tools. *Journal of Marketing Education*, *36*(1), 5–19. https://doi.org/10.1177/0273475313499023

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, J., Ly, A., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Wild, A., Knight, P., Rouder, J. N., Morey, R. D., & Wagenmakers, E.-J. (2019). Jasp: Graphical statistical software for common statistical designs. *Journal of Statistical Software; Vol 1, Issue 2 (2019)*. https://doi.org/10.18637/jss.v088.i02

Lu, J., & Law, N. (2012). Online peer assessment: Effects of cognitive and affective feedback. *Instructional Science*, *40*(2), 257–275. https://doi.org/10.1007/s11251-011-9177-2

Lui, A. M., & Andrade, H. L. (2022). Inside the next black box: Examining students' responses to teacher feedback in a formative assessment context. *Frontiers in Education*, *7*. https://doi.org/10.3389/feduc.2022.751549

Luporini, A. (2020). Implementing an online english linguistics course during the covid-19 emergency in italy: Teacher's and students' perspectives. *ASp*, *78*, 75–88. https://doi.org/10.4000/asp.6682

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

Ma, B., Hettiarachchi, G. P., Fukui, S., & Ando, Y. (2023). Each encounter counts: Modeling language learning and forgetting. https://doi.org/10.1145/3576050.3576062

Majumdar, R., Bakilapadavu, G., Majumder, R., Chen, M.-R. A., Flanagan, B., & Ogata, H. (2021). Learning analytics of humanities course: Reader profiles in critical reading activity. *Research and Practice in Technology Enhanced Learning*, *16*(1), 25. https://doi.org/10.1186/s41039-021-00164-w

Majumdar, R., & Iyer, S. (2016). Isat: A visual learning analytics tool for instructors. *Research and Practice in Technology Enhanced Learning*, *11*(1), 16. https://doi.org/10.1186/s41039-016-0043-3

McDaniel, C., Suffern, C., Joo, J., & Alamuddin, R. (2020). Student and faculty experiences with emergency remote learning in spring 2020. https://doi.org/10.18665/sr.314276

McKenney, S., & Reeves, T. (2018). *Conducting educational design research* (2nd ed.). Routledge.

Means, B., & Neisler, J. (2020). *Suddenly online: A national survey of undergraduates during the covid-19 pandemic* (Report). Digital Promise.

Meek, S. E. M., Blakemore, L., & Marks, L. (2017). Is peer review an appropriate form of assessment in a mooc? student participation and performance in formative peer review. *Assessment & Evaluation in Higher Education*, *42*(6), 1000–1013. https://doi.org/10.1080/02602938.2016.1221052

MEXT. (n.d.). Report on future improvement and enhancement measures for english education-five proposals for reform of english education in response to globalization. https://www.mext.go.jp/b_menu/shingi/chousa/shotou/102/houkoku/attach/1352464.htm

Misiejuk, K., & Wasson, B. (2023). Learning analytics for peer assessment: A scoping review. In O. Noroozi & B. De Wever (Eds.), *The power of peer learning: Fostering students' learning processes and outcomes* (pp. 25–46). Springer International Publishing. https://doi.org/10.1007/978-3-031-29411-2_2

Misiejuk, K., Wasson, B., & Egelandsdal, K. (2021). Using learning analytics to understand student perceptions of peer feedback. *Computers in Human Behavior*, *117*, 106658. https://doi.org/10.1016/j.chb.2020.106658

Moe, R. (2005). Teaching and evaluating conversations. *Journal of Komazawa Junior College*, *33*, 195–210.

Moreno-Marcos, P. M., Alario-Hoyos, C., Muñoz-Merino, P. J., & Kloos, C. D. (2019). Prediction in moocs: A review and future research directions. *IEEE Transactions on Learning Technologies*, *12*(3), 384–401. https://doi.org/10.1109/TLT.2018.2856808

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*(5), 519–533. https://doi.org/10.1016/S0022-5371(77)80016-9

Motallebzadeh, K., Kondori, A., & Kazemi, S. (2020). The effect of peer feedback on efl learners' classroom anxiety. *Language Teaching Research Quarterly*, *18*, 40–52.

Nartiningrum, N., & Nugroho, A. (2020). Online learning amidst global pandemic: Efl students' challenges, suggestions, and needed materials. *ENGLISH FRANCA Academic Journal of English Language and Education*, *4*, 115–140. https://doi.org/10.29240/ef.v4i2.1494

Nicol, D. (2021). The power of internal feedback: Exploiting natural comparison processes. *Assessment & Evaluation in Higher Education*, *46*(5), 756–778. https://doi.org/10.1080/02602938.2020.1823314

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: A peer review perspective. *Assessment & Evaluation in Higher Education*, *39*(1), 102–122. https://doi.org/10.1080/02602938.2013.795518

Ogata, H., Flanagan, B., Takami, K., Dai, Y., Nakamoto, R., & Takii, K. (2024). Exait: Educational explainable artificial intelligent tools for personalized learning. *Research and Practice in Technology Enhanced Learning*, *19*, 019. https://doi.org/10.58459/rptel.2024.19019

Ogata, H., Li, H., Majumdar, R., Toyokawa, Y., Takii, K., & Liang, C. (2023). Towards data and evidence-driven education in the context of language teaching and learning. *JACET International Convention Selected Papers*, *9*, 15–46. https://doi.org/10.50943/jacetselectedpapers.9.0_15

Ogata, H., Majumdar, R., Akçapınar, G., Hasnine, M. N., & Flanagan, B. (2018). Beyond learning analytics: Framework for technology-enhanced evidence-based education and learning. *Procs. of ICCE 2018*, *2*, 486–489.

Ogata, H., Majumdar, R., Yang, S. J. H., & Warriem, J. M. (2022). Learning and evidence analytics framework (leaf): Research and practice in international collaboration. *Information and Technology in Education and Learning*, *2*(1), Inv-p001-Inv–p001. https://doi.org/10.12937/itel.2.1.Inv.p001

Ohland, M. W., Loughry, M. L., Woehr, D. J., Bullard, L. G., Felder, R. M., Finelli, C. J., Layton, R. A., Pomeranz, H. R., & Schmucker, D. G. (2012). The comprehensive assessment of team member effectiveness: Development of a behaviorally anchored rating scale for self- and peer evaluation. *Academy of Management Learning & Education*, *11*(4), 609–630. https://doi.org/10.5465/amle.2010.0177

Ouhaichi, H., Spikol, D., & Vogel, B. (2023). Research trends in multimodal learning analytics: A systematic mapping study. *Computers and Education: Artificial Intelligence*, *4*, 100136. https://doi.org/10.1016/j.caeai.2023.100136

Panadero, E., & Alqassab, M. (2019). An empirical review of anonymity effects in peer assessment, peer feedback, peer review, peer evaluation and peer grading. *Assessment & Evaluation in Higher Education*, *44*(8), 1253–1278. https://doi.org/10.1080/02602938.2019.1600186

Panadero, E., & Lipnevich, A. A. (2022). A review of feedback models and typologies: Towards an integrative model of feedback elements. *Educational Research Review*, *35*, 100416. https://doi.org/10.1016/j.edurev.2021.100416

Patri, M. (2002). The influence of peer feedback on self- and peer-assessment of oral skills. *Language Testing*, *19*(2), 109–131. https://doi.org/10.1191/0265532202lt224oa

Pawlak, M. (2022). Research into individual differences in sla and call: Looking for intersections. *Language Teaching Research Quarterly*, *31*, 200–233.

Philp, J., Walter, S., & Basturkmen, H. (2010). Peer interaction in the foreign language classroom: What factors foster a focus on form? *Language Awareness*, *19*(4), 261–279. https://doi.org/10.1080/09658416.2010.516831

Pipe, J. (2015). Framework to improve spoken discourse in task-based learning. *The Kyoritsu Journal of Arts and Letters*, *61*, 45–52.

ProNoCoders. (2022). Pro nocoders. https://pronocoders.com/

Reith-Hall, E., & Montgomery, P. (2023). Communication skills training for improving the communicative abilities of student social workers: A systematic review. *Campbell Systematic Reviews*, *19*(1), e1309. https://doi.org/10.1002/cl2.1309

Richards, J. C. (2006). *Communicative language teaching today*. Oxford University Press.

Rodríguez-González, E., & Castañeda, M. E. (2016). The effects and perceptions of trained peer feedback in l2 speaking: Impact on revision and speaking quality. *Innovation in Language Learning and Teaching*, *12*(2), 120–136. https://doi.org/10.1080/17501229.2015.1108978

Sánchez-Auñón, E., Férez-Mora, P. A., & Monroy-Hernández, F. (2023). The use of films in the teaching of english as a foreign language: A systematic literature review.

Asian-Pacific Journal of Second and Foreign Language Education, *8*(1), 10. https://doi.org/10.1186/s40862-022-00183-0

Sato, M., & Ballinger, S. (2016). Understanding peer interaction: Research synthesis and directions. In M. Sato & S. Ballinger (Eds.), *Peer interaction and second language learning: Pedagogical potential and research agenda* (pp. 1–30, Vol. 45). John Benjamins.

Sato, M. (2013). Beliefs about peer interaction and peer corrective feedback: Efficacy of classroom intervention. *The Modern Language Journal*, *97*(3), 611–633. https://doi.org/10.1111/j.1540-4781.2013.12035.x

Sato, M. (2017). Oral peer corrective feedback: Multiple theoretical perspectives. In H. Nassaji & E. Kartchava (Eds.), *Corrective feedback in second language teaching and learning: Research, theory, applications, implications* (pp. 19–34). Routledge. https://doi.org/10.4324/9781315621432

Sato, M., & Lyster, R. (2012). Peer interaction and corrective feedback for accuracy and fluency development: Monitoring, practice, and proceduralization. *Studies in Second Language Acquisition*, *34*(4), 591–626. https://doi.org/10.1017/S0272263112000356

Sauvignon, S. (2005). Communicative language teaching: Strategies and goals. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 635–651). Routledge.

Scherer, S., Worsley, M., & Morency, L.-P. (2012). 1st international workshop on multimodal learning analytics: Extended abstract. https://doi.org/10.1145/2388676.2388803

Schneegass, C., Sigethy, S., Mitrevska, T., Eiband, M., & Buschek, D. (2022). Unlocklearning – investigating the integration of vocabulary learning tasks into the smartphone authentication process. *i-com*, *21*(1), 157–174. https://doi.org/doi:10.1515/icom-2021-0037

Shahrokhi Mehr, H., Zoghi, M., & Assadi, N. (2013). Effects of synchronous computer-mediated communication and face-to-face interaction on speaking skill development of iranian efl learners. *2013*, *2*(5), 6. https://doi.org/10.7575/aiac.ijalel.v.2n.5p.36

Siemens, G., & Baker, R. S. J. d. (2012). Learning analytics and educational data mining: Towards communication and collaboration. https://doi.org/10.1145/2330601.2330661

Sluijsmans, D. M. A., Brand-Gruwel, S., van Merriënboer, J. J. G., & Martens, R. L. (2004). Training teachers in peer-assessment skills: Effects on performance and perceptions. *Innovations in Education and Teaching International*, *41*(1), 59–78. https://doi.org/10.1080/1470329032000172720

Sousa, D. A. (2022). *How the brain learns* (6th). Corwin Press.

Sridharan, B., Tai, J., & Boud, D. (2019). Does the use of summative peer assessment in collaborative group work inhibit good judgement? *Higher Education*, *77*(5), 853–870. https://doi.org/10.1007/s10734-018-0305-7

Sutton, P. (2012). Conceptualizing feedback literacy: Knowing, being, and acting. *Innovations in Education and Teaching International*, *49*(1), 31–40. https://doi.org/10.1080/14703297.2012.647781

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*(4), 295–312. https://doi.org/10.1016/0959-4752(94)90003-5

Sweller, J. (2020). Cognitive load theory and educational technology. *Educational Technology Research and Development*, *68*(1), 1–16. https://doi.org/10.1007/s11423-019-09701-3

Takii, K., Flanagan, B., & Ogata, H. (2021). Efl vocabulary learning using a learning analytics-based e-book and recommender platform. *2021 International Conference on Advanced Learning Technologies (ICALT)*, 254–256. https://doi.org/10.1109/ICALT52272.2021.00082

Terhune, N. M. (2016). Language learning going global: Linking teachers and learners via commercial skype-based cmc. *Computer Assisted Language Learning*, *29*(6), 1071–1089. https://doi.org/10.1080/09588221.2015.1061020

The Japan Times. (2021). Tokyo enters fourth covid-19 state of emergency as olympics loom. https://www.japantimes.co.jp/news/2021/07/12/national/fourth-coronavirus-emergency-tokyo/

Thurn, C. M., Edelsbrunner, P. A., Berkowitz, M., Deiglmayr, A., & Schalk, L. (2023). Questioning central assumptions of the icap framework. *npj Science of Learning*, *8*(1), 49. https://doi.org/10.1038/s41539-023-00197-4

To, J., & Panadero, E. (2019). Peer assessment effects on the self-assessment process of first-year undergraduates. *Assessment & Evaluation in Higher Education*, *44*(6), 920–932. https://doi.org/10.1080/02602938.2018.1548559

Topping, K. J. (2013). Peers as a source of formative and summative assessment. In *Sage handbook of research on classroom assessment* (pp. 395–412). Sage Publications.

Toth, P. D. (2008). Teacher- and learner-led discourse in task-based grammar instruction: Providing procedural assistance for l2 morphosyntactic development. *Language Learning*, *58*(2), 237–283. https://doi.org/10.1111/j.1467-9922.2008.00441.x

Trusty, A., & Truong, K. N. (2011). Augmenting the web for second language vocabulary learning. https://doi.org/10.1145/1978942.1979414

van den Bogaard, M. E. D., & Saunders-Smits, G. N. (2007). Peer & self evaluations as means to improve the assessment of project based learning. *2007 37th Annual Frontiers In Education Conference - Global Engineering: Knowledge Without Borders, Opportunities Without Passports*, S1G-12-S1G–18. https://doi.org/10.1109/FIE.2007.4417988

van Zundert, M., Sluijsmans, D., & van Merriënboer, J. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction*, *20*(4), 270–279. https://doi.org/10.1016/j.learninstruc.2009.08.004

Wu, J. G., & Miller, L. (2020). Improving english learners' speaking through mobile-assisted peer feedback. *RELC Journal*, *51*(1), 168–178. https://doi.org/10.1177/0033688219895335

Wu, Y., & Schunn, C. D. (2023). Passive, active, and constructive engagement with peer feedback: A revised model of learning from peer feedback. *Contemporary Educational Psychology*, *73*, 102160. https://doi.org/10.1016/j.cedpsych.2023.102160

Yang, L. (2021). The role of feedback orientation in converting external feedback to learning opportunities for implementing assessment-as-learning in the context of feedback. In Z. Yan & L. Yang (Eds.), *Assessment as learning* (pp. 53–75). Routledge. https://doi.org/10.4324/9781003052081-5

Yeh, H.-C., & Lai, W.-Y. (2019). Speaking progress and meaning negotiation processes in synchronous online tutoring. *System*, *81*, 179–191. https://doi.org/10.1016/j.system.2019.01.001

Yu, Q., & Schunn, C. D. (2023). Understanding the what and when of peer feedback benefits for performance and transfer. *Computers in Human Behavior*, *147*, 107857. https://doi.org/10.1016/j.chb.2023.107857

Zgraggen, E., Galakatos, A., Crotty, A., Fekete, J., & Kraska, T. (2017). How progressive visualizations affect exploratory analysis. *IEEE Transactions on Visualization and Computer Graphics*, *23*(8), 1977–1987. https://doi.org/10.1109/TVCG.2016.2607714

Zhao, H., Liu, J., Wu, J., Yao, K., & Huang, J. (2018). Watch-learning: Using the smartwatch for secondary language vocabulary learning. https://doi.org/10.1145/3202667.3204037

Zhou, Y., Andres-Bray, J. M., Hutt, S., Ostrow, K., & Baker, R. S. (2021). A comparison of hints vs. scaffolding in a mooc with adult learners. In I. Roll, D. McNamara, S. Sosnovsky, R. Luckin, & V. Dimitrova (Eds.), *Artificial intelligence in education. aied 2021. lecture notes in computer science* (Vol. 12749). Springer International Publishing. https://doi.org/10.1007/978-3-030-78270-2_76

Zong, Z., Schunn, C. D., & Wang, Y. (2021). What aspects of online peer feedback robustly predict growth in students' task performance? *Computers in Human Behavior*, *124*, 106924. https://doi.org/10.1016/j.chb.2021.106924