【Research Brief】

Designing a Mobile Application to Track Spoken Fluency Development

Heather Woodward & Charu Gupta

Abstract

English discussion class (EDC) course designers state that the most important language-based objective of EDC is to improve students' spoken fluency (Hurling, 2012). In general, the students do seem to be increasing their speaking speed; however, designers have not developed a method to document their progress across the semester. Meanwhile, mobile-assisted language learning (MALL) applications can enhance personalization and feedback, but these applications have been traditionally designed to build students' second language (L2) vocabulary rather than their L2 fluency (Heil et al., 2016). In Spring 2022, we repurposed an existing mobile application that analyzes speech rate (words per minute) to provide more personalized feedback. The problem with the application is that it does not have features that reflect current spoken fluency research. Therefore, we review the literature on spoken fluency research, then design an application that collects and records students' speech rate (syllables per minute), mid-clause pause frequency, self-assessment, and peer-assessment.

Keywords: *EFL, MALL, speaking, 4/3/2*

Introduction

The goal of language learning for many Japanese students is the ability to speak fluently; however, achieving fluency is challenging in the context of Japan where there are few chances for students to use the foreign language owing in part to the washback effects from university entrance exams, which have historically not included a speaking component (Garside, 2020). Japanese students also strive for accuracy and perfection, so they spend time correcting their syntactic or phonetic mistakes and this extra time can decrease their spoken fluency (Watanabe & Long, 2019). Without incorporating fluency-building activities into the L2 classroom, communicative language teaching is limited in scope (Gatbonton & Segalowitz, 1988), and as a result, students' learning might not be available during the typical demands of real-world communication (Hurling, 2012). Not meeting typical demands of real-world communication means that students are unable to hold listeners' attention or save face (Lennon, 2000).

Applied linguistics researchers analyze three aspects of spoken fluency – speed, breakdown, and repair. The first aspect, speed, equates to the density and flow of speech. The second aspect, breakdown, consists of hesitations and pauses, and the last aspect, repair, comprises corrections, reformations, and repetitions (Skehan, 2003; Tavakoli & Skehan, 2005). For example, researchers might calculate the speed of the L2 spoken performance by using the average number of syllables per minute, then analyze the breakdowns of an L2 spoken performance by calculating the mean length of pauses. The process of transcribing and analyzing students' L2 spoken performances is too time consuming to be widely incorporated by second language (L2) instructors in the classroom. In addition, conditions such as background noises of other students speaking make collecting and analyzing data very challenging.

However, quantitative measures such as the average number of syllables per minute are limited

in describing spoken fluency performance because they are unable to account for other mediating factors such as students' first language (L1) speaking styles, personalities, or socio-pragmatic considerations (Tavakoli & Wright, 2020). One possibility is to create a mobile application that incorporates both quantitative measurements and qualitative judgments while also adding instructional features to enhance students' metalinguistic awareness of spoken fluency to provide more personalized feedback. By using a filter, the background noises of other students in the classroom might be reduced enough to collect information about students' spoken fluency performance. In this research brief, we first review literature on the goals of English discussion class (EDC). Afterward, we investigate research on applied linguistics to determine the most appropriate objective measurements to use to measure spoken fluency. We also discuss gamification and personalization, and then present a mobile application design that includes these features.

Literature Review

English Discussion Class

EDC is required for first-year students at Rikkyo University to improve their communication skills, academic discussion skills, and spoken fluency (Hurling, 2012). Examples of communication skills along with a corresponding formulaic sequence include clarification (Can you repeat that?), comprehension (Do you understand?), and paraphrasing (In other words, do you mean…?) (Kita et al., 2022). Academic discussion skills include joining a discussion (Can I start?), viewpoints (How about from the perspective of...?), and sources of information (How do you know about that?) (Kita et al., 2022). Students are separated into four EDC levels according to their test of English for international communication (TOEIC) scores. The four EDC levels are: Level 1 (TOEIC scores 680 or above; common European framework of reference for languages (CEFR): B2 and above), Level 2 (480-679; B1-B2), Level 3 (280-479; A2-B1), and Level 4 (279 or below; A1-A2). Additionally, the course is for 14 weeks, has 10 students per class, and students meet once a week for 100 minutes.

During each lesson, five students participate in a 20-minute discussion while instructors assess students' participation and use of formulaic sequences (English Discussion Committee Handbook, 2022). On the rubric, students receive a 4, or superior, for academic discussion skills if they use them quickly, appropriately, and without looking in the textbook. Quickness relates to the spoken fluency's aspects of speed and breakdown. They receive a 3, or good, if students use discussion skills, but cannot do so from memory (English Discussion Committee Syllabus, 2022). The other scale descriptors of none, poor, and fair, do not explicitly refer to students' spoken fluency. In addition, the two other constructs assessed—communication skills and participation—also do not explicitly refer to spoken fluency. However, speaking quickly during the discussion performance test gives others time to participate and use discussion and communication skills. The rubric is the same for all EDC levels (i.e., from Level 1 to 4). Although spoken fluency is the most important language-based goal of EDC according to course developers, the only fluency assessment of students consists of the extent to which instructors perceive their fluency of discussion skills in the context of group discussion performance.

4/3/2 Activity and Other Fluency Practices in EDC

A modified version of Maurice's (1983) 4/3/2 activity called 3/2/1 is used during every EDC

lesson. 3/2/1 incorporates task features of repetition and time pressure, and these task features have been shown to boost students' spoken fluency (e.g., Arevat & Nation, 1991; Boers, 2014; Garside, 2020; de Jong & Perfetti, 2011; Molina Barriga & Briesmaster, 2017; Thai & Boers, 2016; Tran & Saito, 2021). For this modified version of 4/3/2, students deliver three iterations of their monologues. Speakers change listeners after each iteration. 3/2/1 takes 15 minutes and with 10 students per class, 5 students speak simultaneously. For theoretical support, course designers cite Schmidt (1992) who uses Anderson's (1989) adaptive control of thought (ACT) theory (Hurling, 2012). In ACT, fluency equates to automatic processing of proceduralized knowledge (i.e., knowing how to use the L2 with effortless efficiency) and after repeated, meaningful practice, in theory, students can retrieve knowledge more quickly and smoothly than before.

Not only do students use 4/3/2 in EDC to improve spoken fluency, but course designers have also adopted other practices to help students boost spoken fluency. EDC students a) use formulaic sequences in communicative contexts; b) spend time pre-task planning to conceptualize and formulate their ideas; c) repeat the task; and d) learn about spoken fluency with metalinguistic awareness-raising activities (See Tavakoli & Hunter, 2018). For a) use formulaic sequences, students incorporate discussion and communication skill phrases into their discussions to boost fluency. For b) pre-tasking planning, students have preparation time before their discussions to think of their ideas. For c) task repetition, students repeat the discussion with slightly different topics for the practice and perform two discussions per lesson. For d) metalinguistic awareness-raising activities, Curran (2019), an EDC instructor, uses self-reflection activities after the 3/2/1. For instance, he asks, "Did you speak more quickly?" and "Did you have to pause?" Awareness-raising activities can help students achieve a greater understanding of spoken fluency and the ways to improve it (Curran, 2019). EDC course designers incorporate these types of activities to help students improve their spoken fluency during the course, but there is no way to measure students' progress across the semester so instructors might find it challenging to provide personalized feedback or support for students. Next, we explain spoken fluency research findings to design a mobile application that can track its development.

Spoken Fluency Research Findings

Spoken fluency is multidimensional so to explain research findings more precisely, Segalowitz (2010, 2016) separates spoken fluency into three interconnected dimensions: cognitive, utterance, and perceived. Cognitive fluency is the fluidity of underlying cognitive mechanisms that cause L2 speech acts (Segalowitz, 2016) and is operationalized as reaction time speed and reaction time stability (Segalowitz & Segalowitz, 1993). Utterance fluency refers to observable speech production (i.e., speed, breakdown, and repair measurements). Perceived fluency is raters' subjective judgments of overall spoken fluency (Segalowitz, 2010, 2016). His model emphasizes that fluency is not only a psycholinguistic construct, but also an interactional one, which means that listeners also play an important role in determining spoken fluency (Tavakoli & Wright, 2020). Fluency is regarded by examiners and raters as the most difficult aspect of L2 spoken performance to judge (Kang et al., 2019). Therefore, determining which aspects of utterance and perceived fluency should be evaluated and which features of utterance fluency should distinguish descriptor levels are key decisions for both human and automated rating systems (Tavakoli & Wright, 2020). In the next sections, we investigate the applied linguistics research to investigate these key decisions for designing our mobile application.

What Aspects of Utterance Fluency Should be Evaluated on the Application?

A meta-analysis of 22 studies by Suzuki et al. (2021) investigates the relationship between aspects of utterance fluency compared with listener-based perceived ratings of monologues. Utterance fluency measurements should be strongly associated with perceived fluency to be a reliable source of information for assessments. Results show that listeners' perceived fluency ratings are strongly associated with pause frequency and speed, moderately associated with pause duration, and weakly associated with repair fluency (Suzuki et al., 2021). Repair fluency (corrections, reformulations, and repetitions) can strongly relate to speaking style preferences (Kahng, 2014; Tavakoli & Skehan, 2005). Meanwhile, composite measurements show the strongest effect sizes compared with any pure speed, repair, or breakdown measures (Suzuki et al., 2021). Composite measurements incorporate two or more pure measurements. For example, one composite measurement is speech rate, which is calculated by dividing the total number of syllables in a given speech sample by the entire time it takes to create the speech sample in seconds, including pause time, and then multiplying by 60 (Kormos & Dénes, 2004). A pure speed measurement is articulation rate, which is similar to speech rate, but does not include pauses.

Pause location strongly affects perceived fluency as well. Suzuki and Kormos (2020) examine L2 argumentative speaking performance judged by 10 native English-speakers inexperienced raters. For this task, perceived fluency is more strongly associated with mid-clause pausing frequency than with other aspects (Suzuki & Kormos, 2020). For instance, between-clause pausing is "She doesn't like doing dishes, (pause) but you don't either," which can indicate topic change or an idea shift. Mid-clause pausing is "She doesn't like doing (pause) dishes." L2 speakers pause mid-clause more frequently as compared with L1 speakers, which suggests that at least some L2 speakers' mid-clause pausing reflects L2 proficiency gaps associated with online planning, reformulation, and replacement (Skehan et al., 2016; Tavakoli, 2010; Tavakoli & Skehan, 2005). In Suzuki et al.'s (2021) meta-analysis, researchers find that mid-clause pausing is even more strongly associated with perceived fluency than pure speed measurements, but slightly less associated than composite fluency measurements such as speech rate. This suggests that speech rate and mid-clause pausing are some of the best utterance fluency measurements for tracking students' spoken fluency progress across the semester.

What Utterance Fluency Features Should Distinguish Descriptor Levels?

EDC does not distinguish spoken fluency features for Levels 1-4, so one question we ask is, if we incorporate descriptor levels on the spoken fluency application, which features should distinguish levels? The problem is as Bradlow et al. (2017) conclude "individual variability in L2 spoken language production may be best understood within the context of individual variability in L1 spoken language production." For instance, repairs (repetition and pausing) can be indicative of L1 speaking style rather than L2-specific disfluencies (Bosker et al., 2012; Duran-Karaoz & Tavakoli, 2020). Additionally, Bradlow et al. (2017) write that in absolute terms, students speak their L2 slower than their L1, and L1 speaking rate can significantly predict L2 speaking rate. That is, faster L1 speakers are also faster L2 speakers. Shrosbree (2020) also finds that participants with TOEIC scores over 900 have positive correlations for 7 of the 10 spoken fluency measures; however, in absolute terms, they speak their L2 slower than L1 (Shrosbree, 2020). Yet, test makers seldomly ask raters to consider students' L1 speaking style when assessing (Segalowitz, 2016).

In addition to internal factors, there are external factors that influence spoken fluency. External

factors include background noise from others, ease or familiarity with topics, and interlocutors (Tavakoli & Wright, 2020). Segalowitz argued for the importance of assessing sources of variability that are not related specifically to L2 disfluencies but that characterize a person's general performance in the given testing conditions (Segalowitz, 2010). For these reasons, zero-stakes assessments, conducted for formative purposes, rather than summative, can include individualized baselines to account for students' speaking style, and this kind of assessment can also take external factors into consideration that might be sources of variability. One example is students journaling about their experiences to discuss different external factors. Therefore, instead of descriptors, an alternative solution is for the mobile application to suggest a small, incremental mid-clause pause frequency and speech rate goals that take into consideration internal factors such as students' L1 speaking performance and external factors such as topic familiarity.

Should Human Raters be Included With a Computer Automated Rating System?

Another design decision for a mobile application that tracks spoken fluency is whether to have the application fully automated or to have an element of human ratings. With regard to human raters, de Jong (2018) states that the specificity and amount of instruction can enable human raters to focus on certain fluency features such as pause frequency, pause location, and speed. Moreover, raters might even be innately sensitive to pause location with an understanding that mid-clause pausing is more likely a reflection of decrease in cognitive fluency (Kahng, 2018). Automated speech evaluation alone might not be suitable for spoken fluency assessments because the goal of measuring students' L2 ability is to determine the quality of test takers' verbal communication with humans, not with machines (Ginther et al., 2010). For these reasons, listener-based responsibilities during 3/2/1 such as tracking mid-clause pausing and peer-evaluation of comprehensibility are also important. Human raters might increase the reliability of fluency judgments because students perceive other reasons for pausing such as socio-pragmatic considerations (e.g., if students are telling a sad story, they might pause for non-L2 specific disfluent reasons).

Saeki et al. (2021) have designed an AI conversational agent called InteLLA that adapts its interview to assess L2 spoken proficiency. InteLLA's fluency aspect detects speakers' pauses, annotates pause locations (e.g., of mid-clause pauses), and also notes disfluency markers such as fillers and false starts. To mimic human ratings, researchers assign different weights to the fluency's temporal features based on associations between utterance fluency and perceived fluency. InteLLA's accuracy in replicating L2 perceived fluency is 60% using the CEFR level classification system (Saeki et al., 2021). Some of these features might be also useful for a mobile application. The technology can note disfluency markers such as fillers, but for the annotations of pauses, we wonder whether listener-based judgments of mid-clause pausing might be more reliable than an automated system because listeners might have greater insight into the reason for the pause than an automated system. As the mobile application's purpose is different from InteLLA's purpose, we can also add other human-rating judgments such as self-assessment and instructor-based assessments.

Gamification and Personalization

We have also reviewed literature on gamification and personalization. Implementing technology-driven play and competition is becoming increasingly popular in the classroom (Chen et al., 2021; Dehghanzadeh et al., 2021; Flores, 2015; Huang & Soman, 2013). Researchers have found that

students are more motivated and engaged, while also feeling less anxious about using their L2 when gaming elements such as points, medals, badges, or progress tracking are introduced (Dehghanzadeh et al., 2021; Flores, 2015; Huang & Soman, 2013). Students are more inclined to compete with their classmates and themselves to earn medals and increase their ranking (Arce & Valdivia, 2020). Rego (2015) specifies important elements of gamification such as having

a. clear goals
b. rules
c. an accessible and aesthetic user interface
d. collaboration and interaction aspects
e. rewards (e.g., medals or ranking)
f. tracking of students' progress
g. performance feedback

We have included features a)–g) in the mobile application design and we discuss these elements in the next section. In addition to Rego's (2015) advantages, gamification has also been lauded for its ability to incorporate personalized learning (Chen et al., 2021), as personalized activities are usually more meaningful, relevant, and self-initiated. For technology-based personalization, Kukulska-Hulme (2016) emphasizes that with continued L2 practice, personalization can help students to a) identify their needs, b) develop greater awareness of the learning process, and c) monitor their progress. For monitoring their progress students can reflect on and the analyze their performance (Dehghanzadeh et al., 2021). Heil et al. (2016) review 50 language learning-related applications and find that while mobile applications track progress and adjust difficulty levels, most adjustments are minor and not always implemented by students. The feedback given also does not provide reasons for incorrect responses or how to improve performance (Heil et al., 2016). To navigate the aforementioned limitations, instructors should include their feedback (Chen et al., 2021).

Application Design

In this section, we describe the functions of the application and connect it to the literature review. To measure students' spoken fluency, they record themselves for each round of 3/2/1. While they are speaking, the application is collecting data on average number of syllables untrimmed, maximum speed, total number of syllable, and fillers. Their partners listen to their monologues while tallying their mid-clause pauses for each round. If students reach or exceed their fluency goal, they have a "Congrats" message. Once speakers finish their monologues, their listening partners rate the speaker on comprehensibility (i.e., what percentage of the speakers' talk could they understand) and rate overall impressions. As students in each class are grouped according to their proficiency, we believe that listeners might be able to provide some useful feedback for speakers on their mid-clause pausing and overall comprehensibility. Below are screenshots of the application login page, homepage, and congrats message.

After students log in screen, the home screen is where students record their speech rate (See Figure 1 above). The green star above the speedometer on the second and third smartphone screens from left represent students' fluency goal. The speedometers have the features of average syllables per minute, maximum speed, total number of syllables, and total number of fillers. The design is inspired by Salomatin's (2019) speech rate screen, but has the addition of a timer, fluency goal, and

Figure 1

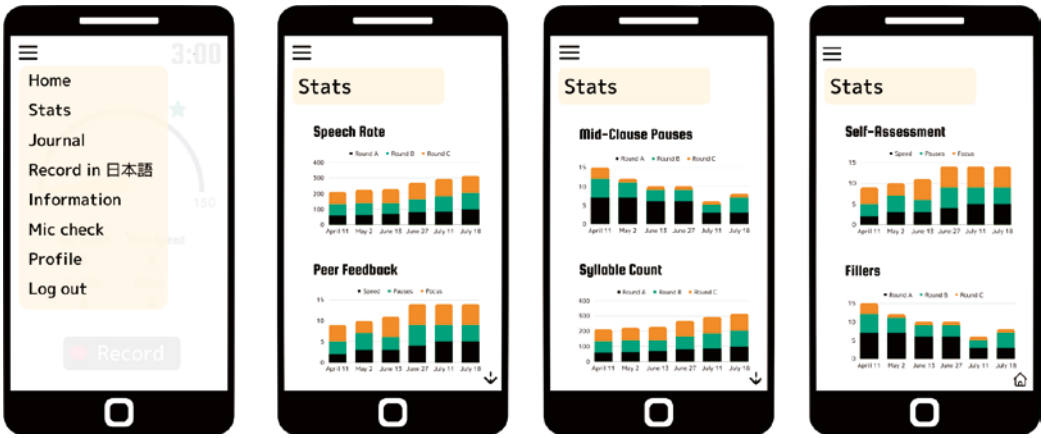
Application Login, Homepage, and Congrats Message

Note. Designed by primary author using Canva Pro with images from Red-Hawk, Jaruka, and Kerismaker

syllable counter. If students receive a speech rate at or above their goal, then the "Congrats" message appears on the fourth screen. The speech rate feature shown in the speedometer is untrimmed which means that disfluencies such as "um" and "uh" have not deleted. Other features include a) a noise suppression filter to reduce background noise so that multiple students can perform the activity at the same time, b) linking students' accounts to Rikkyo University Gmail accounts for safety purposes and ease of access, and c) programming the application for both Android and iOS so that it is accessible to all students with mobile devices. Figure 2 below shows the screens for menu and the statistics.

Figure 2

Application Menu and Stats



Note. Designed by primary author using Canva Pro with images from Red-Hawk, Jaruka, and Kerismaker

The menu screen is the first image on the left of Figure 2. The menu items include home, stats, classmates, journal, record in Japanese, information, mic check, profile, and logout. When students tap on "Stats," the second screen from the left appears, and students can scroll down to view different graphs (See Figure 2 above). These bar graphs include speech rate, peer feedback, mid-clause pauses, syllable count, self-assessment, and fillers. Figure 3 below shows screens for the listener responsibilities.

Figure 3

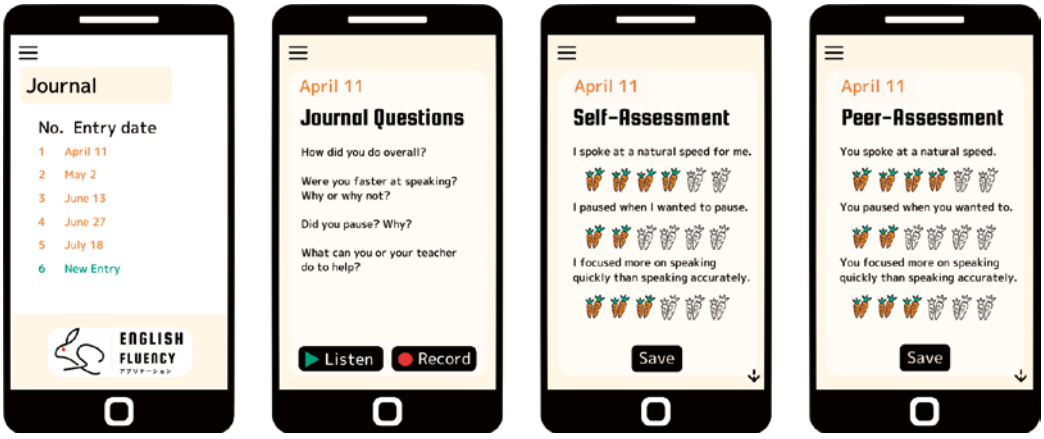
Peer-Assessment Features

Note. Designed by primary author using Canva Pro with images from Red-Hawk, Jaruka, and Kerismaker

On the first screen from the left, there is a list of the students' classmates. Before the 3/2/1 activity, students select their speaking partners. Once they select their partner, they can see the second screen to the left. Listeners press the green plus sign when they hear a mid-clause pause from their speaking partners. If listeners make a mistake, they can press the negative (orange circle) to erase the pause that they mistakenly added to their partners' data. Students need to be taught to detect mid-clause pauses, but we believe that teaching students to be sensitive to mid-clause pausing is feasible given research findings of de Jong (2018) and Kahng (2018). After, listeners can rate their speaking partners on comprehensibility. They can slide the green circle to the percentage that they can comprehend of the speaker's speech. On the last screen from the left, listeners answer the questions about the speakers' performance. The more carrots they select indicates greater agreement with the statements. Figure 4 below shows the journal entry sections for self-assessment and instructors' feedback.

Figure 4

Journal, Self-assessment, and Instructor Feedback Pages



Note. Designed by primary author using Canva Pro with images from Red-Hawk, Jaruka, and Kerismaker

On the first screen to the left of Figure 4, students review previous journal entries or select new journal entries. When they select a date (e.g., April 11), they see the adjacent screen on the right with metalinguistic questions to reflect on their performance. When they scroll down the journal section,

there is a self-assessment screen. There are three self-reflection statements: a) I spoke at a natural speed for me; b) I paused when I wanted to pause; c) I focused more on speaking quickly than speaking accurately. Selecting more carrots equates to a greater agreement with the corresponding statements a)–c). We have decided to use these types of reflective activities so that there is a place for students to discuss other mediating factors such as emotions, topic familiarity, or socio-pragmatic considerations affecting their performance. Figure 5 below shows the L1 mid-clause pause fluency, L1 speech rate, and information pages.

Figure 5

L1 Fluency and Information

Note. Designed by primary author using Canva Pro with images from Red-Hawk, Jaruka, and Kerismaker

On the first screen from the left, students can determine their L1 speech rate. Shrosbree (2020) states that ideally Japanese speech rate is calculated by counting morae for long, spontaneous speeches in part because Japanese is a mora-timed language rather than a syllable-timed one such as English. However, as the application directly compares speech rates, it needs to use the same syllable units. L1 pauses and L1 speech rate data can be used for stats graphs and speedometer as a baseline. For goals, we imagine students continually update their goals to be a little faster than their previous speech rate and they can use the information from their L1 speaking performance to set realistic goals that are personalized. The third screen is the information page, which can help students increase their metalinguistic awareness of spoken fluency. When they tap one of the questions, answers appear. Figure 6 below show the profile, fluency goal, and logout pages.

On the first screen, there is the students' profile page with the language goal. They can change their name, icon, language, and fluency goal. On the second and third screens, students can choose their fluency goal by moving the green circle or the green star. The orange circle and star represent goal recommendations based on students' progress. The last screen is the logout page.

Another aspect is the application design itself. Rego (2015) stresses the importance of an aesthetically pleasing and accessible user interface. Memon (2019) reinforces this idea of designing an application that meets the user's needs, contains consistent elements throughout, and is uncluttered. Although there is some technical terminology that might make the interface less user friendly, we include a language option if students want to use the application in Japanese and we also have a feature whereby students can double tap any technical words to retrieve definition and translation so that they can understand the technical words more easily. For instance, if they double tap on the words "Syllable Count," a definition appears. In the next section, we discuss possible limitations of the application.

Figure 6

Profile, Fluency Goal, Logout Pages

Note. Designed by primary author using Canva Pro with images from Red-Hawk, Jaruka, and Kerismaker

Limitations

Monologic and Dialogic Fluency

Studies have investigated differences between dialogic and monologic fluency. Monologic speaking tasks include Maurice's 4/3/2 and presentations whereas dialogic speaking tasks equate to discussions with another partner. By extension, dialogic speaking tasks share more characteristics in common with group discussions than monologic speaking tasks (e.g., turn-taking, and greater online planning time). Researchers find that students' spoken fluency increases significantly more for dialogic speaking tasks than for monologic speaking tasks (Ferrari, 2012; Michel, 2011; Michel et al., 2007; Tavakoli, 2016); During dialogic speaking tasks, students produce less disfluencies of filled pauses, replacements, reformulations, and repetitions (Michel, 2011; Michel et al., 2007), have less pauses and hesitations (Ferrari, 2012); and have shorter length of pauses, faster articulation rates, longer fluent runs, and higher phonation time ratio (Tavakoli, 2016) than monologic speaking tasks.

Researchers cite Pickering and Garrod's (2004) interactive alignment hypothesis, according to which, interlocutors imitate others' spoken production. They create semi-fixed expressions or routines and reduce cognitive demand of making decisions by streamlining language comprehension and production. Compared with dialogues, monologues have increased cognitive demands for spoken production with no partners to rely on (e.g., less time for online planning because listeners' attention wanes) (Tavakoli & Foster, 2008). The limitation is that applications cannot differentiate speakers' voices during discussions so even though dialogic fluency is more relevant to discussion, we can only assess monologic fluency. Even with differentiation, determining who owns the pauses during dialogic speech is challenging. As EDC only assesses spoken fluency during group discussions, there is a limitation insofar as the mobile application measures only monologic fluency and researchers have shown that students tend to be less fluent for monologic fluency than for dialogic fluency.

Highly Proficient Students

Speed fluency increases with L2 proficiency, but a ceiling effect exists around levels B2 and C1

of CEFR whereby students' speed usually stays constant (Tavakoli et al., 2020). Additionally, fluency judgments no longer become a determining factor of L2 speaking performance assessment for levels C1 and above as raters consider other aspects such as speakers' accuracy or complexity (Tavakoli & Wright, 2020). Therefore, upper EDC Level 1 students might not increase their speed fluency and even if they do increase it, it might not be a determining factor for overall L2 speaking performance ratings. From our own observations, Level 1 students' pausing behavior seems natural rather than a result of possible L2 disfluencies. Building spoken fluency and using a mobile application to track fluency might not be as justifiable for these students; however, other course goals are still important such as building confidence in their ability to use English for communication, to better express their ideas, and respect others' opinions.

Characteristics of Fluency Development

Larsen-Freeman (2020) stresses the importance of acknowledging that language and language learning is complex and dynamic. Spoken fluency, one aspect of language, is also characterized by these attributes as its development is dynamic and nonlinear with complex factors such as individual differences impacting it. For example, influencing factors can be internal such as motivation, personality, language learning aptitude, and L1 speech rate whereas external factors can be background noise from other students or ease or familiarity with topics, as well as the speaker's interlocutor (Tavakoli & Wright, 2020). There are many factors that affect students' fluency, and because of this, students might not always see increases in speed and decreases in pause frequency. Although strong positive correlation exists between objective measurements of L2 overall proficiency and perceived fluency (Bosker et al., 2012; Derwing et al., 2004), EDC is only for one semester (i.e., four months); thus, raising students' awareness of these internal and external factors, and their ability to affect students' spoken fluency, is important because students might not understand why working hard does not always translate into increases in spoken fluency.

Research Project

The application design was preceded by an ongoing study using an existing application, SpeechRate by Yuri Salomatin (2019), to record and measure the words spoken per minute by students during 3/2/1. We chose this application after comparing applications and software; SpeechRate was the most accurate for calculating the words spoken per minute while also having benefits of being accessible offline and free on iOS and Android. Participants were 20 first-year university students at Rikkyo University, and they recorded each round of their 3/2/1 fluency activity in the beginning of the lesson. SpeechRate did not store data, so students had to enter word counts on Google Sheet. To account for the additional time to conduct this activity, 3/2/1 had to be reduced to 2/1.5/1. After the activity, they could write notes about their performance in English or Japanese. At the end of the semester, students were asked to fill out a survey discussing their perceptions of using technology to assess their fluency. From our experience, participants asked more questions about developing spoken fluency and stated that they prefer having specific spoken fluency goals.

Conclusion

EDC course designers write that spoken fluency is the most important language-based objective

for the class (Hurling, 2012). Currently, assessing spoken fluency equates to instructors' perceived quickness of students' academic discussion skill use during group discussion performance. Having an application that can monitor students' monologic spoken fluency progress might help students and instructors to better achieve EDC's language-based objective of improving spoken fluency by increasing metalinguistic awareness and providing feedback on performances. Applied linguists characterize spoken fluency as complex and dynamic insofar as it develops nonlinearly, is multi-dimensional, and has a multitude of factors that influence it such as students' L2 proficiency, L1 speech rate, and topic familiarity. Fluency is also considered to be the most difficult aspect of L2 speaking performance to assess according to Kang et al. (2019), but using mobile applications to track fluency might help. There is more work that needs to be done in designing, developing, and implementing the application, but we look forward to the process and value the feedback that we receive from students and other interested parties.

References

Anderson, J. R. (1989). Practice, working memory, and the ACT* theory of skill acquisition: A comment on Carlson, Sullivan, and Schneider (1989). *Journal of Experimental Psychology: Learning Memory, and Cognition, 15*(3), 527-530. https://doi.org/10.1037/0278-7393.15.3.527

Arce, N. P. H., & Valdivia, A. M. C. (2020). Adapting competitiveness and gamification to a digital platform for foreign language learning. *International Journal of Emerging Technologies, 15*(20), 194-209. https://doi.org/10.3991/ijet.v15i20.16135

Arevat, S., & Nation, I. S. P. (1991). Fluency improvement in second language. *RELC Journal, 22*(1), 84-94. https://doi.org/10.1177/003368829102200106

Boers, F. (2014). A reappraisal of the 4/3/2 activity. *RELC Journal, 45*(3), 221-235. https://doi.org/10.1177/0033688214546964

Bosker, H. R., Pinget, A. F., Quené, H., Sanders, T., & de Jong, N. H. (2013). What makes speech sound fluent? The contributions of pauses, speed, and repairs. *Language Testing, 30*(2), 159-175. https://doi.org/10.1177/0265532212455394

Bradlow, A. R., Kim, M., & Blasingame, M. (2017). Language-independent talker-specificity in first-language and second-language speech production by bilingual talkers: L1 speaking rate predicts L2 speaking rate. *Journal of the Acoustical Society of America, 141*(2), 886-889. https://doi.org/10.1121/1.4976044

Chen, X., Zou, D., Xie, H., & Cheng, G. (2021). Twenty years of personalized language learning: Topic modeling and knowledge mapping. *Educational Technology and Society, 24*(1), 205-222. https://www.jstor.org/stable/26977868

Curran, P. (2019). Raising fluency awareness. *New Directions in Teaching and Learning English Discussion, 7,* 14-18.

Dehghanzadeh, H., Fardanesh, H., Hatami, J., Talaee, E., & Noroozi, O. (2021). Using gamification to support learning English as a second language: A systematic review. *Computer Assisted Language Learning, 34*(7), 934-957. https://doi.org/10.1080/09588221.2019.1648298

Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I. (2004). Second language fluency: Judgments on different tasks. *Language Learning, 54*(4), 655-679. https://doi.org/10.1111/j.1467-9922.2004.00282.x

Duran-Karaoz, Z., & Tavakoli, P. (2020). Predicting L2 fluency from L1 fluency behavior: The case of L1 Turkish and L2 English speakers. *Studies in Second Language Acquisition*, 42(4), 671-695.

https://doi.org/10.1017/S0272263119000755

English Discussion Committee Handbook. (2022). *English discussion class handbook*. Tokyo, Japan: Rikkyo University.

English Discussion Committee Syllabus. (2022). *English discussion class syllabus*. Tokyo, Japan: Rikkyo University.

Ferrari, S. (2012). A longitudinal study of complexity, accuracy, and fluency variation in second language development. In A. Housen, F. Kuiken, and I. Vedder. (Eds), *Dimensions of L2 performance and proficiency: Investigating complexity, accuracy, and fluency in SLA* (pp. 277-298). John Benjamins. https://doi.org/10.1075/lllt.32.12fer

Flores, J. F. F. (2015). Using gamification to enhance second language learning. *Digital Education Review, 27*, 32-54. https://doi.org/10.1344/der.2015.27.32-54

Garside, P. (2020). The effects of time pressure on fluency, complexity, and accuracy: A case study. *Journal of Foreign Language Education and Research, 1*(1), 31-45.

Gatbonton, E., & Segalowitz, N. (1988). Creative automatization: Principles for promoting fluency within a communicative framework. *TESOL Quarterly, 22*, 473-492. https://doi.org/10.2307/3587290

Ginther, A., Dimova, S., & Yang, R. (2010). Conceptual and empirical relationships between temporal measures of fluency and oral English proficiency with implications for automated scoring. *Language Testing, 27*(3), 379-399. https://doi.org/10.1177/0265532210364407

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

Huang, W. H. Y., & Soman, D. (2013). Gamification of education. *Research Report Series: Behavioural Economics in Action. Rotman School of Management*, University of Toronto.

Hurling, S. (2012). Introduction to EDC. *New Directions in Teaching and Learning English Discussion*, 1(1), 2-10.

Jaruka. (2022). Carrots Icon [format]. Canva. https://www.canva.com.

de Jong, N. H. (2018). Fluency in second language testing: Insights from different disciplines. *Language Assessment Quarterly, 15*(3), 237–254. https://doi.org/10.1080/15434303.2018.1477780

de Jong, N., & Perfetti, C. (2011). Fluency training in the ESL classroom: An experimental study of fluency development and proceduralization. *Language Learning, 61*(2), 533-568. https://doi.org/10.1111/j.1467-9922.2010.00620.x

Kahng, J. (2014). Exploring utterance and cognitive fluency of L1 and L2 English speakers: Temporal measures and stimulated recall. *Language Learning, 64*(4), 809-854. https://doi.org/10.1111/lang.12084

Kahng, J. (2018). The effect of pause location on perceived fluency. *Applied Psycholinguistics, 39*(3), 569-591. https://doi.org/10.1017/S0142716417000534

Kang, O., Rubin, D., & Kermad, A. (2019). The effects of training and rater differences on oral proficiency assessment. *Language Testing, 36*(4), 481-504. https://doi.org/10.1177/0265532219849522

Kerismaker. (2022). Spinach [format]. Canva. https://www.canva.com.

Kita, S., Sturges, J. G., Arthurson, D., & Truxel, D. (2022). *What's your opinion? Interactive skills for effective discussion.* Tokyo, Japan: Rikkyo University.

Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of

second language learners. System, 32(2), 145-164. https://doi.org/10.1016/j.system.2004.01.001

Kukulska-Hulme, A. (2016). *Personalization of language learning through mobile technologies.* Cambridge, UK: Cambridge University Press.

Larsen-Freeman, D. (2020). Complexity theory. In B. VanPatten, and J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 248-270). New York, NY: Routledge.

Lennon, P. (2000). The lexical element in spoken second language fluency. In H. Riggenbach (Ed.), Perspectives on fluency (pp. 25-42). Ann Arbor, MI: The University of Michigan Press.

Maurice, K. (1983). The fluency workshop. *TESOL Newsletter, 17*(4), 29.

Memon, M. (2019, February 5). *16 important UX design principles for newcomers*. Springboard. https://www.springboard.com/blog/design/ux-design-principles/

Michel, M. C. (2011). Effects of task complexity and interaction on L2 performance. In P. Robinson (Ed.), *Second language task complexity: Researching the Cognition Hypothesis of language learning and performance*, (pp. 141-173). John Benjamins Publishing Company.

Michel, M. C., Kuiken, F., & Vedder, I. (2007). The influence of complexity in monologic versus dialogic tasks in Dutch L2. *IRAL-International Review of Applied Linguistics in Language Teaching, 45*(3), 241-259. https://doi.org/10.1515/iral.2007.011

Molina Barriga, M., & Briesmaster, M. (2017). The use of the 3/2/1 technique to foster students' speaking fluency, *i.e.: inquiry in education, 9*(2), 1-13.

Pickering, M. J., & Garrod, S. (2004). Why is conversation so easy? *Trends in Cognitive Sciences*, *8*(1), 8-11. https://doi.org/10.1016/j.tics.2003.10.016

Red-Hawk. (2022). Rabbit Line Logo [format]. Canva. https://www.canva.com.

Rego, I. (2015). Mobile language learning: How gamification improves the experience. *Handbook of Mobile Teaching and Learning?* https://doi.org/10.1007/978-3-642-54146-9_76

Saeki, M., Matsuura, R., Suzuki, S., Miyagi, K., Kobayashi, T., & Matsuyama, Y. (2021). InteLLA: A speaking proficiency assessment conversational agent with adaptive interview strategy). 第12回対話システムシンポジウム.

Salomatin, Y. (2019). *SpeechRate* (1.0.3) [Mobile app]. App Store. https://apps.apple.com/us/app/speechrate/id1450633756

Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition, 14*(4), 357-385. https://doi.org/10.1017/S0272263100011189

Segalowitz, N. (2016). Second language fluency and its underlying cognitive and social determinants. *International Review of Applied Linguistics in Language Teaching, 54*(2), 79-95. https://doi.org/10.1515/iral-2016-9991

Segalowitz, N. (2010). *Cognitive Bases of Second Language Fluency.* New York: Routledge. https://doi.org/10.4324/9780203851357

Segalowitz, N. S., & Segalowitz, S. J. (1993). Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics, 14*(3), 369-385.

Shrosbree, M. (2020). The relationship between L1 fluency and L2 fluency among Japanese advanced early learners of English. *Journal of Foreign Language Education and Research, 1*(1), 111-117.

Skehan, P., Foster, P., & Shum, S. (2016). Ladders and Snakes in Second Language Fluency. *International Review of Applied Linguistics in Language Teaching, 54*(2). https://doi.org/10.1515/iral-2016-9992

Skehan, P. (2003). Task-based language instruction. *Language Teaching, 36,* 1-14. https://doi.org/10.1017/S026144480200188X

Suzuki, S., & Kormos, J. (2020). Linguistic dimensions of comprehensibility and perceived fluency: An investigation of complexity, accuracy, and fluency in second language argumentative speech. *Studies in Second Language Acquisition, 42*(1), 143-167. https://doi.org/10.1017/S0272263119000421

Suzuki, S., Kormos, J., & Uchihara, T. (2021). The relationship between utterance and perceived fluency: A meta-analysis of correlational studies. *The Modern Language Journal, 105*(2), 435-463. https://doi.org/10.1111/modl.12706

Tavakoli, P. (2011). Pausing patterns: Differences between L2 learners and native speakers. *ELT Journal, 65*(1), 71-79. https://doi.org/10.1093/elt/ccq020

Tavakoli, P. (2016). Fluency in monologic and dialogic task performance: Challenges in defining and measuring L2 fluency. *International Review of Applied Linguistics in Language Teaching, 54*(2), 133-150. https://doi.org/10.1515/iral-2016-9994

Tavakoli, P., & Foster, P. (2008). Task design and second language performance: The effect of narrative type on learner output. *Language Learning, 58*(2), 439-473. https://doi.org/10.1111/j.1467-9922.2008.00446.x

Tavakoli, P., & Hunter, A. M. (2018). Is fluency being 'neglected' in the classroom? Teacher understanding of fluency and related classroom practices. *Language Teaching Research, 22*(3), 330-349. https://doi.org/10.1177/1362168817708462

Tavakoli, P., Nakatsuhara, F., & Hunter, A. (2020). Aspects of fluency across assessed levels of speaking proficiency. *The Modern Language Journal, 104*(1), 169-191. https://doi.org/10.1111/modl.12620

Tavakoli, P., & Skehan, P. (2005). 9. Strategic planning, task structure and performance testing. *Language Learning and Language Teaching*, 239-273. https://doi.org/10.1075/lllt.11.15tav

Tavakoli, P., & Wright, C. (2020). *Second Language Speech Fluency: From Research to Practice.* London: Cambridge University Press.

Thai, C., & Boers, F. (2016). Repeating a monologue under increasing time pressure: Effects on fluency, complexity, and accuracy. *TESOL Quarterly, 50*(2), 369-393. https://doi.org/10.1002/tesq.232

Tran, M. N., & Saito, K. (2021). Effects of the 4/3/2 activity revisited: Extending Boers (2014) and Thai & Boers (2016). *Language Teaching Research,* https://doi.org/10.1177/1362168821994136

Watanabe, H., & Long, R. (2019). Is there an increase in fluency over levels of proficiency? A study of English L2 usage of Japanese university students. *International Journal of Humanities, Arts and Social Sciences, 5*(6), 215-223. https://doi.org/10.20469/ijhss.5.20001-6