Use of Adaptive Feedback in an App for English Language Spontaneous Speech

Blair Lehman^(✉), Lin Gu, Jing Zhao, Eugene Tsuprun, Christopher Kurzum,
Michael Schiano, Yulin Liu, and G. Tanner Jackson

Educational Testing Service, Princeton, NJ 08541, USA

{blehman, lgu001, jzhao002, etsuprun, ckurzum, mschiano, yliu004,
gtjackson}@ets.org

Abstract. Language learning apps have become increasingly popular. However, most of these apps target the first stages of learning a new language and are limited in the type of feedback that can be provided to users' spontaneous spoken responses. The English Language Artificial Intelligence (ELAi) app was developed to address this gap by providing users with a variety of prompts for spontaneous speech and adaptive, targeted feedback based on the automatic evaluation of spoken responses. Feedback in the ELAi app was presented across multiple pages such that users could choose the amount and depth of feedback that they wanted to receive. The present work evaluates how 94 English language learners interacted with the app. We focused on participants' use of the feedback pages and whether or not performance on spontaneous speech improved over the course of using the app. The findings revealed that users were most likely to access the most shallow feedback page, but use of the feedback pages differed based on the total number of sessions that users completed with the app. Users showed improvement in their response performance over the course of using the app, which suggests that the design of repeated practice and adaptive, targeted feedback in the ELAi app is promising. Patterns of feedback page use are discussed further as well as potential design modifications that could increase the use of feedback and maximize improvement in English language spontaneous speech.

Keywords: MALL · Language learning · Automated speech analysis · Feedback

1 Introduction

Language learning has moved from the traditional classroom-only model to computer-assisted language learning to mobile-assisted language learning (MALL) [1, 2]. MALL apps provide users with flexibility, autonomy, and personalized learning experiences [3]. There are currently over 100 language learning apps in the iOS App Store, and apps have even expanded to smart watches that incorporate exercise into language learning [4]. MALL apps have shown to be an effective method [5, 6]. Duolingo, for example, claims to be as effective as college-level language courses [7], but others report more mixed findings [8]. In this abundance of MALL apps, many have a similar focus in that

they target (a) general language learning and (b) learners at an initially low proficiency level. Thus, there is still a need for the development of apps that provide support to learners at other proficiency levels and with differing goals. For example, recent efforts in MALL app development have focused on the particular language needs of migrants and refugees [9, 10] and low literacy adults [11].

One of the main advantages of MALL apps (and attractions to users) is that they provide immediate, targeted feedback about users' performance on learning activities. This is consistent with years of research that has shown that simply providing feedback is not enough, it must be delivered in a way that is optimally useful for learners [12, 13]. MALL apps are typically able to provide targeted feedback on the quality of selected-response items, grammar and spelling for written responses, and word pronunciation for constrained speaking tasks. However, many MALL apps are limited in the level of detail that can be provided for feedback on speaking tasks [14, 15]. Duolingo, for example, identifies whether or not a user has correctly pronounced a word, but it does not provide feedback about how the user could more accurately pronounce the target word. Given that speaking is often one of the more challenging aspects of learning a language [16–18], it is important for MALL apps to provide feedback in such a way that users feel confident that they can improve their speaking skills.

Given the challenges of providing targeted feedback in real time for speaking tasks, most MALL apps focus only on constrained speaking tasks in which users are provided with a text to read aloud verbatim because automated feedback can be more easily provided. However, our recent user interviews suggested that many language learners would like to practice and receive feedback for spontaneous speaking tasks. Spontaneous speaking tasks involve learners responding to an open-ended prompt (e.g., *Tell me about your favorite vacation.*). This type of task is utilized on many standardized assessments of language skills (e.g., TOEFL®, IELTS™) as it shows an advanced level of speaking proficiency and spontaneous speaking skills are viewed as an important aspect of effective communication [19, 20]. This type of task is often not included in MALL apps because it is difficult to provide immediate, targeted feedback. Spontaneous speaking tasks are typically evaluated by human raters in standardized assessments, which limits the ability to provide feedback to users immediately after responding.

To address the apparent lack of spontaneous speaking practice with immediate feedback for language learners, we have developed the English Language Artificial Intelligence (ELAi) app. The ELAi app was designed to provide users with an opportunity to practice spontaneous speech and receive detailed feedback about the quality of their responses. This learning model is consistent with languaging [21–23] as students are asked to engage in effortful language production that can draw attention to their current weaknesses, but with the added benefit of targeted feedback to help focus efforts for improvement. We utilized an automated speech analysis tool that evaluates spontaneous speech on delivery, language use, and topic development to provide targeted, detailed feedback. However, it is not enough to simply provide feedback [12, 13]. A recent review of research on oral feedback for spoken responses, for example, found that there is a limited understanding of how learners make use of feedback [24]. It is then important that we understand how users interact with the feedback provided. The present work is the first evaluation of the ELAi app and was guided by three research questions: (1)

How do users interact with the app features?, (2) What do users do after viewing feedback?, and (3) Does users' performance improve during app use? We investigated these three research questions with native Mandarin speakers who are learning English for the purpose of attending university in an English-speaking country.

2 ELAi App

The ELAi app was developed to provide an easily accessible resource for English language learners at an intermediate or advanced level, with the goal of attending university in an English-speaking country, to practice spontaneous speech and receive feedback. The development was guided by interviews with potential users from the target audience, which revealed that users were often practicing spontaneous speech on their mobile phone but were unable to receive feedback in the same medium [25]. Users were most interested in feedback that corresponded to standardized English language assessment evaluations. Users also revealed their desire for access to sample responses to compare to their own responses both in terms of delivery and content. The ELAi app was then developed to address the needs of these real-world users.

Users began with the ELAi app by browsing the many prompt options available. Figure 1 shows (from left to right) screenshots of the app splash page as well as the process of selecting a prompt category, responding to a specific prompt (e.g., *Do you think the use of smart watches will increase or decrease in the future? Why?*), and the feedback overview. After completing a new response, users were notified when feedback was available (latency was equivalent to the response length). User responses were evaluated with an automated speech analysis tool that used acoustic and language models to allow for the extraction of acoustic characteristics and creation of a response transcript. The models were based on nonnative English speakers to account for pronunciation differences due to accents. The automated speech analysis tool then evaluated the response on over 100 raw speech features from the acoustic characteristics and transcript. A subset of these features was selected based on their potential for learning feedback and were then combined to provide feedback on six key speech features (filler words, pauses, repeated words, speaking rate, stressed words, vocabulary diversity) to help users improve speaking skills. Users could access feedback on four pages within the app, which allowed for self-selection of the type and amount of feedback provided.

The first feedback page was My History (Fig. 1, rightmost panel), which provided a *Feedback Overview* for each response at a relatively shallow level in that it only identified two speech features that needed improvement (weightlifter icon) and one feature that was done well (thumbs up icon). This was the first instance of feedback that was adaptive to the individual user. For example, in Fig. 1 the user needed to improve on Repeated Words and Vocabulary Diversity, whereas Filler Words was done well in the technology response (top card). Needs work was defined separately for each speech feature with some defined as overuse (filler words, pauses, repeated words, vocabulary diversity), whereas other features had an inverted U-shaped relationship in which too much or too little was problematic (speaking rate, stressed words). However, users were not provided with any explanations or resources to improve future responses on *Overview*. Thus, *Overview* provided minimal feedback on the quality of a response and minimal

Fig. 1. Screenshots of finding and completing a new response in the ELAi app

support for improving future responses but did serve as an organized resource for users to access all of the feedback they had received.

Figure 2 shows the next feedback page that users could access by selecting a specific response card on *Overview* or directly through the feedback ready notification. This next page was *Feedback Summary Report* and was designed to be the main source of feedback for users. On *Summary Report* users could listen to their own response, review explanations for those three speech features that were shown on *Overview*, access additional ideas for how to develop a response to that prompt, and listen to sample responses from both native and nonnative English speakers (from left to right in Fig. 2). The design of *Summary Report* allowed the user to quickly develop an understanding of the quality of their response by focusing on two features that needed improvement and ensured that this feedback was actionable by providing users with additional information and resources to improve their future responses.

Fig. 2. Screenshots of Feedback Summary Report page in the ELAi app

Users could view more detailed feedback on *Feedback Full Report* and *Feedback Details* (see Fig. 3). The *Full Report* provided explanations for all six speech features. For example, in Fig. 3 the two leftmost panels show that the user did well on Filler Words but needed to improve on Repeated Words and Speaking Rate. *Details* provided even more detailed information about four of the six speech features (pauses, repeated words, filler words, vocabulary diversity). *Details* provided a transcript of the response (see second from the right panel in Fig. 3), which highlighted the problematic aspects of the speech feature (e.g., repeated words). *Details* for vocabulary diversity provided suggestions of additional words that could be used to respond to the prompt (see rightmost panel in Fig. 3). *Full Report* and *Details* provided users with a greater amount of and more in-depth feedback, which can be beneficial if users dedicate the time and effort needed to process and apply the information provided [26].

Fig. 3. Screenshots of Feedback Full Report and Details pages in the ELAi app

Users were also able to view information about their app use metrics through the *Me Screen*. Users could see the total amount of time they had recorded responses, total number of responses, the amount of time for recorded responses in the current week, and how many days in a row they had recorded responses. The *Me Screen* also allowed users to *View Badges* that they earned. Users could earn a variety of badges that targeted engagement and performance. Engagement-based badges were designed to encourage persistence and regular practice (e.g., multi-day streaks of recording), whereas performance-based badges allowed users to track their progress over time on a single speech feature (e.g., received “good job” on filler words three times in a row).

3 Method

3.1 Participants

Participants were 94 students from an English language learning program in China that primarily focused on preparation for standardized English language learning assessments. Gender information was obtained from 62 participants: 58% female, 33% male,

and 6% preferred not to respond. Participants completed from 1 to 45 sessions with the ELAi app over a one-month period ($M = 8.62$, $SD = 8.56$). Sessions were a little over five minutes on average ($SD = 4.52$) and included an average of 17.2 user-initiated actions ($SD = 14.8$). Users completed an average of 14.4 spoken responses over the course of using the ELAi app ($SD = 24.2$).

3.2 Procedure

Participants were recruited through their English language learning program. Those participants who were interested then completed an informed consent and were provided with the information needed to access the ELAi app. Participants were free to use the app as they wanted for one month. There were no direct instructions about how users should interact with the app; however, participants were told that they would receive a certificate of participation if they recorded at least five spoken responses.

4 Results and Discussion

4.1 How Do Users Interact with the App Features?

First, we investigated the use of app features in four ways (see Table 1): feature access (proportion of participants), average feature time use (in seconds, avg time per access), proportion of total session time (proportion of time), and proportion of total session actions (proportion of actions) [27]. The proportion of participants that accessed each feature at least once revealed a generally high rate of feature access, with the exception that 60% or less of users accessed the more in depth feedback pages (Full Report, Details) and listened to their own or samples responses, which were features that users specifically requested. This contradiction between what users say they want and how they interact with a MALL app has been found in other apps as well [28].

Repeated measures ANOVAs that compared the average time each feature was accessed [$F(10,930) = 11.4$, $p < .001$, $MSe = 1858$, partial $\eta^2 = .109$], the proportion of time spent on each feature [$F(10,930) = 229$, $p < .001$, $MSe = .008$, partial $\eta^2 = .711$], and the proportion of actions that accessed each feature [$F(10,930) = 474$, $p < .001$, $MSe = .005$, partial $\eta^2 = .836$] were significant. Bonferroni corrections were applied to all post hoc analyses and revealed that users spent the most time viewing prompts and the least time viewing badges. Responding to prompts had one of the highest proportions for both time and actions, with only View Prompt having a higher proportion and FB Overview and Me Screen having similar proportions.

Table 1. Use of ELAi features.

	Proportion of Participants	Avg Time per Access		Proportion of Time		Proportion of Actions	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
View Category	.787	8.97	1.11	.034	.056	.032	.041
Low Engagement	.638	5.37	6.29	.037	.072	.032	.051
High Engagement	.936	12.6	13.0	.031	.034	.032	.028
View Prompt	1.00	17.1	2.79	.493	.187	.542	.156
Low Engagement	1.00	12.0	21.3	.514	.211	.564	.173
High Engagement	1.00	22.3	31.2	.472	.157	.520	.135
New Response	.862	58.7	12.8	.114	.113	.057	.051
Low Engagement	.745	30.6	42.7	.115	.144	.055	.056
High Engagement	.979	86.9	167	.114	.071	.059	.046
FB Overview	.947	23.6	4.75	.164	.115	.150	.078
Low Engagement	.915	12.2	19.9	.159	.142	.134	.077
High Engagement	.979	35.0	60.2	.170	.080	.165	.076
FB Summary Report	.723	17.9	2.44	.053	.067	.040	.041
Low Engagement	.532	14.0	23.2	.042	.071	.026	.040
High Engagement	.915	21.8	23.6	.064	.061	.053	.038
FB Full Report	.574	20.2	3.16	.019	.036	.011	.016
Low Engagement	.404	16.1	34.7	.017	.044	.009	.016
High Engagement	.745	24.4	25.6	.021	.025	.013	.015
FB Details	.606	10.8	1.24	.019	.032	.017	.024
Low Engagement	.362	7.71	13.2	.017	.037	.013	.025
High Engagement	.851	13.8	9.86	.021	.026	.020	.023
Listen Own Response	.436	17.3	2.87	.014	.024	.008	.014
Low Engagement	.468	7.68	17.8	.011	.025	.006	.014
High Engagement	.638	27.0	32.5	.017	.023	.010	.014
Listen Sample Response	.457	11.2	1.90	.013	.026	.011	.021
Low Engagement	.255	5.44	11.1	.008	.022	.006	.015
High Engagement	.660	16.9	22.3	.018	.029	.017	.025
Me Screen Viewed	.936	8.20	1.70	.075	.107	.131	.108
Low Engagement	.894	5.89	17.4	.080	.134	.153	.123
High Engagement	.979	10.5	15.5	.070	.073	.109	.085
View Badge	.223	1.38	.327	.001	.004	.003	.008
Low Engagement	.064	.589	2.75	.001	.004	.001	.009
High Engagement	.383	2.18	3.40	.002	.004	.004	.007

Overall the feature use analyses revealed that users spent the majority of their time interacting with the ELAi app browsing for a prompt, responding to prompts, viewing the shallowest level of feedback, and viewing their overall app usage data. This pattern is both consistent and inconsistent with user requests. Users were frequently practicing their spontaneous speech, but they were not typically utilizing the more detailed feedback

and learning resources that they requested. It is important to note, however, that the more detailed feedback and learning resources were embedded in the app, meaning that users could only access them via another feedback page. Feedback Overview and Summary Report, on the other hand, could be accessed directly. Thus, the lack of access to the more detailed feedback (Full Report, Details) could represent a lack of user interest or lack of feature awareness. In an effort to consider this dependence between actions, we repeated the proportion of actions analysis with instances in which less detailed feedback page views were removed if they immediately preceded a more detailed feedback page view. This was an overly conservative analysis as it assumed that all of these less detailed feedback page views were only in service of accessing more detailed feedback. The pattern of findings remained the same, which suggests that although we cannot target the exact reason for infrequent access of more detailed feedback, we can feel confident that those pages were accessed less frequently.

The previous analyses considered the sample as a whole; however, there was a wide range in the degree to which users engaged with the ELAi app (1 to 45 sessions), which suggests a potential for different use patterns. Users were divided into low (five or less sessions, $n = 47$) and high engagement groups (more than five sessions, $n = 47$) based on a median split to explore potential feature use differences. Table 1 shows the descriptive statistics for each engagement group. Particularly large differences can be seen for accessing more detailed feedback and learning resources, with high engagement users accessing those features at least once at a higher rate than low engagement users.

The two engagement groups were compared with independent samples t -tests for average time spent on each feature, which revealed that the high engagement group spent more time on all features, except Summary Report, Full Report, and Me Screen. Despite this difference in time spent on features, there were no differences in how users in each engagement group distributed their time (proportion of time, p 's $> .05$) and actions within a session (proportion of actions, p 's $> .05$). The comparison of engagement groups revealed that users who had greater engagement with the ELAi app accessed more features and spent more time on those features, particularly those features that provided more in-depth feedback and support for improving future performance.

4.2 What Do Users Do After Viewing Feedback?

The previous findings led us to question if there were particular patterns of behavior after viewing feedback that were indicative of more or less productive behavior. For example, a productive behavior after viewing FB Overview would be to access FB Summary Report to better understand why certain speech features need improvement and access resources for improvement. Thus, we investigated the next action taken after viewing each type of feedback. We combined several actions into action categories that consisted of Browse Behavior (View Category, View Prompt), Feedback Viewed, Me Screen Viewed (Me Screen, View Badge), and Exit App for these analyses.

Repeated measures ANOVAs compared the prevalence of post-feedback actions for each feedback page (see Table 2) and were significant [Overview: $F(4,352) = 33.6$, $p < .001$, $MSe = .044$, partial $\eta^2 = .277$; Summary Report: $F(4,268) = 333$, $p < .001$, $MSe = .024$, partial $\eta^2 = .832$; Full Report: $F(4,212) = 313$, $p < .001$, $MSe = .026$, partial $\eta^2 = .855$; Details: $F(4,224) = 259$, $p < .001$, $MSe = .032$, partial $\eta^2 = .822$]. Bonferroni

Table 2. Proportion of next action type after feedback page view.

Post-Feedback Page Action	Feedback Page							
	Overview		Summary Report		Full Report		Details	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Browse Behavior	.308	.225	.067	.109	.024	.073	.079	.230
Feedback Viewed	.272	.213	.805	.216	.895	.225	.876	.246
Me Screen Viewed	.277	.241	.014	.052	.003	.023	.008	.028
Exit App	.143	.143	.114	.185	.078	.219	.038	.113

corrections were applied to all post hoc comparisons. The pattern for Overview revealed that all action categories were more likely to occur after viewing Overview than exiting the app. A different pattern emerged for the remaining feedback pages. Specifically, viewing feedback was the most likely action to occur after viewing Summary Report, Full Report, and Details, with at least 80% of next actions involving viewing one of the feedback pages. Exit App and Browse Behavior were the next most likely to occur and Me Screen Viewed was the least likely action to occur after viewing those three feedback pages. These findings suggest that if users can go deeper into the feedback than Overview, they may get into a potentially beneficial feedback loop.

4.3 Does Users' Performance Improve During App Use?

Last, we investigated changes in spoken response performance over the course of app interaction. User sessions (visit to app) were divided into thirds (first, middle, last) and we investigated changes in performance from the first third to the last third. Performance was measured as the proportion of spoken responses that received “Needs Work” feedback on each speech feature in each third of sessions. This investigation reduced the number of users to 27 as users were required to have at least three sessions and to have at least one speech in both the first and last third of sessions. All 27 users included in this analysis were in the high engagement group, which means that they made greater use of the app features, in particular the more detailed feedback and resources for response improvement. Table 3 shows the descriptive statistics and paired samples *t*-test comparisons for each of the six speech features.

Table 3. Proportion of spoken responses that need work across session phases.

	Session Phase				<i>t</i>	<i>df</i>	<i>p</i>	<i>d</i>
	First Third		Last Third					
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>				
Filler Words	.520	.509	.110	.320	4.23	26	<.001	-.964
Pauses	.560	.506	.300	.465	2.56	26	.017	-.535
Repeated Words	.670	.480	.150	.362	4.65	26	<.001	-1.22
Speaking Rate	.930	.267	.520	.509	4.23	26	<.001	-1.01
Stressed Words	.560	.506	.070	.267	4.32	26	<.001	-1.21
Vocabulary Diversity	.810	.396	.300	.465	4.65	26	<.001	-1.18

The comparisons revealed a reduction in the proportion of speech features that needed work, which suggests an overall improvement in performance across use of the ELAi app. The effect size differences between the first and last third of sessions were all large ($d > .8$) [29], with the exception of a medium effect size ($.5 < d < .8$) for Pauses. These findings are very promising as they show large improvements in a variety of speech features over a relatively short period of time. However, the findings should be interpreted with a modicum of caution as a small number of participants were included in these analyses (29% of sample), time on task varied across participants, and we were not able to consider additional resources that users may have accessed during this same time period (e.g., language courses, other MALL apps). It is also important to note that this investigation was limited to performance within the app and a more formal investigation of changes in speaking skills is needed (e.g., pre/posttest design) to determine the true effectiveness of the ELAi app as a learning tool [27].

5 Conclusion

There is currently a plethora of language learning apps available to users. However, these apps are often designed for beginning language learners and are limited in their ability to provide feedback to spoken responses. The ELAi app was developed to provide an easily accessible English language learning app for users that want to receive detailed feedback about their speaking skills during spontaneous speech. The present work was the first evaluation of the ELAi app. Overall, the findings revealed that users spent the majority of their time browsing for prompts, completing new responses, and viewing shallow level feedback. This suggests that users are generally not taking advantage of the more in-depth feedback and resources to facilitate improvement, which were requested by users during interviews [28]. Although the prominence of viewing shallow feedback is disappointing, it could represent productive behavior. Feedback Overview is the only page in which users can compare their performance on multiple speech features across individual responses, which could reveal patterns of improvement or persistent issues [30] by leveraging the benefits of open learner models [31]. Future research is needed to determine if this cross-response comparison is occurring and to explore designs to facilitate these comparisons [32, 33] as language learners may not engage in self-regulated learning behaviors on their own in MALL apps [34].

We also investigated changes in user performance. Our preliminary findings were promising in that more engaged users improved their performance on all six speech features from the beginning to the end of their interaction with the ELAi app. However, these findings are only preliminary and a more rigorous investigation of the impact of the ELAi app on speaking skills is needed. Overall our initial findings suggest that the ELAi app is a promising MALL, but there is still room for improvement. The Feedback Summary Report, for example, could be improved by requiring less scrolling for users to access learning resources and explicitly highlighting the availability of more in-depth feedback to reduce any lack of feature awareness. Tailoring the feedback to user characteristics (e.g., cultural background) could also benefit learning [35]. New in-app incentives (e.g., badges) could encourage more frequent use (e.g., more than five sessions) and use of the in-depth feedback pages and learning resources. Users could also

benefit from being shown their improvement over time to implicitly reward continued use of the app. Overall, the ELAi app shows initial promise at creating an easily accessible resource for practicing and receiving feedback on spontaneous speaking tasks, but more research is needed to understand how this app can be the most beneficial to users.

References

1. Kukulska-Hulme, A.: Mobile-assisted language learning. In: Chapelle, C. (ed.) *The Encyclopedia of Applied Linguistics*, pp. 3701–3709. Wiley, New York (2013)
2. Sharples, M., Taylor, J., Vavoula, G.: A theory of learning for the mobile age. In: Andrews, R., Haythornwaite, C. (eds.) *The Sage Handbook of E-Learning Research*, pp. 221–247. Sage, London (2007)
3. Sung, Y.T., Chang, K.-E., Yang, J.-M.: Mobile devices for language learning? A meta-analysis. *Educ. Res. Rev.* **16**, 68–84 (2015)
4. Shadiev, R., Hwang, W.-Y., Shadiev, N., Fayziev, M., Liu, T.-Y., Shen, L.: Smart watches for making EFL learning effective, healthy, and happy. In: Chang, M., et al. (eds.) *Challenges and Solutions in Smart Learning. LNET*, pp. 73–76. Springer, Singapore (2018). https://doi.org/10.1007/978-981-10-8743-1_11
5. Hanson, A.E.S., Brown, C.M.: Enhancing L2 learning through a mobile assisted spaced-repetition tool: an effective but bitter pill? *Comput. Assist. Lang. Learn.* **33**, 133–155 (2019)
6. Lin, J.-J., Lin, H.: Mobile-assisted ESL/EFL vocabulary learning: a systematic review and meta-analysis. *Comput. Assist. Lang. Learn.* **32**, 878–919 (2019)
7. Vesselinov, R., Grego, J.: Duolingo effectiveness study final report (2012). http://static.duolingo.com/s3/DuolingoReport_Final.pdf
8. Loewen, S., Crowther, D., Isbell, D.R., Kim, K.M., Maloney, J., Miller, Z.F.: Mobile-assisted language learning: a Duolingo case study. *ReCALL* **31**, 293–311 (2019)
9. Demmans Epp, C.: Migrants and mobile technology use: gaps in the support provided by current tools. *J. Interact. Media Educ.* **2**, 1–13 (2017)
10. Kukulska-Hulme, A.: Mobile language learning innovation inspired by migrants. *J. Learn. Dev.* **6**, 116–129 (2019)
11. Munteanu, C., et al.: Hidden in plain sight: low-literacy adults in a developed country overcoming social and educational challenges through mobile learning support tools. *Pers. Ubiquit. Comput.* **18**, 1–15 (2013). <https://doi.org/10.1007/s00779-013-0748-x>
12. Kluger, A.N., DeNisi, A.: The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychol. Bull.* **119**, 254–284 (1996)
13. Shute, V.J.: Focus on formative feedback. *Rev. Educ. Res.* **78**, 153–189 (2008)
14. Ahn, T.Y., Lee, S.-M.: User experience of a mobile speaking application with automatic speech recognition for EFL learning. *Br. J. Edu. Technol.* **47**, 778–786 (2016)
15. Kukulska-Hulme, A., Shield, L.: An overview of mobile assisted language learning: can mobile devices support collaborative practice in speaking and listening? Paper presented at EuroCALL 2007 (2008)
16. Horwitz, E.K.: Language anxiety and achievement. *Annu. Rev. Appl. Linguist.* **21**, 112–126 (2001)
17. Pawlak, M., Mystkowska-Wiertelak, A.: Investigating the dynamic nature of L2 willingness to communicate. *System* **50**, 1–9 (2015)
18. Von Wörde, R.: Students' perspectives on foreign language anxiety. *Inquiry* **8**, 1 (2003)
19. Liu, N., Littlewood, W.: Why do many students appear reluctant to participate in classroom learning discourse? *System* **25**, 371–384 (1997)

20. Yanagi, M., Baker, A.A.: Challenges experienced by Japanese students with oral communication skills in Australian Universities. *TESOL J.* **7**, 621–644 (2016)
21. Swain, M.: Communicative competence: some roles of comprehensive input and comprehensible output in its development. In: Glass, S.M., Madden, C. (eds.) *Input in Second Language Acquisition*, pp. 235–253. Newbury House Publishers, Rowley (1985)
22. Swain, M.: Three functions of output in second language learning. In: Cook, G., Seidlhofer, B. (eds.) *Principles and Practice in Applied Linguistics*, pp. 125–144. Oxford University Press, Oxford (1995)
23. Swain, M.: Languageing, agency, and collaboration in advanced second language proficiency. In: Byrnes, H. (ed) *Advanced Language Learning: The Contribution of Halliday and Vygotsky*, pp. 95–108. Continuum, New York (2006)
24. Yu, S., Wang, B., Teo, T.: Understanding linguistic, individual and contextual factors in oral feedback research: a review of empirical studies in L2 classrooms. *Educ. Res. Rev.* **24**, 181–192 (2018)
25. Dourish, P.: The appropriation of interactive technologies: some lessons from placeless documents. *Comput. Support. Coop. Work Three Funct. Output Second Lang. Learn.* **12**, 465–490 (2003). <https://doi.org/10.1023/A:1026149119426>
26. Craik, F.I., Lockhart, R.S.: Levels of processing: a framework for memory research. *J. Verb. Learn. Verb. Behav.* **11**, 671–684 (1972)
27. Demmans Epp, C., Phirangee, K.: Exploring mobile tool integration: design activities carefully or students may not learn. *Contemp. Educ. Psychol.* **59**, 101791 (2019)
28. Botero, G.G., Questier, F., Zhu, C.: Self-directed language learning in a mobile-assisted, out-of-class context: do students walk the talk? *Comput. Assist. Lang. Learn.* **32**, 71–97 (2018)
29. Cohen, J.: A power primer. *Psychol. Bull.* **112**, 155–159 (1992)
30. Demmans Epp, C., McCalla, G.: ProTutor: historic open learner models for pronunciation tutoring. In: Biswas, G., Bull, S., Kay, J., Mitrovic, A. (eds.) *AIED 2011. LNCS (LNAI)*, vol. 6738, pp. 441–443. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-21869-9_63
31. Bull, S., Kay, J.: Open learner models. In: Nkambou, R., Bourdeau, J., Mizoguchi, R. (eds.) *Advances in Intelligent Tutoring Systems*, vol. 308, pp. 301–322. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-14363-2_15
32. Çakir, I.: Mobile-assisted language learning (MALL). In: Yaman, E., Emekçi, E., Şenel, M. (eds.) *Current Trends in ELT*, pp. 170–189. NÜANS Publishing, Ankara (2016)
33. Hug, T.: Mobile learning as ‘microlearning’: conceptual considerations towards enhancements of didactic thinking. *Int. J. Mobile Blend. Learn.* **2**, 47–57 (2010)
34. Viberg, O., Andersson, A.: The role of self-regulation and structuration in mobile learning. *Int. J. Mobile Blend. Learn.* **11**, 42–58 (2019)
35. Yang, J.: Learners’ oral corrective feedback preferences in relation to their cultural background, proficiency level and types of errors. *System* **61**, 75–86 (2016)