

This is the accepted version of the manuscript:

He, X., & Loewen, S. (2022). Stimulating learner engagement in app-based L2 vocabulary self-study: Goals and feedback for effective L2 pedagogy. *System*, 105, 102719. <https://doi.org/10.1016/j.system.2021.102719>

**Stimulating Learner Engagement in App-based L2 Vocabulary Self-study:
Goals and Feedback for Effective L2 Pedagogy**

Xuehong (Stella) He^a and Shawn Loewen^b

^a Corresponding Author. Department of English, Faculty of International Studies, Nagoya University of Commerce and Business, 4-4 Sagamine, Komenoki, Nisshin, Aichi, Japan 470-0193, hexuehon@msu.edu, +81 070-2680-1174.

^b Second Language Studies Program, Department of Linguistics, Languages and Cultures, Michigan State University, B255 Wells Hall, 619 Red Cedar Road, East Lansing, MI, USA 48824, loewens@msu.edu, +1 517-353-9790.

Declaration of Interest: None

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Abstract

Despite its increasing popularity, application-based (app-based) second language (L2) learning faces the persistent issue of low learner engagement. This study explored app-based L2 vocabulary self-study as a course assignment by drawing on Mercer and Dörnyei's (2020) latest proposal of goal-setting with feedback for stimulating learner engagement and motivation. We adopted a quasi-experimental design that consisted of a treatment group who completed goal-setting-and-checking activities ($n = 32$) and a control group who did not ($n = 31$), and compared learner engagement as indexed by the number of words studied weekly and L2 skill development as assessed by TOEIC scores before and after app use. Ratings and comments were also collected from the treatment group on goal-setting-and-checking activities and from both groups on two supplementary feedback tools, Unfinished Lists (lists of students who did not finish app-based self-study assignments) and leaderboards (student rankings based on assignment performance). Quantitative results showed the treatment group studied more words than the control group, although TOEIC scores did not differ between or change within groups significantly. Quantitative ratings and qualitative comments supported overall positive effects of these three pedagogical interventions. Practical suggestions are provided to guide L2 instructors to adapt these pedagogical interventions into their teaching.

Keywords: mobile assisted language learning, language learning apps, computer assisted language learning, second language vocabulary learning, engagement, motivation, goal setting and checking, feedback, gamification, leaderboard

Stimulating Learner Engagement in App-based L2 Vocabulary Self-study: Goals and Feedback for Effective L2 Pedagogy

1. Introduction

Despite over two decades of progress, mobile-assisted language learning (MALL) remains on the fringes of foreign/second language (L2) instruction, as Burston (2014) concluded after reviewing 345 studies in 1994-2012. Specifically, most MALL implementations stood as a voluntary component, lasted for no longer than three hours, involved fewer than ten learners, relied on subjective measures of outcomes, and barely acknowledged their failure to generate positive results (Burston, 2014). In another review by Chwo et al. (2018) of 213 MALL studies from 2009-2017, the instructor's role has emerged as a common theme, but as Reinders and Stockwell (2017) pointed out, it is still unclear how teachers can support L2 learners during MALL. Indeed, empirical studies have revealed L2 teachers' unreadiness to adapt mobile devices into their classrooms, and that they needed further training and support in effectively integrating MALL into their established teaching routines (Tai & Ting, 2011; Wishart, 2008). Within the theme of the instructor's role, one main dimension concerned teacher feedback to L2 learners, with the recommendation that teachers provide feedback regularly throughout the MALL process (Chwo et al., 2018). In addition to the instructor's role, another even more common theme was learner motivation and engagement in MALL (Chwo et al., 2018). More specifically in MALL for L2 English, Elaish et al. (2019) reviewed 69 studies from 2010-2015 and also found learner motivation, such as students' lack of motivation to engage in vocabulary review after class, to be a consistent problem. Given the ubiquity of mobile devices in educational settings, Elaish et al. (2019) advocated continuing to advance the line of L2 research in MALL in order to support both instructors and learners. They also pinpointed the scarcity of

rigorous application of theories in current MALL studies and called for more theory-based, outcome-oriented MALL research (Elaish et al., 2019).

Within MALL, mobile applications (i.e., apps) have become popular in recent years both as a supplement to formal L2 instruction and as a medium of independent L2 study for beginning and advanced learners (Rosell-Aguilar, 2018) in online and face-to-face settings (Levy & Kennedy, 2010). It is therefore important that researchers and teachers continue exploring apps for L2 teaching and learning (Reinders & Pegrum, 2016). One perplexing issue for L2 educators is that although L2 learners often express enthusiasm for MALL, their actual engagement with MALL resources is generally limited (Stockwell, 2016). Actually, quite a few empirical studies have found conspicuously low levels of learner engagement during app-based L2 learning (García Botero et al., 2019; Hanson & Brown, 2020; Kondo et al., 2012; Loewen et al., 2019, 2020; Milliner & Cote, 2015; Nielson, 2011; Read & Kukulska-Hulme, 2015). If apps fail to engage learners, their pedagogical value will never be realized, regardless of their potential effectiveness (see also Bodnar et al., 2016). Hence, it is vital to stimulate learner engagement during app-based L2 learning.

Despite the fact that learner motivation and engagement are often limited in app-based L2 learning (Chwo et al., 2018; Elaish et al., 2019; Hanson & Brown, 2020), so far little evidence has been presented about what teachers can do to effectively motivate and engage L2 learners during MALL. In fact, as Nielson (2011) stressed, in the absence of proper support, even learners with high motivation may fail to engage in MALL-based L2 self-study. Building on Dörnyei's (2001) classic work on teachers' strategies to motivate L2 learners, Mercer and Dörnyei (2020) proposed learner engagement as a closely related, new direction for L2 research on learner motivation and effective pedagogy. Specifically, furthering previous suggestions of goal-setting

motivational strategies (Dörnyei, 2001; Oxford & Shearin, 1994), Mercer and Dörnyei (2020) recommended goal-setting with feedback to stimulate L2 learners' engagement and motivation.

In response to the still marginal role of MALL in L2 education (Burston, 2014), this study aims to expand current research on L2 instructors' role in MALL (Chwo et al., 2018; Reinders & Stockwell, 2017) and to provide ready-to-use activities for L2 teachers to adapt into their classrooms. Our aims stem from Instructed Second Language Acquisition (ISLA) research's goal to bridge research and pedagogy, as L2 teachers often seek accessible research-based instructional practices that they can easily apply to their own educational contexts (Sato & Loewen, 2019). Specifically, we focus on the persistent issue of low learner engagement in app-based L2 learning (Hanson & Brown, 2020; Stockwell, 2016) and draw on Mercer and Dörnyei's (2020) latest proposal of goal-setting with feedback to stimulate learner engagement. By exploring app-based L2 vocabulary self-study as a course assignment, this study joins the endeavors of researchers to support teachers and learners with evidence-based L2 vocabulary instruction (e.g., He & Godfroid, 2019).

2. Literature Review

2.1. Effectiveness of App-based L2 Vocabulary Learning

Regarding the effectiveness of apps for developing L2 vocabulary, results from a meta-analysis of 19 (quasi-)experimental EFL studies showed a large effect size (Hedge's $g = 0.903$) for app-based learning and word retention (Lin & Lin, 2019). Empirical studies on L2s such as Turkish (Loewen et al., 2019) and Spanish (Loewen et al., 2020) also found app-based learning effective for L2 vocabulary development.

In non-app-based settings, research has generally shown close connections between L2 vocabulary and L2 skills (Nation, 2013). Then a question of both theoretical and pedagogical

interest emerges: does app-based L2 vocabulary learning facilitate L2 skill development? Only three studies have examined this question specifically. Bower and Rutson-Griffiths (2016) recorded learners' gain scores in the skill-based TOEIC tests after 10-month use of the Cooori app and found the number of flashcard repetitions predicted TOEIC gain scores. In contrast, Agawa et al. (2011) did not find significant correlations between TOEIC gain scores and 8-week use of the Word Engine app. Finally, Phillips (2011) claimed a correlation trend between 11-week use of Word Engine and TOEIC gain scores. With the current findings still inclusive, our study sets off to examine the relationship between app-based L2 vocabulary learning and L2 skill development as assessed by the TOEIC listening and reading test.

2.2. Learner Engagement in App-based L2 Learning

As a multi-faceted and multi-dimensional construct, learner engagement is widely researched with diverse theoretical underpinnings in educational psychology (Philp & Duchesne, 2016), which further complicates the conceptualization of its relationship with another closely related construct, learner motivation (see Christenson et al., 2012). When reviewing proposals in educational psychology on the motivation-engagement link, Martin (2012) observed a broadly shared idea that “motivation is a basis for subsequent engagement” (p. 305), and he operationalized motivation as cognition and engagement as behavior in his Motivation and Engagement Wheel framework. In L2 research, Dörnyei (2019) highlighted that whereas “motivation only indicates a student’s *potential* for actively pursuing learning” (p. 25), engagement captures the part of motivation that is converted into *action*. Dörnyei (2019) also agreed with educational psychologists Skinner et al. (2008) that behavioral participation in class represents the core of engagement.

Based on previous research, our major focus here is on engagement as a behavioral

manifestation of motivation, i.e., behavioral engagement (see Philp & Duchesne, 2016 for other types of engagement). Specifically, we examine engagement during app-based L2 learning mainly through L2 learners' actual use of apps as recorded by app user logs (see also García Botero et al., 2019). With this working operationalization of learner engagement, next we will review L2 research that investigated learners' L2 app use.

Several studies have reported non-student learners' low engagement with app-based L2 learning outside of course curriculum. Nielson (2011) recruited US government staff to study L2s with two apps (Rosetta Stone; Tell Me More) but found a strikingly high attrition rate: only 22% (65 out of 296) accessed either app for more than 10 hours during five or seven months. The issue of low engagement cannot be avoided even when ISLA researchers are L2 learners themselves, as Loewen et al. (2019) reported only two out of nine (22%) participants achieved the initial goal of studying L2 Turkish for 34 hours within one semester.

García Botero et al. (2019) tracked 118 students' use of Duolingo for out-of-class L2 learning for one year. They found the number of active users dropped quickly to less than 20 (17%) after six months, despite students' approval of Duolingo as an effective complement to L2 classes. In Kondo et al.'s (2012) study, when a mobile app was no longer part of the L2 curriculum (as in a previous semester) but was still available for L2 self-study, six students (40%) did not use the app any more throughout the semester. Loewen et al. (2020) recruited students who were not enrolled in Spanish courses to self-study L2 Spanish with Babbel, and the attrition rate was 32% (from 85 to 58) during three months. Read and Kukulska-Hulme (2015) reported 45 students signed up to practice L2 English with the Audio News Trainer app individually but only nine (20%) accessed the app during the 10-week period.

In addition to persisting in non-curriculum contexts, the issue of low learner engagement

remains even when app-based learning is incorporated into L2 classes. Milliner and Cote (2015) designated the Xreading app as part of an assignment for extensive reading, but found two out of five classes (39% students) were far below the minimum requirement of reading at least 50,000 words within a semester: students in the two classes read on average 9,277 and 11,250 words respectively, fulfilling only 19% and 23% of the expected reading. In a pilot study, Hanson and Brown (2020) added Anki as an optional course activity and found only 21 students (36%) used the app for only about two days during one semester. Such low engagement led Hanson and Brown (2020) to incorporate Anki as an assignment in the main study and to require its use at least five days per week. Even so, the overall average usage ended up being less than two days per week, with only 14 students (27%) accessing the app for more than two days a week. These findings show that simply adding app-based L2 learning as a course requirement may not work as effectively as teachers might hope.

Proactive in stimulating learner engagement and motivation, many L2 apps (e.g., Duolingo, Memrise) have incorporated gamification into their design (see reviews by Boudadi & Gutiérrez-Colón, 2020; Dehghanzadeh et al., 2019; Shortt et al., 2021). Gamification as a pedagogical concept concerns “the integration of design elements or activity patterns traditionally found in games into educational contexts” (Buckley & Doyle, 2016, p. 1163), with the aim to “engage people, motivate action, promote learning, and solve problems” (Kapp, 2012, p. 10). Gamification often involves highlighting competition as an explicit motivator by publicly and instantly recognizing learner efforts, typically in the form of leaderboards which rank learners according to certain criteria (Buckley & Doyle, 2016). In addition to promoting competition with others, gamification can also encourage self-competition via study streaks that require continuous studying to keep them going (e.g., Loewen et al., 2019). Dehghanzadeh et

al.'s (2019) review of 22 studies on gamification in EFL apps revealed that leaderboards were among the most common gamification elements and promoted engagement (e.g., Homer et al., 2018), motivation (e.g., Hung, 2017), or both (e.g., Zarzycka-Piskorz, 2016). Notably, most studies in Dehghanzadeh et al.'s (2019) review adopted questionnaires and interviews to examine the effects of leaderboards, thus reflecting mostly learner perceptions rather than behavioral engagement. Although research has generally supported the benefits of gamification, the low behavioral engagement in the abovementioned app-based L2 studies (Duolingo; García Botero et al., 2019; Loewen et al., 2019) suggests relying on gamification alone may not be sufficient to stimulate learner engagement.

To better understand the issue of low engagement with app-based L2 learning, several studies collected learner perceptions and have concluded that teacher support is crucial. Dashtestani (2016) surveyed over 300 EFL learners in Iran and interviewed over 100 for proposed strategies to implement MALL, and one recurring theme was learners' expectations of teacher support and encouragement. Similarly, Qian and Tang (2018) elicited responses from 148 Chinese EFL learners on their perceptions of MALL, and found almost all respondents expressed the need for teacher support and involvement during out-of-class MALL. American L2 learners in Hanson and Brown's (2020) study also attributed their low engagement with Anki to the lack of teacher support, and they requested more teacher involvement. These findings suggest that regardless of cultural backgrounds (e.g., Iranian, Chinese, American), L2 learners often feel a need for teacher support when engaging in app-based L2 learning.

Overall, the empirical studies discussed above accorded with the review articles summarized earlier, highlighting the persistent issue of low learner engagement (Chwo et al., 2018; Elaish et al., 2019; Stockwell, 2016) and the critical role of L2 instructors (Chwo et al.,

2018; Reinder & Stockwell, 2017) during MALL. As a response, our study draws on Mercer and Dörnyei's (2020) proposal of goal-setting with feedback to explore effective L2 pedagogy for MALL.

2.3. Goals and Feedback to Engage and Motivate L2 Learners

Given the importance of learner motivation for L2 development (see Boo et al., 2015 for review), L2 teachers' motivational strategies have received much attention in research (see Lamb, 2017 for review). Recently, noting the insufficiency of the motivation construct, Mercer and Dörnyei (2020) proposed learner engagement as another direction for exploring effective L2 teaching strategies, and highlighted the dual benefits of stimulating engagement: motivating learners and meanwhile realizing their motivation in action. Based on psychological and educational research, Mercer and Dörnyei (2020) further developed the classic, easily-implemented, yet rarely-adopted motivational strategies of goal-setting (Dörnyei, 2001; Oxford & Shearin, 1994), and they recommended goal-setting with feedback to engage and motivate learners in L2 classrooms. Specifically, psychological research on goal-setting theory (Locke & Latham, 1990) has found that setting specific and challenging goals together with goal-related feedback leads to better performance (see Locke & Latham, 2002, 2015 for review). Reviewing educational research, Hattie and Timperley (2007) suggested teachers give feedback on goal attainment, including direct (e.g., assignment completion) and comparative feedback (e.g., better performance than others), so that learners can identify gaps between their learning goals and their current progress, which may stimulate their motivation and engagement to reduce the gaps.

In L2 research, Tremblay and Gardner (1995) were among the earliest to draw on goal-setting theory (Locke & Latham, 1990) to empirically explore the motivational effects of goals in L2 learning. They found goals that were specific and checked often led to more motivational

behaviors. More recently, Moeller et al. (2012) developed the LinguaFolio goal-setting process based on goal theories and explored its effects on 1,273 learners' L2 achievements during five years. The LinguaFolio goal-setting process consisted of learners setting their own learning goals, creating action plans, and reflecting on the goals at the end of the unit. Results showed better performance in completing the goal-setting process was associated with higher L2 achievements. Moeller et al. (2012) suggested that having learners set their own goals can facilitate learner autonomy and stimulate learner engagement, by allowing them to take “a step forward reaching their own aspirations” (p. 155) instead of just fulfilling course requirements. Another recent study by Shih and Reynolds (2018) also found the facilitative effects of goal-setting activities on motivation and EFL reading development during one year. Notably, some L2 apps (e.g., Memrise) offer goal-setting functions, allowing learners to decide how much to study each day. Well-grounded in goal theories, goal-setting strategies offer much potential to promote learner motivation and engagement (see Lee & Bong, 2019 for review of goal theories in L2 research and education). However, to our knowledge, no studies have applied Mercer and Dörnyei's (2020) latest proposal of goal-setting with feedback to exploring learner engagement in app-based L2 learning, which provides the rationale for our current study.

3. Research Questions

This study makes an initial attempt to investigate and demonstrate how L2 instructors can implement Mercer and Dörnyei's (2020) proposal of goal-setting with feedback during teaching in order to support app-based L2 vocabulary self-study as a course assignment. Specifically, we adopted a quasi-experimental design to compare learner engagement and L2 skill development of a treatment group who completed goal-setting-and-checking activities with a control group who did not. We also explored learner perceptions of two feedback tools, Unfinished Lists (lists of

students who do not finish the app-based self-study assignments) and leaderboards (student rankings based on assignment performance), in addition to the goal-setting-and-checking activities. Lastly, to further guide L2 teachers in adapting the goal-setting-and-checking activities, we examined the treatment group's general patterns in setting and meeting their goals. These research questions (RQs) guided this study:

RQ1. How does learner engagement in app-based L2 vocabulary self-study differ between treatment and control groups?

RQ2. How does L2 skill development differ between treatment and control groups?

RQ3. How do learners perceive the goal-setting-and-checking activities (treatment group) and the two feedback tools, Unfinished Lists and leaderboards (both groups)?

RQ4. How does the treatment group set and meet the weekly study goals?

4. Method

4.1. Participants

Participants were 63 Japanese college learners of English who were enrolled in four sections of the same upper-intermediate EFL course. Two sections were randomly assigned to the treatment group ($n = 32$) and two to the control group ($n = 31$).

4.2. Materials

4.2.1. Memrise

Memrise is a flashcard app that uses spaced repetition algorithms to schedule vocabulary review in order to facilitate word retention (Bower & Rutson-Griffiths, 2016; Hanson & Brown, 2020). For this study, accounts were created for participants to access a free course that was based on Browne and Culligan's (2016) TOEIC Service List 1.1 with Japanese definitions from Raine (2017) (English-Japanese; <https://app.memrise.com/course/2172994/toEIC-service-list-11->

ying-he/). This course was chosen to help participants prepare for TOEIC tests as part of the university curriculum goals. To study a new word in this course, participants started with the L2-L1 paired-associate (see Figure 1a), and then completed recognition (L1-L2, L2-L1; see Figure 1b, 1c) and recall (L1-L2; see Figure 1d) tasks as practice. Review modules were then scheduled by the system to further practice the newly studied words with recall and recognition tasks. Participants received points based on their task performance: the more tasks and the better they did, the more points they received. Notably, Memrise user logs automatically record how many words have been studied and reviewed.

[Insert Figure 1 about here]

4.2.2. Goal-Setting-and-Checking Sheets, Unfinished Lists, and Leaderboards

For the treatment group, a Goal-Setting-and-Checking Sheet (see Figure 2) was designed for participants to set their study goals for the upcoming week(s): the number of words per day, the number of days per week, and the total number of words per week. Participants also checked whether they had met their previous study goals and if not, they needed to explain why not in Japanese.

[Insert Figure 2 about here]

As direct feedback (Hattie & Timperley, 2007) for the treatment and control groups, an Unfinished List (see Figure 3) was compiled with student ID numbers to remind participants who did not complete the assignment. To provide comparative feedback (Hattie & Timperley, 2007), Memrise automatically generates leaderboards by ranking learners within the same group according to their number of points. To offer a separated leaderboard for each section, we used a free account to create four groups for participants to join. We could then show the leaderboard for each section (see Figure 4).

[Insert Figures 3 & 4 about here]

4.2.3. Learner Perception Survey

To collect learner perceptions of the goal-setting-and-checking activities, Unfinished Lists, and leaderboards, we developed an online survey to ask participants to first rate their extent of agreement on a 5-point Likert scale (1-“Strongly Disagree;” 5-“Strongly Agree”) for the statement “I think X is helpful” for each pedagogical intervention. Then they provided reasons for their ratings. Notably, only the treatment group responded to questions on the goal-setting-and-checking activities.

4.3. Procedure

This study adopted a convergent mixed-methods research design that combines quantitative and qualitative analyses to answer the RQs (Creswell & Plano Clark, 2018), and a between-subject pre/post-test design (see Supplementary Materials for details). Participants took TOEIC listening-and-reading tests in September and December as part of the university curriculum requirement, and the two test scores indexed EFL listening and reading proficiency before (pretest) and after using Memrise (posttest). In the week of the September TOEIC test, participants were introduced to Memrise and installed the mobile app. The following week, participants received Memrise training in class, including how to check the number of words studied. Then they started using it for after-class self-study as a course assignment. After using Memrise for eight weeks, participants completed the perception survey and took the December TOEIC test.

For both the control and treatment groups, Unfinished Lists and leaderboards were shown in class weekly (or biweekly when there were no classes in a previous week). The control group was assigned to study five words per day for seven days a week, based on Hanson and Brown’s

(2020) requirement of studying at least five minutes per day for at least five days per week. One day before class, the number of words each participant studied was checked against the 35-word weekly requirement. Differently, the treatment group decided how many words (at least 6) per day and how many days (at least 5) per week as their study goals, with at least 30 words per week to fulfil the assignment requirement. Notably, the number of words participants studied was tracked by Memrise user logs and indexed learner engagement.

The different numbers of words required were intentional: if the control group was able to meet the 35-word requirement and if the treatment group did the required 30-word minimum only, the control group would surpass the treatment group in the number of words studied. Contrarily, if the treatment group studied more words than the control group, the effectiveness of goal-setting-and-checking activities on stimulating learner engagement would be supported.

4.4. Data Analysis

For RQ1, we calculated bootstrapped descriptive statistics with 10,000 samples (LaFlair et al., 2015; Larson-Hall, 2016) for the number of words studied weekly. We also plotted these numbers at group and individual levels. Since the number of words studied came from the same participants (within-subjects) in two different groups (between-subjects), we followed Larson-Hall's (2016) recommendation to conduct robust between-within ANOVA, using the WRS2 package (v1.1-1; Mair & Wilcox, 2020) with R (v4.0.2; R Core Team, 2019). We adopted the bootstrapping method (10,000 samples) to calculate the estimates and *p* values of the between-subjects and within-subjects effects as well as their interaction.

Regarding RQ2, bootstrapped descriptive statistics were calculated for the TOEIC scores. Robust between-within ANOVA was conducted with WRS2 in R to assess bootstrapped main effects for between-subjects (treatment vs. control group) and within-subjects (September vs.

December TOEIC) variables as well as any interaction effect.

For RQ3, quantitative and qualitative analyses were conducted for the ratings and reasons with bootstrapped descriptive statistics and thematic analysis (Braun & Clarke, 2012), respectively. For each pedagogical intervention, the reasons were first grouped into positive, neutral, and negative categories, and recurring themes were summarized and double-checked. Then the number and the percentage of participants for each theme were calculated, with representative examples to illustrate the themes.

To answer RQ4, we calculated bootstrapped descriptive statistics for the raw numbers of weekly study goals and the percentages of meeting them as reported by participants. Summary statistics were also calculated for whether the weekly study goals were different and beyond the required minimums. Lastly, we created plots to contrast the number of words actually studied versus the targeted study goals.

5. Results

5.1. RQ1: Higher Engagement of Treatment over Control Group

Table 1 shows bootstrapped descriptive statistics for the number of words studied weekly, and Figure 5 plots these numbers at group (bolded lines) and individual levels. One notable characteristic is the wide range of numbers (0-341.5) and the large standard deviations (SDs), indicating that learner engagement with Memrise varied widely within groups across weeks. When checking the number of words studied weekly at group level (group means) against the assignment requirement, the control group failed to meet the 35-word target in most weeks except Week 6. In contrast, the treatment group consistently exceeded the 30-word minimum requirement in all weeks and even doubled the target in Weeks 3, 4/5, and 6.

When comparing the two groups, the weekly group mean was larger for the treatment

than the control group in all weeks. According to the interpretation that non-overlapping 95% confidence intervals (CIs) suggest statistically significant differences (Larson-Hall, 2016), the two groups differed significantly in the number of words studied in Weeks 2, 3, and 4/5, because the 95% CIs did not overlap in those weeks. Results of robust between-within ANOVA showed a statistically significant effect of group (Control – Treatment, $estimate = -20.53, p < .001$), but not for week ($p = .771$) or the interaction between group and week ($p = .078$). These results suggest that the treatment group studied more words than the control group generally, despite the lower minimum requirement of the treatment group.

[Insert Table 1 & Figure 5 about here]

5.2. RQ2: Stable and Similar TOEIC Scores Between Groups

Table 2 presents bootstrapped descriptive statistics for the TOEIC scores in September (pretest) and December (posttest). As a preliminary index of group differences, 95% CIs of both groups overlapped considerably, implying that the TOEIC scores did not differ between the two groups in either pretest or posttest. For each group, their TOEIC scores did not change from pretest to posttest, either. Results of robust between-within ANOVA also showed statistically non-significant differences between groups (Control – Treatment, $estimate = -46.33, p = .096$), between the two test times (September – December, $estimate = 8.73, p = .283$), or in group-time interaction ($estimate = -25.86, p = .135$). These findings suggest that both groups had similar EFL listening and reading proficiency before using Memrise, and that their proficiency levels did not change significantly after eight weeks.

[Insert Table 2 about here]

5.3. RQ3: Overall Positive Feedback on All Pedagogical Interventions

Table 3 presents bootstrapped descriptive statistics for the helpfulness ratings on the goal-

setting-and-checking activities, Unfinished Lists, and leaderboards. All pedagogical interventions received an average rating higher than 3 (“Neither Agree or Disagree”), indicating that they were regarded by learners as helpful to some extent. Goal-setting-and-checking activities were rated the most helpful, followed by Unfinished Lists and leaderboards. When checking the 95% CIs for indication of statistical differences, goal-setting-and-checking activities received significantly higher ratings than leaderboards, with no overlap between their 95% CIs. Other comparisons did not have statistically significant differences as shown in the partially overlapping 95% CIs.

[Insert Table 3 about here]

Learners’ comments about their ratings were mostly positive for the goal-setting-and-checking activities. They also gave mainly positive and some neutral feedback on the Unfinished Lists. Regarding the leaderboards, perceptions ranged from positive to negative.

Goal-setting-and-checking activities were commended for boosting motivation by 11 learners (34%) in the treatment group. These activities also offered a sense of obligation and ownership to 7 (22%) and 6 (19%) learners, respectively. However, these activities were not always effective (1, 3%). Excerpt 1 shows goal ownership afforded motivation, whereas Excerpt 2 indicates a lack of motivation.

Excerpt 1) *Since it was the goal I set by myself, I was able to work hard toward that goal.*

Excerpt 2) *[M]aybe it is motivat[ing] for someone but maybe there [is] someone [does]n’t. [I’m] the latter, sorry.*

Unfinished Lists also received mainly positive feedback for boosting motivation (7, 12%), providing a sense of urgency (7, 12%) and obligation (3, 5%), and informing self-performance (11, 19%). Interestingly, whereas some learners tried to avoid being listed (2, 3%),

others did not worry about it (2, 3%) or felt no need to know others' progress (2, 3%). Excerpts 3, 4, and 5 show the decreasing levels of attention paid to the Unfinished Lists.

Excerpt 3) *I always engaged in my study to avoid being listed.*

Excerpt 4) *[S]ometimes I was listed but I did not worry about it that much.*

Excerpt 5) *I thought I didn't need to know others' progress.*

Leaderboards were often commented on positively for boosting motivation (10, 16%), stimulating a sense of competition (11, 18%), and providing information about self (3, 5%) and others' (3, 5%) performance. Nonetheless, some learners did not care about the leaderboards (8, 13%), and others disliked them (2, 3%). Excerpts 6, 7, and 8 show learners' attitudes varying from positive to negative towards the leaderboards.

Excerpt 6) *Since I felt that I was compared with others around me, the sense of competition pushed me to work on it.*

Excerpt 7) *Because I didn't care about the ranking, I stucked with my own learning pace.*

Excerpt 8) *[A]ctually, I don't want to be compared with someone.*

To summarize, findings from both quantitative ratings and qualitative reasons indicate that learners perceived the goal-setting-and-checking activities, Unfinished Lists, and leaderboards as generally helpful, although the leaderboards encountered resistance from some learners.

5.4. RQ4: General Patterns in Setting and Meeting Goals

Table 4 shows bootstrapped descriptive statistics for the treatment group's weekly study goals and the percentages of meeting these goals as reported by learners. On average, learners aimed to study 9 (vs. minimum 6) words per day, 5 (vs. minimum 5) days per week, and totally 46 (vs. minimum 30) words per week. Over 70% reported meeting their weekly study goals.

[Insert Table 4 about here]

Summary statistics are shown in Table 5 to compare the weekly study goals throughout the weeks and against the 30-word minimum. About 50% of learners set the same number of words per day, 63% the same number of days per week, and 44% the same total number of words per week. Over 70% set weekly study goals higher than the 30-word minimum.

[Insert Table 5 about here]

Figure 6 plots the number of words actually studied and targeted as study goals at group (bolded lines) and individual levels. Compared with the studied numbers, the targeted numbers were more stable throughout the weeks, and some learners studied more than the targeted number of words.

[Insert Figure 6 about here]

6. Discussion

6.1. Effectiveness of Goal-setting with Feedback

Findings from both group comparisons and learner perceptions support the effectiveness of goal-setting-and-checking activities on stimulating learner engagement in app-based L2 vocabulary self-study.

For the control group, although app-based L2 vocabulary self-study was incorporated as a course assignment, learners failed to meet the 35-word requirement in all weeks except one. This result echoes previous findings that assigning app-based L2 learning as part of coursework is not sufficient to stimulate learner engagement (Milliner & Cote, 2015; Hanson & Brown, 2020). The current study goes one step further in showing that coupling the assignment requirement with feedback through assignment completion (e.g., Unfinished Lists) and class rankings (e.g., leaderboards) (Hattie & Timperley, 2007) still failed to mitigate low engagement, mirroring

previous results on gamification's (e.g., leaderboards) insufficiency to stimulate learner engagement with app-based L2 learning (García Botero et al., 2019; Loewen et al., 2019).

In contrast to the control group, the treatment group who completed the additional goal-setting-and-checking activities, consistently fulfilled the 30-word minimum requirement. This finding demonstrates the effects of goal-setting-and-checking activities on motivating and engaging L2 learners as previous L2 literature suggested (Dörnyei, 2001; Mercer & Dörnyei, 2020; Oxford & Shearin, 1994; Shih & Reynolds, 2018). The treatment group's continuing engagement may have been stimulated by weekly specifying a target number of words to study and checking those targets each week, echoing Tremblay and Gardner's (1995) finding that motivational behaviors positively correlated with goals that were specific and often checked. Evaluating self-set study goals and providing reasons for not meeting them may have also enabled the treatment group to critically reflect on their previous study behaviors and encouraged them to adjust subsequent behaviors in order to achieve their study goals (Hattie & Timperley, 2007; Moeller et al., 2012). Additionally, the treatment group not only attained the 30-word minimum requirement but also surpassed the control group several times in the number of words studied weekly, even though the required number was slightly higher for the control group. This result shows the power of goal-setting-and-checking activities in motivating and engaging learners to actively work for their own aspirations rather than just passively fulfilling course requirements (Moeller et al., 2012).

Regarding learner perceptions, the goal-setting-and-checking activities were mostly considered as helpful, as learners explained setting their own study goals gave them a sense of ownership and made them feel obliged to attain what they aimed for, which may lead to further development of learner autonomy in regulating L2 learning (Moeller et al., 2012). The

Unfinished Lists and leaderboards, although not always positively commented, were regarded as helpful quite often.

Combining the quantitative and qualitative findings, this study supports the effectiveness of goal-setting-and-checking activities and the benefits of supplementing them with Unfinished Lists and leaderboards for stimulating learner engagement in app-based L2 vocabulary self-study. This study constitutes the first attempt to apply Mercer and Dörnyei's (2020) latest proposal of goal-setting with feedback to practical classroom teaching, and provides L2 instructors with ready-to-use activities that can easily be adapted into their own classes (see 6.3. for detailed suggestions), advancing research on goal theories in L2 educational contexts (Lee & Bong, 2019).

6.2. App-based L2 Vocabulary Learning and L2 Skill Development

This study did not find statistically significant differences in the TOEIC scores from pretest to posttest between the control and treatment groups or within each group. One possible explanation for this result might be the short duration (eight weeks) of app-based L2 learning in this study. For example, Bower and Rutson-Griffiths (2016) detected increased TOEIC scores after ten months of app-based L2 vocabulary learning. Another reason could be that TOEIC listening-and-reading tests may not have been sensitive enough to capture any incremental L2 vocabulary development resulting from app-based learning in this study. Additionally, the December TOEIC test encountered some logistical issues: the test was originally planned to take place face-to-face, but due to worsening coronavirus pandemic conditions, the test was moved online just two weeks before the test dates, which created scheduling problems. Some learners reported less than ideal test-taking conditions, such as taking the test in a noisy cafeteria or joining the test after it started. These factors may have had a negative impact on test

performance.

6.3. Integrating App-based L2 Learning into L2 Teaching

Our study affords an example of integrating app-based L2 learning into L2 teaching and here we would like to offer some guidance to teachers. Given the potential of app-based learning to extend formal classroom instruction with informal self-study (Levy & Kennedy, 2010), we incorporated the activity of studying TOEIC words with Memrise as an after-class assignment, which was in line with the curriculum goal to prepare learners for TOEIC tests. In considering app-based L2 learning, teachers are recommended to choose materials that are closely connected to curriculum goals (Chwo et al., 2018). Also, we designed Goal-Setting-and-Checking Sheets for learners to self-set and self-check their study goals and to explain why, if they did not meet their goals. Teachers may decide minimum targets for study goals and then encourage learners to set challenging goals beyond the minimum (Dörnyei, 2001; Mercer & Dörnyei, 2020). However, even if learners set easily attainable study goals throughout the semester, our results indicate that learners might exceed these goals and that they might set easier goals to provide a cushion for meeting them.

Providing feedback in the form of Unfinished Lists and leaderboards was generally viewed as helpful for engaging learners in app-based L2 learning. Our results suggest that whereas most learners paid attention to the Unfinished Lists, some did not care, so teachers may consider giving additional reminders to those who appear on the Unfinished Lists several times. Leaderboards are often automatically generated by apps, and teachers may also consider creating by themselves. When choosing ranking criteria for leaderboards, effort indices (e.g., the number of study sessions) would be better than result indices (e.g., test score) in encouraging continuous engagement (Dörnyei, 2001). Although competition associated with leaderboards is generally

beneficial (Buckley & Doyle, 2016), we found some learners did not like or care about being compared to others. Teachers are recommended to emphasize in class that leaderboards are not merely about winning or losing but more importantly, about friendly competition to improve L2 learning. Additionally, showing Unfinished Lists and leaderboards publicly in class may negatively affect self-worth of learners who do not perform as well (Covington, 1992). Teachers are advised to carefully assess learners' feelings about these feedback tools and adopt culturally appropriate alternatives, such as displaying less identifiable information (e.g., student ID number) or giving individual feedback privately.

We believe goal-setting-and-checking activities, along with feedback in the form of Unfinished Lists and leaderboards, can support teachers in stimulating learner engagement in app-based L2 learning. With the abovementioned guidance, these pedagogical interventions can be adapted to both face-to-face and online classes by slightly modifying the delivery method (via classroom projectors or online screen sharing). Notably, every teacher may face a unique group of learners in a different teaching setting, so they are strongly recommended to pay close attention to the characteristics of their local contexts and make proper adjustments to these pedagogical interventions. Additionally, teachers may collect learner feedback at the end of the course in the form of ratings and comments, so as to better understand learners' needs and enhance the positive influence of these pedagogical interventions.

6.4. Limitations and Future Directions

Despite the meaningful findings, this study has a few limitations. Due to the university's TOEIC test schedule, app-based L2 learning lasted only eight weeks, which is relatively short. Future research can investigate longer learning periods. Additionally, this study did not measure L2 vocabulary knowledge before or after app-based learning, so we were not able to investigate

the relationships between vocabulary gains and learner engagement. This study did not explore the relationships between TOEIC scores and app usage statistics (e.g., study time), either. Future studies can measure vocabulary gains by employing lexical tests to examine how they relate to learner engagement, and they can also investigate how TOEIC scores are associated with app usage statistics.

7. Conclusion

Focusing on app-based L2 vocabulary self-study as a course assignment, this study investigated the effectiveness of goal-setting-and-checking activities and the benefits of Unfinished Lists and leaderboards as supplementary feedback tools for stimulating learner engagement. Results from comparing the number of words studied weekly by the control and the treatment groups showed goal-setting-and-checking activities were effective, although TOEIC scores did not differ between or change within groups significantly. Findings from learner ratings and comments supported goal-setting-and-checking activities, Unfinished Lists, and leaderboards as generally helpful. Through this study, we made the first attempt to apply Mercer and Dörnyei's (2020) latest proposal of goal-setting with feedback to L2 classroom instruction, and provided ready-to-use activities and practical suggestions for L2 teachers to adapt the three pedagogical interventions into their own teaching. More generally, this study responds to the call for bringing MALL from the fringes towards the core of L2 education (Burston, 2014) and for exploring instructors' role in supporting MALL (Chwo et al., 2018; Reinders & Stockwell, 2017), and it also enriches the lively investigation on motivation and engagement in MALL (Chwo et al., 2018; Elaish et al., 2019; Stockwell, 2016). Last but not least, this study further contributes to evidence-based L2 vocabulary instruction (He & Godfroid, 2019) and ISLA research's goal of strengthening the research-pedagogy dialogue (Sato & Loewen, 2019).

References

- Agawa, G., Black, G., & Herriman, M. (2011). Effects of web-based vocabulary training for TOEIC. In A. Stewart (Ed.), *JALT2010 Conference Proceedings*. JALT: Tokyo.
- Bodnar, S., Cucchiarini, C., Strik, H., & van Hout, R. (2016). Evaluating the motivational impact of CALL systems: Current practices and future directions. *Computer Assisted Language Learning*, 29(1), 186-212.
- Boo, Z., Dörnyei, Z., & Ryan, S. (2015). L2 motivation research 2005–2014: Understanding a publication surge and a changing landscape. *System*, 55, 145-157.
- Boudadi, N. A., & Gutiérrez-Colón, M. (2020). Effect of Gamification on students' motivation and learning achievement in second language acquisition within higher education: A literature review 2011-2019. *The EuroCALL Review*, 28(1), 57-69.
- Bower, J. V., & Rutson-Griffiths, A. (2016). The relationship between the use of spaced repetition software with a TOEIC word list and TOEIC score gains. *Computer Assisted Language Learning*, 29(7), 1238-1248.
- Braun, V., & Clarke, V. (2012). Thematic analysis. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological* (pp. 57–71). Washington, DC: American Psychological Association.
- Browne, C. & Culligan, B., (2016). TOEIC service list 1.1. Retrieved from <http://www.newgeneralservicelist.org/toeic-list>
- Buckley, P., & Doyle, E. (2016). Gamification and student motivation. *Interactive Learning Environments*, 24(6), 1162-1175.
- Burston, J. (2014). The reality of MALL: Still on the fringes. *CALICO Journal*, 31(1), 103-125.

- Christenson, S. L., Reschly, A. L., & Wylie, C. (2012). *Handbook of research on student engagement*. Springer.
- Chwo, G. S. M., Marek, M. W., & Wu, W. C. V. (2018). Meta-analysis of MALL research and design. *System, 74*, 62-72.
- Covington, M. (1992). *Making the grade: A self-worth perspective on motivation and school reform*. Cambridge: Cambridge University Press.
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE.
- Dashtestani, R. (2016). Moving bravely towards mobile learning: Iranian students' use of mobile devices for learning English as a foreign language. *Computer Assisted Language Learning, 29*(4), 815-832.
- Dehghanzadeh, H., Fardanesh, H., Hatami, J., Talaei, E., & Noroozi, O. (2019). Using gamification to support learning English as a second language: A systematic review. *Computer Assisted Language Learning, 1-24*.
- Dörnyei, Z. (2001). *Motivational strategies in the language classroom*. Cambridge: Cambridge University Press.
- Dörnyei, Z. (2019). Towards a better understanding of the L2 Learning Experience, the Cinderella of the L2 Motivational Self System. *Studies in Second Language Learning and Teaching, 9*(1), 19-30.
- Elaish, M. M., Shuib, L., Ghani, N. A., & Yadegaridehkordi, E. (2019). Mobile English language learning (MELL): A literature review. *Educational Review, 71*(2), 257-276.
- García Botero, G., Questier, F., & Zhu, C. (2019). Self-directed language learning in a mobile-assisted, out-of-class context: Do students walk the talk? *Computer Assisted Language*

- Learning*, 32(1-2), 71-97.
- Hanson, A. E. S., & Brown, C. M. (2020). Enhancing L2 learning through a mobile assisted spaced-repetition tool: An effective but bitter pill? *Computer Assisted Language Learning*, 33(1-2), 133-155.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.
- He, X., & Godfroid, A. (2019). Choosing words to teach: A novel method for vocabulary selection and its practical application. *TESOL Quarterly*, 53(2), 348-371.
- Homer, R., Hew, K. F., & Tan, C. Y. (2018). Comparing digital badges-and-points with classroom token systems: Effects on elementary school ESL students' classroom behavior and English learning. *Journal of Educational Technology & Society*, 21(1), 137–151.
- Hung, H. T. (2017). Clickers in the flipped classroom: Bring your own device (BYOD) to promote student learning. *Interactive Learning Environments*, 25(8), 983–995.
- Kapp, K. M. (2012). *The gamification of learning and instruction: Game-based methods and strategies for training and education*. San Francisco, CA: John Wiley & Sons.
- Kondo, M., Ishikawa, Y., Smith, C., Sakamoto, K., Shimomura, H., & Wada, N. (2012). Mobile assisted language learning in university EFL courses in Japan: Developing attitudes and skills for self-regulated learning. *ReCALL*, 24(2), 169-187.
- LaFlair, G., Egbert, J., & Plonsky, L. (2015). A practical guide to bootstrapping descriptive statistics, correlations, t tests, and ANOVAs. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 46–77). New York, NY: Routledge.
- Lamb, M. (2017). The motivational dimension of language teaching. *Language Teaching*, 50(3),

301-346.

- Larson-Hall, J. (2016). *A guide to doing statistics in second language research using SPSS and R*. London, England: Routledge.
- Lee, M., & Bong, M. (2019). Relevance of goal theories to language learning research. *System*.
- Levy, M., & Kennedy, C. (2010). Materials development in three Italian CALL projects: Seeking an optimal mix between in-class and out-of-class learning. *CALICO Journal*, 27(3), 529-539.
- Lin, J. J., & Lin, H. (2019). Mobile-assisted ESL/EFL vocabulary learning: A systematic review and meta-analysis. *Computer Assisted Language Learning*, 32(8), 878-919.
- Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting and task performance*. Prentice-Hall, Inc.
- Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, 57(9), 705-717.
- Locke, E. A., & Latham, G. P. (2015). Breaking the rules: a historical overview of goal-setting theory. In *Advances in motivation science* (Vol. 2, pp. 99-126). Elsevier.
- Loewen, S., Crowther, D., Isbell, D. R., Kim, K. M., Maloney, J., Miller, Z. F., & Rawal, H. (2019). Mobile-assisted language learning: A Duolingo case study. *ReCALL*, 31(3), 293-311.
- Loewen, S., Isbell, D. R., & Sporn, Z. (2020). The effectiveness of app-based language instruction for developing receptive linguistic knowledge and oral communicative ability. *Foreign Language Annals*, 53(2), 209-233.
- Mair, P., & Wilcox, R. (2020). Robust statistical methods in R using the WRS2 package. *Behavior Research Methods*, 52, 464-488.

- Martin, A. J. (2012). Part II commentary: Motivation and engagement: Conceptual, operational, and empirical clarity. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 303-314). Springer.
- Mercer, S., & Dörnyei, Z. (2020). *Engaging language learners in contemporary classrooms*. Cambridge University Press.
- Milliner, B., & Cote, T. (2015). Mobile-based extensive reading: An investigation into reluctant readers. *International Journal of Computer-Assisted Language Learning and Teaching*, 5(4), 1-15.
- Moeller, A. J., Theiler, J. M., & Wu, C. (2012). Goal setting and student achievement: A longitudinal study. *The Modern Language Journal*, 96(2), 153-169.
- Nation, I. S. P. (2013). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nielson, K. B. (2011). Self-study with language learning software in the workplace: What happens? *Language Learning & Technology*, 15(3), 110-129.
- Oxford, R., & Shearin, J. (1994). Language learning motivation: Expanding the theoretical framework. *The Modern Language Journal*, 78(1), 12-28.
- Phillips, J. (2011). An investigation into the effect of targeted vocabulary learning using a spaced repetition flashcard system on TOEIC scores. *Aoyamagakuin Joshi Tankidaigaku Bulletin*, 65, 55-61.
- Philp, J., & Duchesne, S. (2016). Exploring engagement in tasks in the language classroom. *Annual Review of Applied Linguistics*, 36, 50-72.
- Qian, K., & Tang, J. (2018). Researching mobile-assisted English language learning among adult distance learners in China: Emerging practices and learner perception of teacher

- role. *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT)*, 8(3), 1-28.
- R Core Team. (2019). R: A language and environment for statistical computing [software]. Vienna, Austria: R Foundation for Statistical Computing.
- Raine, P. (2017). Rankings, definitions, pronunciations and additional data for NGSL, NAWL, TSL, and BSL. Retrieved from <https://paulsensei.com/2017/04/04/rankings-definitions-pronunciations-and-additional-data-for-ngsl-nawl-tsl-and-bsl/>
- Read, T., & Kukulska-Hulme, A. (2015). The role of a mobile app for listening comprehension training in distance learning to sustain student motivation. *Journal of Universal Computer Science*, 21(10), 1327-1338.
- Reinders, H., & Pegrum, M. (2016). Supporting language learning in the move: An evaluative framework for mobile language learning resources. In B. Tomlinson (Ed.), *SLA research and materials development for language learning* (pp. 221-233). Routledge.
- Reinders, H., & Stockwell, G. (2017). Computer-assisted SLA. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 361-375). Routledge.
- Rosell-Aguilar, F. (2018). Autonomous language learning through a mobile application: A user evaluation of the busuu app. *Computer Assisted Language Learning*, 31(8), 854-881.
- Sato, M., & Loewen, S. (2019). Do teachers care about research? The research–pedagogy dialogue. *ELT Journal*, 73(1), 1-10.
- Shih, Y. C., & Reynolds, B. L. (2018). The effects of integrating goal setting and reading strategy instruction on English reading proficiency and learning motivation: A quasi-experimental study. *Applied Linguistics Review*, 9(1), 35-62.

- Shortt, M., Tilak, S., Kuznetcova, I., Martens, B., & Akinkuolie, B. (2021). Gamification in mobile-assisted language learning: A systematic review of Duolingo literature from public release of 2012 to early 2020. *Computer Assisted Language Learning*, 1-38.
- Skinner, E. A., Furrer, C., Marchand, G., & Kindermann, T. (2008). Engagement and disaffection in the classroom: Part of a larger motivational dynamic? *Journal of Educational Psychology*, 100(4), 765-781.
- Stockwell, G. (2016). Mobile language learning. In F. Farr & L. Murray (Eds.), *The Routledge handbook of language learning and technology*. Routledge.
- Tai, Y., & Ting, Y. L. (2011). Adoption of mobile technology for language learning: Teacher attitudes and challenges. *The JALT CALL Journal*, 7(1), 3-18.
- Tremblay, P. F., & Gardner, R. C. (1995). Expanding the motivation construct in language learning. *The Modern Language Journal*, 79(4), 505-518.
- Wishart, J. (2008). Challenges faced by modern foreign language teacher trainees in using handheld pocket PCs (Personal Digital Assistants) to support their teaching and learning. *ReCALL*, 20(3), 348-360.
- Zarzycka-Piskorz, E. (2016). Kahoot it or not? Can games be motivating in learning grammar? *Teaching English with Technology*, 16(3), 17-36.



a

b

c

d

Figure 1. Memrsie screenshots.

Student No.	*****	Name	****	Last Week	226		
Term & Week	Words/Day	Met? (Yes/No)	Days/Week	Met (Yes/No)	Words/Week	Met (Yes/No)	If no, why? (Please write in Japanese.)
T3 · W4	6	Yes		5	Yes	30	Yes
T3 · W5	8	No		5	No	40	No
T3 · W6	7	Yes		5	Yes	35	Yes
T3 · W7	8	No		5	No	40	No
T4 · W8	8	No		5	No	40	No
T4 · W1	7	No		5	No	35	No
T4 · W2	7			7		49	
T4 · W3	7			7		49	

Figure 2. Goal-Setting-and-Checking Sheet screenshot.

Unfinished List: less than 35 words EACH week

Student #	Words Learned	This Week	Last Week
*****	29	66	37
*****	17	47	30
*****	12	42	30
*****	19	41	22
*****	17	36	19
*****	0	8	8
*****	0	0	0

Figure 3. Unfinished List screenshot.





Students	Study time (In the last 30 days)	Words learned	Words healthy in LTM	Difficult Words	Course progress	Points (In the last 30 days)
 *****	1h20m	226	158	15		141384
 *****	6h05m	111	1	8		60770
 *****	44m	60	48	7		39588
 *****	31m	51	45	2		34242
 *****	19m	37	20	14		27200

Figure 4. Leaderboard screenshot.

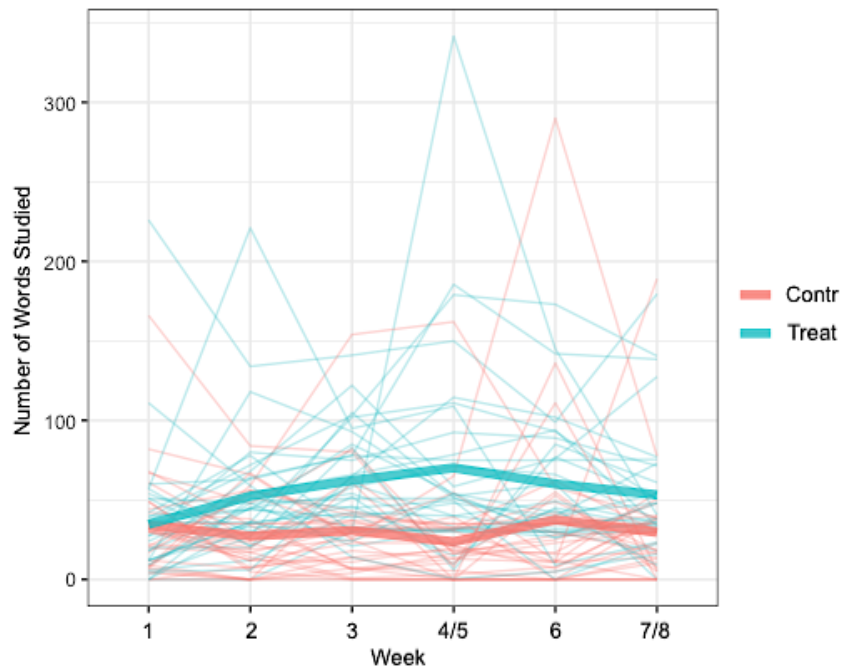


Figure 5. Numbers of words studied weekly. Contr = Control Group. Treat = Treatment Group.
Bolded lines represent group means.

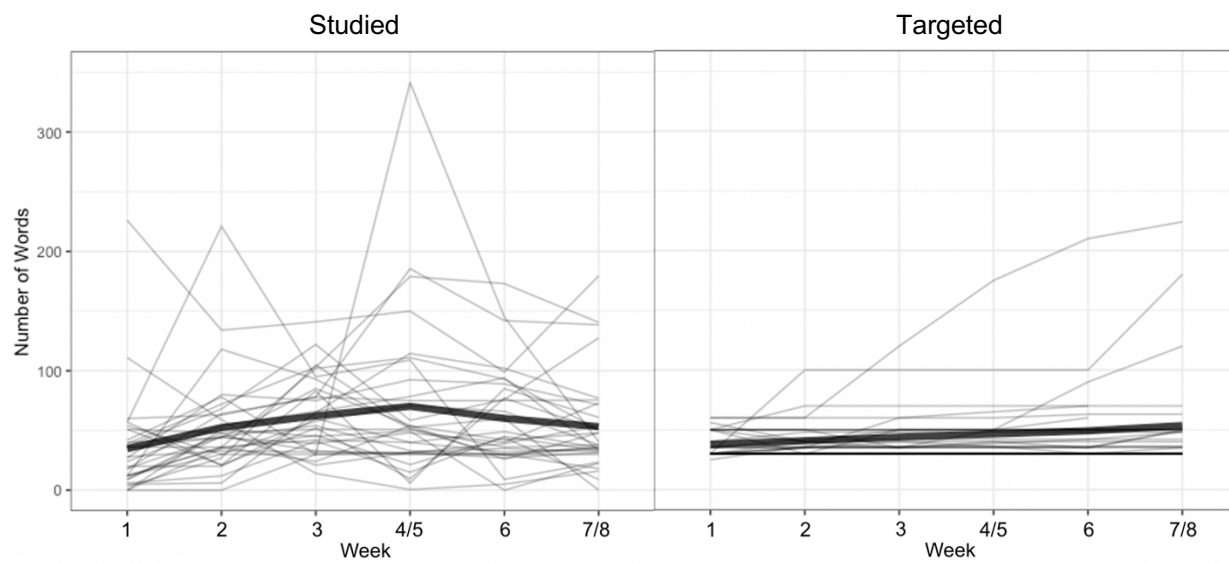


Figure 6. Numbers of words studied vs. targeted weekly. Bolded lines represent group means.

Table 1. *Bootstrapped Descriptive Statistics for Numbers of Words Studied Weekly*

Week	Group	<i>n</i>	<i>Mean</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>BCa 95% CI</i>	
							<i>Lower</i>	<i>Upper</i>
1	Contr	31	34.06	32.40	0	166	24.43	45.57
	Treat	32	34.75	42.41	0	226	23.09	49.36
2	Contr	31	27.26	23.62	0	84	19.55	35.44
	Treat	32	52.47	42.01	0	221	40.34	66.45
3	Contr	31	30.84	32.00	0	154	21.13	42.33
	Treat	32	62.28	31.70	14	141	51.42	73.80
4/5 ¹	Contr	31	23.73	30.63	0	162	15.33	34.50
	Treat	32	70.22	67.78	0.5	341.5	50.06	94.97
6	Contr	31	37.65	56.58	0	290	22.17	58.14
	Treat	32	60.13	41.51	0	173	47.00	75.08
7/8 ¹	Contr	31	29.68	35.21	0	189	19.81	42.65
	Treat	32	53.22	40.82	0	179.5	40.85	67.44
All	Contr	31	30.53	20.60	2	88	23.83	37.99
	Treat	32	55.51	29.99	25.33	154.92	45.93	66.22

Note. Contr = Control Group. Treat = Treatment Group.

¹ Average numbers of words studied weekly were calculated for two weeks.

Table 2. *Bootstrapped Descriptive Statistics for TOEIC Scores*

Time	Group	<i>n</i>	<i>Mean</i>	<i>SD</i>	<i>Min</i>	<i>Max</i> ¹	<i>BCa 95% CI</i>	
							<i>Lower</i>	<i>Upper</i>
Sept	Contr	31	488.87	126.27	245	920	447.76	532.50
	Treat	32	538.91	101.18	350	715	501.86	576.50
Dec	Contr	31	491.29	123.09	265	935	445.96	529.79
	Treat	32	519.38	111.10	300	785	477.76	559.19

Note. Sept = September. Dec = December. Contr = Control Group. Treat = Treatment Group.

¹ The maximum score of TOEIC listening-and-reading tests is 990.

Table 3. *Bootstrapped Descriptive Statistics for Helpfulness Ratings*

	<i>n</i>	<i>Mean</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>BCa 95% CI</i>	
						<i>Lower</i>	<i>Upper</i>
Goal-Setting-and-Checking ¹	32	3.91	1.12	1	5	3.53	4.25
Unfinished List	59	3.44	1.01	1	5	3.20	3.66
Leaderboard	61	3.23	1.16	1	5	2.95	3.49

Note. ¹ This question was shown to the treatment group only.

Table 4. *Bootstrapped Descriptive Statistics for Weekly Goals and Percentages of Meeting Them*

		<i>n</i>	<i>Mean</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>BCa 95% CI</i>	
							<i>Lower</i>	<i>Upper</i>
Goal	Words/Day	32	8.53	3.89	5.88	23	7.42	9.89
	Days/Week	32	5.33	0.56	4.38	6.71	5.15	5.54
	Words/Week	32	46.48	25.96	30	156	39.23	55.43
Met % ¹	Words/Day	32	76.35	23.47	0	100	68.02	84.06
	Days/Week	32	76.46	23.49	0	100	68.13	84.06
	Words/Week	32	73.33	17.52	33.33	100	67.19	79.38

Note. ¹ Goal-checking data was missing for Weeks 7 & 8.

Table 5. *Summary Statistics for Weekly Goals*

	<i>Count</i>	<i>%</i>
Same Words/Day	16	50.00
Same Days/Week	20	62.50
Same Words/Week	14	43.75
Over 30 Words/Week	23	71.88