



ARTICLE

# Effects of a mobile game-based English vocabulary learning app on learners' perceptions and learning performance: A case study of Taiwanese EFL learners

Chih-Ming Chen

Graduate Institute of Library, Information and Archival Studies, National Chengchi University, Taipei City, Taiwan, ROC (chencm@nccu.edu.tw)

Huimei Liu

Department of Statistics, National Chengchi University, Taipei City, Taiwan, ROC (amy1988liu@gmail.com)

Hong-Bin Huang

Graduate Institute of Library, Information and Archival Studies, National Chengchi University, Taipei City, Taiwan, ROC (endless\_1222@hotmail.com)

## Abstract

Many studies have demonstrated that vocabulary size plays a key role in learning English as a foreign language (EFL). In recent years, mobile game-based learning (MGBL) has been considered a promising scheme for successful acquisition and retention of knowledge. Thus, this study applies a mixed methodology that combines quantitative and qualitative approaches to assess the effects of PHONE Words, a novel mobile English vocabulary learning app (application) designed with game-related functions (MEVLA-GF) and without game-related functions (MEVLA-NGF), on learners' perceptions and learning performance. During a four-week experiment, 20 sophomore students were randomly assigned to the experimental group with MEVLA-GF support or the control group with MEVLA-NGF support for English vocabulary learning. Analytical results show that performance in vocabulary acquisition and retention by the experimental group was significantly higher than that of the control group. Moreover, questionnaire results confirm that MEVLA-GF is more effective and satisfying for English vocabulary learning than MEVLA-NGF. Spearman rank correlation results show that involvement and dependence on gamified functions were positively correlated with vocabulary learning performance.

**Keywords:** game-based learning; mobile learning; English vocabulary learning; second language learning; app

## 1. Introduction

As English is the dominant international language and greatly affects the competitiveness of a country around the world, a critical issue in non-English speaking countries is how to enhance English-language skills. In English as a second language (ESL) and English as a foreign language (EFL), learning vocabulary items plays an essential role in all language skills, including listening, speaking, reading, and writing (Nation, 2001). More importantly, Alqahtani (2015) indicated that a limited vocabulary in a second language impedes successful communication. Laufer and Sim

---

Cite this article: Chen, C.M., Liu, H. & Huang, H.B. (2019). Effects of a mobile game-based English vocabulary learning app on learners' perceptions and learning performance: A case study of Taiwanese EFL learners. *ReCALL* 31(2): 170–188, doi:10.1017/S0958344018000228

(1985) emphasized that the most pressing need of foreign-language learners is vocabulary. Wilkins (1972) also claimed that “without grammar very little can be conveyed, without vocabulary nothing at all can be conveyed” (p. 111). Additionally, Read (2000) asserted that vocabulary is the foundation for general language comprehension. Nation (2001) argued that a successful language learner should master roughly 98% of the running words in texts. To read English articles fluently, a learner must understand at least 2,000 commonly used English words (Nation, 1990). It is clear that vocabulary learning is central to English-language acquisition.

Vocabulary can be defined as the words of a language or a lexical unit, including single items and phrases or chunks of several words, which convey a particular meaning (Alfaki, 2015). Namely, vocabulary not only addresses single lexical items – words with specific meaning(s) – but also includes lexical phrases or chunks. Vocabulary acquisition should focus on the process of pronouncing a word correctly, understanding the meaning of a word in a specific context, and remembering the spelling of a word (Ehri & Rosenthal, 2007). According to Nation (1990), the vast gulf separating Chinese characters and English words may impede vocabulary learning for Chinese EFL learners. Moreover, Schmitt (2000) concluded that the meanings of a given word may vary in different contexts, leading to problems in vocabulary acquisition. Liang (1996) pointed out that the age-old patterns of traditional teaching methods may make learners passive or indifferent to vocabulary learning. Thus, developing effective methods to make learners more active and interested in vocabulary learning is an urgent issue, particularly for EFL learners.

Furthermore, the role that vocabulary retention plays in language acquisition has long been a valuable research direction (Pimsleur, 1967). Pimsleur (1967) claimed that when effective review processes are absent, acquired words inevitably fade from a learner’s memory. Examining learners’ ability to retain vocabulary is the most effective means of evaluating vocabulary learning outcomes (Jensen, 2005). However, retaining vocabulary in EFL environments, to some extent, can be challenging as learners rarely have opportunities to practice. Nation (2001) suggested that EFL learners generally forget words easily when retrieval processes are not frequent. Thus, language instructors have devoted considerable effort to developing effective and efficient pedagogies or computer-assisted learning tools that facilitate vocabulary acquisition and retention.

In addition, rapid development of mobile technologies has forced researchers and instructors to consider the potential of mobile-assisted language learning (MALL) (Kukulka-Hulme & Shield, 2008). According to Roschelle (2003), MALL, when compared with traditional learning approaches in fixed contexts, provides learners with opportunities to study anytime and anywhere. Song and Fox (2008) further argued that MALL can support seamless, personalized, authentic, spontaneous learning. Their study confirmed that learners generate positive perceptions of mobile-assisted vocabulary learning. Motallebzadeh and Ganjali (2011) found that MALL can improve vocabulary retention and reading comprehension. Learners can also achieve better vocabulary learning outcomes with MALL (Mostafa & Zahra, 2014), and learners utilizing mobile-assisted vocabulary learning had a high tendency to participate in vocabulary learning activities.

The positive role of digital game-based learning (DGBL) has been confirmed in the domain of learning technology. According to Prensky (2001), learning through games is the way of the future. Hogle (1996) emphasized that typical game attributes, including challenges, control, and fantasy can increase learners’ interests, motivation, and knowledge retention. In the language-learning domain, game-based learning has long been used to help students review material. For example, Sukstrienwong and Vongsumedh (2013) concluded that learners had positive perceptions of learning vocabulary through game play on mobile devices, and they confirmed that learners utilizing mobile devices for vocabulary acquisition show significant improvement in terms of vocabulary learning. To meet the needs of learners as digital natives, DGBL as support for language learning is regarded as a promising approach.

Most schools in Taiwan employ traditional instruction approaches when teaching foreign languages (Liao, 2004); that is, Taiwanese EFL learners tend to memorize and regurgitate what has been taught. Several empirical studies of young adult EFL learners in Taiwan indicated that

repetition is still the most common strategy (Li, Haggard & Chen, 2010; Liao, 2004; Wu, 2002). However, repetition may render learners passive or indifferent (Liang, 1996). Further, Butler (2015) mentioned that language learners today are digital natives. Therefore, their needs and preferences differ from those of previous generations. Jones, Ramanau, Cross and Healing's study (2010) explored the experiences of digital native first-year university students as they encountered e-learning at five universities in England. Their study found that although the laptop and the mobile phone are not yet universal, the vast majority of students make extensive use of mobile technologies and computing facilities for communication and for access to course materials and resources. Moreover, their study also indicated that first-year university students are active users of technology and that in general they use technologies more than they believe they are required to. Goodman, Bradley, Paras, Williamson and Bizzochi (2006) claimed that mobile game-based learning (MGBL) can increase learners' motivation, knowledge acquisition, and learning effectiveness. Hence, to meet the needs of today's learners, MGBL is a promising technique for language acquisition. For example, Wang (2017) developed a mobile English vocabulary learning app that adopts the New General Service List (NGSL) as the main learning content. Moreover, busuu is a social language-learning app that can support 12 languages (Rosell-Aguilar, 2018). However, few studies have explored the effects of MALL embedded with game functions on vocabulary acquisition, vocabulary retention, and learning perceptions. Thus, the study recruited 20 sophomores (second-year university students) and randomly assigned them to the experimental group with MEVLA-GF support or the control group with MEVLA-NGF support for English vocabulary learning. As the PHONE Words app developed by the Alice English Education Studio simultaneously provides the gamified assessment with competition mechanism (i.e. MEVLA-GF) and traditional assessment (i.e. MEVLA-NGF) that can be optionally selected to support learning English vocabulary, it was respectively used as the English vocabulary learning tool for the learners in the experimental and control groups. The primary research questions of the study are whether significant differences exist in vocabulary acquisition, vocabulary retention, and learners' perceptions between the experimental group (MEVLA-GF) and the control group (MEVLA-NGF), and whether learners' involvement in gamified functions is correlated with vocabulary learning performance.

## 2. Research methodology

### 2.1 Research design

Learners in the experimental group used MEVLA-GF to learn English vocabulary, whereas the learners in the control group used MEVLA-NGF. Learning performance based on vocabulary acquisition and vocabulary retention was assessed using three vocabulary tests, a pre-test, a post-test, and a delayed post-test obtained randomly from the New TOEIC Official Test-Preparation Guide III published by the Educational Testing Service (ETS; 2011), an organization devoted to educational measurements and research in educational policy. In addition, a questionnaire and a semi-structured interview were used to identify learners' perceptions of using MEVLA-GF for vocabulary learning. Also, log files of usage in the experimental and control groups also supported data analysis. In short, this study adopted a mixed methodology of quantitative and qualitative approaches to answer the research questions. The methodology used to organize and anonymize data for protecting the personal privacy of the study's research participants was to assign code names composed of a letter and number instead of personal names for the gathered quantitative and qualitative data, including the questionnaire, pre-test, post-test, delayed post-test, log files of usage, and interview.

The experimental treatment did not use any formal classroom instruction, as learners can learn vocabulary anytime and anywhere via MEVLA-GF or MEVLA-NGF in the mobile context. Namely, learners in both groups utilized the MEVLA-GF and MEVLA-NGF as portable learning tools in an autonomous learning context during the four-week experimental period. As no strict

time restrictions within the four-week experimental period were imposed, learners in both groups could make their own learning schedules; however, they were expected to use MEVLA-GF or MEVLA-NGF for at least five hours per week. The app has a function allowing a learner to monitor the length of time spent on the app. Also, to evaluate learning performance, a vocabulary post-test was performed immediately at the end of the four-week experimental period, and was followed by a vocabulary delayed post-test two weeks later to evaluate vocabulary retention. Both the MEVLA-GF and MEVLA-NGF had the same ETS TOEIC word list.

## 2.2 Research participants

To recruit the research participants, this study issued a call via the Internet for volunteers who were willing to pass the TOEIC test. The research participants were 20 EFL learners who were Taiwanese sophomores studying at the College of Liberal Arts at National Chengchi University (NCCU). Before performing the formal experiment, a vocabulary pre-test was undertaken by all the participants in both groups for the purpose of evaluating their original vocabulary levels. The pre-test results were further adopted as the group-assigning criterion, which was anticipated to be the prerequisite for establishing two evenly distributed groups with the same original vocabulary level. In addition, gender balance in both groups was also considered. That is, the participants were divided into two groups based on as equal a distribution of gender and vocabulary levels as possible. Based on these considerations, 10 students (five males, five females) were assigned to the experimental group (i.e. MEVLA-GF), and the remaining 10 students (five males, five females) were assigned to the control group (i.e. MELVA-NGF). The attempt to distribute gender and vocabulary levels evenly was aimed at minimizing possible interference.

To consider research ethics and to offset any disadvantage of the designed experiment in which all the research participants were divided into two groups with different learning treatments, the research participants in both groups were invited to experience the treatment that had not been offered them (either MEVLA-GF or MELVA-NGF) after ending the formal instruction experiment. Participants were free to accept or reject the invitation themselves.

## 2.3 Research instruments

### 2.3.1 The MEVLA-GF

This study adopted the PHONE Words app, developed by the Alice English Education Studio in Taiwan, as the vocabulary learning tool because it is a novel mobile English vocabulary learning app designed for learning English vocabulary on such formal vocabulary tests as TOEIC and TOEFL. Moreover, the app simultaneously provides traditional assessment and gamified assessment with competition mechanism that can be optionally selected to support learning English vocabulary. Since passing the TOEIC test is a graduation requirement at several universities in Taiwan – including NCCU – this study thus chose the PHONE Words app as the research instrument to conduct the experiment.

The six main functions provided by the PHONE Words app are the word list, customized word list, pre-established learning path, traditional assessment, gamified assessment, and ranking among learning peers. The main difference in the functions provided between the MEVLA-GF and MEVLA-NGF is that the MEVLA-NGF does not have the gamified assessment with ranking among learning peers, but the other functions provided by the two apps are the same. Learners using MEVLA-GF or MEVLA-NGF learn the words in the list in an alphabetical manner (see Figure 1). Detailed information on each word, such as definitions, pronunciations, and example sentences with audio clips, is presented after clicking on a word. Moreover, learners can mark words and put them into a customized list for later review. The words archived in this list were also served as the question pool for traditional assessment or gamified assessment.

Figure 2 shows the pre-established learning path in MEVLA-GF and MEVLA-NGF. In using this pre-established path, learners must finish the task assigned each day and pass the related

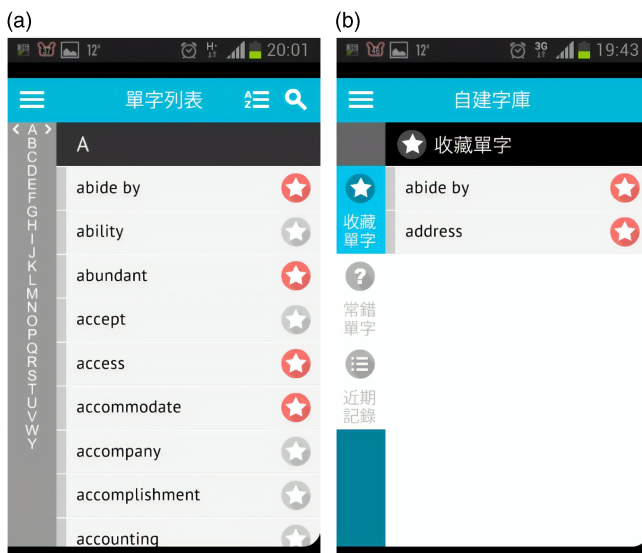


Figure 1. Word list and customized word list of MEVLA-GF and MEVLA-NGF. (a) Word list; (b) Customized word list



Figure 2. Pre-established learning path of MEVLA-GF and MEVLA-NGF. (a) Pre-established learning path; (b) Acquired vocabularies in Day 1

assessment – either traditional or gamified. After finishing a task, learners are awarded a course badge, allowing them to move to the next task.

Furthermore, two self-assessments – the traditional (see Figure 3) and gamified (see Figure 4) – are simultaneously provided by the MEVLA-GF, but the MEVLA-NGF only has the traditional assessment function. Learners can choose either the archived words in their customized word list or frequently incorrectly answered questions (FIAQ) for the assessment. The learners in the control group with MEVLA-NGF support had to learn vocabulary words on their own without a ranking and competitive mechanism based on the traditional assessment results. The types of traditional assessment provided by the MEVLA-NGF include multiple choice, true/false, and



Figure 3. Traditional assessment of MEVLA-GF and MEVLA-NGF



Figure 4. Gamified assessment of MEVLA-GF. (a) Tic-Tac-Toe; (b) Tug-of-VOC; (c) Star VOC

matching tests. Moreover, three vocabulary games, Tic-Tac-Toe (see Figure 4(a)), Tug-of-VOC (see Figure 4(b)), and Star VOC (see Figure 4(c)), are provided by MEVLA-GF. Tic-Tac-Toe is for two players. One player uses circles and the other uses crosses, and they take turns marking the spaces on a three-by-three grid. A player is then tested using multiple-choice questions. When the player successfully answers a vocabulary question, he/she marks a space on the grid. To win, a player must succeed in placing three marks in a horizontal, vertical, or diagonal line. Tug-of-VOC – its name is adapted from tug-of-war – is a mission-based game in which players must spell a tested word successfully to save sheep from a black bear. This blank-filling vocabulary test can assess learners' ability to spell learned words. Star VOC – adapted from the name Star Trek – is also a mission-based game in which a user pilots a spacecraft and must answer



**Figure 5.** Ranking among learning peers of MEVLA-GF. (a) Ranking among learning peers for showing vocabulary learning outcomes (b) Looking at acquired vocabulary stickers

vocabulary questions along the journey. When a vocabulary question appears, the player must choose which answer best represents the word being tested. Hung and Young (2007) would classify the three kinds of vocabulary games as quiz-based games. Dondi and Moretti (2007) argued that quiz games are an effective approach to fostering factual knowledge, assisting learners in recalling acquired knowledge, thereby promoting knowledge retention.

Moreover, MEVLA-GF also has a game competition mechanism that ranks learning peers according to their performance (see Figure 5). Ranking is based on number of vocabulary stickers, familiarity with words, and number of course badges granted from a pre-established learning path. When a word question is answered correctly by a learner on the gamified test, the learner receives a vocabulary sticker. When a word question is answered three times in a row, the word is added to the learner's list of familiar words. When viewing the ranking of learning peers, learners can see how many vocabulary stickers, familiar words, and course badges other learners have received. This function can provide learners with a sense of competition and extrinsic learning motivation (Huang, Huang & Tschopp, 2010; Prensky, 2001). Importantly, the PHONE Words app can record detailed information of learners' usage behaviors in the database of the server.

### 2.3.2 Vocabulary assessment

To assess the effects of MEVLA-GF as a tool for vocabulary acquisition, this study adopted questions from the New TOEIC Official Test-Preparation Guide III (ETS, 2011), an official test-preparation guide, to evaluate participants' vocabulary. The TOEIC test was developed in 1979. The adopted questions consisted of incomplete sentences that the student must complete. The vocabulary tests in this study were the vocabulary pre-test, immediate vocabulary post-test, and the vocabulary delayed post-test. As mentioned, the pre-test aimed to evaluate the initial vocabulary abilities of the learners in both the groups before performing the experiment. The vocabulary post-test, which was assigned to the research participants on the last day of the experiment, aimed to assess the difference in vocabulary learning performance between both groups. To evaluate the participants' vocabulary retention, this study conducted a vocabulary delayed post-test two weeks later that concluded the learning activity, but in order to avoid affecting the assessment results, all the learners in both groups had not been told that their vocabulary retention would be examined by the delayed post-test. The contents of the delayed

post-test were identical to that of the earlier post-test; however, the order of questions and answer set of each question were rearranged to avoid memory effects.

### 2.3.3 Questionnaire survey

The study referred to several previous studies (Brooke, 1996; Chu, Hwang, Tsai & Tseng, 2010; Evans, 2008) to design a questionnaire with a total of 15 items that consisted of potential effectiveness, usability, and satisfaction dimensions to identify participants' learning perceptions toward the MEVLA-GF and MEVLA-NGF (see Appendix). The three dimensions had five questions, respectively. The potential effectiveness dimension aimed to evaluate whether a learner thinks that using the MEVLA-GF or MEVLA-NGF to support vocabulary learning is effective; the usability dimension aimed to evaluate whether the MEVLA-GF or MEVLA-NGF provides a friendly and useful interface; the satisfaction dimension aimed to evaluate whether the MEVLA-GF or MEVLA-NGF provides satisfied functions in assisting vocabulary learning. Participants responded to items on a 5-point Likert scale. To confirm the validity and reliability of the developed questionnaire, this study recruited 42 non-experimental participants who were Taiwanese sophomores studying at the College of Liberal Arts at National Chengchi University to perform the pre-test of the questionnaire after they experienced the MEVLA-GF for English vocabulary learning. Exploratory factor analysis was applied to verify the construct validity of the questionnaire, in which Kaiser–Meyer–Olkin (KMO) and factor loadings were employed. KMO values of these three dimensions are above 0.7 and factor loadings of each item are all above 0.7, indicating that the questionnaire has sufficient construct validity. Moreover, the questionnaire's reliability was confirmed by Cronbach's alpha (overall = .883, potential effectiveness = .857, usability = .803, satisfaction = .887). These analytical results confirm that the questionnaire has satisfactory reliability, with Cronbach's alpha values exceeding 0.7.

### 2.3.4 Log files gathered from the MEVLA-GF

PHONE Words app can record detailed information of learners' usage behaviors in the database of the server regardless of MEVLA-GF or MEVLA-NGF support. However, this study only adopted the number of clicks and amount of time to confirm whether significant correlations existed among the learning performance and usage behaviors of the six functions, including the pre-established learning path, word list, customized word list, traditional assessment, gamified assessment, and ranking of learning peers respectively provided by the MEVLA-GF or MEVLA-NGF.

### 2.3.5 Interview procedure

As this study mainly focused on examining the effects of the game-related features provided by the MEVLA-GF, semi-structured interviews were conducted only for the learners in the experimental group. All 10 learners in the experimental group were invited to take part in the interview, but only eight of them agreed to participate in the interview. Before performing the interview, the eight interviewees were invited to experience MEVLA-NGF that they have not been distributed in the formal instruction experiment. With the flexibility inherent in a semi-structured interview, the interviewer reused or repurposed questions to obtain in-depth information on the perspectives and personal experiences of each interviewee. The interview questions contained explanations of the difference between MEVLA-GF and MEVLA-NGF for English vocabulary learning, and how to learn English vocabulary with MEVLA-GF or MEVLA-NGF support, and so on.

## 3. Experimental results

### 3.1 Vocabulary learning performance

As the sample was small and not normally distributed, nonparametric statistics were employed to assess the learning performance of participants in each group for statistical significance.



### 3.1.1 Comparisons of pre-test scores, post-test scores, and delayed post-test scores

Since the experiment is a between-by-within design of repeated measurement, the nonparametric equivalent of factorial ANOVA, which is a between-within design by Wilcoxon, was used to perform the learning performance assessment (Wilcox, 2012). The statistical model can be written as  $y_{ijk} = \mu + G_i + S_{j(i)} + T_k + (GT)_{ik} + \varepsilon_{k(ij)}$ , where  $y_{ijk}$  is the  $j$ th subject's test score of the  $k$ th time's test in the  $i$ th group,  $i = 1, 2$ ,  $j = 1, \dots, 10$ ,  $k = 1, 2, 3$ ;  $G_i$  is the effect of the  $i$ th group;  $S_{j(i)}$  is the effect of the  $j$ th subject within the  $i$ th group;  $T_k$  is the effect of the  $k$ th time's test;  $(GT)_{ik}$  is the interaction between the  $i$ th group and the  $k$ th time's test;  $\varepsilon_{k(ij)}$  is an error term. Tables 1 and 2 show the results of average rank and rank-based ANOVA for pre-test, post-test, and delayed post-test of both groups, respectively. The plot of relative effects between Time and Group is shown in Figure 6. The results show that two segments of relative effects are not parallel and the  $p$  value ( $< 0.05$ ) of the interaction between Group and Time is very small (see Table 2). Both the results show that strong interactions between Group and Time exist. Therefore, this study fixed the variables of Group and Time to conduct multiple comparisons, respectively. Analytical results show that the mean pre-test score for the experimental group was not significantly different from that of the control group ( $p$  value = .9993  $> 0.05$ ), indicating that learners in the two groups had similar vocabulary skills before performing the experiment. On the other hand, the mean post-test score and delayed post-test score of the experimental group are strongly significantly higher than those of the control group ( $p$  value = .000  $< 0.05$ ). Analytical results confirm that the vocabulary learning performance of the experimental group was significantly superior to that of the control group in both the post-test and the delayed post-test.

### 3.1.2 Vocabulary acquisition and retention

Vocabulary acquisition (post-test scores minus pre-test scores) and vocabulary retention (delayed post-test scores minus pre-test scores) of students in the two groups were analyzed. Tables 3 and 4 show the results of average rank and rank-based ANOVA for vocabulary acquisition and vocabulary retention of both groups, respectively. The plot of relative effects between Time and Group is shown in Figure 7. Analytical results show that the interaction between Time and Group is not significant ( $p$  value = .1184  $> 0.05$ ). The mean scores of vocabulary acquisition and vocabulary retention of the experimental group were significantly higher than those of the control group ( $p$  value = .000  $< 0.05$ ).

**Table 1.** Result of average rank for pre-test, post-test, and delayed post-test of both groups

Group	Pre-test	Post-test	Delayed post-test
Experimental group	14.7	48.9	41.75
Control group	14.7	39.3	23.65

**Table 2.** Result of rank-based ANOVA for pre-test, post-test, and delayed post-test of both groups

Item	Source	$F$	$p$ value
Between	Group (G)	3.6798	.071
Within	Time (T)	251.7079	.000
	T $\times$ G	23.4854	.000

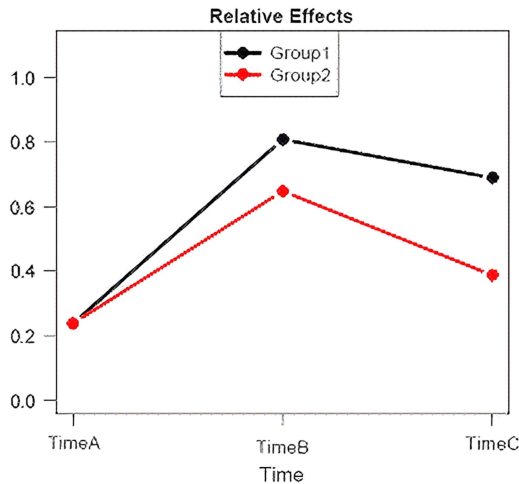


Figure 6. Plot of relative effects between Time and Group

Table 3. Result of average rank for acquisition and retention of both groups

Group	Acquisition	Retention
Experimental group	27.90	26.55
Control group	13.10	14.45

Table 4. Result of rank-based ANOVA for acquisition and retention of both groups

Item	Source	<i>F</i>	<i>p</i> value
Between	Group (G)	38.5936	.000
Within	Time (T)	71.6525	.000
	T × G	1.7668	.1184

### 3.2 Questionnaire results

The questionnaire, as mentioned, was used to investigate participants' perceptions and attitudes towards the use of the MEVLA-GF as a vocabulary learning tool. Table 5 shows the descriptive statistics and Wilcoxon–Mann–Whitney *U* test results of the two groups. The mean scores of the experimental group for the potential effectiveness and satisfaction dimensions were significantly higher than that of control group ( $Z = -3.78$ , Sig. = .000;  $Z = -3.82$ , Sig. = .000). However, the mean score of the experimental group for the usability dimension was not significantly different from that of the control group ( $Z = -.533$ , Sig. = .594).

Based on the questionnaire results, the study argued that the significant results of the English vocabulary learning performance in post-test and delayed post-test between the experimental and control groups may be due to the significant difference found in potential effectiveness and satisfaction variables. Therefore, the study respectively considered the post-test and delayed post-test in both groups as dependent variables and the potential effectiveness and satisfaction as the covariate variables to test whether the covariate variables explain the dependent variables by using Quade's ranking-based covariance analysis (Quade, 1967). Table 6 shows the results.

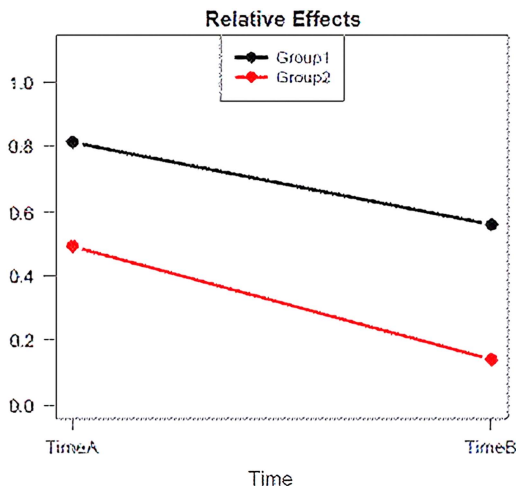


Figure 7. Plot of relative effects between Time and Group

Table 5. Results of descriptive statistics and Wilcoxon–Mann–Whitney *U* test of questionnaire assessment between the two groups

Item	Group (N)	M	SD	Mann–Whitney <i>U</i>	Z	<i>p</i> value (2-tailed)
Potential effectiveness	Experimental group	4.34	0.21	0.5	-3.78***	.000
	Control group	3.42	0.32			
Usability	Experimental group	3.60	0.33	43.5	-.53	.594
	Control group	3.54	0.32			
Satisfaction	Experimental group	4.36	0.25	0.00	-3.82***	.000
	Control group	3.28	0.23			

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

Analytical results showed that the mean scores of the experimental group for the post-test and delayed post-test were significantly higher than that of the control group, while simultaneously considering the potential effectiveness and satisfaction as the covariate variables (*p* value = .0069 < 0.05; *p* value = .0268 < 0.05). The results confirmed that the significantly different results of the English vocabulary learning performance in post-test and delayed post-test between the experimental and control groups were not due to the significant difference found in potential effectiveness and satisfaction variables. Namely, although the learners in the experimental group using the MEVLA-GF had significantly higher perceptions in potential effectiveness and satisfaction than the learners in the control group using the MEVLA-NGF, the differences in the game-related functions between the MEVLA-GF and MEVLA-NGF are the main factors affecting the post-test and delayed post-test.

### 3.3 Correlations between MEVLA-GF usage behaviors and vocabulary acquisition performance

#### 3.3.1 Descriptive statistics for MEVLA-GF usage behaviors

Table 7 shows the descriptive statistics of MEVLA-GF usage behaviors in the experimental group. Among the six functions, the pre-established learning path had the highest mean for number of clicks (*M* = 282.10, *SD* = 37.71), followed by gamified assessment (*M* = 136.00, *SD* = 15.64), and customized word list (*M* = 61.70, *SD* = 17.34). The mean of total clicks of all the

**Table 6.** Results of Quade test for post-test and delayed post-test while simultaneously considering potential effectiveness and satisfaction as covariate variables

Item	Covariate variable	$p$ value	Quade test (Q)	$p$ value (Q)
Post-test	Effectiveness	.1655	12.115	.0069
	Satisfaction	.1632		
Delayed post-test	Effectiveness	.9194	6.9751	.0268
	Satisfaction	.0620		

**Table 7.** Descriptive statistics of usage behaviors of MEVLA-GF in the experimental group

Used function	Number of learners	$M$ (click times)	$SD$ (click times)
Pre-established learning path	10	282.10	37.71
Word list	10	39.80	5.81
Customized word list	10	61.70	17.34
Traditional assessment	10	48.50	6.52
Gamified assessment	10	136.00	15.64
Ranking among learning peers	10	48.90	6.12
<b>Total click times of all the above functions</b>	10	617.00	17.80
Used time	Number of learners	$M$ (hr)	$SD$ (hr)
<b>Total usage time</b>	10	23.10	0.91

functions provided by the MEVLA-GF during the four weeks was 617 times, and the mean of total usage time was as high as 23.10 hours.

### 3.3.2 Spearman rank correlation results for MEVLA-GF usage behaviors and vocabulary acquisition performance

This study used Spearman rank correlation analysis to confirm whether significant correlations between several considered usage behaviors and vocabulary acquisition performance existed or not. Analytical results of Spearman rank correlation show that significant positive correlations between several considered usage behaviors and vocabulary acquisition performance existed, including usage time ( $r = 0.703$ , Sig. = .023), total number of clicks ( $r = 0.871$ , Sig. = .001), pre-established learning path ( $r = 0.724$ , Sig. = .018), gamified assessment ( $r = 0.663$ , Sig. = .037), and ranking of learning peers ( $r = 0.770$ , Sig. = .009). The results are very encouraging because the game-related features provided by the MEVLA-GF indeed generated the positive effects on vocabulary acquisition performance. Moreover, a significant negative correlation existed between traditional assessment and vocabulary acquisition performance ( $r = -0.756$ , Sig. = .011). That is, spending much time on traditional assessment is negative to the learners in the experimental group. Also, neither the word list ( $r = -0.179$ , Sig. = .621) nor customized word list ( $r = -0.24$ , Sig. = .948) was strongly correlated with vocabulary acquisition performance.

In terms of correlations among vocabulary retention and usage behaviors of MEVLA-GF, this study found that significant positive correlations existed, including usage time ( $r = .682$ , Sig. = .030), total click times ( $r = .681$ , Sig. = .030), pre-established learning path ( $r = .782$ , Sig. = .008), gamified assessment ( $r = .708$ , Sig. = .022), and ranking among friends ( $r = .637$ , Sig. = .047). The

**Table 8.** Descriptive statistics of usage behaviors of MEVLA-NGF in the control group

Used function	Number of learners	<i>M</i> (click times)	<i>SD</i> (click times)
Word list	10	147.40	15.69
Customized word list	10	161.40	14.05
Traditional assessment	10	222.40	28.19
Total click times of all the above functions	10	531.20	49.06
Used time	Number of learners	<i>M</i> (hr)	<i>SD</i> (hr)
Total usage time	10	21.33	0.55

results demonstrate that the game-related features provided by the MEVLA-GF are helpful in improving learners' vocabulary retention.

### 3.4 Correlations between MEVLA-NGF usage behaviors and vocabulary acquisition performance

#### 3.4.1 Descriptive statistics for MEVLA-NGF usage behaviors

Table 8 shows the descriptive statistics of MEVLA-NGF usage behaviors in the control group. Among the three functions, the traditional assessment had the highest mean for the number of clicks ( $M = 222.40$ ,  $SD = 28.19$ ), followed by the customized word list ( $M = 161.40$ ,  $SD = 14.05$ ), and word list ( $M = 147.40$ ,  $SD = 15.69$ ). The mean of total clicks of all the functions provided by the MEVLA-NGF during the four weeks was 531.20 times, and the mean of total usage time was as high as 21.33 hours. This study found that the mean of total clicks of all the functions and the mean of total usage time for the learners in the experimental group during the four weeks were higher than those of the learners in the control group. Compared to the learners in the control group, it is obvious that the game-related functions motivated the learners in the experimental group to engage themselves more actively and invest much more time in English vocabulary learning activities.

#### 3.4.2 Spearman rank correlation results for MEVLA-NGF usage behaviors and vocabulary acquisition performance

Analytical results of Spearman rank correlation show that significant positive correlations between several considered usage behaviors and vocabulary acquisition performance existed, including usage time ( $r = .759$ ,  $\text{Sig.} = .011$ ), total number of clicks ( $r = .863$ ,  $\text{Sig.} = .001$ ), word list ( $r = .725$ ,  $\text{Sig.} = .018$ ), customized word list ( $r = .809$ ,  $\text{Sig.} = .005$ ), and traditional assessment ( $r = .695$ ,  $\text{Sig.} = .026$ ). However, in terms of correlations among vocabulary retention and usage behaviors of MEVLA-NGF, this study found that there was only a significant positive correlation in traditional assessment ( $r = .707$ ,  $\text{Sig.} = .022$ ). The result demonstrates that the traditional assessment provided by the MEVLA-NGF is also helpful in improving learners' vocabulary retention.

## 4. Discussion

This study confirmed that the learners using MEVLA-GF, the experimental group, as an assistive tool markedly outperformed learners using MEVLA-NGF, the control group, in terms of vocabulary acquisition and vocabulary retention. The analytical results echo those obtained by several other empirical studies (Dolati & Mikaili, 2011; Fotouhi-Ghazvini, Earnshaw, Robison & Excell, 2009; Ke, 2008; Lim, 2008; Sukstrienwong & Vongsumedh, 2013; Uzun, Çetinavcı, Korkmaz & Salihoğlu, 2013), confirming benefits and effectiveness of MGBL for vocabulary acquisition. Findings also support Hogle's (1996) assertion that games can help learners memorize

and retain acquired knowledge. The semi-structured interviews of students in the experimental group indicate that MEVLA-GF was a favorable learning approach, and improved motivation and involvement. The interview results were consistent with findings in previous studies – that game attributes can facilitate and sustain motivation (Huang *et al.*, 2010) and that games are conducive to satisfactory learning outcomes (Uzun *et al.*, 2013). With regard to vocabulary retention, most interviewees in the experimental group agreed that MEVLA-GF, compared with traditional learning methods, was challenging, favorable, interesting, interactive, authentic, and exciting. Such learning experiences can assist students in retaining words, leading to improved learning outcomes. This interview result is consistent with those obtained by several studies, suggesting that game-based learning is beneficial in terms of having clear, challenging goals (Prensky, 2001), triggering flow, cultivating interest (Garris, Ahlers & Driskell, 2002), excitement, fun (Schwabe & Göth, 2005), and sensory stimuli (Wilson *et al.*, 2009). Importantly, the MEVLA-GF can assist learners to pronounce words correctly, spell out all the letters, and better understand the meaning of the learned English vocabularies because the app provides the gamified vocabulary assessment with a competition mechanism and detailed information on each vocabulary, such as definitions, pronunciations, and example sentences with audio clips for learners.

In particular, MEVLA-GF has a game competition mechanism that ranks learning peers on a leaderboard according to their vocabulary learning performance. It is obvious that this kind of peer competitive learning strategy can facilitate the peer-to-peer zone of proximal development (ZPD) on English vocabulary learning despite performing the vocabulary learning activity without physical help from peers (Rezaee & Azizi, 2012). Obviously, this game competition mechanism stimulated the learners in the experimental group to reduce the gap of vocabulary size between what they had already mastered and what their peers with higher ranking on a leaderboard had already achieved. The findings of this study imply that the game-based mobile English vocabulary learning app is an effective learning tool that is appropriate for learners' autonomous and spontaneous learning by using the available time. Teachers therefore could make use of this kind of game as an educational tool to support classroom language learning processes. Additionally, the game-based mobile English vocabulary learning app opens up a way for teachers to direct their students in language learning in a more interesting, fun, and active way compared to traditional teaching methods like mechanical drilling, which often make for a daunting learning experience.

In addition, questionnaire results show that learners using MEVLA-GF rated the learning tool as more effective and more satisfactory than those using MEVLA-NGF. This analytical result is consistent with those acquired by several studies (Franciosi, 2011; Kang, 2012; Neville, Shelton & McInnis, 2009; Uzun *et al.*, 2013), asserting that learners have positive attitudes toward learning tools with game functions. On the other hand, most interviewees agreed that MEVLA-GF increases effectiveness and satisfaction by offering varied, appealing learning approaches; corresponds to learners' needs and expectations; and overcomes limitations to learning, such as boredom and repetition, associated with traditional learning approaches. Thus, study findings indicate that MEVLA-GF was more effective and satisfying than MEVLA-NGF. Additionally, analytical results show that no significant difference existed between the two groups in terms of usability. This analytical result may derive from the fact that the usability dimension was concerned primarily with users' experiences operating MEVLA-GF or MEVLA-NGF. However, the two assigned mobile vocabulary learning apps differed only in game-related functions. Further, the study confirmed that the significant differences of the English vocabulary learning performance in acquisition and retention between MEVLA-GF and MEVLA-NGF were not due to the significant difference found in the potential effectiveness and satisfaction dimensions based on Quade's ranking-based covariance analysis. The differences in the game-related functions between the MEVLA-GF and MEVLA-NGF are the main factors affecting acquisition and retention.

Analytical results show that usage time of gamified functions and vocabulary learning outcomes were strongly and positively correlated. This finding is very encouraging and means that well-designed mobile games not only increase the amount of time that learners spend learning

vocabulary in their available time, but also leads to the strong correlation between learning outcomes with usage time of gamified functions, thus improving vocabulary learning performance. Franciosi (2011) indicated that games can enhance learners' engagement, and Kang (2012) indicated that games can improve learners' motivation. Moreover, vocabulary learning performance and several gamified functions, such as the gamified assessment, and ranking of learning peers, were also strongly correlated. This correlation result echoes the study of Van Eck (2006), indicating that games can accelerate learning performance. Moreover, one particularly interesting fact revealed by usage behavior analysis is that traditional assessment was negatively correlated with learning performance, whereas gamified assessment was positive correlated with learning performance. Also, vocabulary learning performance and the two word lists were not correlated. The semi-structured interviews show that most interviewees agreed that edutainment via the gamified functions of MEVLA-GF can effectively overcome limitations related to traditional learning approaches, thereby improving learning outcomes. Moreover, the involvement of learners may be unintentionally improved by addiction, a feature sometimes related to gamification. Some interviewees proposed that the game-related functions were conducive to vocabulary learning. Thus, it is reasonable that strong correlations existed between vocabulary learning performance and several gamified functions.

Furthermore, all interviewees utilized word lists organized either by different publishing companies or by themselves to expand their vocabulary. This interview result is consistent with those in several past studies (Li *et al.*, 2010; Liao, 2004; Wu, 2002); that is, the repetition strategy is still widely used for Taiwanese EFL learners. However, half of them claimed that a word list is not an effective learning method because words are easily forgotten. This finding confirmed the graduated interval recall hypothesis by Pimsleur (1967), claiming that acquired words would inevitably fade from learners' memory if there were no effective reviewing processes. In addition, most interviewees relied on the gamified functions of MEVLA-GF. Only some utilized word lists and customized word lists for reviewing purposes. Unanimously, traditional assessment was considered the least conducive to learning. Also, most interviewees agreed that MEVLA-GF and MEVLA-NGF can accomplish ubiquitous learning and fragmented learning, provide convenient, diversified learning methods, and offer instant assessments. Most importantly, all interviewees preferred MEVLA-GF more than MEVLA-NGF, and agreed that the gamified features improved interaction, reduced boredom and repetition, enhanced motivation, and were conducive to active involvement. Last, with regard to the factors influencing the interviewees' perceptions of MEVLA-GF for vocabulary learning, interface design, gamified design, and content design were identified. Based on interviewees' perceptions of MEVLA-GF for vocabulary learning, this study suggests that (1) the small screen size of mobile devices should be used efficiently, and interface design should lead to high interaction, satisfactory stability, and intuitive operation; (2) the gamified features should be thoughtfully designed, interesting, challenging, and engaging, and updated frequently; (3) content design should satisfy the needs of learners at different levels and/or with different purposes, should offer authentic context-related vocabulary because MEVLA-GF currently only supports learning the words in the list in an alphabetical manner, and should offer well-organized learning plans, varied learning methods, and expandability.

## 5. Conclusions and future work

First, compared with MEVLA-NGF, this study confirmed that MEVLA-GF achieved its educational goal and effectively assists learners in improving their vocabulary size. Analytical results show that MEVLA-GF positively influenced learners' vocabulary acquisition and was helpful in augmenting the learners' ability to dispel the graduated interval recall hypothesis (Pimsleur, 1967), thus effectively assisting learners in retaining vocabulary. Based on the interview results, this study logically inferred that the main reasons that generated the positive effects on promoting learners' vocabulary acquisition performance are probably due to the features of challenging, favorable, interesting, interactive, authentic, and exciting gamified assessment and game competition mechanisms

provided by the MEVLA-GF. Such learning experiences can assist learners in retaining words, thus leading to improved learning outcomes. Also, questionnaire findings indicate that learners in the experimental group were more satisfied than learners in the control group. Additionally, the interview results show that MEVLA-GF, with its edutainment elements and game attributes, provides a favorable learning approach to learners and offers additional benefits in promoting motivation and active involvement. This study also confirmed that strong correlations existed between the gamified functions of MEVLA-GF and learning outcomes. Notably, traditional assessment and learning outcomes were negatively correlated. More importantly, interview results reveal that learners paid attention to several aspects of MEVLA-GF, including its interface design, gamified design, and content. Hence, this study suggests that these primary factors should be taken into account. The implication of the study is that the mobile learning game with well-designed gamified assessment and game competition mechanism for vocabulary learning is indeed very helpful in promoting learners' motivation of vocabulary learning, thus generating strongly significant positive correlations between vocabulary acquisition performance and several usage behaviors associated with gamified functions, such as gamified assessment and ranking of learning peers.

Additional studies are warranted. First, examining the roles that learners' characteristics and prior knowledge play in the use of MEVLA-GF for vocabulary learning is essential. Accordingly, our future research should consider determining which learners benefit from MEVLA-GF. In particular, age groups, backgrounds, characteristics, prior knowledge, learning styles (Coffield, Moseley, Hall & Ecclestone, 2004), and preferences may affect learning performance and attitudes toward MEVLA-GF. Second, our future research will consider whether the rural–urban divide, digital divide, or regional differences influence learning performance and attitudes toward MEVLA-GF.

**Ethical statement.** To consider the research ethics of the designed experiment, which involved recording usage behaviors of the research participants, written informed consent was obtained from the participants after the experiment was explained in full. The informed consent letter detailed the specific nature of the research. This included explaining that the data that were collected from participants was only for the research, that participants' names will never appear on any data collected, and that instead a unique identification number will be assigned to their data; the information that participants provide will remain secure such that only the principal investigator of this study will have access to it; the collected data that are no longer needed will be destroyed; and how participation will make a contribution to the study's goals.

## References

- Alfaki, I. M. (2015) Vocabulary input in English language teaching: Assessing the vocabulary load in spine five. *International Journal of English Language and Linguistics Research*, 3(1): 1–14.
- Alqahtani, M. (2015) The importance of vocabulary in language learning and how to be taught. *International Journal of Teaching & Education*, III(3): 21–34. <https://doi.org/10.20472/TE.2015.3.3.002>
- Brooke, J. (1996) SUS: A 'quick and dirty' usability scale. In Jordan, P. W., Thomas, B., Weerdmeester, B. A. & McClelland, I. L. (eds.), *Usability evaluation in industry*. London: Taylor & Francis, 189–194.
- Butler, Y. G. (2015) The use of computer games as foreign language learning tasks for digital natives. *System*, 54: 91–102.
- Chu, H.-C., Hwang, G.-J., Tsai, C.-C. & Tseng, J. C. R. (2010) A two-tier test approach to developing location-aware mobile learning systems for natural science courses. *Computers & Education*, 55(4): 1618–1627. <https://doi.org/10.1016/j.compedu.2010.07.004>
- Coffield, F., Moseley, D., Hall, E. & Ecclestone, K. (2004) *Learning styles and pedagogy in post-16 learning: A systematic and critical review*. London: Learning and Skills Research Centre.
- Dolati, I. & Mikaili, P. (2011) Effects of instructional games on facilitating of students' vocabulary learning. *Australian Journal of Basic and Applied Sciences*, 5(11): 1218–1224.
- Dondi, C. & Moretti, M. (2007) A methodological proposal for learning games selection and quality assessment. *British Journal of Educational Technology*, 38(3): 502–512. <https://doi.org/10.1111/j.1467-8535.2007.00713.x>
- Ehri, L. C. & Rosenthal, J. (2007) Spellings of words: A neglected facilitator of vocabulary learning. *Journal of Literacy Research*, 39(4): 389–409. <https://doi.org/10.1080/10862960701675341>
- Educational Testing Service (ETS) (2011) *New TOEIC Official Test-Preparation Guide III*. Taipei: Chun Shin.
- Evans, C. (2008) The effectiveness of m-learning in the form of podcast revision lectures in higher education. *Computers & Education*, 50(2): 491–498. <https://doi.org/10.1016/j.compedu.2007.09.016>



- Fotouhi-Ghazvini, F., Earnshaw, R. A., Robison, D. & Excell, P. (2009) The MOBO city: A mobile game package for technical language learning. *International Journal of Interactive Mobile Technologies*, 3(2): 19–24. <https://doi.org/10.3991/ijim.v3i2.757>
- Franciosi, S. J. (2011) A comparison of computer game and language-learning task design using flow theory. *Computer Assisted Language Learning-Electronic Journal*, 12(1): 11–25.
- Garris, R., Ahlers, R. & Driskell, J. E. (2002) Games, motivation, and learning: A research and practice model. *Simulation & Gaming*, 33(4): 441–467. <https://doi.org/10.1177/1046878102238607>
- Goodman, D., Bradley, N. L., Paras, B., Williamson, I. J. & Bizzochi, J. (2006) Video gaming promotes concussion knowledge acquisition in youth hockey players. *Journal of Adolescence*, 29(3): 351–360. <https://doi.org/10.1016/j.adolescence.2005.07.004>
- Hogle, J. G. (1996) *Considering games as cognitive tools: In search of effective "Edutainment"*. <https://files.eric.ed.gov/fulltext/ED425737.pdf>
- Huang, W.-H., Huang, W.-Y. & Tschopp, J. (2010) Sustaining iterative game playing processes in DGBL: The relationship between motivational processing and outcome processing. *Computers & Education*, 55(2): 789–797. <https://doi.org/10.1016/j.compedu.2010.03.011>
- Hung, H.-C. & Young, S. S.-C. (2007) Constructing the game-based learning environment on handheld devices to facilitate English vocabulary building. *Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007)*: 348–350. <https://doi.org/10.1109/ICALT.2007.105>
- Jensen, E. (2005) *Teaching with the brain in mind* (2nd ed.). Alexandria: ASCD.
- Jones, C., Ramanau, R., Cross, S. & Healing, G. (2010) Net generation or digital natives: Is there a distinct new generation entering university? *Computers & Education*, 54(3): 722–732. <https://doi.org/10.1016/j.compedu.2009.09.022>
- Kang, J. (2012) *Attributes and motivation in game-based learning: A review of the literature*. The University of Texas at Austin, unpublished master's thesis.
- Ke, F. (2008) Computer games application within alternative classroom goal structures: Cognitive, metacognitive, and affective evaluation. *Educational Technology Research & Development*, 56(5–6): 539–556. <https://doi.org/10.1007/s11423-008-9086-5>
- Kukulka-Hulme, A. & Shield, L. (2008) An overview of mobile assisted language learning: From content delivery to supported collaboration and interaction. *ReCALL*, 20(3): 271–289. <https://doi.org/10.1017/S0958344008000335>
- Laufer, B. & Sim, D. D. (1985) Measuring and explaining the reading threshold needed for English for academic purposes texts. *Foreign Language Annals*, 18(5): 405–411. <https://doi.org/10.1111/j.1944-9720.1985.tb00973.x>
- Li, C.-L., Haggard, S. & Chen, S.-H. (2010) A study of Taiwan's technical college students' English learning strategies. *Proceedings of the 2010 International Conference on ELT Technological Industry and Book Fair*, 136–146. Taiwan: National Pingtung University of Science and Technology.
- Liang, T. (1996) Cooperative learning in English education. *Proceedings of the 5th International Symposium on English Education*. Taipei, Taiwan: Crane Publishing Co., Ltd: 65–71.
- Liao, Y. F. (2004) A survey study of Taiwan EFL freshmen's vocabulary learning strategies. *Journal of Pingtung Teachers College*, 21: 271–288.
- Lim, C. P. (2008) Global citizenship education, school curriculum and games: Learning mathematics, English and science as a global citizen. *Computers & Education*, 51(3): 1073–1093. <https://doi.org/10.1016/j.compedu.2007.10.005>
- Mostafa, Z. & Zahra, A. (2014) Exploring the effect of Java mobile dictionaries on Iranian EFL students' vocabulary learning. *International Journal of Research Studies in Educational Technology*, 3(1): 23–35. <https://doi.org/10.5861/ijrset.2013.535>
- Motallebzadeh, K. & Ganjali, R. (2011) SMS: Tool for L2 vocabulary retention and reading comprehension ability. *Journal of Language Teaching and Research*, 2(5): 1111–1115. <https://doi.org/10.4304/jltr.2.5.1111-1115>
- Nation, I. S. P. (1990) *Teaching and learning vocabulary*. New York: Newbury House.
- Nation, I. S. P. (2001) *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Neville, D. O., Shelton, B. E. & McInnis, B. (2009) Cybertext redux: Using digital game-based learning to teach L2 vocabulary, reading, and culture. *Computer Assisted Language Learning*, 22(5): 409–424. <https://doi.org/10.1080/09588220903345168>
- Pimsleur, P. (1967) A memory schedule. *The Modern Language Journal*, 51(2): 73–75. <https://doi.org/10.1111/j.1540-4781.1967.tb06700.x>
- Prensky, M. (2001) *Digital game-based learning*. New York: McGraw-Hill.
- Quade, D. (1967) Rank analysis of covariance. *Journal of the American Statistical Association*, 62(320): 1187–1200. <https://doi.org/10.1080/01621459.1967.10500925>
- Read, J. (2000) *Assessing vocabulary*. Cambridge: Cambridge University Press <https://doi.org/10.1017/CBO9780511732942>
- Rezaee, A. A. & Azizi, Z. (2012) The role of zone of proximal development in the students' learning of English adverbs. *Journal of Language Teaching and Research*, 3(1): 51–57. <https://doi.org/10.4304/jltr.3.1.51-57>
- Roschelle, J. (2003) Unlocking the learning value of wireless mobile devices. *Journal of Computer Assisted Learning*, 19(3): 260–272. <https://doi.org/10.1046/j.0266-4909.2003.00028.x>
- Rosell-Aguilar, F. (2018) Autonomous language learning through a mobile application: A user evaluation of the *busuu* app. *Computer Assisted Language Learning*. Advance online publication <https://doi.org/10.1080/09588221.2018.1456465>

- Schmitt, N. (2000) *Vocabulary in language teaching*. Cambridge: Cambridge University Press.
- Schwabe, G. & Göth, C. (2005) Mobile learning with a mobile game: Design and motivational effects. *Journal of Computer Assisted Learning*, 21(3): 204–216. <https://doi.org/10.1111/j.1365-2729.2005.00128.x>
- Song, Y. & Fox, R. (2008) Using PDA for undergraduate student incidental vocabulary testing. *ReCALL*, 20(3): 290–314. <https://doi.org/10.1017/S0958344008000438>
- Sukstrienwong, A. & Vongsumedh, P. (2013) Software development of word search game on smart phones in English vocabulary learning. In Zhang, L. (ed.), *The Proceedings of the 2013 International Conference on Education and Modern Educational Technologies*, 39: 37–42.
- Uzun, L., Çetinavcı, U. R., Korkmaz, S. & Salihoğlu, U. M. (2013) Developing and applying a foreign language vocabulary learning and practice game: The effect of VocaWord. *Digital Culture & Education*, 5(1): 48–70.
- Van Eck, R. (2006) Digital game-based learning: It's not just the digital natives who are restless. *EDUCAUSE Review*, 41(2): 16–30.
- Wang, B.-T. (2017) Designing mobile apps for English vocabulary learning. *International Journal of Information and Education Technology*, 7(4): 279–283. <https://doi.org/10.18178/ijiet.2017.7.4.881>
- Wilcox, R. (2012) *Introduction to robust estimation and hypothesis testing* (3rd ed). Amsterdam: Academic Press.
- Wilkins, D. A. (1972) *Linguistics in language teaching*. London: Edward Arnold.
- Wilson, K. A., Bedwell, W. L., Lazzara, E. H., Salas, E., Burke, C. S., Estock, J. L., Orvis, K. L. & Conkey, C. (2009) Relationships between game attributes and learning outcomes: Review and research proposals. *Simulation & Gaming*, 40(2): 217–266. <https://doi.org/10.1177/1046878108321866>
- Wu, S.-C. (2002) English vocabulary learning strategies for junior college students in Taiwan. *The Journal of Hsing Wu College*, 31: 269–288.

## Appendix

*Questionnaire for identifying participants' learning perceptions in potential effectiveness, usability, and satisfaction*


Potential effectiveness	The used app is effective in terms of learning vocabulary
	The used app motivates me to learn English vocabularies
	The used app allows me to accomplish learning tasks more efficiently
	The used app helps me enlarge my vocabulary size
	The used app entitles me to greater control over my learning process
Usability	The resolution of mobile devices is adequate in terms of overall use
	The pixel of mobile devices is adequate in terms of overall use
	Learning to operate the app for learning vocabularies is easy for me
	Interaction with the app requires little mental effort
	Generally speaking, the used app is easy to use
Satisfaction	I think it is interesting while using the app for vocabulary acquisition
	I am satisfied with the functions offered by the used app
	I am satisfied with the used app as a vocabulary learning tool
	I am willing to keep using mobile devices for English vocabulary learning
	I am willing to keep using the app for vocabulary acquisition

**About the authors**


Chih-Ming Chen is a distinguished professor in the Graduate Institute of Library, Information and Archival Studies at National Chengchi University, Taiwan, ROC. He received a PhD degree from the Department of Electronic Engineering at National Taiwan University of Science and Technology in 2002. His research interests include computer-assisted language learning, digital libraries, e-learning, data mining, and intelligent agents on the web.

Huimei Liu is a professor in the Department of Statistics at National Chengchi University, Taiwan, ROC. She received a PhD degree from the Department of Statistics at North Carolina State University in 1993. Her research interests include application of importance sampling, multivariate analysis, and tests for hypotheses about linear inequalities.

Hong-Bin Huang is currently a graduate student in the Graduate Institute of Library, Information and Archival Studies at National Chengchi University, Taiwan, ROC. His research interests include computer-assisted language learning and e-learning.

Author ORCID.  Chih-Ming Chen, <http://orcid.org/0000-0002-7088-5516>

Author ORCID.  Huimei Liu, <http://orcid.org/0000-0001-6604-5509>

Author ORCID.  Hong-Bin Huang, <http://orcid.org/0000-0002-3078-3574>