

GAMIFIED MOBILE-ASSISTED FORMATIVE ASSESSMENT FOR REVIVING UNDERGRADUATE LEARNERS' OVERALL LANGUAGE PROFICIENCY: A QUASI-EXPERIMENTAL STUDY

by **Baderaddin Yassin** and **Mohammed Abdulgalil Abugohar**

Alghad International Colleges for Applied Medical Sciences, Saudi Arabia

byassin @ gc.edu.sa

Abstract

The lack of opportunities to practice the English language outside the English as a Foreign Language (EFL) classroom can prevent English language learners (ELLs) from promoting their language proficiency to high standards. This lack makes the progress from one level of English to the next one a hard mission for Arab students. Subsequently, students' mastery of the English language is often not expected to occur without frequent practice and organized formative assessment. To enrich such methods of assessment, a mobile-assisted language learning (MALL) technique was adopted in a classroom formative assessment for holistic language proficiency of listening, vocabulary, and grammar to determine whether it would make a difference in results. For fourteen weeks, this quasi-experiment consisting of a pre-post-test one group design was carried out over two cycles of seven weeks each, with 598 participants. During the experiment, students practiced formative assessment conventionally during the first cycle, while this assessment was done with the medium of two mobile apps: Kahoot! and Quizizz during the second cycle. The results of the Oxford Placement Test (OPT), replicated as the pre-test and post-test, indicated that mobile-assisted formative assessment resulted in a statistically-significant positive influence of using mobile apps on students' overall language proficiency.

Keywords: English Language Learners (ELLs); Mobile-Assisted Language Learning (MALL); Formative Assessment; Oxford Placement Test (OPT); Adult Learning

1. Introduction

There is a vital need for learning the English language in the Arab Gulf countries. These countries are like a melting pot where a common language is essential, and that language is often English. The Arab Gulf countries include a large number of people from different parts of the world. Foreign experts, workers and expatriates from various nationalities come to Arab Gulf countries to work for financial reasons or various other reasons. There is a need for

communication and interaction in the society and working environment; however, the native language that is used in Arab Gulf countries is Arabic, which is not an easy language to learn. Therefore, many international people struggle with the language barrier that prevents them from communicating professionally in the national language. To overcome this language barrier, many foreign nationals use the English language since most visitors and expatriates speak English or at least have some background knowledge of the language. Furthermore, English is the most widely spoken language in the world, with 1.5 billion speakers. As we become increasingly interconnected, English has assumed a position as the dominant language of global communication, notably on the Internet.

Most universities in Saudi Arabia (an Arab Gulf country) teach curricula in English. Therefore, Saudi students are required to learn the English language to succeed in their university studies. Applied Medical colleges follow the same path and learning English for medical purposes (EMP) is compulsory in the first year of study in Saudi medical colleges. Students are required to attend intensive English language classes during the first and the second semester of the first year. The courses that are taught in the first year concentrate on English for medical purposes in addition to practicing other English language skills. Although students are required to successfully pass the first year to proceed with their majors, they often struggle to convey messages in English accurately. Based on these facts, many teachers have attempted to help students by using MALL teaching strategies in their classes. This increased use of MALL has led many English language teachers to wonder if students will be more motivated to learn English using such strategies and if their overall language proficiency will be improved as a result.

The lack of English practice opportunities has its impact on every English language learner in the Arab Gulf countries. Still, technology-based tools can motivate learners and give them the drive to effectively utilize the opportunities that exist and create practice opportunities as well. Assessing whether and to what extent learning techniques and teaching strategies employed in certain contexts influence student engagement is essential to delivering effective language courses. In other words, if English language learners receive appropriate and updated techniques of feedback covering instant and delayed error corrections in a motivating, engaging environment, their language awareness and skills are expected to be improved. Consequently, mobile-assisted formative assessment, which is one of the offered solutions, was empirically examined in this study.

2. Literature review

2.1. Gamification

In recent years, educators have invested considerable efforts in incorporating digital technologies into English as a Foreign Language (EFL) pedagogy to motivate students in new ways (Jalili, Khalaji & Ahmadi, 2020). Over the past two decades, many scholars have investigated the influence of technology on EFL learners. Researchers have also scrutinized the effects of using technological development on learners' proficiency, performance, attitudes and motivation. The integration of technology in EFL learning is of great importance because English is the language of technology and most of the new generation of students are frequent users of technology, especially mobile applications (Rashid et al., 2021). The use of technology as a means of delivering EFL learning includes various applications, some designed for the sole aim of EFL learning, while others utilized for social networking with potential use in the classroom.

The reviewed literature in this section can be divided into two domains – exploring and investigating attitudes and perceptions and examining actual implementation. Each domain seeks to clarify the position of the current study in the literature.

In terms of investigating attitudes and/or perceptions, Jebur (2020) discussed the attitudes of Iraqi English as Foreign Language learners towards mobile-assisted language learning (MALL) usage. The study concluded that Iraqi students showed positive attitudes and were engaged users of MALL applications. This suggests that learners integrated MALL in the process of learning English to improve their learning. The study recommended including mobile applications in the learning-teaching process to enhance written and oral skills as it could increase the learners' motivation.

In a similar vein but with a different objective, Abugohar, Yunus and Rashid (2019) investigated EFL teachers' perceptions and their current actual practices of utilizing a package of three categories of smartphone applications in language classrooms for fostering Saudi tertiary students' speaking skills. Applying a mixed-mode approach, the research findings revealed that most participants had high, inspiring positive perceptions of using smartphone applications in teaching speaking. However, classroom practices also revealed weaknesses and insufficient real-world experiences.

Zou and Li (2015) investigated how mobile apps could be integrated into the teaching-learning of the English language and what can be utilized to support the process of learning for EFL learners. They found that most of the participants had a positive attitude. The majority of

the students would conduct a variety of practices for learning on their devices and most participants enjoyed the learning process using apps. The authors noted that this created an impact on student motivation to learn English.

Moreno and Vermeulen (2015) highlighted the availability of mobile-assisted language learning (MALL) apps and their role in practicing English oral skills. The population sample selected for the research study was a group of Spanish and Belgian students. The study concluded that a pedagogical standard must be established for MALL apps as these need to be linked to a setting of blended learning and to only used as a form of support. There were signs of enhancements for the Spanish students in terms of motivation and curiosity, and the Belgian students obtained good learning outcomes as well.

Nowbattula, Devi, and Nimmala (2016) shed new light on the use of social media and mobile applications to improve students' language skills. Their study concluded that learners' motivation and confidence was enhanced through the use of applications. They noted that these apps can be flexible, personalized, and customized to each individual according to his/her preference. Another factor noted in the study was the accessibility of these apps and the possibility to choose a suitable time and place to carry out the learning process. They also noted that these apps seemed to encourage learners to create life-long habits of learning.

Regarding actual practices and experimenting with the effectiveness of MALL, Trust (2020) reviewed the top five trends in educational technology (EdTech) during the ISTE 2017 Conference. These top five trends were identified as Google, Tools and apps, Global education, Making technology, and 3D. Among the tools and apps she listed were Kahoot!, Popplet, Padlet, Quizziz, Google Street View, Screencast-O-Matic, and Adobe Spark.

Chen and Yeh (2019) explored the implications of student-generated questioning (SGQ) using Kahoot! for facilitating English learning in the flipped Foreign Language classroom. In a flipped classroom, the out-of-class instructional videos familiarized learners with what they would be asked to do when they were assigned in-class collaborative or individual work to perform (Noroozi, Rezvani, & Ameri-Golestan, 2021). During Chen and Yeh's (2019) study, seventy-seven university students were subdivided into an experimental group labeled SGQ_FL and a control group determined by teacher-generated questioning (TGQ), labeled TGQ_FL. The results demonstrated that the SGQ was effective in enhancing students' English performance. However, there was no significant difference in the mental load between groups, yet SGQ induced a higher mental effort than TGQ.

Meanwhile, Nurhalim, Saputra, and Pujasari (2019) explained the importance of speaking English as a means of communication with the world and how mobile apps such as

Opentalk can motivate speaking for EFL students. The results illustrated that the app gave made students more confident, especially those who had anxiety and fear of making mistakes. The app also motivated students to speak and improve their performance. Another factor noted was the acquisition improvement that the students gained from the Opentalk app.

Hidayati and Diana (2019) investigated the use of Duolingo and Hello English apps in motivating students' learning of the English language. The researchers found that students were using both apps actively, and enjoyed the flexibility of the apps. The researchers concluded that the mobile application could assist learners and allowed students to learn independently using such apps due to their particularity and flexibility. The authors also noted that the variety of resources using mobile technology provides an additional teaching resource in the classroom beyond the teacher.

Similarly, Hamad, Metwally, and Alfaruque (2019) investigated the impact of using YATI (YouTube videos and Listening Audio Tracks Imitation) on enhancing the speaking skills of EFL learners. They claimed that the problem of the non-native speakers and their teachers is related to the learners' inability to pick the right tone and intonation and their weaknesses in vocabulary, which can be easily gained by practicing listening and speaking. They concluded that the imitation has a slow but steady enhancement in listening and speaking skills. In contrast to earlier research which suggested that using imitation limits the talents of students (e.g., Finke, Ward, & Smith, 1992; Jansson & Smith, 1991; Lowenfeld, 1957), the researchers proved that using such methods improved students' competence and helped them discover their abilities.

Meanwhile, Al-Garawi (2019) shed light upon the usage of Instagram social networking application as a MALL tool. Due to the popularity of this application, it can work as a means of language acquisition through linking the learners to native speakers. The study concluded that using the Instagram application has many benefits, such as providing learners with an enjoyable experience while enhancing their oral skills, improving writing, reading and listening skills, and developing learners' vocabulary storage.

Mustafa (2018) investigated the improvement of speaking skills for EFL learners through Skype, YouTube, and WhatsApp. He concluded that the English spoken language could be improved by using these apps as teaching tools. Each of the apps had its proof-based benefits. He noted that YouTube improved listening while WhatsApp enhanced speaking. On the other hand, Mustafa observed that Skype enhanced both as it led to improvement of speaking and listening skills via real conversations. The study recommended that teachers

should consider utilizing apps to provide learners with a means of communication with native speakers.

Similarly, Ramos and Valderruten (2017) studied the use of mobile applications to develop listening skills and other linguistic skills. For the aim of the study, the researchers designed an app to carry out their investigation. Their results demonstrated that the use of the mobile app influenced the development of listening skills and the linguistic competence of English learners. The authors suggested that this was due to the apps being popular with younger generations, which allowed them to practice more.

Mindog (2016) studied the use of smartphone apps and their role in learning English as a Foreign Language (EFL). The apps included social networking apps and others. The sample used for this study was a group of university students in Japan. The study findings showed that intermediate learners use apps to participate in social networking. The study supported that these apps can be used for learning. The large number of apps also helped learners enjoy the freedom of choice and the frequency of usage as it allowed learners to enjoy learning without worrying about tests or any other type of pressure.

In a 2016 study, Moghaddas and Bashirmezhad studied MALL (Mobile-assisted Language Learning) apps in enhancing the EFL (English as a Foreign Language) of Iranian learners' oral skills. The study results showed applying mobile apps to be effective for improving learners' performance in speaking skills. The study also concluded that mobile applications motivated practitioners and had a great impact on the fluency of language learners.

Meanwhile, Read and Kukulska-Hulme (2015) shed light on the importance of the audio recordings of the Audio News Trainer (ANT) app in listening comprehension and the use of social media to reflect on the news. They indicated that two factors must be available in the MALL (Mobile Assisted Language Learning) apps: the domain and the possible interaction. ANT appeared to be effective for listening practice; however, the students' reactions were typically limited for various reasons, such as the naturalness of social media apps and the topic.

Baniabdelrahman (2013) studied the speaking proficiency of Saudi university students mediating online diaries in the context of EFL. The study results showed that the students' motivation and participation increased, and they showed motivation towards learning. The study recommended that teachers utilize various techniques, including oral diaries, in teaching the English language as it can be used by students outside classrooms because it is available for mobile devices.

From the literature briefed in the previous section, it can be said that theories such as behaviourism, constructivism and situated learning theories have been applied to mobile-

assisted language learning (MALL) strategies. MALL strategies deal with the practices that deal with developmental process by simultaneously interlinking learners and social level of learning. Learning can be effectively achieved when learning is reinforced with particular response and stimulus (Smith and Ragan, 2005). The same theoretical approach functions as a tool in the process of analysing the benefit of integrating MALL strategies in EFL classrooms. Thus, this study was designed to examine the effectiveness of formative assessment in language classrooms with the intervention of MALL with a large sample of participants to provide comprehensive conclusions about the effectiveness of MALL in the teaching of EFL.

The intervention was the medium of formative assessment, which is the type of treatment implemented using mobile apps in formative assessment through which researchers thought proposed that it would create changes in participants' results on the post-test. However, the gender effect was also examined at a later stage of the study. This study's findings were expected to expose the experience using mobile-mediated techniques implemented in classroom formative assessment. One more objective included in the study was the training of respective teachers on using mobile-assisted formative assessment to promote learners' overall language proficiency.

3. The study

Although current literature proved the idea that game-based mobile applications could motivate EFL learners and provide flexible, accessible means for acquiring and learning second and foreign languages, it is still not clear whether MALL-assisted learning strategies should play a primary role in educating EFL learners or should remain as a secondary supportive educational tool. In this research paper, the authors tried to unveil the promising potential of mobile-assisted formative assessment that make incorporating MALL teaching strategies a must for educators and curriculum planners.

Most study findings so far were either surveying attitudes and/or perception of teachers and/or students or experimenting with a small number of participants. Consequently, this research was aimed at either supporting or contradicting the findings of previous studies on one hand, and, on the other, hoped to be of a distinguished value with its large number of participants while experimenting mediating mobile apps in language formative assessment. This study was also in response to those who claimed the lack of actual practices of MALL in the EFL context.

3.1. The aim of the study

This study aimed to examine the effect of using mobile apps in classroom formative assessments on Saudi First Year learners' overall language proficiency compared to the traditional methods of formative assessment. The study addressed the following questions:

- 1) Are there significant improvements in the students' mean scores in the post-test due to the employment of mobile apps in formative assessment?
- 2) Are the proficiency scores obtained by participants affected by gender?

3.2. Participants

The study's target population were university students enrolled in their first semester in a higher learning medical institution in Saudi Arabia. After being admitted by the institution, students were required to join the Preparatory Year Program (PYP) for one academic year of two semesters. They were required to successfully complete the foundation program which includes English and additional science courses. The population number reached 658 (male = 303, female = 355) enrolled in their first semester.

The research was first intended to be a census studying the entire population since all students enrolled had to sit for the tests. However, after collecting the data of the post-test, it was found that some students missed the post-test. Consequently, the final number of participants to complete the study were 598 (N= 598; male = 299, female = 299). This large sample size was aimed to guarantee comprehensive conclusive findings.

3.3. Design and procedure

Researchers generally opt for the hypothesis-testing experimental design (Kothari, 2004) to determine the cause-effect outcomes of the treatment (Mackey & Gass, 2005). This study was designed to test the effect of using mobile apps in formative assessments on the university learners' overall English language proficiency. This study was a quasi-experiment with only one group of students, in other words, a one group pre-test/post-test experimental design. All participants were treated as one group with two sets of scores prior to and after the treatment. Participants' overall language proficiency was assessed before and after the intervention of gamified mobile-assisted formative assessment, which lasted 14 weeks. However, no control group made this experiment quasi, which in return limits the generalization of the findings.

The study's main independent variable was the assessment method, which was mobile-assisted formative assessment, while the dependent variable was the students' scores in the proficiency post-test. However, any proficiency level differences due to gender were examined

at a later stage when gender was tested as a moderator categorical variable to determine whether participants' gender influenced the scores obtained.

The study was carried out in two cycles. During each cycle, a different type of assessment was implemented. Moreover, each cycle was followed by a language proficiency test. More details are given below:

Cycle One

Language instructors deliver part of the syllabus to the target population using the conventional teaching kit provided by the English Language Center and utilizing traditional methods of assessment and using paper-based quizzes on the video clips attached to the course kit for the listening skill, as well as paper quizzes and worksheets for assessing learners' awareness of meaning and forms. This cycle covered a period of seven weeks. Then, the online pre-test assessed students' overall language proficiency was implemented to determine how the regular ways of formative assessment reflect teaching-learning process outcomes.

Cycle Two

At the end of the first cycle, all ESL instructors were trained intensively by a Pearson Middle East team on using the Kahoot! and Quizziz mobile apps. Both their attitudes and perceptions and their hands-on implementation have been regularly evaluated. After the training was completed, the teachers started using the apps as the media of formative assessment for seven weeks. Teachers' and learners' classroom practices were rigorously followed, and unannounced visits were performed to ensure the application of the target apps. Once again, students' overall language proficiency was assessed using the post-test after the treatment time had expired.

3.4. Instruments

3.4.1. Characteristics of the mobile apps set

There are several websites and mobile apps that are devoted to assessment. These can be used for summative or formative assessment (Edwards, 2020). For the purpose of this research, available mobile apps' properties were first previewed to determine which ones can work for classroom formative assessment. The initial list included Slido, Socrative, Flipgrid, Quizziz, Polleverywhere, Canva, Adobe Spark, Kahoot, Edpuzzle, Peergrade, Gimkit, Sutori, and Trello.

Then out of the listed apps, some were selected on the basis of a set of inclusion criteria of convenience to use, attractiveness, and availability on both Google Play and Apple Store for

smartphone users. The Pearson team, assigned by the general administration for the teaching staff coaching and drilling programs, trained ESL instructors on the use of the recommended apps, provided motivation for the teachers to use them, and then coordinated with the English Language Center to prepare them for actual classroom implementation. After the training was completed, the final list of mobile apps was closely tested, and after a constructive debate, two apps were selected to utilize for the research: Kahoot! and Quizizz.

3.4.2. Pre-test and post-test

Oxford Placement Test (OPT) was used for both the pre-test and the post-test in this research. No interference was allowed in the selection of the test items; it is 100% online and the sole decision regarding the selected test question falls to the Oxford English Testing services at Oxford University Press, English Language Teaching (ELT). The test format is presented in Table 1, as it exists on the Oxford University Press (OUP) ELT Website (2020).

Table 1. Oxford Online Placement Test Format (OUP, 2020)

SECTION		Part 1	Part 2	Part 3
Use of English	Description	Complete a short dialogue with the appropriate grammatical form or lexical item	Read a short dialogue and identify what a speaker means	Type the meaning word in gapped text with the appropriate grammar or lexis
	Testing focus	Grammatical form, lexical precision	Pragmatics: understanding explicit and implied meaning	Grammatical form, lexical precision
	Response type	Four-option multiple choice	Three-option multiple choice	Gapped text. For A1–C2, test takers type their answers. For Pre–A1, test takers select their answers.
	Questions	Ten tasks	Ten tasks, each with one or two questions	One task, seven questions
Listening	Description	Listen to a short dialogue and identify what the speaker means	Listen to a longer dialogue and identify what the speaker means	Listen to a monologue and identify what the speaker means
	Testing focus	Pragmatics: understanding explicit and implied meaning		
	Response type	Four-option multiple choice		

	Questions	Five tasks per part, each with one or two questions
--	------------------	---

As shown in Table 1, the test consists of two sections that identify test takers' use of English in terms of grammatical forms and meaning, as well as their listening skill. The two sections are used in this paper for overall language proficiency. The items test either language systems or listening comprehension, either explicitly or implicitly.

After taking the test, scores appeared to the organizer on the website and could be downloaded. Moreover, information about what a learner that achieved a particular CEFR level could perform was listed. Scores out of 120 were recorded and reported in light of CEFR by Oxford University Press ELT (2020) on a continuous numerical scale as follows:

Table 2. Test results reference (OUP, 2020)

Level	Score range
Pre-A1	0.1 – 0.9
A1	1 – 20
A2	21 – 40
B1	41 – 60
B2	61 – 80
C1	81 – 100
C2	101 – 120

Table 2 tabulates the reference of the scores based on CEFR. There are seven sets of scores with a level category referred to for each. In this paper, the standardized scores were used and tabulated as a CEFR level for the test as a whole to find overall proficiency.

3.5. Data analysis

This experimental research aims to investigate the impact of mediating mobile apps in formative assessment on participants' proficiency. The scores for both the pre-test and the post-test were recorded automatically on the respective part of Oxford website. All data were quantitatively analyzed and then statistically interpreted. The collected data were inserted and manipulated using SPSS (V 24.0).

Criterion validity was ensured since the assessor was an external party, Oxford English Testing Domain. Moreover, the reliability of the results recorded was tested by calculating Cronbach's alpha. Using SPSS, the result of internal stability was found at 0.841, which is considered a very good indicator of the results reliability, as shown in Table 3.

Table 3. Reliability Statistics

Cronbach's Alpha	N of Items
.841	2

4. Results and findings

4.1. Change in students' level

Data obtained during the first cycle were obtained from the pre-test scores. The participants' CEFR levels varied, as shown in Table 4.

Table 4. Pretest CEFR

	Frequency	Percent	Valid Percent	Cumulative Percent
Pre-A1	112	18.7	18.7	18.7
A1	337	56.4	56.4	75.1
A2	112	18.7	18.7	93.8
B1	32	5.4	5.4	99.2
B2	4	.7	.7	99.8
C1	1	.2	.2	100.0
Total	598	100.0	100.0	

The pre-test results showed that the students (N=598) who sat for the test scored varied results, but mainly left-shifted. As illustrated in Table 4, 112 students (18.7%) were at the Pre-A1 level, 337 participants (56.4%) were within the A1 range, and 112 (18.7%) were computed to have the features of A2 reference, while a few participants, (32 – 5.4%), could reach level B1. Even fewer ones (only 4 students – 0.7%) were seated in level B2, and only one participant (0.2%) could score higher to be placed at the C1 level. No participant (0%) was placed at the C2 level.

Data obtained during the first cycle were obtained from the pre-test scores. The participants' CEFR levels varied, as shown in Table 5.

Table 5. Posttest CEFR

	Frequency	Percent	Valid Percent	Cumulative Percent
Pre-A1	89	14.9	14.9	14.9
A1	307	51.3	51.3	66.2
A2	157	26.3	26.3	92.5
B1	40	6.7	6.7	99.2
B2	4	.7	.7	99.8
C1	1	.2	.2	100.0
Total	598	100.0	100.0	

The post-test results showed that the students ($N=598$) who sat for the test scored varied results. As illustrated in Table 5, 89 (14.9%) were placed at the Pre-A1 level, 307 (51.3%) were within the A1 range, 157 (26.3%) were computed to have the features of A2 reference, and 40 (6.7%) could reach the B1 level. Very few participants (4 – 0.7%) proved to be at B2 level B2, only one participant (0.2%) could score higher at C1; whereas none (0%) was placed at the C2 level.

To deeply analyze the results obtained for the pre-test and the post-test, the paired-samples t -test was run using SPSS (V 24.0). The statistics can be seen in Table 6.

Table 6. Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	PRETEST	15.2916	598	13.12613	.53677
	POSTTEST	21.3329	598	15.21680	.62226

As depicted in Table 6, the mean scores ($N=598$) obtained for the pre-test was around 15.3 ($M = 15.29$, $SD= 13.13$), while the post-test mean score ($N=598$) was 21.33 ($M = 21.33$, $SD= 15.22$). The mean score of the posttest ($M= 21.33$) is higher than that of the pretest ($M= 15.29$). To check this finding statistically, Table 7 presents the calculation done.

Table 7. Paired Samples *t*-Test

		Paired Differences							
		Mean	Std. Deviation	Std. Error	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
					Lower	Upper			
Pair 1	PRE-TEST – POST-TEST	-6.04130	6.36541	.26030	-6.55252	-5.53009	-23.209	597	.002

It has been stated earlier that all participants were treated as one group at the onset of data collection and analysis. Table 7 shows the result of running the paired-samples *t*-test to compare the pre-test's mean score to the post-test. As shown, it was found statistically significant ($T=23.209$, $DF=597$, $p=0.002$).

These calculations conclude that the participants' mean scores are not similar in the two tests. There is a significant difference between the participants' results in the pretest and their results after the posttest. The paired-samples *t*-test, as in Table 7, reveals that 95 percent of the students' proficiency scores in the post-test are better than their scores on the pre-test.

4.2. Gender influence

At this stage, the participants and their scores were divided into two groups based on gender ($N=598$, Male=299, Female=299). Then, their proficiency in the pre-test and the post-test were compared to see if there was an effect of the gender variable on the results recorded in each test.

The true difference between the mean scores of the two groups according to the variable of gender (Male & Female) was determined by running the independent samples *t*-test in SPSS (V 24.0) for both the pre-test and the post-test. The summary of each gender group elements is tabulated in Table 8.

Table 8. Group statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Pretest	Male	299	15.6291	13.37160	.77330
	Female	299	14.9542	12.88954	.74542
Posttest	Male	299	21.1967	16.23306	.93878
	Female	299	21.4692	14.15377	.81853

Table 8 compares the data obtained for the two groups according to gender. In the two sets of scores, males represented half of the total number ($N=598$) of the samples; (Males: $N=299$, 50%), and the females were the other half; (Females: $N=299$, 50%). In the pre-test scores, the mean score for male students was 15.62 and 14.95 for females. For the post-test, the mean score was 21.19 for the male and 21.46 for the female participants. It can be elicited from Table 8 that the mean scores are close to each other for male and female groups. Table 9 explores the difference statistically.

Table 9. Independent samples *t*-test

Test	Variable	Levene's Test		t-test	
		F	Sig.	df	Sig. (2-tailed)
Pretest	Male	1.790	.181	596	.530
	Female				
Posttest	Male	2.744	.098	596	.827
	Female				

The data depicted in Table 9 shows the outcomes of the independent samples *t*-test for the two sets of scores for each gender. First, the Levene's Test for the equality of variances (Levene, 1960) of the variable for the pretest ($F=1.790$, $Sig.=0.181$), and the posttest ($F=2.744$, $Sig.=0.098$) concluded that it is safe to report the *t*-test result.

The *t*-test is run to check for the equality of means. For the pre-test, the *t*-test was not significant ($DF=569$, $p=0.530$); thus, this reveals no significant difference between the mean scores of males and females in the pre-test scores. Moreover, the *t*-test for the post-test scores states that there is again no significant difference between the two sets of scores based on the variable of gender ($DF=569$, $p=0.827$). These calculations justify rejecting the hypothesis that there is an effect of gender on proficiency scores in either the pre-test or the post-test.

5. Discussion

In this quasi-experiment of a pre-test/post-test design, researchers experimentally investigated the influence of mobile-assisted formative assessment on university students' overall language proficiency. Students' holistic proficiency was assessed in two cycles; once after using conventional classroom formative assessment, and the other after utilizing mobile apps using the two target apps: Kahoot! and Quizizz, during which Kahoot! was used for assessing the

listening comprehension by formulating quizzes based on listening while watching through YouTube links, whilst Quizizz was used for checking language systems awareness; meaning and forms. Despite this classification, the two apps were treated as one package used for overall language proficiency, and all participants were examined as one group. Figure 1 visualizes a summary of the two sets of results by comparing the scores recorded in the two cycles.

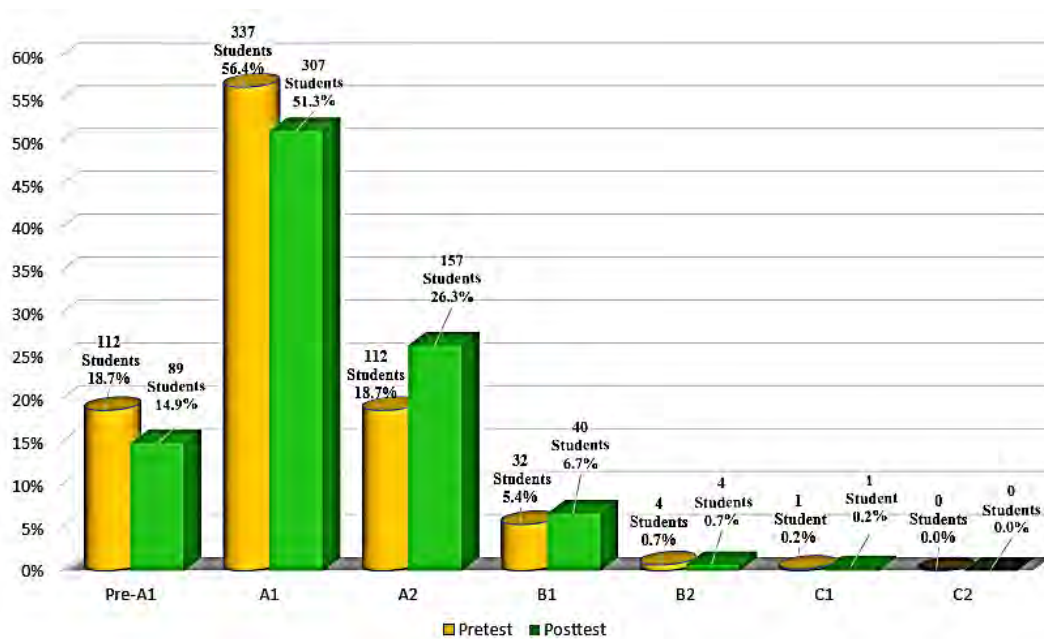


Figure 1. Summary of results

Based on the data presented in Figure 1, aided by those in Tables 4 and 5, it can be said that the number of students with Pre-A1 level decreased by 23 students (3.8%), and A1 dramatically dropped by 30 participants (5.1%). The reduction that occurred in the levels of Pre-A and A1 was mainly in favour of the A2 level, which went up by 45 students (7.6%), while B1 had a small share of that increase by 8 students (1.3%). On the other hand, levels B2 and C1 did not witness any change in number. However, when referring to the detailed scores of males and females, it was determined that a female was the one who scored C1 in the pre-test, whereas it was a male who scored C1 in the post-test.

Furthermore, the calculation of the paired-samples t-test, as presented in Tables 6 and 7 showed that there was a statistically significant difference ($p=0.002$) between the participants' proficiency scores in the pre-test and theirs in the post-test in favor of the post-test which was administered after the treatment. This conclusion indicates a significantly positive impact of mediating the mobile apps used in formative assessment; Kahoot! and Quizizz.

These apps helped students effectively practice language, complete homework, and participate in the classroom for on-the-go assessment while the course was in progress. The mobile-assisted formative assessment could also encourage a motivating learning environment evident in their post-test results compared to the pre-test scores. MALL shows the potential to encourage learners to become more engaged in the target language and evoke less test anxiety among participants. Moreover, the challenging and competitive gamified formative assessment helped to reinforce learning experiences and transition those experiences to long-term memories since they are not only about answering or ticking but also communicating, competing, and racing. This conclusion aligned with findings from Balchin and Wild (2020), Baniabdelrahman (2013), Chen and Yeh (2019), Nurhalim, Saputra, and Pujasari (2019), Ramos and Valderruten (2017), and Mindog (2016), among others. This agreement may have resulted from appealing mobile apps to language learners who already accustomed to utilizing apps in their daily lives.

Referring to the hypothesis of having gender interference, the independent samples *t*-test excluded that expectation. The non-significant effect of gender between participants was also supported by Baniabdelrahman (2013), which can be justified by having a similar study context, that of Saudi university students. However, this finding contrasted with a considerable amount of literature which concluded that females learn and acquire a second language better and faster than males due to biological growth in favour of females (Al-Saadi, 2020; Bermúdez & Prater, 1994; Medina, 1993; Wyk & Mostert, 2016). This inconsistency between this study and previous studies in terms of gender influence might be due to some observed assumed cultural properties of most Arab countries and Muslims' cultures where female students have fewer opportunities to be in contact with foreigners and native speakers before they join the university than male students often have. In return, this situation might have enabled male participants to bridge that gap and strike a mean score similar to that for females.

In brief, mobile devices facilitate anytime-anywhere learning and offer entirely new ways for students to learn. In higher education, mobile learning is a way to add innovation in learning strategies. Benefits of mobile learning include providing flexible, accessible, authentic, personalized, ubiquitous, and seamless language proficiency. Nonetheless, some mobile learning issues can also add technical problems, cognitive load issues, distraction, equity, and safety issues as well (Bower, 2017).

6. Conclusion

In conclusion, teachers are highly encouraged to infuse modern technology in the respective areas (Edwards, 2020), especially language teachers who can, aided by mobile apps, help students promote their second and foreign languages proficiency. Gamified formative assessment can help students become more engaged in the lesson materials, provide in-the-moment feedback, promote autonomy, and facilitate interactive assessment methods. Despite the obstacles that might be encountered, this research proves that it may be worth the extra effort. Language teachers need to spotlight the outcomes and attempt to overcome the shortcomings of apps in language learning. Moreover, curriculum designers are urged to incorporate various types of assessment activities with game-based utilities that reduce foreign language learners' anxiety. Finally, further studies are invited to examine the same apps or different ones in various learning contexts.

Although MALL techniques help EFL students enhance their English language proficiency and motivate them to study in classrooms and outside classrooms, it is desirable to implement MALL teaching techniques according to students' needs, and to deliver multiple language skills in real learning environments. In addition, since the present study has been carried out in a private higher educational institution, the results may vary in a public university setting, thus, further future research is urged in different contexts.

References

- Abugohar, M, Yunus, K., & Rashid, R. (2019). Smartphone applications as a teaching technique for enhancing tertiary learners' speaking skills: Perceptions and practices. *International Journal of Emerging Technologies in Learning*, 14(9), 47-92. <https://doi.org/10.3991/ijet.v14i09.10375>
- Al-Garawi, A. (2019). Investigating the use of Instagram as a MALL tool: The case of Al-Imam University EFL students. *Arab World English Journal Theses ID*, 233.
- Al-Saadi, Z. (2020). Gender differences in writing: The mediating effect of language proficiency and writing fluency in text quality. *Cogent Education*, 7(1), 1770923. <https://doi.org/10.1080/2331186X.2020.1770923>
- Balchin, K., & Wild, C. (2020). Exploring the role of context and collaboration in normalising technology use in English language teaching in secondary schools in Malaysia. *Computer Assisted Language Learning*. <https://doi.org/10.1080/09588221.2020.1803360>
- Baniabdelrahman, A. A. (2013). Effect of using internet tools on enhancing EFL students' speaking skill. *Journal of Contemporary Research*, 3(6), 79-87.
- Bermúdez, A. B., & Prater, D. L. (1994). Examining the effects of gender and second language proficiency on Hispanic writers' persuasive discourse. *Bilingual Research Journal*, 18(3-4), 47-62. <https://doi.org/10.1080/15235882.1994.10162667>

- Black, P., & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 92(1), 81-90. <https://doi.org/10.1177/003172171009200119>
- Bower, M. (2017). Designing for mobile learning. In M. Bower (ed.), *Design of technology-enhanced learning* (pp. 261-304). Emerald Publishing Limited. <https://doi.org/10.1108/978-1-78714-182-720171011>
- Chen, C., & Yeh, H. (2019). Effects of integrating a questioning strategy with game-based learning on students' language learning performances in flipped classrooms. *Technology, Pedagogy and Education*, 28(3), 347-361. <https://doi.org/10.1080/1475939X.2019.1618901>
- Coatney, S. (2006). The importance of background knowledge. *Teacher Librarian*, 34(1), 60. Retrieved from Academic Search Premier database.
- Corpus, J. H., Iyengar, S. S., & Lepper, M. R. (2005). Intrinsic and extrinsic motivational orientations in the classroom: Age differences and academic correlates. *Journal of Educational Psychology*, 97(2), 184-196.
- Edwards, V. (2020). It is time to make use of technology. *Journal of Physical Education, Recreation & Dance*, 91(2), 56-57. <https://doi.org/10.1080/07303084.2020.1702395>
- Finke, R. A., Ward, T. B., & Smith, S. M. (1992). *Creative cognition: Theory, research, and applications*. Cambridge, MA: MIT Press.
- Furrer, C., & Skinner, E. (2003). Sense of relatedness as a factor in children's academic engagement and performance. *Journal of Educational Psychology*, 95(1), 147-159.
- Gibbs, G. (1998). Marking and giving feedback. In Graham Gibbs (Ed.), *Teaching in higher education: Theory and evidence* (pp. 3-37). Open University.
- Glauser, E.C. (2004). Summary of the Book by Alfie Kohn: Punished by Rewards. *The Swiss Deming Institute*, 47(8), 1-6.
- Hamad, M. M., Metwally, A. A., & Alfaruque, S. Y. (2019). The impact of using YouTubes and audio tracks imitation YATI on improving speaking skills of EFL learners. *English Language Teaching*, 12(6), 191-198.
- Hidayati, T., & Diana, S. (2019). Students' motivation to learn English using mobile applications: The case of Duolingo and Hello English. *Journal of English Education and Linguistics Studies*, 6(2), 189-213. <https://doi.org/10.30762/jeels.v6i2.1233>
- Higgins, M., & Grant, F., & Thompson, P. (2010). Formative assessment: Balancing educational effectiveness and resource efficiency. *Journal for Education in the Built Environment*, 5(2), 4-24. <https://doi.org/10.11120/jebe.2010.05020004>
- Jalili, S., Khalaji, H., & Ahmadi, H. (2020). Vocabulary learning in the mobile-assisted flipped classroom in an Iranian EFL context. *Teaching English with Technology*, 20(4), 82-95. <https://tewtjournal.org/download/6-vocabulary-learning-in-the-mobile-assisted-flipped-classroom-in-an-iranian-efl-context-by-shahriar-jalili-hamidreza-khalaji-and-hosseini-ahmadi/>
- Jang, H., & Reeve, J. (2006). What teachers say and do to support students' autonomy during a learning activity. *Journal of Educational Psychology*, 98(1), 209-218.
- Jebur, M. S. (2020). The use of mobile-assisted language learning (MALL) by Iraqi EFL college students' and their attitudes towards it. *Journal of University of Babylon for Humanities*, 28(2), 150-160.
- Jansson, D. G., & Smith, S. M. (1991). Design fixation. *Design Studies*, 12, 3-11. [https://doi.org/10.1016/0142-694X\(91\)90003-F](https://doi.org/10.1016/0142-694X(91)90003-F)

- Kan, A. (2007). An alternative method in the new educational program from the point of performance-based assessment: Rubric scoring scales. *Educational Sciences: Theory & Practice*, 7(1), 144-152.
- Kothari, C. R. (2004). *Research methodology: Methods and techniques*. New Age International.
- Levene, H. (1960). Robust tests for equality of variances. In Ingram Olkin; Hotelling; et al. (Eds.), *Contributors to probability and statistics: Essays in Honor of Harold Hotelling* (pp. 278-292). Stanford University Press.
- Lowenfeld, V. (1957). *Creative and mental growth* (3rd ed.). MacMillan.
- Mackey, A., & Gass, S. M. (2005). *Second language research: Methodology and design*. Lawrence Erlbaum Associates.
- Medina, M. (1993). Spanish achievement in a maintenance bilingual education program: Language proficiency, grade and gender comparisons, *Bilingual Research Journal*, 17(1-2), 57-81. <https://doi.org/10.1080/15235882.1993.10162648>
- Mindog, E. (2016). Apps and EFL: A case study on the use of smartphone apps to learn English by four Japanese university students. *JALT CALL Journal*, 12(1), 3-22.
- Moghaddas, B., & Bashimezhad, H. (2016). The pedagogical applications of mobile-assisted language learning (MALL) in improving the Iranian EFL learners' oral performance. *International Journal of Applied Linguistics Translation*, 2(1), 8-14.
- Moreno, A. I., & Vermeulen, A. (2015). Profiling a MALL app for English oral practice: A case study. *Journal of Universal Computer Science*, 21(10), 1339-1361.
- Mustafa, E. N. E. (2018). The impact of YouTube, Skype and WhatsApp in improving EFL learners' speaking skill. *International Journal of Contemporary Applied Researches*, 5(5), 18-31.
- Noroozi, A., Rezvan, E., & Ameri-Golestan, A. (2021). Students' perceptions of the incorporation of flipped learning into L2 grammar lessons. *Teaching English with Technology*, 21(1), 112-130. <https://tewtjournal.org/download/8-students-perceptions-of-the-incorporation-of-flipped-learning-into-l2-grammar-lessons-by-ahmad-noroozi-ehsan-rezvani-and-ahmad-ameri-golestan/>
- Nowbattula, P. K., Devi, V. R., & Nimmala, T. (2016). Teaching listening and speaking skills to the pupils of higher education through mobile applications and social networks. *Scholar Critic*, 3(1), 351-360.
- Nurhalim, G. A., Saputra, Y., & Pujasari, R. S. (2019). Encouraging students' speaking through Opentalk app: Participatory action research. *Journal of Universitas Siliwangi*. <https://doi.org/10.13140/RG.2.2.14738.81608>
- Oxford University Press (OUP) ELT Website (2020). <https://elt.oup.com/feature/global/oxford-online-placement/>
- Ramos, L., & Valderruten, A. (2017). Development of listening and linguistic skills through the use of a mobile application. *English Language Teaching*, 10(9), 95-107.
- Rashid, R., Annamalali, N., Saed, H., Yassin, B., Alsmadi, O. (2021). Developing an interactive university orientation app: Potential users' feedback. *International Journal of Interactive Mobile Technologies*, 15(22), 165-171.
- 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.
- Ruenzel, D. (2000). Gold star junkies. *Teacher magazine: News and information for teacher leaders*, 11(5), 6-24.
- Trust, T. (2017) The top five trends in Edtech according to ISTE 2017. *Journal of Digital Learning in Teacher Education*, 33(4), 126-127. <https://doi.org/10.1080/21532974.2017.1350082>

- Wang, J. H., & Guthrie, J. (2004). Modeling the effects of intrinsic motivation, extrinsic motivation, amount of reading, and past reading achievement on text comprehension between U.S. and Chinese students. *Reading Research Quarterly*, 39(2), 162-186.
- Watkins, C. (2005). Classrooms as learning communities: A review of research. *London Review of Education*, 3(1), 47-64.
- Wiggins, G. P., & McTighe, J. (2005). *Understanding by design* (2nd ed.). Association for Supervision and Curriculum Development ASCD.
- Wilbourne, M. (2006). Intrinsic and extrinsic motivational orientations in the classroom: Age differences and academic correlates. *Childhood Education*, 82(5), 310-311. Retrieved from Academic Search Premier database.
- Wooten, J. J., Acchiardo, C., & G. Dirk Mateer, G. D. (2020). Economics is a Kahoot! *The Journal of Economic Education*, 51(3-4), 380-380. <https://doi.org/10.1080/00220485.2020.1804499>
- Wyk, J. V., & Mostert, M. L. (2016). The influence of mother tongue and gender on the acquisition of English (L2): The case of Afrikaans in Windhoek schools, Namibia. *Cogent Education*, 3(1), 1210997. <https://doi.org/10.1080/2331186X.2016.1210997>
- Zou, B., & Li, J. (2015). Exploring mobile apps for English language teaching and learning. In F. Helm, L. Bradley, M. Guarda, & S. Thouëсны (Eds.), *Critical CALL – Proceedings of the 2015 EUROCALL Conference, Padova, Italy* (pp. 564-568). Research-publishing.net. <https://doi.org/10.14705/rpnet.2015.000394>