LEARN Journal: Language Education and Acquisition Research Network
(ISSN: 2630-0672 (Print) | ISSN: 2672-9431 (Online)
Volume: 15, No: 2, July – December 2022
Language Institute, Thammasat University
https://so04.tci-thaijo.org/index.php/LEARN/index

Usability and motivation study of mobile application for English language proficiency test preparation in Thailand: A case study of TU-GET CBT

Sichon Koowuttayakorn[a,*], Pimsiri Taylor[b]

[a] sichon.k@litu.tu.ac.th, Language Institute, Thammasat University, Thailand
[b] pimsiri.t@litu.tu.ac.th, Language Institute, Thammasat University, Thailand
[*] Corresponding author, sichon.k@litu.tu.ac.th

APA Citation:
Koowuttayakorn, S. & Taylor, P. (2022). Usability and motivation study of mobile application for English language proficiency test preparation in Thailand: A case study of TU-GET CBT. *LEARN Journal: Language Education and Acquisition Research Network, 15*(2), 625-648.

Received
20/04/2022

Received in revised form
02/06/2022

Accepted
07/06/2022

Keywords
usability; motivation; mobile application; English language proficiency test; test preparation

Abstract

Usability testing is a method that determines how user-friendly a website or a mobile application is. The present study investigated the usability of a newly launched mobile application called TU-GET CBT or Thammasat University General English Test (Computer-Based Test). It also strived to understand the extent to which the application contributes to users' motivation. Twenty-one target users were involved in three focus group sessions; they also completed a usability and intrinsic motivation questionnaire. In addition, five experts in the field of language education and technology performed the heuristic evaluation that measured the application's usability. The findings showed that in general, all participants were satisfied with the application in terms of its usefulness for English language learning and test preparation. Nonetheless, several technical usability issues were identified including accessibility, navigability, and aesthetics. Pedagogical usability of consistency and feedback were also raised across findings. In

	regard to motivation, the findings suggested that the users' motivation can increase if they perceive the application as valuable and enjoyable. These findings provide valuable implications for the next phase of the TU-GET CBT application development.

Introduction

The Language Institute of Thammasat University (LITU) has been offering the Thammasat University General English Test or TU-GET for more than 20 years. The original TU-GET is a paper-based test (also known as TU-GET PBT) consisting of three parts: structure, vocabulary, and reading. It is a test of English language proficiency required for those who wish to enroll in Thammasat University at both undergraduate and graduate levels. Public candidates may also take the test to measure their competency in English language use. In 2019, LITU launched a computer-based format of TU-GET, also known as TU-GET CBT. Unlike the paper-based format, candidates are required to take TU-GET CBT on the computerized system. The format of TU-GET CBT is based on the four skills of reading, listening, speaking and writing. By launching this test format, LITU had an objective to create a standardized test which assesses candidates' communicative competence rather than only the English grammar and vocabulary knowledge or reading skills.

In order to introduce the new test to the public and help them become familiarized with the test, LITU launched the pilot version of TU-GET CBT application at the end of 2020. The application acts as an alternative platform to test practice textbooks and it is available for free download on Android, and iOS for iPhone/iPad. Registration is required for first-time users. Once registered, users will be able to navigate through the application whose home page consists of six icons: Listening, Reading, Speaking, Writing, Vocabulary, Grammar, and Apply for TU-GET (See Appendix 1 for some screenshots of the application). The four language skills are parts that appear in the actual test, while vocabulary and grammar icons are for additional practice. The Apply for TU-GET button leads to an external URL which provides information about how to apply for the test. All of these are designed based on what candidates need in order to complete the test.

While factoring in the test requirements during design, how users experience the actual use of the application, i.e., how usable the

application is to them, must be prioritized. To investigate the usability of the application, this study is conducted amongst the prospective target users of 1) Thammasat university undergraduate students, 2) Thammasat university postgraduate students, and 3) the public who are interested in TU-GET CBT, as well as experts in the field of educational technology or language teaching and learning. Through data triangulation, the results of the study can inform developers on how to improve the quality of the TU-GET CBT mobile application.

The research questions of the present study are as follows:

1. What are the perceptions of target users of the TU-GET CBT mobile application in terms of its usability, and its contribution to users' motivation for English language learning and test preparation?

2. What are the usability issues pertaining to the TU-GET CBT mobile application as identified by expert participants?

Literature Review

Mobile Assisted Language Learning (MALL) and Test Preparation

MALL, or mobile assisted language learning, has gained more attention in recent years due to the mobility and distributed nature of mobile technology which enables users to access the learning content conveniently across time, locations, and contexts (Kukulska-Hulme, 2009). Aside from being a tool to encourage collaborative and personalized learning (Elaish et al., 2017), MALL is also found to increase learners' motivation by increasing their engagement, persistence, and effort for task completion (Carrera et al., 2018).

A few recent studies (Lestary, 2020; Saritepeci et al., 2019) have extended the use of MALL to exam and standardized test preparation and explored users' attitudes towards using certain MALL technologies for learning and preparing for such tests. Lestary (2020), for example, found that the participants held positive attitudes towards using mobile learning for IELTS preparation because of its convenience and flexibility. The mobility and portability of the mobile devices facilitates casual and personalized learning where users can customize learning to suit their individual needs, learning styles, and proficiency levels (Elaish et al., 2017). Similarly, Saritepeci et al.'s (2019) study, which explored the use of Whatsapp, an instant messaging application, for preparation of foreign

language proficiency exams in Turkey, reported several positive findings. These included the fact that MALL enhanced active learning, fostered learner-instructor interaction, and boosted learner motivation among the participants.

Despite various advantages, challenging aspects were also raised in both Lestary's (2020) and Saritepeci et al.'s (2019) studies. The main shortcomings reported included the lack of real-time direct feedback and self-motivation when learning on mobile devices. Several participants also mentioned that the knowledge gained from MALL interactions was superficial and hence, it cannot replace face-to-face learning. These findings, therefore, have led to the conclusion that while MALL offers diverse benefits for test preparation students, it might not be able to satisfy learners' need for direct feedback and deep learning gained from active and live interaction with instructors or tutors.

Usability

In order for MALL to become successful, content designers and system developers must fulfill the 'usability' criteria. Drawn from the International Standards Organization's (ISO) definition of usability, the term can be understood as the degree to which a mobile application can be used by specific users to achieve specific goals in a specific context with effectiveness, efficiency, and satisfaction (Hoehle & Venkatesh, 2015). In measuring a mobile application's usability, usability testing is a method commonly used in various disciplines including engineering, human-computer interaction disciplines, health care (Georgsson & Staggers, 2015), as well as education (Nokelainen, 2006). The Post-Study System User Questionnaire (PSSUQ) is one of the most widely used tools in measuring usability and perceived user satisfaction. It consists of 16-standardized items with three subsets which reflect System Usefulness (SYSUSE), Information Quality (INFOQUAL), and Interface Quality (INTEQUAL) (Lewis, 2002). Despite its comprehensive framework, PSSUQ may lack the perspective of pedagogical usability which is significant for MALL.

For educational mobile applications, usability can be investigated not only in terms of technical usability, but also pedagogical usability. Technical usability involves the ease of use, the efficiency of functions, and the convenience of mobile applications to avoid cognitive overload. In

designing mobile applications, consistency, or consistent visual design, is one of the meta principles which must be taken into account to yield good user experience (Schlatter & Levinson, 2013). In addition to technical usability, pedagogical usability places importance on the design of learning platforms or the system functions to facilitate users' learning (Nokelainen, 2006). To develop a useful digital learning platform for learners, both technical usability and pedagogical usability are crucial. While learning contexts in defining usability may vary, certain criteria in assessing technical usability and pedagogical usability have been proposed. Hadjerrouit (2012), for instance, summarized technical usability and pedagogical usability criteria in digital learning platforms based on Nielson and Molich's (1990) and Nokelainen's (2006) works as illustrated in Table 1. It is interesting to note that some pedagogical usability criteria, i.e., motivation and feedback, are closely linked with users' attitudes and perceptions towards MALL.

Table 1

Technical Usability and Pedagogical Usability Criteria in Digital Learning Platform (Hadjerrouit, 2012, pp. 49-50)

Technical usability	Pedagogical usability
• Ease-of-use. • Efficiency. • Technical design. • Accessibility and navigability.	• Added value. • Motivation. • Differentiation. • Collaboration. • Discussion. • Assessment. • Peer-review and feedback.

Learner Motivation and MALL

Due to the unprecedented growth in digital and internet technologies, the advent of MALL, especially in the form of smartphone applications for English language learning, has increased exponentially over the past decade. It has become normalized for people in this generation to access, learn, exchange information, or entertain themselves via online resources in their second and additional languages (Godwin-Jones, 2017). The emergence of said MALL applications has led to the exploration of whether these technologies can motivate learners

and facilitate autonomous learning habits outside of the classroom context. A number of scholars and language educators (See, for example, Botero et al., 2019; Kukulska-Hulme, 2018; Underwood, 2014), have seen MALL as a promising tool in shaping independent learners who are self-motivated and are responsible for one's own learning over the course of a lifetime.

Motivation affects the whole learning process. Thus, from the perspective of an application developer, motivation is one important aspect that contributes to the success or failure of a given MALL application. That is because it determines whether a learner will be willing to use and continue using the said application (Zaharias & Poylymenakou, 2009). Understanding motivation through the use of MALL involves exploring the degree to which an application motivates users/learners on two levels, intrinsically and extrinsically (Hadjerrouit, 2012). Learners with intrinsic motivation use an application because it is naturally satisfying, valuable, or interesting. On the other hand, extrinsic motivation involves external rewards gained from using an application, such as passing the test, getting higher grades, or avoiding punishment (Nokelainen, 2006). Intrinsic motivation may derive from the perceived value placed on the application, and the amount of effort a learner is willing to invest in it. The motivation can increase if learners found the application enjoyable and contain useful information that has a high value to them (Hadjerrouit, 2012). They are also driven to learn more if they think they perform tasks on the application well. On the other hand, motivation decreases if too much tension and pressure is felt while using the application. These concepts parallel with the IMI questionnaire items (Carrera et al., 2018) that focus on five subscales, namely value/usefulness, interest/enjoyment, effort/importance, perceived competence, and pressure/tension.

A learner's perceived usefulness of MALL may derive from encouraging and immediate feedback, which in turn, helps increase motivation (Hadjerrouit, 2012). In fact, a study by Tsourounis and Demmans Epp (2016) confirmed that corrective feedback can help increase language learners' motivation in MALL. In their study, learning dashboards are used to provide personalized feedback where language learners can track their knowledge, strengths, and weaknesses. This helps learners in terms of reflection, monitoring, and self-regulation, resulting in increased intrinsic motivation. It is apparent from the literature that learners will be more driven to select and use a MALL application that they

perceive as useful and that satisfies their needs, goals, and learning motivation.

Research Participants

In this study, the research participants included 21 target users and five experts in the field of educational technology or language teaching and learning. The number of users and experts was based on Nielson's (1994) suggestion of five participants in usability testing. Purposive sampling was used in recruiting the research participants. For prospective target users of the TU-GET CBT mobile application, there were three groups of participants: 1) six Thammasat undergraduate students, 2) eight Thammasat postgraduate students, and 3) seven public users who were interested in TU-GET CBT. (See Table 2 for a detailed demographic profile.) In terms of experts, five were recruited to perform heuristics evaluation of the mobile application. They were all lecturers in different universities in Thailand and graduated with a Ph.D. in a field related to second language teaching.

Table 2

Target Users' Demographic Information

Baseline characteristics	Public		Grads		Undergrads		Full sample	
	n	%	*n*	%	*n*	%	*n*	%
Gender								
Male	3	43	4	50	2	33	9	43
Female	4	57	4	50	4	67	12	57
Age								
18-25	2	29	2	25	6	100	10	48
26-35	4	57	6	75	0	0	10	48
36-45	1	14	0	0	0	0	1	5
Occupation								
Student	3	43	3	38	6	100	12	57
Employment - Full time	3	43	4	50	0	0	7	33
Freelancer	1	14	0	0	0	0	1	5
Unemployed	0	0	1	13	0	0	1	5
Mobile phone system								

Android	4	57	2	25	2	33	8	38
iOS	3	43	6	75	4	67	13	62
Prior experience with TU-GET PBT								
Yes	5	71	7	75	5	83	17	76
No	2	29	2	25	1	17	5	24
Prior experience with TU-GET CBT								
Yes	5	71	1	13	3	50	9	24
No	2	29	7	88	3	50	12	76

Research Instruments

The study employed the mixed method approach of three instruments: 1) User's questionnaire 2) Users' focus group interviews, and 3) expert's questionnaire. For prospective target users, two instruments of questionnaire and focus groups were employed for data triangulation purposes. The user's questionnaire contains three main parts: demographic profile, usability, and motivation. The questionnaire items were adapted from Lewis's (2002) Post-Study System Usability Questionnaire (PSSUQ) and from Carrera et al.'s (2018) Intrinsic Motivation Inventory (IMI) questionnaire. The PSSUQ applied the 7-point Likert scale to determine the participants' agreement with each statement from 1 (strongly disagree) to 7 (strongly agree), whereas the IMI questionnaire used the 7-point Likert scale where it ranged from 1 (not true at all) to 7 (very true).

In terms of focus group interviews, the target users were invited to join focus groups of 150 minutes where they were required to perform specific tasks on the TU-GET CBT mobile application such as registering, logging in, and navigating around various tasks on the app. After the participants finished performing each task, they were asked by the researchers/moderators about issues of task performance. All the interviews were audio-recorded and saved as digital files for transcription purposes.

Another research instrument was an experts' questionnaire which experts used to perform heuristic evaluation, a method introduced by Nielson and Molich (1990) to further identify usability of the TU-GET CBT mobile learning application. There are ten 1-5 Likert Scale items and one open-ended question where experts can provide qualitative comments on the usability and issues pertaining to the application.

Data Collection

Prospective research participants were recruited to participate in the study. Once agreed to participate in the study, both experts and target users were informed of the research protocol. They then signed a consent form to confirm their participation in the study.

For target users, moderated usability testing was conducted in three testing sessions among three groups of target participants (6-8 participants per group). Each session lasted approximately 150 minutes. During each testing session, the researchers acted as the moderators who observed, guided participants as they completed the tasks, as well as asked them questions to gain insights into their behavior and perceptions towards usability issues. The sessions were audio-recorded using a digital recorder. Audio files were digitally saved and transcribed for data analysis purposes. At the end of each testing session, the participants filled out the questionnaire consisting of PSSUQ and IMI items.

In terms of experts, they had to perform specified tasks and fill in the expert's questionnaire. These heuristic evaluations contained both quantitative and qualitative data, which will later be used for data triangulation purposes.

Data Analysis

Data analysis in this study involved both quantitative and qualitative data, and two intercoders were involved in all stages of the data analysis. First, data from experts' questionnaires were analyzed. Descriptive statistical analysis was employed for quantitative data whereas thematic coding was used for analyzing qualitative data. Second, the PSSUQ and IMI questionnaire data obtained from the target users were analyzed for its descriptive statistics. Then, the interview data was transcribed and coded for emerging themes. These themes were then triangulated with quantitative data to confirm the results of the study.

Findings

Target Users' Questionnaire Results

PSSUQ Results

Table 3 demonstrates the results of the PSSUQ used to measure the target users' satisfaction towards the usability of the application. The questionnaire used the Likert scale ranging from 1 (strongly agree) to 7 (strongly disagree), with lower scores indicating better user experience than higher scores. The results showed that the participants were moderately satisfied with all subscales of PSSUQ, namely Information Quality (M=4.00, SD=1.33), System Usefulness (M=3.60, SD=1.10), and Interface Quality (M=3.54, SD=1.30).

Table 3

Results of Post-Study System Usability Questionnaire (PSSUQ)

PSSUQ	Public (*n*=7)		Grads (*n*=8)		Undergrads (*n*=6)		Full sample (*n*=21)	
	M	SD	M	SD	M	SD	M	SD
Overall	3.58	1.38	3.37	0.95	4.21	1.31	3.68	1.21
System usefulness	3.48	1.34	3.31	0.79	4.00	1.17	3.60	1.10
Informational quality	3.69	1.39	3.65	1.19	4.67	1.41	4.00	1.33
Interface quality	3.62	1.38	3.04	0.95	3.94	1.57	3.54	1.30

Note. The usability statements ranged from 1 (strongly agree) to 7 (strongly disagree).

IMI Questionnaire Results

Table 4 shows the results of the IMI questionnaire used to explore learner motivation derived from using the application in five subscales. The 7-point Likert scale was used to identify the participants' opinions towards each statement from 1 (not true at all) to 7 (very true). On average, the participants rated the statements in the Value/Usefulness subscale the highest (M=5.40, SD=1.30), suggesting that they found the application useful to them. Ranked second and third were Effort/Importance and

Perceived Competence subscales (M=4.90, SD=1.49 and M=4.89, SD=1.21, respectively). This means that overall, the participants put effort in performing the tasks, and they were satisfied with their performances. On the other hand, they did not enjoy using the application as much, and thus the Interest/Enjoyment subscale was rated slightly lower (M=4.32, SD=1.47). Last but not least, the Pressure/Tension subscale received the lowest mean score (M=3.49, SD=1.42). This yielded a positive outcome since it indicated that the tasks and the application did not cause too much tension or pressure on users.

Table 4

Results of Intrinsic Motivation (IMI) Questionnaire

IMI	Public (*n*=7)		Grads (*n*=8)		Undergrads (*n*=6)		Full sample (*n*=21)	
	M	SD	M	SD	M	SD	M	SD
Interest/Enjoyment	4.14	1.77	4.29	1.45	4.56	1.18	4.32	1.47
Perceived Competence	5.10	1.36	5.04	0.78	4.44	1.50	4.89	1.21
Effort/Importance	4.79	1.48	4.69	1.13	5.33	1.86	4.90	1.49
Pressure/Tension	3.57	0.64	3.42	1.69	3.50	1.93	3.49	1.42
Value/Usefulness	4.88	1.41	5.50	1.19	5.89	1.31	5.40	1.30

Note. The IMI statements ranged from 1 (Not true at all) to 7 (Very true).

It can be seen from these results that, while the PSSUQ questionnaire results showed participant's moderate satisfaction towards the mobile application's usability, IMI questionnaire statements were rated relatively highly, especially on the Value/Usefulness subscale.

Target Users' Focus Group Interview Findings

Accessibility, Navigability, and Aesthetics

In general, three groups of participants managed to navigate through the TU-GET CBT application smoothly. They were able to locate information despite their few attempts to complete specific given tasks. Nonetheless, a few issues were mentioned regarding the accessibility and navigability of the application, i.e., the lack of instructions or information

which facilitates task completion. For example, Graduate Student 8 suggested that "I personally would like to see short instruction videos before I start using the app. For those who have never used the app before or don't really know how to navigate around, they may get confused".

A similar idea was echoed among undergraduate students. When asked if they would like to have an additional help menu with the instructions, Undergraduate Student 4 answered, "I agree with having instructions, but I prefer images". This idea was supported by Undergraduate Student 7 who mentioned, "I agree. The first time we use the application, we don't want to read lengthy instructions and see where to press and what to do". These findings suggest that instructions are necessary in helping first-time users navigate through the application with ease; however, such instructions should be in a form of image or video rather than long, descriptive text to aid understanding.

It is also interesting to note that the younger participants pointed out the importance of aesthetics. Two undergraduate students agreed that the application was 'easy to navigate around', but it needed to be more 'attractive' (Undergraduate Student 2, 3). 'Color' appeared to be one important element in defining the aesthetics of the application. As one Undergraduate Student commented, "I think the colors are quite nice. They are a bit too dark but very nice" (Undergraduate Student 2). As can be seen, the aesthetics of the user interface, such as page layout or color, was another crucial factor in mobile application design.

Value and User Motivation

All participants agreed the application helped familiarize themselves with the TU-GET CBT test and to practice the four skills which appear on the test. For example, Public User 3 mentioned that she liked the reading section because "The application helps me to practice this skill well". Similarly, another participant found that the script in the listening section helps her 'really practice' it (Graduate Student 2). For one of the undergraduate students, Undergraduate Student 2, he personally found grammar and vocabulary sections useful as he expressed, "Among all the parts I've practiced, these parts are truly useful. I really feel like I'm preparing for the TU-GET".

Additionally, the application offers an added value of enjoyment when practicing their grammar and vocabulary skills. This added value is

closely linked with their motivation as it helped them prepare for the test in a more relaxed environment. Graduate Student 7 mentioned, "I really like these sections. It's like I get to relax after practicing so hard with the four skills. When I reached these sections, it was fun". In a similar vein, another participant found the grammar and vocabulary sections less intimidating. As Undergraduate Student 7 stated, "If I'm scared to do the actual test practice [i.e., four skills], I will complete these sections first". Their comments suggested that enjoyment, as well as benefits of using the app (i.e., to pass the test), can add to the value of the application, which in turn, contribute to their increased motivation to continue using the application.

Corrective feedback

Corrective feedback is a very important factor in defining the usability of the application. On digital devices where human interaction is minimal, users are still looking for corrective feedback that will help them improve their performance. Most of the participants expressed the need to receive feedback; however, the timing of the feedback varied between participants. Some participants preferred instant feedback right after each question. When practicing the test, it is preferred to 'do one item at a time and see the answer' to 'understand every question one by one' (Undergraduate Student 3). Additionally, instant feedback can help participants to recall information, as one participant mentioned 'I'll forget what I answered in the first place because I have a rather bad short-term memory' (Graduate Student 3). On the contrary, others preferred finishing the entire section before receiving the feedback because "it's like an actual test. You've got only one chance" (Graduate Student 5).

Whichever the format of corrective feedback the participants prefer, consistency is key. In the current version of the application, there are various test formats, and the corrective feedback shown in each practice is different. Therefore, several participants pointed out the inconsistent feedback format within and across sections. For example, Undergraduate Student 6 found inconsistent format confusing as she said, "This part provides explanations, so I think they simply give examples. But actually, it is the practice test. This is not like the previous section. They're different". The inconsistent feedback format results in the participants preferring some practice sections over others. Public User 7, for example,

preferred the writing section because it provides an 'answer key'. On the contrary, he disliked the listening section because "The answer key isn't well designed. Yes, they tell you whether the answer is correct or not, but the question and the answer should appear side by side". All in all, the participants seemed to prefer corrective feedback that is consistent across sections because "it's easy to understand and saves time" (Undergraduate Student 2).

Another issue raised about corrective feedback was the need for personalized feedback, especially for speaking and writing skills. However, such personalized feedback might not be necessarily effective when offered through the mobile application, and therefore, some participants, especially the graduate students, were willing to pay for additional service. Graduate Student 4 said, "If I really need to use the TU-GET scores, obtaining feedback from the test writers [i.e., LITU] will be an advantage. We can be certain that they write the test questions and own the test, so it's an attractive deal". Similarly, Graduate Student 1 explained, "We don't have to hire a teacher or anyone. Also, it is checked by the institute who writes the test items, so we'll know the guidelines of how to write".

On the other hand, it should be noted that many public users and undergraduate students felt reluctant to pay for personalized feedback because they were unsure if it was worthwhile. Almost all of the public users mentioned that they would not pay for the feedback service due to other available alternatives including 'taking a course' (Public User 1), 'self-practicing' (Public User 2), 'buying a book' (Public User 4), and 'investing in a more difficult test like IELTS' (Public User 5). Moreover, paid feedback service may only be suitable for users who are already proficient users of English but need advice to improve their test scores. For some users, having fundamental language skills is necessary. As one participant stated, "If I know nothing, I won't send my writing for the service in the first place" (Undergraduate Student 1).

These findings showed that the participants had varying needs, goals, and preferences when it comes to the format of feedback, whether it be paid or free. Despite their differences, the rich discussion on this topic suggested that all of them considered corrective feedback necessary for their English language skill development and test preparation.

Experts' Questionnaire Results

To obtain the data from different perspectives, the usability heuristic evaluation was conducted with five experts in the field of language teaching and educational technology. The evaluation form contained 10 statements involving 10 general principles for user interface design. The rating scales ranged from 0 (No usability problem) to 4 (Usability catastrophe). Table 5 shows the results of the heuristic evaluation.

Table 5

Results of Experts' Heuristic Evaluation Arranged by the Highest Means

Item#	Usability Issue	Ex1	Ex2	Ex3	Ex4	Ex5	M	SD
6	Recognition rather than recall	1	1	1	1	4	1.60	1.34
4	Consistency and standards	2	1	1	1	2	1.40	0.55
5	Error prevention	2	0	1	2	1	1.20	0.84
9	Help users recognize, diagnose and recover from errors	1	1	0	1	3	1.20	1.10
3	User control and freedom	1	0	0	1	3	1.00	1.22
10	Help and documentation	1	0	0	1	3	1.00	1.22
1	Visibility of system status	0	0	0	2	2	0.80	1.10
2	Match between system and real world	1	0	0	1	2	0.80	0.84
7	Flexibility and efficiency of use	0	0	0	1	0	0.20	0.45
8	Aesthetic and minimalist design	0	0	0	1	0	0.20	0.45

Note. The rating scales ranged from 0 (I don't agree that this is a usability problem at all), 1 (Cosmetic problem only: need not be fixed unless extra time is available on project), 2 (Minor usability problem: fixing this should be given low priority), 3 (Major usability problem: important to fix, so should be given high priority), and 4 (Usability catastrophe: imperative to fix this in the next version).

It was found that the most urgent usability issue to fix was recognition rather than recall (M=1.6, SD=1.34), which means that the application should be improved to allow users to recall information instead of having to remember it, by making actions and options visible or easily retrievable whenever appropriate. The consistency and standard of the application also showed some usability problems (M=1.40, SD=0.55). This suggests that the look and function of the application might not be as consistent throughout or might not follow mobile platform conventions. The experts were also concerned about error prevention (i.e., preventing

problems from occurring in the first place (M=1.20, SD=0.84), as well as helping users recognize, diagnose and recover from errors by showing error messages that precisely indicate a problem and suggest a constructive solution (M=1.20, SD=1.10). Ranked fifth and sixth were the issues of user control and freedom (i.e., the application that is easier for users to fix mistakes, undo, and redo) (M=1.00, SD=1.22), and help and documentation that are easy to search and focus on the user's task, list concrete steps to be carried out, and are concise) (M=1.00, SD=1.22). The other four usability items appeared to be nonexistent, or they were cosmetic problems that can be fixed or improved if time is available.

Experts' Additional Comments

Overall, the application is considered 'practical', 'useful', and 'informative'. As one expert wrote, "UX/UI of this application is quite good, suitable, and very practical for learners at all ages." and "Overall, this is a very informative application especially for one who plans to take a TU-GET exam" (Expert 1). This is also agreed with by another participant who stated, "The UX UI of this application is quite good, suitable, and very practical for learners at all ages" (Expert 4). Despite the benefits of the application for language learners, the experts pointed out some technical usability issues including recognition rather than recall. For example, Expert 5 mentioned that "The application should remember my profile and the activities I have done and the answers I have selected." and "'Forgotten your password' should be provided. I often forget my password so I have to make new registrations every time I use the application."

Another key issue mentioned by several experts was the inconsistency of the application, from the design to the test items and the format of the feedback. They wrote: "It should be consistent in every item when the answer is incorrect." (Expert 4); "The issue of consistency in each skill is not that parallel, including the feedback system." (Expert 3); and "Very comprehensive and balanced but a finish button should be inserted for the consistent design platform" (Expert 4). These additional comments from the experts suggested that while the application has various benefits to users/learners, there are some usability issues, especially the consistency, that needs to be fixed.

Discussion and Implications

With regard to the usability of the TU-GET application, the results of PSSUQ suggested that the target users were moderately satisfied with the application. However, when compared to preliminary norms published by Lewis (2002) as shown in Table 6, the means from the present study were much higher. This means that the usability of the application should be improved in all three subscales, including system usefulness, information quality, and interface quality. Moreover, the application performed worst for item 7, which is "The system gave error messages that clearly told me how to fix problems". This implies that more effort should be put into designing a good error message. When comparing across target users' groups, it was found that the undergraduate students had the most problems locating information they needed to complete the tasks and to recover from mistakes while using the application.

Table 6

Means Comparison with Lewis (2002)

PSSUQ	Full sample (*n*=21)	Lewis (2002)
	M	M
Overall	3.68	2.82
System usefulness	3.60	2.80
Informational quality	4.00	3.02
Interface quality	3.54	2.49

The qualitative findings further revealed that the participants were able to navigate through the application without many problems. This could be due to the fact that the application's design and behavior follows standard conventions that the users are already familiar with (Nielson, 1994). Even though they encountered some technical difficulties, they were able to manage it using their prior experience with similar existing applications. Nonetheless, because the content of the application is specific to the TU-GET CBT test formats, some participants expressed the need for tutorials and help documentation that can facilitate accessibility and navigation.

Despite some usability issues, the results of IMI questionnaires were relatively high, especially in the Value/Usefulness subscale, suggesting a positive correlation between perceived usefulness and motivation. This was supported by the interview findings, with many participants mentioning the usefulness of the application for TU-GET CBT preparation as well as English language skill practices. Truly, when a learner perceives that a technological tool is beneficial for their language learning processes and is enjoyable to some extent, they tend to be more motivated to learn with that tool (Hadjerrouit, 2012). Another important factor contributing to motivation in this study was enjoyment. Even though the items in Interest/Enjoyment subscale were rated lower than others due to the application's serious content, many participants expressed that they appreciated the grammar and vocabulary sections because they were enjoyable and relaxing. These findings shed light on the importance of balancing between heavy and light content to drive and maintain users' interest and motivation via MALL applications.

Beside the usefulness and enjoyment, corrective feedback was the topic widely discussed by the target users. They expressed the need for a more consistent feedback system across different sections of the application. Admittedly, the current version of the application contains several types of practice test items in varying formats, which were potentially confusing. In designing websites and mobile applications, consistency is crucial because it allows users to get things done as quickly and easily as possible. When consistency is achieved, users do not have to constantly learn new things, and as a result, it reduces their cognitive load and leads to a better user experience (Schlatter & Levinson, 2013).

In addition to consistent feedback format, the issue regarding personalized feedback was also raised across findings. Due to the limitation of mobile platforms, some participants proposed the idea of paying for additional services to obtain personalized feedback for their writing and speaking performances. Such paid personalized feedback, however, might not be of value to all participants. While some were willing to pursue this option, others remained hesitant, uncertain of its benefit. Despite their differing needs, the researchers found that corrective feedback is tightly intertwined with users' motivation. That is, if providing feedback, especially personalized one, language learners' or application users' motivation level can be increased (Tsourounis & Demmans Epp,

2016). The option of paid personalized feedback, therefore, is something to be considered for the next phase of the application development.

The findings from the experts' heuristic evaluation generally echoed what was found among the target users. For one thing, the experts considered the application a practical and useful tool for helping learners develop English skills and prepare them for the test. As for the drawbacks, several of them mentioned the issue of consistency in both the design and feedback system. While the results of the heuristic evaluation did not indicate any major usability problems that require immediate attention, because the primary objective of heuristic evaluation is problem discovery rather than users' general opinions (Nielson, 1994), the scores from individual raters should also be taken into consideration. Because some of the experts rated the usability item numbers 3-6, 9 and 10 as major or catastrophic problems, these are the issues that the researchers recommend fixing in the future.

Several implications can be made from the rich findings of the present study. First of all, while the application already provides tangible benefits for its users, the usability of the application can be further improved for better user experience and increased satisfaction. Attention should be paid to all aspects of the usability including system quality, information quality, and interface quality. The application should also perform better at preventing errors, offering help, and providing documentation – preferably in image and video formats rather than text – on how to complete tasks and overcome problems. The second improvement can be made with the application's internal consistency. This can be achieved by having consistent visual usability tools, such as color, layout, typeface, and controls, across all sections of the application (Schlatter & Levinson, 2013). More consistent test and feedback formats are also highly recommended. Moreover, the developer may consider adding paid services that are tailored to learners' individual needs and language proficiency levels. A solid plan of action, however, is crucial before embarking on the project. Finally, the present study confirmed that in exploring the usability of a MALL application, one should take into account not only the technical aspects (i.e., accessibility, navigability, aesthetics), but also pedagogical aspects (i.e., usefulness, enjoyment, motivation, and feedback) of the application. These lenses are tightly intertwined, and they can potentially contribute to the users' satisfaction and a successful application.

Conclusions and Future Research

The purpose of this study was to investigate the usability of the TU-GET CBT application and its contribution to learner motivation. The findings that have been reported in this paper demonstrate that the use of a new technology contain both benefits and drawbacks. User experience towards a MALL application can never be straightforward, and thus, a usability study that combines several research instruments and tested with different client sections is necessary. While the findings suggested positive user experience because of the application's practicality and usefulness, some usability issues, both the technical and pedagogical ones, need to be addressed. The researchers believed that the findings of the present study were not only useful in informing the future development of the TU-GET CBT application, but they also revealed a complex interplay between technical and pedagogical usability for a MALL development. While the process of designing standard applications focuses mainly on achieving technical usability to facilitate ease-of-use, a MALL application must also take into account pedagogical aspects that foster and motivate learning. Future research, therefore, can explore and refine the usability criteria in these two lenses in order to develop more comprehensive instruments for assessing students' perceptions of and satisfaction towards the application. In addition, future work can be undertaken with larger user groups to ensure more reliability and validity.

Acknowledgements

We would like to express our sincere gratitude to the Language Institute, Thammasat University for providing us with this research opportunity. This research was funded by the Language Institute, Thammasat University research grant.

About the Authors

Sichon Koowuttayakorn: A faculty member and a director of the M.A. in Career English for International Communication at the Language Institute, Thammasat University. Her research interests include technology-

enhanced English language teaching and learning and multimodal discourse analysis.

Pimsiri Taylor: A faculty member at the Language Institute, Thammasat University. Apart from teaching and supervising postgraduate research projects, Dr. Taylor is also involved in teacher training programs. Her research interests include English-medium instruction, English as a lingua franca, English for specific purposes, internationalization of education, and interculturality.

References

Botero, G. G., Questier, F. & Zhu, C. (2019). Self-directed language learning in a mobile-assisted, out-of-class context: Do students walk the talk? *Computer Assisted Language Learning, 32*(1-2), 71-97. DOI: 10.1080/09588221.2018.1485707

Carrera, C. C., Perez, J. L. S., & Cantero, J. T. (2018). Teaching with AR as a tool for relief visualization: Usability and motivation study. *International Research in Geographical and Environmental Education, 27*(1). https://doi.org/10.1080/10382046.2017.1285135

Elaish, M. M., Shuib, L., Ghani, N. A., Yadegaridehkordi, E. & Alaa, M. (2017). Mobile learning for English language acquisition: Taxonomy, challenges, and recommendations. *IEEE Access, 5*, 19033-19047. DOI: 10.1109/ACCESS.2017.2749541

Georgsson, M. & Staggers, N. (2015). Quantifying usability: An evaluation of a diabetes mHealth system on effectiveness, efficiency, and satisfaction metrics with associated user characteristics. *Journal of the American Medical Informatics Association, 23*(1), 5-11. DOI: 10.1093/jamia/ocv099

Godwin-Jones, R. (2017). Smartphones and language learning. *Language, Learning and Technology, 21*(2), 3-17. Retrieved from http://llt.msu.edu/issues/june2017/emerging.pdf

Hadjerrouit, S. (2012). Investigating Technical and Pedagogical Usability Issues of Collaborative Learning with Wikis. *Informatics in Education, 11*(1), 45–64.

Hoehle, H. & Venkatesh, V. (2015). Mobile application usability: Conceptualization and instrument development. *MIS Quarterly, 39*(2), 435-472.

Kukulska-Hulme, A. (2009). Will mobile learning change language learning? *ReCALL, 21*(2), 157–165. DOI: 10.1017/S0958344009000202

Kukulska-Hulme, A. (2018). Mobile-assisted language learning [Revised and updated version]. In A. C. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Wiley. https://doi.org/10.1002/9781405198431.wbeal0768

Lestary, S. (2020). Perceptions and experiences of mobile-assisted language learning for IELTS preparation: A case study of Indonesian learners. *International Journal of Information and Education Technology, 10*(1), 67-73. DOI: 10.18178/ijiet.2020.10.1.1341

Lewis, J. R. (2002). Psychometric evaluation of the PSSUQ using data from five years of usability studies. *International Journal of Human-Computer Interaction, 14*(3-4), 463-488. doi: 10.1080/10447318.2002.9669130

Nielson, J. (1994, April 24). *10 usability heuristics for user interface design.* Nielsen Norman Group. https://www.nngroup.com/articles/ten-usability-heuristics/

Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces. *Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 249-256. https://doi.org/10.1145/97243.97281

Nokelainen, P. (2006). An empirical assessment of pedagogical usability criteria for digital learning material with elementary school students. *Journal of Educational Technology & Society, 9*(2), 178-197. https://www.jstor.org/stable/10.2307/jeductechsoci.9.2.178

Saritepeci, M., Duran, A. & Ermis, U. F. (2019). A new trend in preparing for foreign language exam (YDS) in Turkey: Case of WhatsApp in mobile learning. *Education and Information Technologies, 24*, 2677-2699. https://doi.org/10.1007/s10639-019-09893-4

Schlatter, T. & Levison, D. (2013). *Visual usability: Principles and practices for designing digital applications*. Morgan Kaufmann.

Tsourounis, S. & Demmans Epp, C. (2016). Learning dashboards and gamification in MALL: Design guidelines in practice. In A. P. Palalas, M. Allies (Eds), *The international handbook of mobile-assisted*

language learning (pp. 370–398). China Central Radio & TV University Press Co., Ltd.

Underwood, J. (2014). Using iPads to help teens design their own activities. In S. Jager, L. Bradley, E. J. Meima, & S. Thouësny (Eds.), *CALL Design: Principles and Practice. Proceedings of the 2014 EUROCALL Conference, Groningen, The Netherlands* (pp. 385-390). Dublin: Research-publishing.net. DOI:10.14705/rpnet.2014.000250

Zaharias, P. & Poylymenakou, A. (2009). Developing a usability evaluation method for e-learning applications: Beyond functional usability. *International Journal of Human–Computer Interaction, 25*(1), 75-98. doi: 10.1080/10447310802546716

Appendix A

Screenshots of the TU-GET CBT Application