Expert Review on Mobile Augmented Reality Applications for Language Learning

Nur Asylah Suwadi[1], Nazatul Aini Abd Majid[2]
Meng Chun Lam[3]
Mixed Reality and Pervasive Computing Lab, Center for Artificial Intelligence Technology, Faculty of Information Science & Technology, Universiti Kebangsaan Malaysia 43600 Bangi, Malaysia

Junaini Kasdan[5]
Institut Alam & Tamadun Melayu (ATMA), Universiti Kebangsaan Malaysia, 43600 Bangi

Nor Hashimah Jalaluddin[4],Aznur Aisyah Abdullah[6]
Center for Research in Language and Linguistics
Faculty of Social Sciences and Humanities
Universiti Kebangsaan Malaysia, 43600 Bangi, Malaysia

Afifuddin Husairi Hussain[7], Azlan Ahmad[8]
Daing Zairi Ma'arof[9]
Pusat Citra Universiti, Universiti Kebangsaan Malaysia
43600 Bangi, Malaysia

Abstract—**Many mobile applications that can increase user engagement and promote self-learning have been developed to date. Nevertheless, mobile applications specific to Malay language learning for non-native speakers with relevant materials are still lacking. Moreover, expert reviews are needed to identify usability issues and check whether such applications can meet the learning goal with relevant materials features. This study developed an augmented reality (AR)-based mobile application called *RakanBM* for learning the Malay language (i.e. the language officially spoken in Malaysia), and then performed an expert review on the application contents, text presentations, learning outcomes, assessments, effectiveness, efficiency, and satisfaction. The expert review was conducted by a panel of six experts from two specific fields, namely the Malay language and Human-Computer Interaction (HCI), using methods such as cognitive walkthrough (CW), semi-structured interviews, think-aloud protocols, and survey. The results from CW, semi-structured interviews, think-aloud protocols shows that enhancement was needed on user interface and user experience in term of aesthetic and interactivity. The survey results were classified into two levels: high (mean > 4.0) and satisfied (mean > 3.5). Application factors that were recorded as satisfied were the application contents, text presentations, and satisfaction, while the factors recorded as high were the learning outcomes, assessments, effectiveness, and efficiency. The comments or suggestions for improvement were mainly around the contents of the application. Nevertheless, the application received good comments on its usefulness and the topics covered, which were suitable and best for non-native speakers. The findings of this study can guide developers and researchers in the development of future applications that can support language learning for non-native speakers in particular.**

Keywords—*AR; expert review; HCI; language learning; mobile application; self-learning*

I. INTRODUCTION

Mobile phone device is one of the most popular types of information and communication technology (ICT) as digital learning has been an evolution in the present days[1]. The use of mobile applications for language learning has shown promising outcomes, especially for language improvement, learner interaction [2, 3], and learner motivation [4]. The use of a mobile phone in and out of the classroom can help students learn more effectively and create a conducive environment for effective teaching and learning [5]. Similarly, flexible learning would be less costly, adaptable, require little instrument, and is easy to use. These advantages allow the implementation of mobile learning in almost all learning environments. Furthermore, such technological advancements can make learning more efficient and enjoyable for students [6].Previously, several language learning using the mobile application has been invented and conveyed positive feedback from learners, for instance, Mayo language [7], Spanish language [8], Malay language [6] and etc.

In the context of the Malay language, not only the written materials, in fact, the audio and video in electronic forms are far from adequate. It is challenging to learn Malay because it has some elements of modification from other languages such as English and Arabic [9] and was influenced by Indonesian and English pronunciation as well [10]. In another study, [6] introduced the Malay Language Mobile Learning (M-Lang) system using the near field communication (NFC) technology to utilise mobile learning in all learning environments as this technological improvement can make learning more effective and fun for students. However, the system was only in the early stages and excluded AR technologies. The study realised that it was challenging to study and teach foreign languages, which the fundamental principle is to ensure that learners are introduced to the desired foreign language accordingly. At the same time, most of the new applications are limited to elementary-level students[11], and there is also a lack of attractive functionality [12].

Previous studies have highlighted that the Malay language learning deficiency, especially in the mobile application, has become an obstacle for learners to learn the language. Therefore, a better mobile application for Malay language learning are beneficial in ensuring learners, especially foreign students, have a basic knowledge of the language, which could

help them communicate with the locals. This study aimed to validate and ensure that the low-fidelity *RakanBM* prototype was relevant and suitable to support Malay language mobile learning. An expert review was conducted to identify usability issues and check whether the prototype can meet the learning goal with relevant materials features. The remainder of this paper is organised in the following manner: Section II describes the related works and expert reviews methods; Section III presents the evaluation procedure for the developed mobile application; Section IV provides a detailed description of the results and discusses the main features of the application that bring positive results, especially for learning outcomes, assessments, effectiveness, and efficiency; and Section V concludes the findings.

II. RELATED WORK

The related works highlighted previous studies, including mobile application, language learning, and review method to identify the potential of RakanBM application towards users.

A. Mobile Applications on Languages Learning

Many approaches have been investigated on the use of mobile applications for language learning. For instance, Bojórquez, Villegas [7] aimed to identify the perceptions of undergraduates in the task of learning the Mayo language (i.e. language spoken in north-western Mexico) through a specific mobile AR (MAR) program used by a group of students.The game named Loter´ıa Mayo use was based on images and audio using MAR through a card game.A survey was collected from students with items concerning the use and technology acceptance and culturaldimensions.Resultantly, although the MAR system could be easily used, there was still room for improvement, especially on the students' ability to continue their learning experience both within and outside the classroom that should be enhanced and made more effective. Similarly, [8] developed a mobile application having game-like activities with a grammar-focused mobile application to learn Spanish Languages. The study aims to determine the motivational and affective engagement of students towards mobile applications.Results revealed that students' inspiration and affective involvement were growing. The mobile apps allowed learners to work collaboratively and learn about conjugating verbs from each other. In another study, [3] found that *Busuu,* one of the most popular apps for language learning in the market, boosted the use of mobile application technology to a higher level. The app is part of the *Busuu* network for language learning and offers 12 different languages organised in levels. The majority of users evaluated were beginners and learning for personal reasons. The *Busuu* app helped users to better understand the language studied, especially in terms of vocabulary.

Other studies have investigated the learning impact of such applications. For instance, (Ng, Bakri, & Rahman, 2016) investigated the mobile courseware application's impact on children with special needs. The study found that the application was suitable for children with learning difficulties and had positively impacted their learning performance compared to the traditional teaching method. The application, which consists of basic Malay language syllables, was developed to assist educators and parents in teaching children

with special needs to improve their learning progress from time to time.

B. Mobile Augmented Reality Applications

The ability of AR technology has been investigated by researchers because it has been proven to boost learning performance [13-16]. The AR technology enables a view of reality to be modified by the addition of digital information like three-dimensional (3D) model, two-dimensional (2D) image, video, and audio. The modification improves a person's perception of reality. This technology consists of four main components: (1) marker which is the target object in the reality; (2) a camera of a mobile phone to capture the target object, which is the marker; (3) a mobile phone to store and process information, which is the captured image or the target object (marker); and (4) digital content that will be displayed on the mobile phone screen, where the tracked image by the phone camera is the marker. Fig. 1 shows an example of an AR application where the marker is an image of fried noodles. When the camera phone has tracked the marker, digital information corresponding to the name of the food in Malay which is 'Mi goreng' will be displayed on the mobile screen.
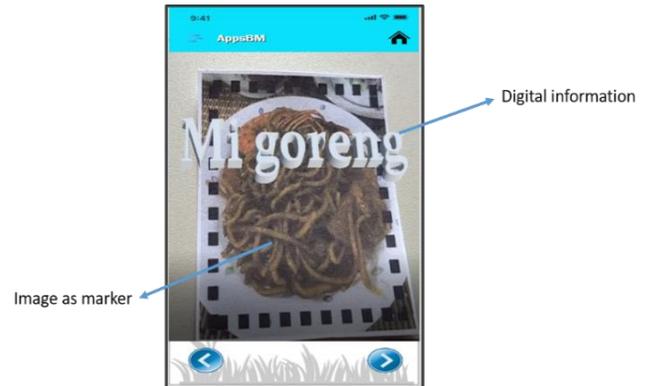

Fig. 1.   Addition of Word 'Mi goreng' with Audio on Top of the Marker via AR.

AR provides simulated access to learning methods and has proved to be an effective 3D visualisation technology in educational areas [17, 18]. Ali and Azmi [12] investigated the use of AR for Malay language learning by developing a MAR Malay language application for international students to learn the language. Although the study found that students were satisfied with the developed application, the features and contents could be further extended. Moreover, the application was only in the early stages and not yet published publicly [12]. Based on the review of relevant studies, it can be concluded that studies on Malay language learning utilising AR are still limited, and further research is needed to deepen the knowledge, especially in supporting language learning. Therefore, this study introduced a Malay language mobile application known as *RakanBM* with comprehensive content embedded with AR technology to fulfill the current limitation of the domain.

C. Expert Review Method

In this section, the expert review method is further described. An expert review is a form of testing that can be used to ensure that students are not exposed to any

inappropriate procedure that could affect their learning [19]. The selected expert evaluation method was appropriate because only a small sample of feedback was required to evaluate and support the application development. As such, the value of this expert review method has been recognised in software quality assessment [20]. This method can be classified into four categories: CW, semi-structured interviews, thinks-aloud protocols, and surveys.

1) CW: Walkthroughs take place where the activities as carried out in the user testing process are conducted by experts. Cognitive and pluralistic walkthroughs are the two primary walkthrough forms. CW seeks to simulate the problem-solving skill of users in each separate task, and it also relies on assessing how learnable the method is [21]. In addition, CW determines whether the end-users background knowledge and the technological cues embedded in the computer system interface are adequate to help the end-users accomplish a task. Furthermore, CW focuses on cognitive behaviour like recognising icons and behavioural or physical actions like mouse clicks necessary to complete tasks. This approach is extremely helpful when designing technologies for inexperienced users because it can recognise usability challenges that can impede task completion [22, 23]. A CW technique is a test that uses a sequence of actions or steps on a user interface to achieve a specific goal, and it is often used to test usability based on actions and ways of thinking. Besides seeking to identify problems and measure the level of complexity of each task scenario, another goal is to get an idea of the extent to which the interface supports a positive level of user experience or identify any areas that may cause difficulties [24]. CW is also a platform to detect misunderstandings between the user and the designer about a certain task involved [25]. In this study, the expert review using CW employed experts from the Malay language and HCI fields of expertise.

2) Semi-structured interviews: Typically, semi-structured interviews are in the form of a face-to-face interview on a relevant platform (e.g. digital) that focuses on usability-related topics such as the application's functionality, navigability, and ease of use. Such engagements are helpful to collect knowledge about a person's behaviours, values, practices, and experiences [26]. In this study, semi-structured interviews were conducted with six experts on the application prototype. The focus was on the content and interface of the materials and potential improvements based on their recommendations. The interviews determined that one of the main goals of the interactive experience should be to foster user experience while using the application [27].

3) Think-aloud protocols: Many studies have used the think-aloud protocol to evaluate the websites and mobile applications concerning their usability [28]. The method is a usability assessment that gathers information on functionality, navigability, and ease of use while end-users interact with the application. As part of a heuristic evaluation, users (i.e. end-users and/or experts) communicate their opinions and concerns in real-time as they perform tasks, enabling observers to see the

cognitive process involved. During the think-aloud assessment, the researcher usually uses software to record the users' responses, e.g. verbal comments and physical responses such as eye movements. This approach examines real issues faced by end-users as well as the factors causing those problems, offering insight into usability deficiencies [22, 29].

4) Survey: A questionnaire used in survey is a set of questions or items in form of writing. It is a method designed specifically to collect information for analytical purposes that can answer research questions. Researchers use several standardised and validated questionnaires to quantitatively evaluate the questionnaire [30]. Typically, a survey involves participants scoring items in questionnaires using a predetermined scale. This study obtained feedback from expert evaluators (respondents) on the mobile application design using a set of questions, which helped evaluate the low-fidelity application in terms of contents, text presentation, learning outcome, assessment, effectiveness, efficiency, and satisfaction. In other words, the questionnaire measured the level of user experience on the prototype. The questionnaire technique consisted of open and closed structured questions. In principle, closed-ended questionnaire questions offer a set of answer choices and are used to obtain uniform answers. On the contrary, open-ended questions provide an opportunity for respondents to give opinions or suggestions, and they complement closed-ended questions [31].

III. METHODOLOGY

In this section, the design of the application and evaluation procedure for the developed mobile application is presented through Phase I, Phase II and Phase III. There are seven main chapters in the RakanBM application: vowel and consonant systems, greetings, numbers and currency, time, day, direction, and food. Furthermore, some elements (e.g. videos, listening exercises, quizzes, vocabulary, and practices using AR technology) diversify the approach and attract users to learn the Malay language. Interactivity is also offered by providing users with scores for each exercise (i.e. a set of tasks) completed. At the same time, the application also comes with feedback (e.g. "excellent", "very good", "keep trying", and "try again") interface, displayed according to the users' achievements after completing a task as a medium to increase user motivation and continue using the mobile learning application.

A. Phase I: The Development of RakanBM

The *RakanBM* prototype was developed in collaboration with experts in the Malay language and then evaluated using CW, semi-structured interviews, think-aloud protocols, and surveys (i.e. open and closed structure questions) methods. Fig. 2 illustrates the application developed using Adobe Xd software based on digital prototyping. The hardware required included a notebook and a smartphone. In addition, the related software and programming language for the application development were such as 3D Viewer - for AR model, Adobe Photoshop - for editing marker, Vuforia - for generating marker, Filmora - for video and sound editing, Unity - as the engine, and C# - for coding.
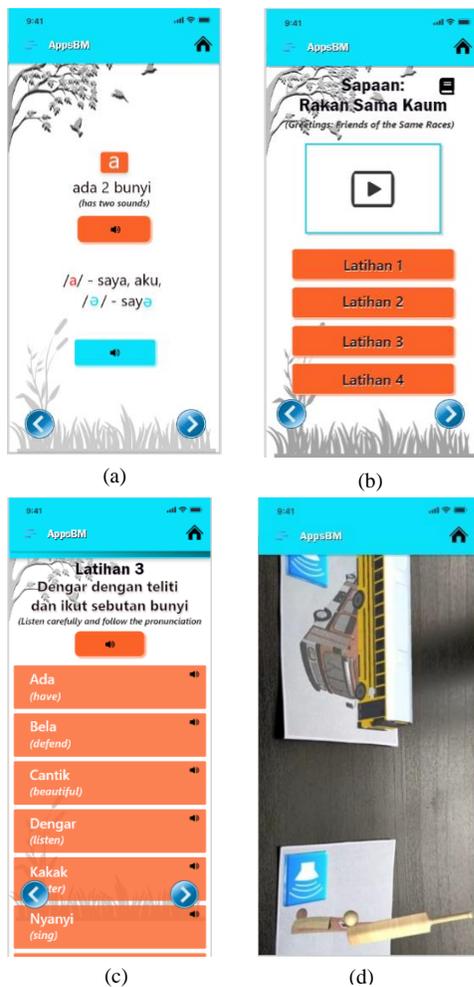
Fig. 2. Screenshots of the RakanBM Prototype. (a) Text Content. (b) Video Screen. (c) Audio Screen. (d) AR Content.

Fig. 2 shows four screenshots of the RakanBM application: Fig. 2(a) illustrates the text content used to present the language learning basics on vowels and consonants. Each vowel and consonant is supported with audio features for users to learn reading and at the same time to listen to the correct pronunciation of the particular Malay words; Fig. 2(b) illustrates the video screen concerning the daily routine conversation. From this approach, users can apply the language in their surroundings. Fig. 2(c) illustrates the audio screen for exercise. Users can learn the pronunciation of the listed words by clicking the sound button. Users also can repeat the exercise several times to ensure mastery of words, and Fig. 2(d) presents the advanced feature of the application, the AR-based content, designed to attract users' attention towards learning via the latest technology. The AR screen shows a 3D model used in several topics in the applications, such as number, currency, and time. The markers shown are images of a bus and that of a bat. Apart from a 3D model (i.e. the digital content), an audio icon appeared on the same screen when the phone camera tracked the markers. Therefore, users can relate the static images with better visualisation of the object with accurate pronunciation repeatedly.

B. Phase II: Evaluation Procedure for CW, Semi-Structured Interviews and Think-Aloud Protocols

One of the advantages of an expert review is that it requires only a small number of experts to identify the flaw of the mobile application (Beecham et al., 2005). This study performed an expert evaluation to ensure the relevance of the RakanBM application prototype. The expert evaluation involved the perspective of contents and interface of the prototype. Table I provides the background of the six experts recruited for the review. The selected experts were from two fields, namely the HCI and the Malay language. The demographics of the experts are as follows:

- Malay language: The experts consisted of educators who specialise in the field of the Malay language and have more than five years of work experience in related fields.

- HCI: The experts consisted of individuals with more than five years of experience in the mobile application development field.

TABLE I. DEMOGRAPHIC OF EXPERTS (I.E. EVALUATORS)

Expert	Specialisation	Year of Experience	Gender
A	Malay language expert	25	Female
B	Malay language expert	11	Female
C	Malay language expert	31	Female
D	Malay language expert	28	Male
E	HCI expert	20	Female
F	HCI expert	20	Female

The Fig. 3 illustrates the evaluation procedures involved. This study started the testing by explaining the evaluation objectives to the experts. The experts were then introduced to the mobile application and explained the concept of the developed application prototype. Next, the RakanBM application workflow was described in detail to the experts for clarity on the activity workflow. Afterward, activities such as prototype testing using CW techniques, semi-structured interviews, and think-aloud protocols were also carried out.
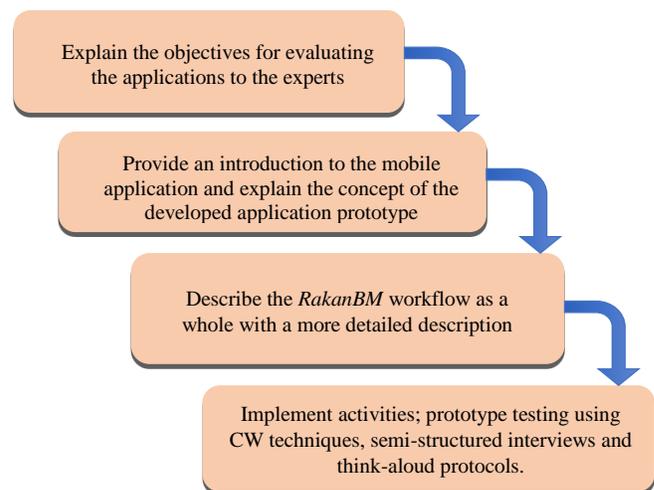

Fig. 3. Evaluation Procedure for the RakanBM Prototype using CW Techniques, Semi-Structured Interviews and Think-aloud Protocols.

For the CW, the experts evaluated the developed application for two weeks. Guidance was given to the experts on the needed steps to undergo a CW test on the prototype. During the testing process, the experts provided views and comments. Subsequently, semi-structured interviews were conducted to get feedback on the prototype from the experts. During the semi-structured interview, the experts gave their opinion on the application interface and contents as a whole. The experts' user experience was an extra knowledge that should be considered [27] to improve the application features and make it suitable for non-native speakers. The problems and requirements stated by the experts were recorded and then summarised in Table IV, which clearly shows that most of the feedback on the RakanBM prototype was related to its interface. Under the think-aloud protocol, the experts evaluated the prototype alongside researchers. During this session, the experts asked the researchers any questions or portrayed their feelings verbally about the current prototype. The session was screen recorded to allow researchers to refer again to the responses and behaviours of the experts. At the same time, the usability of the prototype was evaluated using the think-aloud protocol, whereby the experts directly gave their comments and suggestions to meet the aim of the developed application.

C. Phase III: Evaluation Procedure for Survey

Questionnaires used for the survey in this research were adapted from Goal Question Metric (GQM) considering effectiveness, efficiency, and satisfaction as a construct[32-34]. Besides, construct named application content, text presentation, learning outcome, and assessment were built based on the studies requirement discussed by the Malay language teachers. This questionnaire was adapted into current research to describe the research objectives with the required data. It will be conducted at the end of the study. The developed questionnaire was given to the experts to identify their background and obtain their expectations on the application. Fig. 4 illustrates the evaluation procedures involved in survey techniques. In addition, the questionnaire technique was used to measure the level of user experience on the prototype.
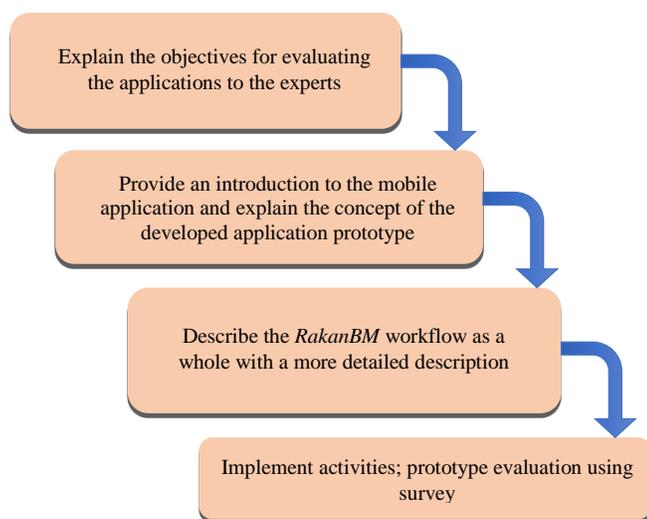

Fig. 4.   Evaluation Procedure for the RakanBM Prototype using Survey.

1) Produce questionnaire questions to users: The questionnaire was developed in Malay language. The experthas no problem understanding the real meaning conveyed by the questionnaire. The questionnaire for this study consists of seven constructs, namely application content, text presentation, learning outcome, assessment, effectiveness, efficiency, and satisfaction. Application content was a question regarding user feedback on the content of the application. Text presentation was a question that required users to give their feedback towards the user interface available on the apps. The learning outcome was a question about the outcome that users may gain from the particular apps. The assessment was about scoring and interactive feedback provided by the apps. Last but not least, three constructs adapting GQM to gain information of effectiveness, efficiency, and satisfaction from a user by using *RakanBM* apps. All the constructs used a Likert five-point scale.

2) Questionnaire reliability analysis: The questionnaire was tested with Cronbach's alpha values to test the reliability of the questionnaire. Data should be unidimensional and consistent for purposes of internal consistency through Cronbach's alpha analysis [35]. Internal consistency can be measured in a number of ways. The most frequently used statistic was Cronbach's alpha [36]. In this study, a questionnaire was answered by six experts. The application contents constructs consisted of 10 items (α = .89), text presentations constructs consisted of 5 items (α = .88), learning outcome constructs consisted of 5 items (α = .86), assessment constructs consisted of 5 items (α = .78), effectiveness constructs consisted of 5 items (α = .87), efficiency constructs consisted of 5 items (α = .89) and satisfaction constructs consisted of 5 items (α = .97). The Cronbach's alpha values shows an excellent ($\alpha \geq 0.9$), good ($0.9 > \alpha \geq 0.8$) and accepted ($0.8 > \alpha \geq 0.7$) internal consistency according to Tavakol and Dennick [37] benchmark.

IV. RESULTS AND DISCUSSION

In this section, the results of the low-fidelity testing and evaluation process are further elaborated and discussed, especially concerning the field of language learning with apps. Furthermore, the gathered results provided substantial data in response to the experts' review on the *RakanBM* application.

A. CW, Semi-Structured Interviews and Think-Aloud Protocol Results

The collected data were analysed and the results are as shown in Table II, specifically in terms of comments and suggestions. During this process, things that were noted were the feasibility of the experts on the given task, the accuracy of the process flow, the information developed, and the solution of the task scenario provided. Experts were required to provide feedback on whether they could perform the task or not. The experts were also free to comment on new ideas and suggestions for improvement. Open-ended questions were posed to get more complex feedbacks. Each question was analysed using a coding scheme by category adopted from Moustakas [38].

TABLE II.   PROTOTYPE EVALUATION ANALYSIS RESULTS BASED ON CW, SEMI-STRUCTURED INTERVIEWS, AND THINK-ALOUD PROTOCOL

No.	Comments/Suggestions	No. of Feedback	Percentage (%)
1	Improvement on affix used	3	50
2	Improvement on the video content	2	33
3	Increase the interactivity of the application content	1	16
4	Restructure the content	4	66
5	Applying the application with others help	1	16
6	Introduce level in application	4	66
7	Improvement on aesthetic	4	66
8	Uniformity of the design	2	33

The results revealed that the task scenario completion level was 6/6, which indicates that the experts were able to complete the given task. In other words, this means that the experts completed 100% of the tasks, they were able to understand the application flow, and were aware of and knew the actions taken during the testing process. From Table IV, it can be concluded that the application's content and interface needed enhancement to increase user engagement in the self-learning applications. The results also confirmed that the function modelling and flow of the RakanBM application process were correct, but some improvisation is needed. For instance, the use of buttons, information displays, and contents needed some modification. Nonetheless, the experts agreed that the developed prototype was able to meet the needs of the user, but with some enhancements needed.

It should be noted that the open-ended questions supported the results of the questionnaire. For instance, most of the experts suggested an improvement in the content structure by adding more English translations on the earlier topics in the application. The experts also agreed on the need to introduce levels in the application. In short, the suggestions by experts to increase the aesthetic value on the current prototype supported (validated) the lowest mean value obtained by the text presentation factor of the questionnaire (see Table II).

The overall results revealed that the materials of the *RakanBM* mobile application for Malay Language learning for non-native speakers were relevant, validated by the factors that recorded a high mean: learning outcomes, assessments, effectiveness, and efficiency. The results were similar to previous studies where the mobile application was able to support language learning [6-8, 12]. In general, the *RakanBM* application is focused on Malay language learning with validated benefits. For instance, it is a novel creation for Malay language learning with complete topics suitable and best for non-native speakers. At the same time, the application is student-friendly, and users can use it easily, anytime, and anywhere. Furthermore, AR technology is also integrated into the mobile application to increase students' engagement and relate their surroundings to the learning activity.

B. Survey Results

Descriptive statistical analysis was carried out to identify the mean, the percentage, and the standard deviation (SD) for each item of the application factors: contents, text presentations, learning outcomes, assessments, effectiveness, efficiency, and satisfaction. Table III shows the mean for each of the items and factors, respectively. The calculated mean for each factor ranged from 3.53 to 4.50, while the average mean (for all factors) is 3.96. The results indicate that, in general, the evaluators tend to agree that the developed low-fidelity prototype was able to satisfy the factors' requirements. The results were then classified into two levels: high (mean > 4.0) and satisfied (mean > 3.5).

TABLE III.   PERCENTAGE (SCALE 1 TO 3) AND (SCALE 4 TO 5) FOR THE SEVEN FACTORS OF USING THE RAKANBM APPLICATION

No.	Factor/ Item	Percentage (%)(scale of 1 to 3)	Percentage (%)(scale of 4 to 5)	Mean (SD)
1.0	**Application contents**			**3.80 (1.05)**
1.1	The contents provided are suitable to the novice level	16.7	83.3	4.17 (1.17)
1.2	The contents provided have a clear continuity	33.4	66.6	3.83 (1.17)
1.3	The theme of the contents provided is appropriate to the novice level of learning	16.7	83.3	4.00 (1.10)
1.4	The chapters provided are adequate for the novice level	33.4	66.6	3.83 (1.17)
1.5	The language used is appropriate to the novice level of the student	50.0	50.0	3.17 (1.33)
1.6	The exercise provided is adequate	16.7	83.3	4.17 (0.75)
1.7	The exercise provided corresponds to the level of learning	16.7	83.3	4.00 (0.63)
1.8	Listening exercise through audio is adequate	16.7	83.3	4.17 (0.75)
1.9	The audio is clear and helps students listen well	16.7	83.3	3.83 (0.98)
1.10	Translation into English is limited to chapter 1 only	66.6	33.4	2.83 (1.47)
2.0	**Text presentations**			**3.53 (0.92)**
2.1	Presentation of contents using attractive graphics	66.6	33.4	2.83 (0.98)
2.2	The graphics used can help students understand the situation	16.7	83.3	4.33 (0.82)
2.3	The colours in the text are interesting	50.0	50.0	2.67 (0.82)
2.4	The length of the text for one chapter is sufficient	16.7	83.3	4.00 (0.63)
2.5	The presentation of the content is appropriate to the age level of the target students	50.0	50.0	3.83 (1.33)
3.0	**Learning outcomes**			**4.03 (1.00)**

No.	Factor/Item	Percentage (%)(scale of 1 to 3)	Percentage (%)(scale of 4 to 5)	Mean (SD)
3.1	Students will learn something new from this contents	0.0	100.0	4.67 (0.52)
3.2	Students are able to understand the Malay language is the basis of this contents	16.7	83.3	4.00 (1.55)
3.3	Students can also learn cultural information from this contents	33.4	66.6	4.00 (0.89)
3.4	Students will be excited to continue to level 2	50.0	50.0	3.83 (0.98)
3.5	These contents are sufficient for novice students	33.4	66.6	3.67 (1.03)
4.0	**Assessments**			**4.03 (0.79)**
4.1	The scoring for each exercise is balanced	16.7	83.3	4.00 (1.10)
4.2	Assessment is appropriate to the student's level of competence	0.0	100.0	4.17 (0.41)
4.3	The emojis used are appropriate to the student's level of competence	33.4	66.6	4.00 (1.26)
4.4	The assessment rubric is appropriate to the student's level of competence	16.7	83.3	4.17 (0.75)
4.5	Written, graphic/picture and oral tests can measure students' level of competence	33.4	66.6	3.83 (0.41)
5.0	**Effectiveness**			**3.97 (0.93)**
5.1	The learning contents displayed by this application are accurate and complete	33.4	66.6	3.67 (1.03)
5.2	Functions and learning contents on the application produce accurate output	33.4	66.6	3.83 (0.75)
5.3	Use of this application will increase the effectiveness in learning Malay	33.4	66.6	4.17 (0.98)
5.4	This application provides audio assistance in learning the Malay language	16.7	83.3	4.50 (0.84)
5.5	The terms contained in this application are clear and not confusing	33.4	66.6	3.67 (1.03)
6.0	**Efficiency**			**4.50 (0.91)**
6.1	This application helps learning the Malay language quickly	0.0	100.0	4.50 (0.55)
6.2	This application is able to increase student achievement in Malay language learning	16.7	83.3	4.33 (0.82)
6.3	Navigating this application is easy	33.4	66.6	4.00 (1.26)

No.	Factor/Item	Percentage (%)(scale of 1 to 3)	Percentage (%)(scale of 4 to 5)	Mean (SD)
6.4	This application allows for effective learning	16.7	83.3	4.17 (0.75)
6.5	The information available on this application is organised	33.4	66.6	3.83 (1.17)
7.0	**Satisfaction**			**3.90 (0.97)**
7.1	I am satisfied with this application	50.0	50.0	3.50 (1.05)
7.2	I would recommend my friends to use this application	50.0	50.0	4.00 (1.10)
7.3	I enjoy using this application	50.0	50.0	4.00 (1.10)
7.4	This application works just as well as I want it to	50.0	50.0	3.67 (0.82)
7.5	I want to use this application	16.7	83.3	4.33 (0.82)

Consequently, the factors that were recorded as high were the learning outcomes, assessments, effectiveness, and efficiency, while the factors that were recorded as satisfied were the application contents, text presentations, and satisfaction. The efficiency factor obtained the highest mean value of 4.50 with an SD of 0.91, which indicates that the experts (evaluators) tend to agree that the application was very efficient and necessary to help non-native speakers improve their Malay language basics, especially with the assistance of teachers. Meanwhile, the assessments (mean = 4.03, SD = 0.97) and learning outcomes factors (mean = 4.03, SD = 1.00) both obtained a mean value of 4.03, the second-highest mean value. The rest of the factors in the order of descending mean value are effectiveness (mean = 3.97, SD = 0.93), satisfaction (mean = 3.90, SD = 0.97), application content (mean = 3.80, SD = 1.05), and text presentation (mean = 3.53, SD = 0.92). The lowest mean value for the text presentation factor was likely due to feedback from some evaluators stating that the RakanBM interface was less aesthetic. Nevertheless, the positive responses were still more than the negative ones in which the percentage value for factors having a scale of more than 4 was 53% (i.e. agreed by 53% of experts).

The experts' opinions on the positive and negative aspects of the RakanBM prototype application were also collected and listed in Table IV. Following the prototype evaluation, the experts tend to agree with its interface features, elements, and functions. The experts also agreed that the application achieved a positive level of user experience. Nevertheless, analysis of the experts' opinions (i.e. related to the positive and negative aspects) and comments revealed that some improvements were needed.

Feedback on the positive aspects indicated that the application was student-friendly, a more compact learning material, and provided a more accessible and interactive Malay language learning experience. On the other hand, the negative aspects highlighted the difficulty of certain words, more interactive aspects needed, and more effective technique for

assessment required. The items that need to be improved, as indicated in Table III and Table IV will be addressed in the next prototype development to develop a more effective Malay language learning application.

TABLE IV.   FEEDBACK ON POSITIVE AND NEGATIVE ASPECTS OF RAKANBM APPLICATION

Expert	Positive Aspect	Negative Aspect
A	• "This application is very student-friendly, the content and learning materials are also presented in a very simple, easy, and compact way."	• None
B	• "The first complete application was produced, very useful and easy for international students and can be applied on Android and iPhone". • "Topics are very suitable and best for non-native speakers."	• None
C	• "Content. Use of different languages (pronouns) according to the conversation situation as shown in Chapter 1. Various exercises such as True / False, Multiple Choice and Vocabulary that tests listening and viewing skills and the application of those skills."	• "As a novice student, it is quite difficult to comprehend the rich words. Words exist in conversation."
D	• "An effort made to develop Malay language application for foreign speakers use."	• "Need to add more interactive aspects."
E	• "Malay language learning is more accessible and interactive even on its own."	• None
F	• "There is an AR element. AR can be used for collaborative teaching."	• "Assessment items should be based on audio not surrounding objects."

Furthermore, the study finding corresponded to past research where mobile application utilization enhanced user performance and language learning [3, 7, 8]. The implementation of gamification in education has increased engagement and achieved learning more effectively[39]. For instance, the study of Sorrentino and Sorrentino and Spano [40] and Fang, Yeh [41]. Both of these studies prove how the use of mobile apps with the mobile application has really helped improve the English proficiency of non-native students. Among the focuses given were mastery of vocabulary, pronunciation practices, and the construction of simple sentences. Besides, Wang and Han [42] have used English language learning modules for native Chinese speaking students to further improve their mastery of communication strategies in English. This learning module was based on digital games using the smartphone application "Liulishou" (fluent in English). This study shows that this digital learning material has been able to help improve the level of pronunciation accuracy, fluency in communication, and the production of complex sentence structures orally among students. Another interesting mobile application was developed by Yamamoto, Rodriguez [43], who sought to innovate FinDo, which is a smartphone app capable of listing vocabulary related

to where users are located. For example, when a consumer is in a supermarket, the consumer can learn all the words related to the objects around it. Specifically, it turns out that all studies on language learning using mobile devices improved students' language proficiency well.

However, there were some challenges that lie within this expert review on mobile augmented reality applications for language learning studies. Since the whole process is majorly held by online meetings, it becomes boundaries in communication involving the experts. Nevertheless, the researchers' team has planned the method so well to lessen the impact as the situation was unavoidable due to pandemic times. Further guidelines can be taken for future research methods, especially in order to make progress more effective and efficient.

This whole chapter discusses the expert review ranging from prototype testing using CW techniques, semi-structured interviews, think-aloud protocols, and prototype evaluation using surveys upon completion of the test. It has been proven that RakanBM application was relevant and suitable to support Malay language mobile learning in line with previous studies where mobile learning has increased user's performance, especially in education [44, 45]. The analysis was conducted on seven construct and open-ended questionnaires to obtain expert feedback on the positive and negative aspects of the application as well. The analysis results show that the usefulness and the topics covered were suitable and best for foreign speakers.

V.   CONCLUSION

In this study, a review panel of six experts from the Malay language and HCI fields evaluated *RakanBM*, a newly developed mobile application prototype for Malay language learning for first speakers and foreign speakers. The evaluation methods consisted of CW, semi-structured interviews, think-aloud protocols, and surveys. The findings were analysed and summarised to determine the usability of the application and have contributed to new knowledge. In this study, it has been found that the expert review was in line with the previous literature where the mobile application was able to support learning language. The *RakanBM* application aimed to assist foreign users, especially novice user learning in classroom environments. The mobile application proved to be interesting, and the experts saw high potential in the application, especially if the needed changes could be implemented. The experts' recommendations were presented for the benefit of researchers and developers for the future development of Malay language learnings. Further research involving usability testing will be done for the target user - the non-native speakers – taking into consideration the findings of this present study.

REFERENCES

[1] Poultsakis, S., et al., The management of Digital Learning Objects of Natural Sciences and Digital Experiment Simulation Tools by teachers. Advances in Mobile Learning Educational Research, 2021. 1(2): p. 58-71.

[2] Gang, B., et al., A Bahasa Malaysia Interactive Book App as A Speech-Language Therapy Tool for Children with Language Delay. Asia-Pacific Journal of Information Technology and Multimedia, 2017. Vol. 6 No. 1, June 2017: 23 - 37.

[3] Rosell-Aguilar, F., Autonomous language learning through a mobile application: a user evaluation of the busuu app. Computer Assisted Language Learning, 2018. 31(8): p. 854-881.

[4] Huynh, D. and H. Iida, An Analysis of Winning Streak's Effects in Language Course of" Duolingo". Asia-Pacific Journal of Information Technology and Multimedia, 2017. Vol. 6 No. 2.

[5] Razak, N.A., H. Saeed, and H. Alakrash, Pedagogical issues of using ICT applications in Iraq. Asia-Pacific Journal of Information Technology and Multimedia, 2019. 7(2-2): p. 158-168.

[6] Shawai, Y.G. and M.A. Almaiah, Malay language mobile learning system (MLMLS) using NFC technology. International Journal of Education and Management Engineering, 2018. 8(2): p. 1.

[7] Bojórquez, E.M., et al., study on mobile augmented reality adoption for Mayo language learning. Mobile Information Systems, 2016.

[8] Cho, M.-H. and D.A. Castañeda, Motivational and affective engagement in learning Spanish with a mobile application. System, 2019. 81: p. 90-99.

[9] Halid, N.A. and N. Omar, Malay Part Of Speech Tagging Using Ruled-Based Approach. Asia-Pacific Journal of Information Technology and Multimedia, 2017. 6(2): p. 91-107.

[10] Tiun, S., A.L. Salikin, and S.K. Muhammad, Using FOSS TTS Developer Tool to build Malay TTS System. Asia-Pacific Journal of Information Technology and Multimedia, 2016. Vol. 4 No. 2.

[11] Shadiev, R., W.-Y. Hwang, and Y.-M. Huang, Review of research on mobile language learning in authentic environments. Computer Assisted Language Learning, 2017. 30(3-4): p. 284-303.

[12] Ali, S.K. and N.S. Azmi. Augmented Reality in learning Malay Language. in 2019 2nd International Conference on Applied Engineering (ICAE). 2019. IEEE.

[13] Abd Majid, N.A. and N. Abd Majid, Augmented reality to promote guided discovery learning for STEM learning. Int. J. on Advanced Science, Engineering and Information Technology, 2018. 8(4-2): p. 1494-1500.

[14] Che Hashim, N., et al., User satisfaction for an augmented reality application to support productive vocabulary using speech recognition. Advances in Multimedia, 2018.

[15] Lam, M.C., M.J. Sadik, and N.F. Elias, The effect of paper-based manual and stereoscopic-based mobile augmented reality systems on knowledge retention. Virtual Reality, 2020: p. 1-16.

[16] Lam, M.C., et al., Interactive Augmented Reality with Natural Action for Chemistry Experiment Learning. TEM Journal, 2020. 9(1): p. 351.

[17] Hanafi, H.F., et al. Improving students' motivation in learning ict course with the use of a mobile augmented reality learning environment. in IOP Conference Series: Materials Science and Engineering. 2017. IOP Publishing.

[18] Nordin, N.A.A., N.A. Abd Majid, and N.F.A. Zainal, Mobile augmented reality using 3D ruler in a robotic educational module to promote STEM learning. Bulletin of Electrical Engineering and Informatics, 2020. 9(6): p. 2499-2506.

[19] Botha, B.S., L. de Wet, and Y. Botma, Experts' review of a virtual environment for virtual clinical simulation in South Africa. Computer Animation and Virtual Worlds, 2020: p. e1983.

[20] Beecham, S., et al., Using an expert panel to validate a requirements process improvement model. Journal of Systems and Software, 2005. 76(3): p. 251-275.

[21] Lazar, J., J.H. Feng, and H. Hochheiser, Research methods in human-computer interaction. 2017: Morgan Kaufmann.

[22] Jaspers, M.W., A comparison of usability methods for testing interactive health technologies: methodological aspects and empirical evidence. International journal of medical informatics, 2009. 78(5): p. 340-353.

[23] Kaufman, D.R., et al., Usability in the real world: assessing medical information technologies in patients' homes. Journal of biomedical informatics, 2003. 36(1-2): p. 45-60.

[24] Zimmermann, P.G., Beyond usability: measuring aspects of user experience. 2008, ETH Zurich.

[25] Mourouzis, A., I. Chouvarda, and N. Maglaveras. Mhealth: common usability and user experience practices and flaws. in European, Mediterranean & Middle Eastern Conference on Information Systems 2015. 2015.

[26] DiCicco-Bloom, B. and B.F. Crabtree, The qualitative research interview. Medical education, 2006. 40(4): p. 314-321.

[27] Schaper, M.-M., et al., Learning about the past through situatedness, embodied exploration and digital augmentation of cultural heritage sites. International Journal of Human-Computer Studies, 2018. 114: p. 36-50.

[28] Beauchemin, M., et al., A multi-step usability evaluation of a self-management app to support medication adherence in persons living with HIV. International journal of medical informatics, 2019. 122: p. 37-44.

[29] Yen, P.-Y. and S. Bakken, Review of health information technology usability study methodologies. Journal of the American Medical Informatics Association, 2012. 19(3): p. 413-422.

[30] 30.Gruenstein, A., I. McGraw, and I. Badr. The WAMI toolkit for developing, deploying, and evaluating web-accessible multimodal interfaces. in Proceedings of the 10th international conference on Multimodal interfaces. 2008.

[31] Ssemugabi, S. and M. De Villiers, Effectiveness of heuristic evaluation in usability evaluation of e-learning applications in higher education. South African computer journal, 2010. 2010(45): p. 26-39.

[32] 32.Fabil, N.B., A. Saleh, and R.B. Isamil, Extension of PACMAD model for usability evaluation metrics using Goal Question Metrics (GQM) Approach. Journal of Theoretical and Applied Information Technology, 2015.

[33] Hussain, A., et al., A metric-based evaluation model for applications on mobile phones. Journal of Information and Communication Technology, 2013. 12: p. 55-71.

[34] Weichbroth, P., Usability of mobile applications: a systematic literature study. IEEE Access, 2020. 8: p. 55563-55577.

[35] Talib, O., SPSS: Analisis data kuantitatif untuk penyelidik muda. 2015: MPWS Rich Publication Sdn. Bhd.

[36] Pallant, J., SPSS survival manual. 2013: McGraw-hill education (UK).

[37] Tavakol, M. and R. Dennick, Making sense of Cronbach's alpha. International journal of medical education, 2011. 2: p. 53.

[38] Moustakas, C., Phenomenological research methods. 1994: Sage publications.

[39] Kalogiannakis, M., S. Papadakis, and A.-I. Zourmpakis, Gamification in science education. A systematic review of the literature. Education Sciences, 2021. 11(1): p. 22.

[40] Sorrentino, F. and L.D. Spano, Post-it notes: supporting teachers in authoring vocabulary game contents. Multimedia Tools and Applications, 2019. 78(16): p. 23049-23074.

[41] Fang, W.-C., et al., Effects of mobile-supported task-based language teaching on EFL students' linguistic achievement and conversational interaction. ReCALL, 2021. 33(1): p. 71-87.

[42] Wang, Z. and F. Han, Developing English language learners' oral production with a digital game-based mobile application. Plos one, 2021. 16(1): p. e0232671.

[43] Yamamoto, K., J. Rodriguez, and Y. Tsujino. FinDo: A Foreign Language Vocabulary Learning System Based on Location-Context. in 2019 20th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD). 2019. IEEE.

[44] Papadakis, S., et al., Developing and exploring an evaluation tool for educational apps (ETEA) targeting kindergarten children. Sustainability, 2020. 12(10): p. 4201.

[45] Papadakis, S., M. Kalogiannakis, and N. Zaranis, Designing and creating an educational app rubric for preschool teachers. Education and Information Technologies, 2017. 22(6): p. 3147-3165.