Software Quality Testing in Mobile Application (ArabEasy) Based on the PACMAD Model

https://doi.org/10.3991/ijim.v16i10.28433

Safitra Aditya Bagaskara, Achmad Teguh Wibowo(✉), Muhammad Andik Izzuddin, Mohammad Khusnu Milad, Rr. Diah Nugraheni Setyowati, Taufik, Mujib Ridwan, Sirajul Arifin
UIN Sunan Ampel, Surabaya, Indonesia
atw@uinsby.ac.id

Abstract—ArabEasy is an application to learn and understand basic concepts of Arabic. This app has a problem with the user interface (UI); based on the comment from users, UI this app is confusing to use. This research focuses on measuring, testing, and improving UI usability using the PACMAD model. First, data collection was obtained from the Performance Measurement, SUS Questionnaire, NASA-TLX Questionnaire, and RTA Questionnaire. The result from this process is measured using usability metrics, statistical testing, and calculation from the questionnaire, respectively. The following process, developing a new UI to improve the app, became easy for users from a prototype based on The four Golden Rules of User Interface Design. This process showed increased effectiveness 37.22%, efficiency 46.08%, satisfaction 24.25. Furthermore, the aspect of the error decreased 13.4%, and the cognitive load 18.54, so that an improvement using the PACMAD model is effective in this case.

Keywords—ArabEasy application, PACMAD usability model, the four golden rules of user interface design, usability evaluation

1 Introduction

Arabic is one of the official languages approved by the United Nation's in 1974 because Arabic has unique language characteristics, and many science books are written using the Arabic language [1]. Based on data from the https://worldpopulationreview.com/countries/indonesia-population, Indonesia has the most Muslim population in 2021, which is 231 million people [2]. In Indonesia, Arabic is applied in formal and non-formal education institutions, and this language is taught from elementary school to university. However, this does not guarantee success in learning Arabic because Arabic focuses on textbooks, which is difficult to understand, so that developing a mobile app is needed to increase interest in users in studying the Arabic language.

The increasing emergence of mobile applications makes many people think that smart mobile devices (mobile apps) can be one way to provide learning services for preschools and schools based on their age [3][4][5]. Many research uses mobile apps because this technology grows faster and helps humans in all aspects [6]. In 2011, more than 300 million apps were developed based on Android and iOS [7]. Mobile apps can be implemented in digital learning to give a new user experience and simplify teaching and learning processes. Many opportunities for the younger generation to develop knowledge and perspectives to increase learning outside of school in the digital world [8][9][10]. This technology is called mobile learning [11]. Mobile learning develops using resources by mobile devices and mobile apps. They are specifically designed and adapted to education [12]. Which mobile learning impacts students to improve access to learning anytime and anywhere [13].

This research improves mobile apps, namely ArabEasy. ArabEasy is a mobile app used to learn the Arabic language. This app was released on Jul. 21, 2020, and can be a free download via Google Playstore based on Android. ArabEasy is more effective than conventional learning. However, users still have suggestions, especially UI, which is still confusing, so usability testing needs to be done.

Usability testing in the application is essential because usability measures requirements and success apps based on user recognition [14]. Because when apps fail to show the user expected, the app is left by the user [15]. The usability level not only measures the satisfaction and learnability of users but also can identify hidden problems in the application, which can prevent potential the app [16][17]. The Nielsen and ISO model are typical for desktop-based applications. Mobile technology brings challenges in terms of usability testing. Mobile apps have many features and menu levels that are a measurement aspect more detailed and spacious. Therefore People At The Center Mobile Application Development (PACMAD) Model is suitable for usability testing in this research [18].

The PACMAD model aims to extend existing usability models, such as Nielsen and ISO, into the context of mobile applications. With the addition of aspects of cognitive load as an essential contribution of this model [18]. This model has seven aspects in testing. Aspects of effectiveness, efficiency, learnability, memorability, errors and cognitive load were tested using performance measurement techniques [19]. Aspects of cognitive load were tested subjectively using the NASA-TLX questionnaire [20]. This aspect was also tested using a Retrospective Think Aloud (RTA) questionnaire to record the cognitive load experienced by users during the test [21]. The satisfaction aspect was tested using the System Usability Scale (SUS) questionnaire [22].

This research aims usability testing in login, creating an account, learning, hearing, writing, and practice questions. The training feature consists of two steps: 1. Pre, and 2. Post-Evaluation. Step 1 to measure effectiveness, efficiency, errors, and cognitive load aspect.

Furthermore, step 2 measures learnability and memorability. Questionnaire SUS was used to measure satisfaction, and cognitive load was tested again using NASA-TLX and RTA. Finally, all aspects tested using statistical to know significant differences all aspects tested to a user.

Based on usability testing in pre-evaluation. Obtained the result of a user fail when performing login, creating an account, and practice questions. Furthermore, users also have trouble showing sound and learning, hearing, and writing features. So that recommendation to improvement and post-evaluation is necessary. This process results in user success in logging in, creating an account, practicing questions, showing sound in learning, hearing, and writing menu with a new UI developed.

The benefit of this research is improving the ArabEasy apps, especially the UI, to make it easier for users to use, especially on features that can bring up voices and practice questions. Then the UI improvements in this research can then be implemented for UI the ArabEasy apps in the future.

This research is arranged as follows. Section II describes the related work, flowchart, data collection and calculation techniques, and data analysis. The results and evaluation are shown in sections III, and the discussion and conclusions we show in section IV.

2 Material and method

This section describes a literature review of relevant previous research and explains several retrieving and calculation data in the PACMAD usability model.

2.1 Related work

The related work in this research refers to some previous papers about usability testing. Usability testing for mobile apps uses a specific usability model, namely the PACMAD model. This model was developed to focus on mobile apps with the increased cognitive load as an essential role. Usability PACMAD model consists of effectiveness, efficiency, satisfaction, learnability, memorability, errors, and cognitive load [18].

The following research for usability testing uses the PACMAD model and the GQM approach for evaluating mobile apps [19]. This approach is carried out in several stages: 1) conducting a literature or journal study regarding the evaluation of mobile apps. 2) develop usability metrics for mobile apps. 3) usability metrics for the usability attributes of the PACMAD model will be applied to the GQM approach. 4). evaluation experiments to ensure the model results are valid and effective. 5) The data collected is divided into two methods: objective data by observation and subjective data using the CSUQ questionnaire with a scale of 1–5.

References support usability testing based on the PACMAD model like health and fitness applications [20]. Using calculation usability metrics, this app measured effectiveness, efficiency, learnability, memorability, and errors. While cognitive load was measured using the NASA-TLX questionnaire. The result in this research is that health and fitness application has low usability because this app was developed without aspect and usability factor considered.

The model developed to capture usability attributes in mHealth applications is a hierarchical usability model based on the Integrated Measurement Model (IMM) and People at the Center of Mobile Application Development (PACMAD). This model aims to assess the usability of the mHealth application effectively, minimize resource development, and improve usability. A literature review on usability testing using the PACMAD model was carried out on a mobile health application to design a specific usability model for mHealth application assessment [21].

Another literature review that uses the PACMAD model to test usability is the PSAU mobile application [23]. This research aims to measure the extent to which the usability attributes of effectiveness, efficiency, satisfaction, learnability, memorability, error, and cognitive load on the PSAU application based on the point of view of students enrolled in the College of Business Administration (CBA) at Prince Sattam Bin Abdulaziz. University for the 2019–2020 school year. The results obtained are that the PSAU mobile application can be accepted with an overall high average of 2.8, which reflects that this application is feasible to use.

Literature review on advances in Mobile Learning Education Research (AMLER): Mobile learning as an educational reform [24]. This research aims to discuss digital technology in the world of education, such as natural mobile devices (touch, trial, and error) as teaching and learning media for students. Mobile devices make it easy for students to interact directly with the system; for example, smart devices accompanying apps can be helpful for effective and engaging learning processes in educational environments.

Support for other literature reviews the use of interactive touch screen devices among young people at this time has grown [25]. However, the role of parents is still doubtful about the appropriateness of the apps used to deliver educational content. However, mobile technology is increasingly becoming important in the lives of young people to facilitate the process of remote learning because it is more effective and portable.

Based on the literature review, the research that has been carried out can be used as a reference to show that touch screen mobile devices have an essential role in young people's learning process in the educational environment. Mobile devices such as the ArabEasy apps make it easy for young people to understand basic Arabic concepts. These apps need to be tested for usability using the PACMAD model. Then make recommendations to improve the UI, becoming a prototype based on The Four Golden Rules of User Interface Design. This research used performance measurement, SUS questionnaire, RTA questionnaire, and NASA-TLX questionnaire for data collection and measured using a usability matrix with questionnaire calculations and statistical tests.

2.2 Flowchart this research

Flowchart this research explains preparation, data collection, calculation techniques, data analysis, design recommendations for improvement, post-evaluation, conclusions, and suggestions. Figure 1 shows the flowchart as follow:

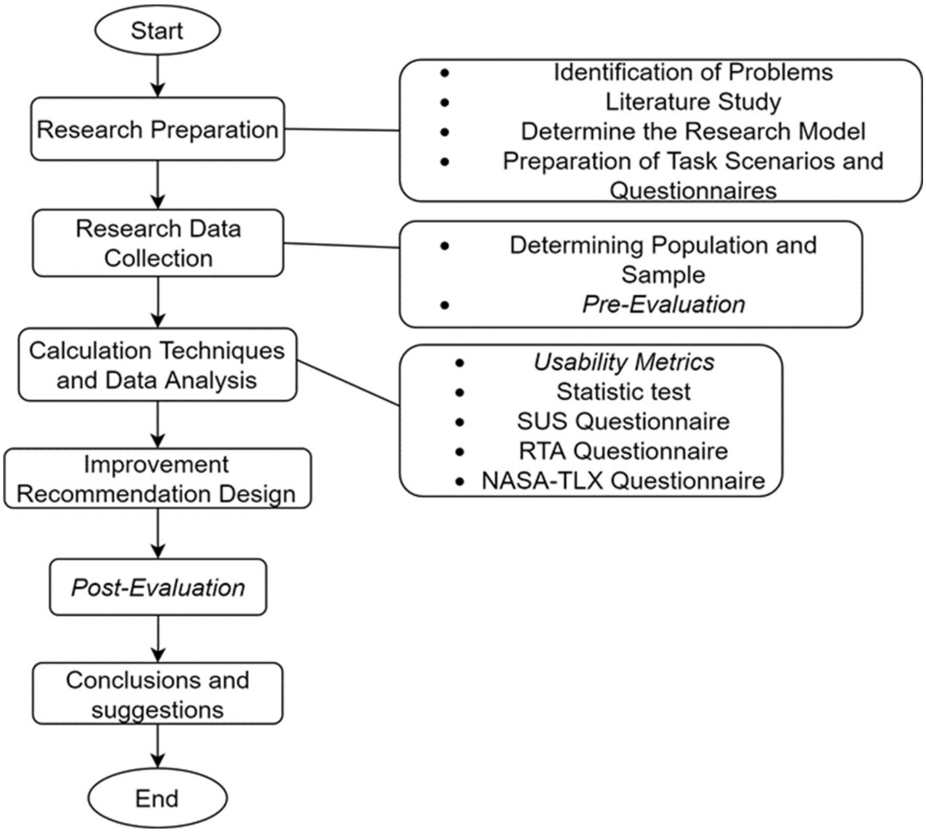

Fig. 1. The flowchart in this research

2.3 Research data collection

Data collection in this research begins with determining the population and sample. The people used are high school students and undergraduate students aged 15–24 years who are in the scope of Surabaya city. In determining the amount, the sample usability testing of this study refers to the research entitled "Beyond the five-user assumption: Benefits of increased sample sizes in usability testing." The researchers assume that to do usability testing with five users only can produce problems at a usability level of 55%. To achieve usability problems, 80% needed ten users later to reach the problem usability 95% required as many as 20 users [26]. So in this research uses a sample of 20 samples divided into two groups, namely high school student respondents and undergraduate student respondents.

The Pre-valuation process for users is carried out using four techniques. The tests are Performance Measurement, SUS Questionnaire, RTA Questionnaire, and Questionnaire NASA-TLX. The PM technique collects data on the number of successful tasks performed, the number of clicks, the duration of time worked, and the number of user errors during testing via a screen recorder. This technique was carried out two times (PM Phase 1 and PM Phase 2). PM Phase 1 collects data on the number of tasks successfully executed, the number of clicks, duration of processing time, and user errors during testing. PM Phase 1 assesses the effectiveness, efficiency, errors, and cognitive load. PM Phase 2 testing is the same as PM Phase 1 but only used to collect data on the number of completed tasks and the time duration quality compared to PM Phase 1. So this is to obtain an assessment on aspects of learnability and memorability. Each PM Phase 1 and 2 contains 27 tasks based on frequently used application features in the same scenario.

Meanwhile, the SUS, RTA, and NASA-TLX questionnaires were distributed after completing the performance measurement technique. Users are requested to fill in distributed questionnaires. The SUS and NASA-TLX questionnaires were used to obtain quantitative data on satisfaction and cognitive load after using the application. Furthermore, the RTA questionnaire was used to obtain qualitative data regarding the experience felt by the respondent during the test, so this is related to cognitive load or aspects of cognitive load. In addition, the RTA questionnaire was used to get criticism, suggestions, and input from users whose results are used as a reference in designing user interface improvement design recommendations.

2.4 Calculation techniques and data analysis

This stage is carried out to assess each aspect of the model PACMAD with each formula whose data is obtained from quantitative data (performance measurement, SUS questionnaire, and NASA-TLX questionnaire) and qualitative data (RTA questionnaire).

Quantitative Data This method uses usability testing with performance techniques measurement. There are seven aspects to model usability: 1) Effectiveness is calculated in two methods. A) the data is measured using SPSS software to determine the calculation of the nonparametric statistical test of the Mann Whitney U test. B) collects metrics for the number of tasks completed in PM Phase 1 (Completion Rate). The equation process is as the following:

$$Completion\,Rate = \frac{\text{Number of tasks completed the successfully}}{\text{total number of tasks undertaken}}\,100\% \tag{1}$$

The result of this process completing a minimum task on usability testing is 78%, but if the test result is below 49%, it can be placed in the lower quartile [29]. 2) Efficiency is calculated in two methods. A) the data is measured using SPSS software to determine the Independent T-Test parametric statistical test. B) collect the duration metric for the time a task is completed in PM Phase 1 (Overall Relative Efficiency). The equation process is as the following:

$$Overall\ Relative\ Efficiency = \frac{\sum_{j=1}^{R}\sum_{i=1}^{N} nijtij}{\sum_{j=1}^{R}\sum_{i=1}^{N} tij}\, x100\% \tag{2}$$

Where N is the total number of assignments, and R is the total number of respondents. Then nij is the result of respondent j's task. If the respondent completes the job, the value is one, and if it fails, it means 0. While the time will be completed by respondent *j* to complete the task, if the job is completed, the time will be counted until the respondent quits the task.

3) Satisfaction is calculated in two methods. A) the data is measured using SPSS software to determine the size of the Independent T-Test parametric statistical test. B) collect the SUS metric total score questionnaire. For each odd-numbered question (1,3,5,7,9), the value of the respondent is reduced by 1.

$$odd\ SUS\ value = \sum Px - 1 \quad (3)$$

For each question with an even number (2,4,6,8,10), the respondent's score is used to subtract 5.

$$SUS\ value\ even = 5 - \sum Px \quad (4)$$

Furthermore, the results of each question in each respondent are added up and then divided by 2.5.

$$(\sum \text{odd value} + \sum \text{even value}) \times 2.5 \quad (5)$$

The last step is to average the results of all respondents. Add up all scores, then divide by the total number of respondents.

$$X^{-} = \sum \times \text{n} \quad (6)$$

The minimum value of SUS to produce competent usability is 68 [30].

4) Learnability is calculated in two methods. A) the data is measured using SPSS software to determine the calculation of the nonparametric statistical test of the Mann Whitney U test. B) compare the effectiveness values observed from PM Phase 1 and PM Phase 2. If there is an increase in the comparison value, then the level of learnability aspect can be said to be good. 5) Aspect Memorability is calculated in two methods. A) the data is measured using SPSS software to determine the Independent T-Test parametric statistical test calculation. B) compare the overall relative effectiveness and efficiency values obtained from PM Phase 1 and PM Phase 2. If both values increase, then the memory level can be good, but if one of the two values or both decreases, it could be less good.

6) Errors are calculated in two methods. A) measured using SPSS software to determine the calculation of the nonparametric statistical test of the Mann Whitney U test. B) collect the number of user error metrics in PM Phase 1 (Error rate). The equation process is as the following:

$$Error\ Rate = \frac{Total\ Defects}{Total\ Opportunities} \quad (7)$$

Where: Error Rate = Number of error rates, Total Defects = The total number of wrong tasks, and Total Opportunities = Total job opportunities. 7) Cognitive Load is calculated in two methods. A) measured using SPSS software to determine the Independent T-Test parametric statistical test calculation. B) collect the number of NASA-TLX questionnaire values based on the six aspects. Namely, the user has perceived mental need, and the user has perceived physical condition, the user is essential for the amount of time, the user's performance, the level of effort made by the user, and the level of frustration felt.

The mental workload of each respondent is calculated based on the six aspects. The equation process is as the following:

$$\text{Workload} = \frac{\sum(\text{weight} \times \text{rating})}{15} \quad (8)$$

Where Σ(weight × rating)is the number of selected indicators. Then the average final result of the respondents' mental workload is as the following:

$$\text{Total final result} = \frac{\sum \text{mental work load of each user}}{\text{number of Users}} \quad (9)$$

Where Σ the mental workload of each user is the total mental workload per user. Obtaining a mental workload score can be shown as follow [31], And the pseudocode of this process is as follow:

```
1. Score value > 80, which is to state that the workload
   is heavy.
2. A score of 50-70 states that the workload is moderate
3. , and a score of <50 indicates that the workload is
   relatively light.
```

Qualitative Data is carried out by collecting data specifically from the RTA questionnaire regarding suggestions, input, feelings, and any obstacles respondents face when using the ArabEasy application during the testing process and as a reference for designing recommendations for improvement.

3 Material and method

This research is based on seven aspects of the PACMAD model. This aspect depends on statistical nonparametric test results that the Mann Whitney U Test shows in Tables 1–3. Furthermore, this aspect also depends on the statistical parametric test result Independent T-Test shown in Tables 4–7.

Table 1 shows the result from the statistical test Mann Whitney U test analysis on the effectiveness aspect. Before testing, first, determine the hypothesis with the equation H0: (p-value > 0.05) there is no difference in the number of completed tasks and H1: (p-value < 0.05) there is a difference in the number of completed tasks in PM Phase 1. Based on this hypothesis, the p-value is .075, which means that the p-value > 0.05 decides to reject H1. So it can be concluded that there is no significant difference in the number of tasks completed in PM Phase 1 to respondents.

Table 1. Statistical test results Mann Whitney U Test

Statistics Test	
	Number of Completed Tasks in PM Phase 1
Mann-Whitney U	31.000
Wilcoxon W	86.000
Z	−1.782
Asymp. Sig. (2-tailed)	.075
Exact Sig. [2*(1-tailed Sig.)]	.165

Table 2. Statistical test results Mann Whitney U Test

Statistics Test	
	Number of Completed Tasks in PM Phase 2
Mann-Whitney U	37.500
Wilcoxon W	92.500
Z	−1.129
Asymp. Sig. (2-tailed)	.259
Exact Sig. [2*(1-tailed Sig.)]	.353

Table 3. Statistical test results Mann Whitney U Test

Statistics Test	
	Number of Mistakes Tasks in PM Phase 1
Mann-Whitney U	31.000
Wilcoxon W	86.000
Z	−1.782
Asymp. Sig. (2-tailed)	.075
Exact Sig. [2*(1-tailed Sig.)]	.165

Table 4. Statistical test results independent t-test

Independent Samples Test				
		t-test for Equality of Means		
		Sig. (2-tailed)	**Mean Difference**	**Std. Error Difference**
Total time duration completed in PM Phase 1	Equal variances assumed	.019	320.60000	124.86844
	Equal variances not assumed	.020	320.60000	124.86844

Table 5. Statistical test results independent t-test

Independent Samples Test				
		t-test for Equality of Means		
		Sig. (2-tailed)	**Mean Difference**	**Std. Error Difference**
The total score of SUS Questionnaire	Equal variances assumed	.011	11.00000	3.87657
	Equal variances not assumed	.014	11.00000	3.87657

Table 6. Statistical test results independent t-test

Independent Samples Test				
		t-test for Equality of Means		
		Sig. (2-tailed)	**Mean Difference**	**Std. Error Difference**
Total time duration completed in PM Phase 2	Equal variances assumed	.018	293.80000	113.17781
	Equal variances not assumed	.021	293.80000	113.17781

Table 7. Statistical test results independent t-test

Independent Samples Test				
		t-test for Equality of Means		
		Sig. (2-tailed)	**Mean Difference**	**Std. Error Difference**
Cognitive Load	Equal variances assumed	.019	107.66667	41.97772
	Equal variances not assumed	.020	107.66667	41.97772

Table 2 shows the Mann Whitney u test analysis on the learnability aspect. Before testing, first, determine the hypothesis with the equation H0: (p-value > 0.05) there is no difference in the number of completed tasks and H1: (p-value < 0.05) there is a difference in the number of completed tasks in PM Phase 2. Based on this hypothesis, the p-value is .259, which means that the p-value > 0.05 decides to reject H1. So it can be concluded that there is no significant difference in the number of tasks completed in PM Phase 2 to respondents.

Table 3 shows the statistical test Mann Whitney u test analysis on errors aspect. Before testing, first, determine the hypothesis with the equation H0: (p-value > 0.05) there is no difference in the number of mistakes tasks and H1: (p-value < 0.05) there is a difference in the number of mistakes tasks in PM Phase 1. Based on this hypothesis, the p-value is .075, which means that the p-value > 0.05 decides to reject H1. So it can be concluded that there is no significant difference in the number of mistakes tasks in PM Phase 1 to respondents.

Table 4 shows the result from the statistical test independent t-test analysis on the efficiency aspect. Before an independent t-test was conducted, the main requirement was to test for normality and homogeneity because it is a parametric statistical test. After testing, the normality test results were .524, and the homogeneity test was .632. Based on these results, data on the efficiency aspect is normality distributed and homogeneity. The next step is to perform an independent t-test and perform a hypothesis first with the equation H0: (p-value > 0.05); there is no difference in the amount of time completed, and H1: (p-value < 0.05) there is a difference in the time completed in PM Phase 1. Based on this hypothesis, the p-value is .019 and .020, which means the p-value< 0.05 decides to reject H0. So it can be concluded that there is a significant difference in the amount of time completed in PM Phase 1 to respondents.

Table 5 shows the result from the statistical test independent t-test analysis on the satisfaction aspect. Before an independent t-test was conducted, the main requirement was to test for normality and homogeneity because it is a parametric statistical test. After testing, the normality test results were .780, and the homogeneity test was .106. Based on these results, data on the satisfaction aspect is normality distributed and homogeneity. The next step is to perform an independent t-test and perform a hypothesis first with the equation H0: (p-value > 0.05). There is no difference in the total score of SUS Questionnaire and H1: (p-value < 0.05); there is a difference in the total score of SUS Questionnaire. Based on this hypothesis, the p-value is .011 and .014, which means the p-value< 0.05 decides to reject H0. So it can be concluded that there is a significant difference in the total score of SUS Questionnaire to respondents.

Table 6 shows the result from the statistical test independent t-test analysis on the memorability aspect. Before an independent t-test was conducted, the main requirement was to test for normality and homogeneity because it is a parametric statistical test. After testing, the normality test results were .273, and the homogeneity test was .263. Based on these results, data on the memorability aspect is normality distributed and homogeneity. The next step is to perform an independent t-test and perform a hypothesis first with the equation H0: (p-value > 0.05); there is no difference in the amount of time completed, and H1: (p-value < 0.05) there is a difference in the time completed in PM Phase 2. Based on this hypothesis, the p-value is .018 and .021, which means the p-value < 0.05 decides to reject H0. So it can be concluded that there is a significant difference in the amount of time completed in PM Phase 2 to respondents.

Table 7 shows the statistical test independent t-test analysis on the cognitive load aspect. Before an independent t-test was conducted, the main requirement was to test for normality and homogeneity because it is a parametric statistical test. After testing, the normality test results were .366, and the homogeneity test was .730. Based on these results, data on the cognitive load aspect is normality distributed and homogeneity. The next step is to perform an independent t-test and perform a hypothesis first with the equation H0: (p-value > 0.05). There is no difference in the cognitive load, and H1: (p-value < 0.05) there is a difference in the cognitive load. Based on this hypothesis, the p-value is .019 and .020, which means the p-value < 0.05 decides to reject H0. So it can be concluded that there is a significant difference in the cognitive load to respondents. The effectiveness, efficiency, satisfaction, learnability, memorability, and cognitive load are shown in Figure 2.

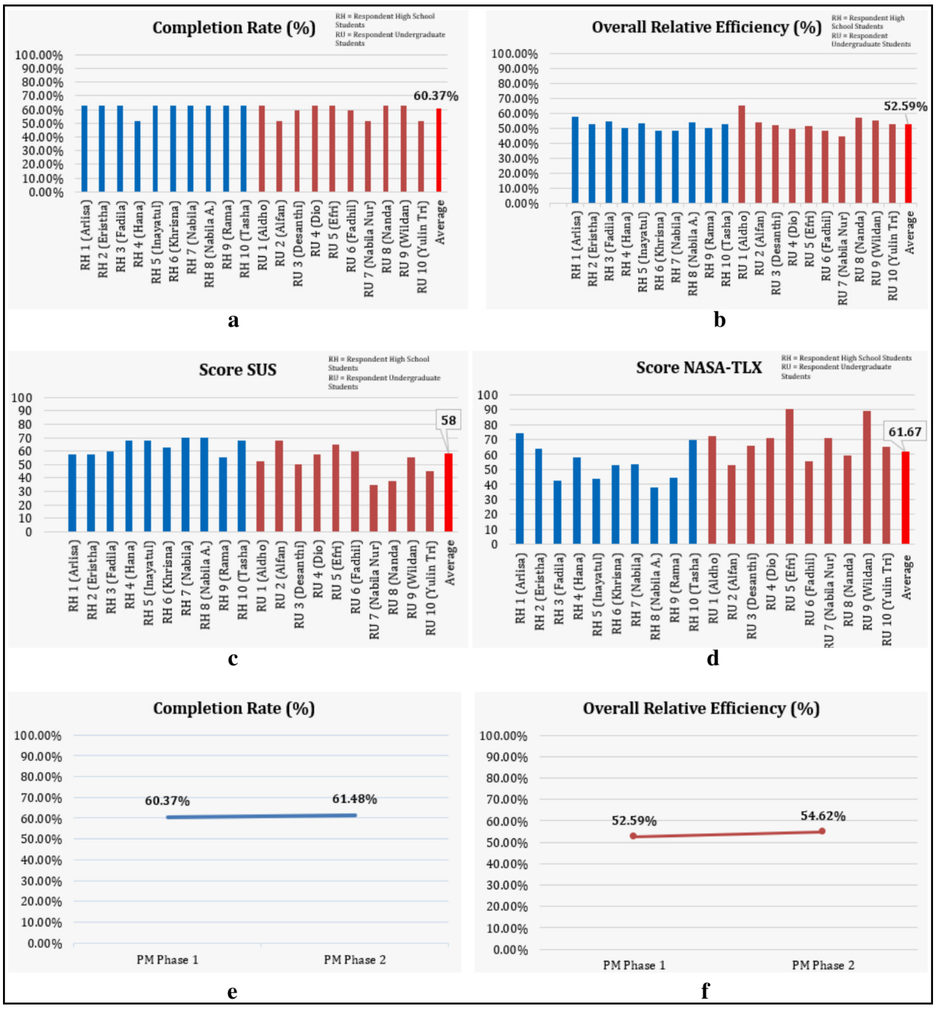

Fig. 2. The result of this research (a) Effectiveness, (b) Efficiency, (c) Satisfaction, (d) Cognitive Load, (e) Learnability, (f) Memorability

Figure 2(a) shows results obtained from 20 respondents using the completion rate calculation in 60.37%, so the effectiveness aspect is below standard [29]. Figure 2(b) shows a result obtained from 20 respondents using the calculation of overall relative efficiency in 52.59%. Figure 2(c) shows a result obtained from 20 respondents using the calculation of SUS questionnaire in 58, so the satisfaction aspect is below standard [30]. Figure 2(d) shows a result obtained from 20 respondents using the calculation of NASA-TLX questionnaire in 61.67, so the cognitive load aspect is moderate [31]. Figure 2(e) shows results obtained from 20 respondents by comparing the observed effectiveness values of PM Phase 1 and PM Phase 2 in 60.37% and 61.48%, respectively. So the learnability aspect has increased well. Figure 2(f) shows results obtained from 20 respondents by comparing the overall relative effectiveness and efficiency values obtained from PM Phase 1 and PM Phase 2 in 52.59% and 54.62%, respectively. So the memorability aspect has increased well.

Furthermore, the error aspect is calculated using the error rate equation based on the number of errors in PM Phase 1. The results obtained from 20 respondents are 14.27%.

Based on usability testing using the PACMAD model carried out at the Pre-Evaluation step, it can be concluded that the ArabEasy application has good values in the aspects of learnability and memorability. However, effectiveness, efficiency, satisfaction, errors, and cognitive load are still far below the standard. So based on the results of the Pre-Evaluation, the ArabEasy application still needs UI improvements on some features to make it easier for users to use.

3.1 Mapping problem solved

Table 8 shows the results of the data recapitulation in the RTA questionnaire. In general, respondents stated that several features did not work correctly: login, create an account, learning, hearing, writing, and practice questions features. This research conducted a mapping to overcome the existing problems based on these problems.

Table 8. Mapping problem solved

No.	Pages or Features	Problems or Suggestions
1.	Login	Failure often occurs when a user inputs the password because the login page does not have a visible eye icon.
2.	Create an account	The interface design looks less attractive and stiff. Can add background color and enlarge user data entry fields for easier understanding.
		Failure often occurs when the user inputs the password because the create account page has no visible eye icon.
3.	Learning	The problem lies in the font box being too small, so it is difficult for users to understand.
		Failures often occur when the user is unable to make a sound. This failure is because when the user wants to make a sound, the user must first click allow.
4.	Hearing	The problem lies in the font box being too small, so it is difficult for users to understand.
		Failures often occur when the user is unable to make a sound. This failure is because when the user wants to make a sound, the user must first click allow.
5.	Writing	The problem lies in the font box being too small, so it is difficult for users to understand.
		Failures often occur when the user is unable to make a sound. This failure is because when the user wants to make a sound, the user must first click allow.
6.	Practice questions	The problem lies in the font box being too small, so it is difficult for users to understand.
		In the practice questions feature, it is better to add the correct answer to determine where the user's response is wrong.

3.2 Improvement recommendations

The UI design improvement recommendations from the login page, create an account, learning, hearing, writing, practice questions show in Figures 3 and 4. Figure 3(a,b) shows the login page before and after improvements in ArabEasy apps. There was a problem on the login page with some users failing to log in because the user entered the wrong password. This issue can waste user time on login, affect efficiency and errors using the ArabEasy app. The solution that can solve this problem is to implement the 4 Golden Rules of UI Design (rule number 1): placing the user as the controller of the interface in providing an informative feedback section [32]. By adding a visible option or an eye icon on the login page, the user is no longer entering a password error when logging in.

Figure 3(c,d) shows creating an account page before and after improvements in ArabEasy apps. There was a problem with the create an account page. Namely, some users were confused because the navigation title on the create an account page was unclear, so users did not know where they were, and the password input was not substantial. So this impacts user error and the level of efficiency in using the ArabEasy apps. The solution that can solve this problem is to implement the 4 Golden Rules of UI Design (rule number 1): placing the user as the controller of the interface in creating an easy-to-navigate interface and providing an informative feedback section [32]. By changing the navigation title on the create an account page and adding an option or eye icon visible on the create an account page. So that users are no longer confused to fill in data on the create an account page.

Figure 3(e,f,g,h) shows the learning and hearing pages before and after improvements in ArabEasy apps. There is a problem with the learning and hearing pages. Namely, frequent failures when the user is unable to make a sound. This failure is because when the user wants to make a sound, the user must first click allow, and the Arabic font box is too small. The solution that can solve this problem is implementing the 4 Golden Rules of UI Design (rules number 2, 3, and 4): making it comfortable for a user to interact with a product, reduce cognitive load, and create consistent user interfaces [32]. Changing the size of the Arabic font box to be significant makes it easier to click. Then add audio into the application without having to click allow to bring up the sound to make it easier for users to listen to Arabic sentences.

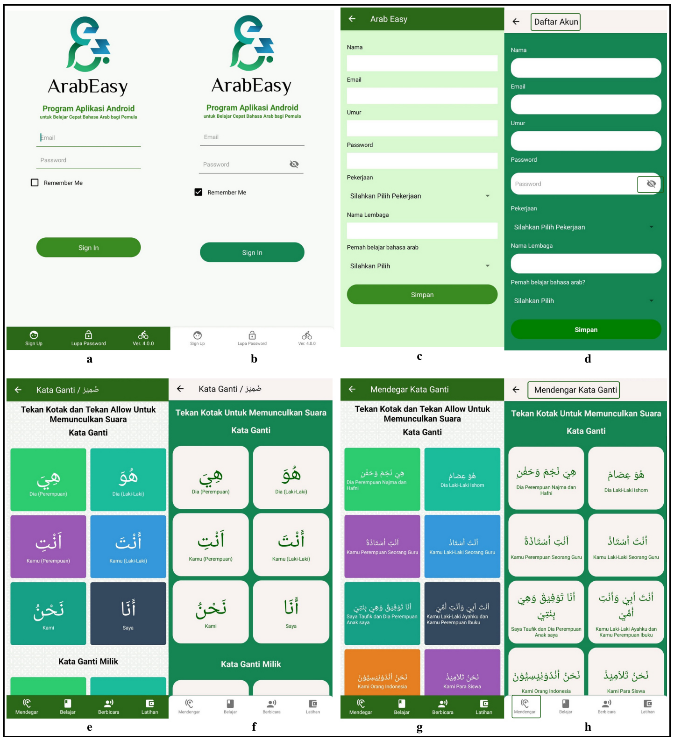

Fig. 3. The result of this research (a) before improvement login page, (b) after improvement, (c) before improvement create an account page, (d) after improvement, (e) before improvement learning page, (f) after improvement, (g) before improvement hearing page, (h) after improvement

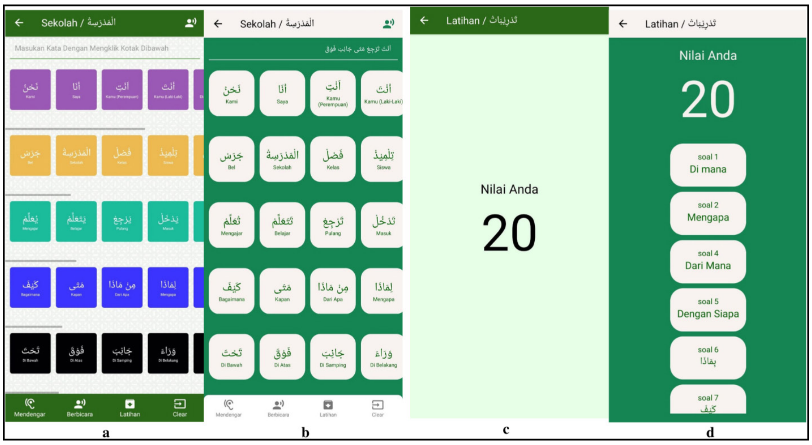

Fig. 4. The result of this research (a) before improvement writing page, (b) after improvement, (c) before improvement practice questions page, (d) after improvement

Figure 4(a,b) shows the writing page before and after improvements in ArabEasy apps. There is a problem with the writing page. Namely, frequent failures when the user is unable to make a sound. This failure is because when the user wants to make a sound, the user must first click allow, and the Arabic font box is too small. The solution that can solve this problem is implementing the 4 Golden Rules of UI Design (rules number 2, 3, and 4): making it comfortable for a user to interact with a product, reducing cognitive load, and making user interfaces consistent [32]. Changing the size of the Arabic font box to be significant makes it easier to click. Then add audio into the application without having to click allow to bring up the sound to make it easier for users to listen to Arabic sentences. Figure 4(c,d) shows the practice questions page before and after improvements in ArabEasy apps. There is a problem with the practice questions page. Some users are confused because the navigation title on the practice question page is unclear, so users do not know where they are. In addition, users recommend adding the correct answer to determine where the user's response is wrong in the practice questions feature. The solution that can solve this problem is to implement the 4 Golden Rules of UI Design (rules number 1, 2, and 4): placing the user as the controller of the interface, making it comfortable for a user to interact with a product, and create user interfaces consistent [32]. By changing the navigation title on the practice questions page and changing the size of the Arabic font box to be significant, it is easier to click. Then add the correct answer to determine where the user's response is wrong in the practice questions feature.

3.3 Post-evaluation

Table 9 shows the results of the post-evaluation. The flow of this test is carried out as in the pre-evaluation using the calculation technique Usability metric, SUS questionnaire, and NASA-TLX questionnaire. After post-evaluation, there is a significant increase in seven aspects of the PACMAD model.

Table 9. Post-evaluation

Usability Factor	Aspect Usability Model PACMAD	Calculation Techniques	Data Analysis Results
User	Effectiveness	Completion Rate	97.59%
	Efficiency	Overall Relative Efficiency	98.67%
Task	Satisfaction	SUS Questionnaire	82.25
	Learnability	Compare the effectiveness values (PM Phase 1 and PM Phase 2)	Good
	Memorability	Compare the overall relative effectiveness and efficiency values (PM Phase 1 and PM Phase 2)	Good
Context of Use	Errors	Error Rate	0.87%
	Cognitive Load	NASA-TLX Questionnaire	43.13

3.4 Pre-evaluation and post-evaluation comparison result

Table 10 shows the comparison between pre-evaluation and post-evaluation. This process showed increased effectiveness 37.22%, efficiency 46.08%, satisfaction 24.25. Furthermore, aspect error decreased 13.4%, and the cognitive load 18.54.

Table 10. Pre-evaluation and post-evaluation comparison result

Usability Factor	Aspect Usability Model PACMAD	Pre-Evaluation	Post-Evaluation
User	Effectiveness	60.37% (below standard)	97.59% (satisfy standard)
	Efficiency	52.59%	98.67%
Task	Satisfaction	58 (below standard)	82.25 (satisfy standard)
	Learnability	Good	Good
	Memorability	Good	Good
Context of Use	Errors	14.27%	0.87%
	Cognitive Load	61.67 (moderate load)	43.13 (relatively light load)

4 Discussion and conclusion

Performing a PACMAD usability model on a mobile app has an important role. This model focuses on the importance of cognitive load as usefulness. The aspects of effectiveness, efficiency, satisfaction, and error in the PACMAD model play an important role in measuring the effectiveness and interest of users when using mobile applications. In addition, the Learnability and Memorability aspects are helpful to find out how much the user's ability and memory power in interacting with mobile applications.

So in this research, testing the mobile application (ArabEasy) using the PACMAD model. This study carried out two stages to test the ArabEasy application, namely Pre-Evaluation and Post-Evaluation. These two stages use the same data collection techniques, calculation techniques, and data analysis. The results obtained in the Pre-Evaluation are 60.37% effectiveness, 52.59% efficiency, 58 satisfaction, 14.27% error, 61.67 cognitive loads, learnability and memorability are good. While the post-evaluation effectiveness is 97.59%, efficiency is 98.67%, satisfaction is 82.25, an error is 0.87%, cognitive load is 43.13, learnability and memorability are good.

Based on the results obtained, it can be concluded that there is a significant improvement after making recommendations for UI improvement, becoming a prototype on the login page, creating an account, learning, hearing, writing, and practicing questions. This process showed an increase in the effectiveness of 37.22%, efficiency 46.08%, satisfaction 24.25%. Furthermore, the error aspect decreased by 13.4%, and cognitive load by 18.54%.

Furthermore, in the statistical tests on pre-evaluation results, there are significant differences between respondents regarding aspect efficiency, memorability, satisfaction, and cognitive load. In addition, there is no significant difference between respondents in terms of effectiveness, learnability, and errors. While in the post-evaluation, there is no significant difference between the respondents on the seven aspects.

Based on this explanation, after recommendations for UI improvements were made, becoming a prototype using The 4 Golden Rules of User Interface Design, there were significant changes to the ArabEasy application UI. Previously users had trouble logging in, creating accounts, and clicking sounds. After the recommended UI improvements, the user has no problem logging in, creating an account, and clicking sound so that the UI improvement recommendations can be considered for improving the UI of ArabEasy apps users in the future.

5 References

[1] I. Al-huri, "Arabic Language : Historic and Sociolinguistic Characteristics," *English Lit. Lang. Rev.*, vol. 1, no. 4, pp. 28–36, 2015, http://dx.doi.org/10.13140/RG.2.2.16163.66089/1

[2] "Indonesia Population 2021 (Demographics, Maps, Graphs)," Worldpopulationreview, 2021. https://worldpopulationreview.com/countries/indonesia-population (accessed Dec. 21, 2021).

[3] F. Aladé, A. R. Lauricella, L. Beaudoin-Ryan, and E. Wartella, "Measuring with Murray: Touchscreen technology and preschoolers' STEM learning," Comput. Human Behav., vol. 62, pp. 433–441, 2016, https://doi.org/10.1016/j.chb.2016.03.080

[4] S. Papadakis and M. Kalogiannakis, "Mobile educational applications for children. What educators and parents need to know," Int. J. Mob. Learn. Organ., vol. 11, no. 2, p. 1, 2017, https://doi.org/10.1504/IJMLO.2017.10003925

[5] S. Papadakis, M. Kalogiannakis, and N. Zaranis, "Educational apps from the Android Google Play for Greek preschoolers: A systematic review," Comput. Educ., vol. 116, pp. 139–160, 2018, https://doi.org/10.1016/j.compedu.2017.09.007

[6] H. Asrohah, M. Khusnu Milad, A. T. Wibowo, and E. I. Rhofita, "Improvement of Academic Services using Mobile Technology based on Single Page Application," Telfor J., vol. 12, no. 1, pp. 62–66, 2020, https://doi.org/10.5937/telfor2001062A

[7] N. Hamzah, N. Dayana, A. Halim, M. H. Hassan, and A. Ariffin, "Android Application for Children to Learn Basic Solat The Approach of Behaviourism Learning Theory," vol. 13, no. 7, pp. 69–79, 2019, https://doi.org/10.3991/ijim.v13i07.10758

[8] J. Boilevin and K. Ravanis, "L ' éducation scientifique et technologique à l ' école obligatoire face à la désaffection: recherches en didactique, dispositifs et références," no. January, 2007.

[9] T. Mooij, "Design of educational and ICT conditions to integrate differences in learning: Contextual learning theory and a first transformation step in early education," Comput. Human Behav., vol. 23, no. 3, pp. 1499–1530, 2007, https://doi.org/10.1016/j.chb.2005.07.004

[10] S. Papadakis, M. Kalogiannakis, and N. Zaranis, "Teaching mathematics with mobile devices and the Realistic Mathematical Education (RME) approach in kindergarten," Adv. Mob. Learn. Educ. Res., vol. 1, no. 1, pp. 5–18, 2021, https://doi.org/10.25082/AMLER.2021.01.002

[11] C. Quinn, "mLearning: Mobile, Wireless, In-Your-Pocket Learning," 2001.

[12] E. Dolzhich, S. Dmitrichenkova, and M. K. Ibrahim, "Using M-Learning Technology in Teaching Foreign Languages: A Panacea during COVID-19 Pandemic Era," Int. J. Interact. Mob. Technol., vol. 15, no. 15, pp. 20–34, 2021. https://doi.org/10.3991/ijim.v15i15.22895

[13] S. Eltalhi, H. Kutrani, R. Imsallim, and M. Elrfadi, "The Usability of BenKids Mobile Learning App in Vocabulary Teaching for Preschool," Int. J. Interact. Mob. Technol., vol. 15, no. 24, pp. 4–18, 2021, https://doi.org/10.3991/ijim.v15i24.22237

[14] J. Nielsen, Usability Engineering. USA: United States of America: Academic Press, 1993. https://doi.org/10.1016/B978-0-08-052029-2.50007-3
[15] J. Nielsen, "Usability 101: Introduction to Usability," Nielsen Norman Group, 2012. https://www.nngroup.com/articles/usability-101-introduction-to-usability/ (accessed Feb. 27, 2021).
[16] V. Saswadkar, V. Paygude, and P. D. | P. G. | R. Gaikwad, "Pet Care System Based On Android Application," Int. J. Trend Sci. Res. Dev., vol. Volume-2, no. Issue-4, pp. 1488–1491, 2018, https://doi.org/10.22214/ijraset.2018.3296
[17] N. A. N. Ahmad, N. I. M. Hamid, and A. M. Lokman, "Performing Usability Evaluation on Multi-Platform Based Application for Efficiency, Effectiveness and Satisfaction Enhancement," Int. J. Interact. Mob. Technol., vol. 15, no. 10, pp. 103–117, 2021, https://doi.org/10.3991/ijim.v15i10.20429
[18] R. Harrison, D. Flood, and D. Duce, "Usability of mobile applications : literature review and rationale for a new usability model," pp. 1–16, 2013, https://doi.org/10.1186/2194-0827-1-1
[19] A. Saleh, R. B. Isamil, and N. B. Fabil, "EXTENSION OF PACMAD MODEL FOR USABILITY EVALUATION METRICS USING GOAL QUESTION METRICS (GQM) APPROACH," vol. 79, no. 1, pp. 90–100, 2015.
[20] R. Alturki and V. Gay, "Usability Testing of Fitness Mobile Application : Methodology and Quantitative Results," pp. 97–114, 2017, https://doi.org/10.5121/csit.2017.71108
[21] F. A. Kasali, O. O. Taiwo, I. O. Akinyemi, and O. B. Alaba, "An Enhanced Usability Model for Mobile Health Application," Int. J. Comput. Sci. Inf. Secur., vol. 17, no. 2, pp. 20–29, 2019.
[22] Z. Sharfina and H. B. Santoso, "An Indonesian adaptation of the System Usability Scale (SUS)," 2016 Int. Conf. Adv. Comput. Sci. Inf. Syst. ICACSIS 2016, pp. 145–148, 2016, https://doi.org/10.1109/ICACSIS.2016.7872776
[23] M. H. Afif, "Evaluating PSAU Mobile Application Based on People at the Center of Mobile Application Development (PACMAD) Usability Model: Empirical Investigation," J. Comput. Sci., vol. 17, no. 3, pp. 275–283, 2021, https://doi.org/10.3844/jcssp.2021.275.283
[24] S. Papadakis, "Advances in Mobile Learning Educational Research (A.M.L.E.R.): Mobile learning as an educational reform," Adv. Mob. Learn. Educ. Res., vol. 1, no. 1, pp. 1–4, 2021, https://doi.org/10.25082/AMLER.2021.01.001
[25] P. Stamatios, A. Foteini, and Z. Nikolaos, Mobile device use among preschool-aged children in Greece, no. 0123456789. Springer US, 2021.
[26] L. Faulkner, "Beyond the five-user assumption: Benefits of increased sample sizes in usability testing," Behav. Res. Methods, Instruments Comput., vol. 35, no. 3, pp. 379–383, 2003. https://doi.org/10.3758/BF03195514
[27] J. Mifsud, "Usability Metrics - A Guide To Quantify The Usability Of Anys System," 2011, https://usabilitygeek.com/usability-metrics-a-guide-to-quantify-system-usability/ (accessed Dec. 21, 2021).
[28] N. Thomas, "How To Use The System Usability Scale (SUS) To Evaluate The Usability Of Your Website," 2016, https://usabilitygeek.com/how-to-use-the-system-usability-scale-sus-to-evaluate-the-usability-of-your-website/ (accessed Dec. 21, 2021).
[29] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Result of Empirical and Theoretical Research," Hum. Ment. Workload. Adv. Psychol., no. 52, pp. 139–183, 1988, https://doi.org/10.1016/S0166-4115(08)62386-9
[30] N. Babich, "The 4 Golden Rules of UI Design," Adobe, 2019. https://xd.adobe.com/ideas/process/ui-design/4-golden-rules-ui-design/ (accessed Mar. 03, 2021).

6 Authors

Safitra Aditya Bagaskara is a bachelor's degree from the Islamic State University of Sunan Ampel Surabaya, Indonesia. His current interest in Computer Science and Technology concentrates more on Designing User interfaces and User Experience Designs.

Achmad Teguh Wibowo received a bachelor's degree in Information systems from STIKOM, Surabaya, Indonesia, in 2010. He received his Master of Electrical Engineering from Brawijaya University, Malang, Indonesia, in 2013. He joined the Information System Department as a lecturer at the Islamic State University of Sunan Ampel, Surabaya, Indonesia, in 2014. His current interest is Blockchain and Intelligent systems. He is currently pursuing a Ph.D. degree at Institut Teknologi Sepuluh November (ITS), Surabaya, Indonesia, since 2019.

Muhammad Andik Izzuddin received a bachelor's degree in Information and Technology Education from University Negeri Malang (UM), Malang, Indonesia, in 2011. He received his Master of Electrical Engineering from the Insitute of Technology Bandung (ITC), Bandung, Indonesia, in 2013. He joined the Information System Department as a lecturer at the Islamic State University of Sunan Ampel, Surabaya, Indonesia, in 2014. His current interests are Computer Networks and the Internet of Things.

Mohammad Khusnu Milad received a bachelor's degree in Electrical Engineering, Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia, in 2003. He received his Master of Industrial Managament from Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indone-sia, in 2006. He joined the Information System Department as a lecturer at the Islamic State University of Sunan Ampel, Surabaya. Indonesia since 2014. His current interests are Organization and Management. He is currently pursuing a Ph.D. degree at Management Science Sekolah Tinggi Ilmu Manajemen Indonesia (STIESIA), since 2020.

Rr. Diah Nugraheni Setyowati received a bachelor's degree in Civil Engineering from Diponegoro University, Semarang, Indonesia, in 2006. She received her Master of Civil Engineering from Diponegoro University, Semarang, Indonesia, in 2012. She joined the Environmental Engineering Department as a lecturer at the Islamic State University of Sunan Ampel, Surabaya. Indonesia since 2014. Her current interests are Water Resources.

Taufik, Lecturer at the Faculty of Education and Teacher Training of state Islamic university sunan ampel Surabaya. The bachelor's degree was obtained from the Department of Arabic education in 2000 and continued his studies in his first master's in Islamic education graduated in 2004, and the second master's study in teacher education Madrasah Ibtidaiyah graduated in 2010. The Doctorate is obtained in 2021 in media and Arabic language learning at the same campus. He is also a consultant and developer of the ArabicEasy application to learn Arabic fast for beginners.

Mujib Ridwan received a bachelor's degree in Information Technology from Islamic State University of Maulana Malik Ibrahim, Malang, Indonesia, in 2009. He received his Master of Electrical Engineering from Brawijaya University, Malang, Indonesia, in 2013. He joined the Information System Department as a lecturer at the Islamic State

University of Sunan Ampel, Surabaya. Indonesia since 2014. His current interest is in Data Mining and Deep Learning.

Sirajul Arifin is an Associate Professor in the Faculty of Islamic Economics and Business UIN Sunan Ampel Surabaya. He completed both bachelor's degrees at IAI Ibrahimy Situbondo and the University of Indonesia Jakarta. The first degree is in Islamic Economic Law, whereas the second is in Information Sciences. In 2002 he obtained a Master of Islamic Economics from IAIN Sunan Ampel Surabaya. Six years later, he pursued a Ph.D. degree in Islamic Economics at UIN Sunan Kalijaga Yogyakarta and completed it in 2014. Now he serves as Vice Dean for Academic and Institutional Affairs at the Faculty of Islamic Economics and Business UIN Sunan Ampel Surabaya.

Article submitted 2021-11-29. Resubmitted 2021-12-29. Final acceptance 2021-12-29. Final version published as submitted by the authors.