

The Evaluation of AR Mobile App as a Learning Media for Children

Adhistya Erna Permanasari¹, Indriana Hidayah², Faizal M. Priyowibowo³, M. Arifin Hidayat⁴, Fachrul Budi Prayoga⁵, Intan Sulistyanningrum Sakkinah⁶

Abstract—A user evaluation stage is an integral part of designing an application. A practical evaluation can provide an overview of the problems that arise in the application and improve the user experience. The Kupuku application is an augmented reality (AR)-based game application for learning about butterflies. The Kupuku application is specifically intended for children aged 6-13 years. The user sample was selected using a purposive sampling method with the criteria for users of elementary school-age children for the child user segment and their companions as the adult user segment. This study aims to evaluate the usability of the Kupuku game application to users. User evaluation was carried out to measure the application's usability. The evaluation process was conducted on two user segments, namely 20 child users and 16 adult users. Assessment of children employed the Fun Toolkit and usability factor-based question - Nielsen method. The obtained results showed positive feedbacks. In contrast, the assessment for adult users utilized the system usability scale (SUS) and the user experience questionnaire (UEQ). The SUS score of 76 was included in the good category, and the UEQ score produced an excellent average. The test results indicate that this application can be accepted by users, both children, and adults.

Keywords—Augmented Reality, Learning Media, Fun Toolkit, Usability, User Experience.

I. INTRODUCTION

The development of mobile applications is currently proliferating. It needs to go together with the need to evaluate existing applications. Measurement of the mobile application utilization from the user's point of view is essential to obtain the advantages of the developed applications [1], [2]. The obtained results can describe how well users can accept an application.

The Kupuku application is an augmented reality (AR)-based game application for learning about butterflies. Many interactive learning media have been developed to support the learning process, including the AR technology support. AR is a technology that combines two-dimensional (2D) or three-dimensional (3D) virtual objects and then applies them to real objects. The Kupuku application is developed on the Android platform. This application was created to help the learning

process at the Alian Kebumen Butterfly Park. This place is a butterfly-breeding-education-based garden. The Kupuku application is intended as a learning medium about butterflies, especially for elementary school children aged 6-13 years. The focus of this study is the testing of application users to measure the application's usability.

The evaluation method for Kupuku users was divided into two segments, namely, child users and adult users. Testing on child users was conducted using the Fun Toolkit and usability factor-based questions. The Fun Toolkit is a survey instrument designed to help researchers and developers collect children's opinions about using technology from children [3]. Testing for adult users was conducted using system usability scale (SUS) and user experience questionnaire (UEQ). The use of SUS and UEQ was based on the need for product evaluation in terms of user experience and usability [4], [5]. The user sample was selected using a purposive sampling method. The criteria included users of elementary school-age children as the child user segment, while their companions as the adult user segment.

This paper is structured in the following way. Part II reviews some supporting literature studies. Part III presents the Kupuku Application. The next part, Part IV, presents Research Methods. Part V reports Research Results. Furthermore, the Conclusions are presented in Part VI.

II. USER EVALUATION

Evaluation is something important needed by an application to determine the effectiveness of the application or media that has been produced. Evaluation can take various forms and aspects. Reference [6] developed a game application that was evaluated using evidence-collecting evaluation or heuristic evaluation. This evaluation was developed to assess the effectiveness level of an application, which was mainly on the usability aspect [6].

Usability testing is defined as a systematic testing method that observes actual users who try a product and collect information whether the product used is easy or complex [7]. The roots of usability testing lie firmly in experimental methods of psychology (particularly cognitive and applied psychology) and human factors engineering; in addition, they are closely related to the concept of iterative design [8].

Evaluation using SUS and UEQ is a usability evaluation that can be applied to various platforms. Reference [9] implemented a set of usability tests on the website and obtained reliable results. In practice, there are many ways to collect data, including interviews and distributing questionnaires, combined with scenarios described by the team.

^{1,2,3,4,5,6} Department of Electrical and Information Engineering, Faculty of Engineering, Universitas Gadjah Mada, Jln. Grafika No. 2, Kampus UGM, Yogyakarta 55281 INDONESIA (phone: 0274-552305; email: ¹adhistya@ugm.ac.id, ²indriana.h@ugm.ac.id, ³faizal.m.priyowibowo@mail.ugm.ac.id, ⁴m.arifin@mail.ugm.ac.id, ⁵fachrul.b@mail.ugm.ac.id, ⁶intan.sakkinah@mail.ugm.ac.id)



Fig. 1 Interface of the Kupuku application.

Furthermore, this session discusses the evaluation of games for children using a comparison of two application methods, namely Fun Toolkit and This or That. Reference [10] discussed the use of subjective user experience evaluation and cannot be measured by traditional measurement methods. In [10], there was a summary of the use and comparison of the Fun Toolkit, a group of tools used to evaluate the children's experience as the user during, before, and after using the application, compared to the This or That evaluation method, which would test and focus on the reliability of the collected responses.

Reference [11] evaluated the web-based academic progress information system (SIsKA), which was carried out using the UEQ. With an average result of 1.59, the benchmark results showed that the SIsKA was categorized into a good category, and of course, there had to be improvements to make the system more perfect.

Reference [12] conducted a study to evaluate the web's usability, namely clinical guidelines published on the web. Stages in conducting the research included selecting participants and studying scenarios and implementing this usability test itself so that when the application was implemented, it would run smoothly. In addition, this test employed SUS as an evaluation tool.

The use of evaluation can also be applied in assessing information systems. Reference [13] discussed the usability evaluation of the labor market information system. This evaluation employed the SUS to assess the usability value in the information system [13]. The research instrument of this evaluation method was using the Google Form, and the prospective respondents were users in the job market who obtained information about this distribution via email and cellphone numbers. The results showed that there was a part of usability that had a low value. Therefore, it can be concluded that there is ineffectiveness in the system's usability value.

Reference [14] conducted a study on usability evaluation conducted on the computer brain interface. This study explained that several aspects became a benchmark in a usability assessment, including ease of use, consistency, accuracy. This study assessed an interface system, which explained that it was not only applications that could be assessed for system usability.

Reference [15] in 2017 explained that SUS could be comprehensively utilized to evaluate the use of a ticket vending

machine from a commuter line (C-VIM) in Jakarta at Pondok Cina Station, Depok. This test obtained that plus and minus results were still contained in a system that had even been used massively and en masse. One of which was used was SUS to test the usability function of C-VIM.

In previous studies, in evaluating a system, the most widely used instruments were SUS, UEQ to measure the system's usability. In contrast, the Fun Toolkit and This or That instruments were widely used considering that the instrument filling was easier to understand if the study used children as respondents. Of the various uses of this evaluation instrument, no one has applied it to evaluate game applications with AR technology. Thus, in this study, these instruments to evaluate the usability of the Kupuku application as an AR-based game application were employed.

III. THE KUPUKU APPLICATION

The Kupuku application is an educational game application that aims to introduce butterflies to children. The introduced butterflies are packed into the game. Fig. 1 shows the main menu display in the Kupuku application. In the application, there are three main menus, namely Quest, Quiz, and Card. The Quest menu is a game menu to collect butterfly stickers that have been placed in several locations in the Alian Butterfly Park. The Quiz menu will display questions related to butterflies, such as butterfly metamorphosis, which is packaged in the form of a quiz. The third menu is Card. The Card menu is a menu displaying a collection of butterfly cards that have been successfully collected. In this menu, users can see pictures of butterflies and their scientific names in 3D with AR technology.

IV. RESEARCH METHODS

A. Evaluation Scenario

The testing of the Kupuku application's users was carried out online due to limited interaction restrictions during the COVID-19 pandemic. Fig. 2 is a scenario from live testing, presented in the form of a flow chart that have been conducted, starting with communication through the companion (parent/guardian of the child). Subsequently, the companion and the child tried the Butterfly application together, and finally, they assessed it through the questionnaire provided. Companion represented adult users.

In this study, the target respondents were children aged 6-13 years and adults as children's companions. The selection of children respondents took into account the purpose of developing the Kupuku application, which was developed to educate children and introduce butterflies. By adding adult respondents, it would increase the perspective of application evaluation.

B. Fun Toolkit

The Fun Toolkit consists of four specialized tools, Smileyometer, Funometer, Again-Again Table, and Fun Sorter. It also supports the idea of measuring memory and using video footage to assess engagement. In this research, the utilized tools were Smileyometer and Again-Again Table.

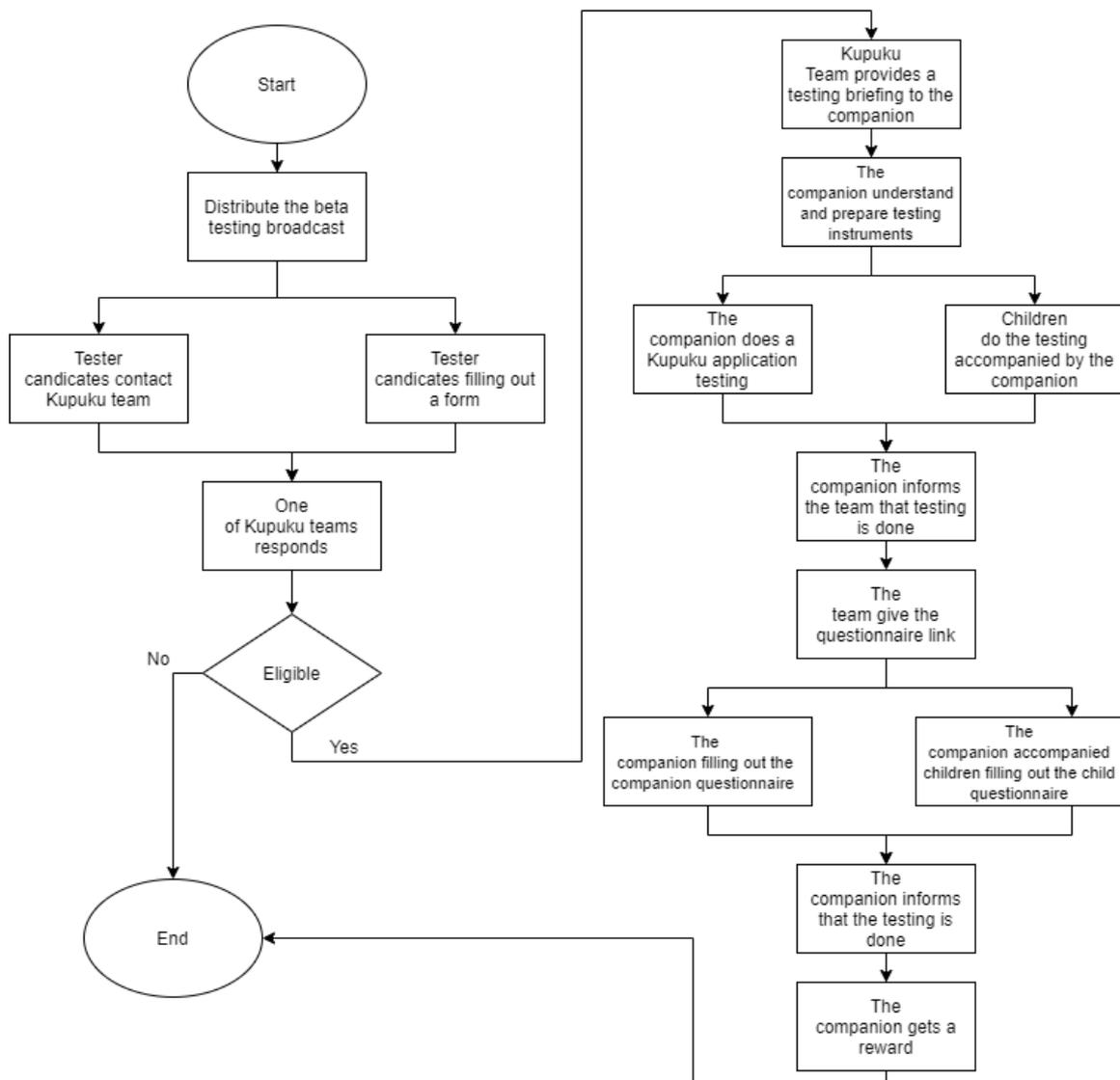


Fig. 2 Kupuku application evaluation scenario flowchart.



Fig. 3 A Smileyometer.

1) *Smileyometer*: It is one of the methods developed to measure usability, directly measured to children. In the Smileyometer, the measured rate is mood change, and it is measured numerically [16] as seen in Fig. 3. Number 1 for very sad, 2 for sad, 3 for normal, 4 for happy, and 5 for very happy.

2) *Usability Factor-Based Question*: In the Nielsen usability check, the testing on children with questions reflecting five usability factors were utilized. A Likert scale was employed, this scale used five levels including a strongly agree answer was scored 5, an agree answer was scored 4, a neutral answer was scored 3, a disagree answer was scored 2, and a

strongly disagree answer was scored 1. Questions asked to children were for the followings.

- Measuring whether the Kupuku game was easy to learn (learnability) for the user so that there were no obstacles when playing. This aspect was represented in questions number 1-2.
- Measuring whether the child played efficiently, i.e., giving less effort, and producing the expected results (efficiency). This aspect was represented in question number 3.
- Measuring how easy the application was to remember (memorability), including features, feature locations, or available buttons. This aspect was represented in questions number 4-5.
- Measuring whether there was an error in the Kupuku system and an error that did not occur as it should (error). This aspect was represented in question number 6.

TABLE I
SUS QUESTIONS

| No | Questions |
|----|--|
| 1 | I think I will use this system again. |
| 2 | I find this system is complicated to use. |
| 3 | I find this system is easy to use. |
| 4 | I need help from someone else or a technician in using this system. |
| 5 | I feel the features of this system are working properly. |
| 6 | I feel there are a lot of things inconsistent (incompatible with this system). |
| 7 | I feel like other people will understand how to use this system quickly. |
| 8 | I find this system confusing. |
| 9 | I feel there are no obstacles in using this system. |
| 10 | I need to get used to it. |

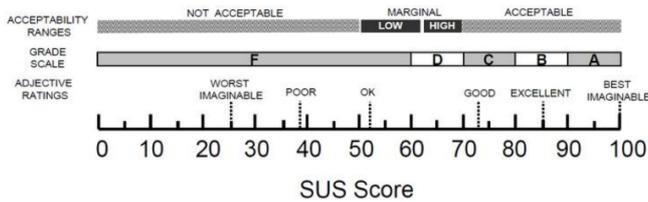


Fig. 4 Basis of SUS assessment.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | |
|--------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|----------------------------|----|
| annoying | <input type="radio"/> | enjoyable | 1 |
| not understandable | <input type="radio"/> | understandable | 2 |
| creative | <input type="radio"/> | dull | 3 |
| easy to learn | <input type="radio"/> | difficult to learn | 4 |
| valuable | <input type="radio"/> | inferior | 5 |
| boring | <input type="radio"/> | exciting | 6 |
| not interesting | <input type="radio"/> | interesting | 7 |
| unpredictable | <input type="radio"/> | predictable | 8 |
| fast | <input type="radio"/> | slow | 9 |
| inventive | <input type="radio"/> | conventional | 10 |
| obstructive | <input type="radio"/> | supportive | 11 |
| good | <input type="radio"/> | bad | 12 |
| complicated | <input type="radio"/> | easy | 13 |
| unlikable | <input type="radio"/> | pleasing | 14 |
| usual | <input type="radio"/> | leading edge | 15 |
| unpleasant | <input type="radio"/> | pleasant | 16 |
| secure | <input type="radio"/> | not secure | 17 |
| motivating | <input type="radio"/> | demotivating | 18 |
| meets expectations | <input type="radio"/> | does not meet expectations | 19 |
| inefficient | <input type="radio"/> | efficient | 20 |
| clear | <input type="radio"/> | confusing | 21 |
| impractical | <input type="radio"/> | practical | 22 |
| organized | <input type="radio"/> | cluttered | 23 |
| attractive | <input type="radio"/> | unattractive | 24 |
| friendly | <input type="radio"/> | unfriendly | 25 |
| conservative | <input type="radio"/> | innovative | 26 |

Fig. 5 Components of UEQ questions.

- Measuring whether the Kupuku application generated user satisfaction, this aspect was represented in questions 7-9.

3) *Again-Again Table*: Again-Again Table was one of the instruments in the Fun Toolkit that was used to find out the

TABLE II
AR GAME SMILEYOMETER TEST RESULTS

| Smileyometer | Before Playing | After Playing |
|--------------|----------------|---------------|
| Very Sad | 0 | 0 |
| Sad | 0 | 0 |
| Moderate | 12 | 1 |
| Happy | 5 | 12 |
| Very Happy | 3 | 7 |

TABLE III
KUPUKU APP SMILEYOMETER TESTING ANALYSIS

| Smileyometer | Before Playing | | After Playing | |
|--------------|----------------|-------------|---------------|-------------|
| | Mean | SD | Mean | SD |
| Kupuku | 3.55 | 0.759154655 | 4.3 | 0.571240571 |

TABLE IV
ANALYSIS OF USABILITY FACTOR-BASED QUESTION METHOD TESTING OF KUPUKU APPLICATION

| Usability Test Results for the Child | | | | |
|--------------------------------------|-------------------|-----------|--------------------|--------|
| Aspect | Number of Samples | Average | Standard Deviation | Median |
| Learnability | 20 | 4.525 | 0.640012 | 5 |
| Efficiency | 20 | 4.45 | 0.7591547 | 5 |
| Memorability | 20 | 4.45 | 0.7493587 | 5 |
| Error | 20 | 3.9 | 1.11192102 | 4 |
| Satisfaction | 20 | 4.5666667 | 0.620734 | 5 |
| Overall Measurement | 20 | 4.3783333 | 0.7565153 | 5 |

response from users. Again-Again Table focused on the question “do you want to use the feature again?” where each feature was asked.

C. *System Usability Scale (SUS)*

The SUS assessment component included ten questions consisting of five points ranging from agree to disagree, and the maximum score is 100 [17]. The questions from the SUS are as shown in Table I [18] with the basic assessment in Fig. 4.

D. *User Experience Questionnaire (UEQ)*

UEQ contains six basic aspects: attractiveness, clarity, efficiency, accuracy, stimulation, and novelty [19]. UEQ has 26 question components and seven answer choices. Fig. 5 is a component of UEQ questions [20].

V. RESEARCH RESULTS

Testing the Kupuku application was divided into two target users for testing, namely testing for elementary school children as the main target of our application and testing for companions from elementary school children to provide more detailed feedback.

In testing for elementary school children, who were accompanied by a companion, a data collection method in the form of questionnaire was used. The employed sampling was a purposive sampling method with the user criteria of 6-13 years, participated by 20 children. While testing for companions, 16 people taken as samples. It was because two companions were accompanying two children. The test results are as follows.

TABLE V
TESTING ANALYSIS OF AR GAME AGAIN-AGAIN TABLE

| Again-Again Table | | | |
|-------------------|--------|-------|---------------|
| Features | I Want | Maybe | I Do not Want |
| AR Game | 13 | 7 | 0 |
| Quiz | 15 | 5 | 0 |
| Card Collection | 16 | 4 | 0 |

TABLE VI
AR GAME SUS TEST RESULTS

| Respondents | Total | SUS Score |
|-------------|-------|-----------|
| 1 | 38 | 95 |
| 2 | 30 | 75 |
| 3 | 38 | 95 |
| 4 | 32 | 80 |
| 5 | 28 | 70 |
| 6 | 32 | 80 |
| 7 | 32 | 80 |
| 8 | 34 | 85 |
| 9 | 27 | 68 |
| 10 | 32 | 80 |
| 11 | 33 | 83 |
| 12 | 25 | 63 |
| 13 | 24 | 60 |
| 14 | 40 | 100 |
| 15 | 19 | 48 |
| 16 | 25 | 63 |
| Average | | 76 |

A. Fun Toolkit

1) *Smileyometer*: The Smileyometer test results in Table II were measured by filling out a questionnaire before and after playing the Kupuku application. The analysis of the Smileyometer test results based on the mean and standard deviation (SD) in Table III shows a change in the user's level of pleasure before and after using the application. Results show that there is an increase from an average of 3.55 (between somewhat likely and happy) to 4.3 (between happy and very happy).

2) *Usability Factor-Based Question*: The list of statements and the basis for the questions are as follow.

- The Kupuku app is relatively easy to play. This statement in the learnability aspect was made to measure how easy children learned the application.
- The buttons are pretty easy to find. This statement in the learnability aspect was made to measure how easy it was for children to memorize and learn the button locations.
- Child users are relatively fast in completing application commands one by one. The statement in the efficiency aspect was made to measure whether the application was efficient enough to be played by children.
- Child users do not forget where the button was. This statement was intended to assess the aspect of memorability, how far the child remembered the location of the existing buttons.

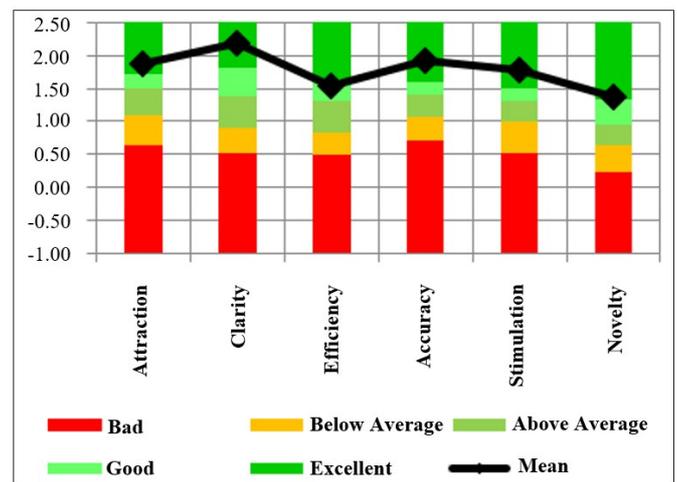


Fig. 6 Basics of Kupuku UEQ testing.

TABLE VII
AR GAME UEQ TESTING RESULTS

| Scale | Mean | Comparison to Benchmark |
|-------------|------------|-------------------------|
| Attraction | 1.86458333 | Excellent |
| Clarity | 2.171875 | Excellent |
| Efficiency | 1.53125 | Good |
| Accuracy | 1.921875 | Excellent |
| Stimulation | 1.765625 | Excellent |
| Novelty | 1.359375 | Excellent |

- Child users easily distinguish one feature from another. This question assessed the aspect of memorability, which assessed a child's memory to distinguish between features.
- There are no errors in the Kupuku application. This statement was to measure the error rate in the Kupuku application.
- Child users are satisfied. This statement was intended to measure the usability aspect number five, namely satisfaction.
- Child users benefit from AR games. This statement was intended to measure the usability aspect number five, namely satisfaction. Satisfaction can be seen from the benefits obtained, and it was measured by how satisfied with the benefits of AR game features.
- Child users obtain knowledge from the game quiz. This statement was intended to measure the usability aspect number five, namely satisfaction. Satisfaction can be seen from the benefits obtained, and it was measured by how satisfied with the benefits of AR game features.

In addition, the usability factor-based question method test results which is shown in Table IV show positive results in all aspects.

3) *Again-Again Table*: The results of the Again-Again Table test in Table V show that the feature receiving the most "want" responses to be played again is the card collection (Educard) feature, followed by the Quiz and AR Game (Quest) feature. Additionally, the test results prove that most users want to reuse it in the future, and no one does not want to reuse it.

B. System Usability Scale (SUS)

The SUS test, which was participated by 16 companions, give an average SUS score in Table VI of 76, which indicates that the application is in a good category. The visual results are shown in Fig. 6.

C. User Experience Questionnaire (UEQ)

UEQ respondents consisted of 16 people. The results of the UEQ test in Table VII and Fig. 5 show that the average is in the excellent category. However, there is one aspect with the good category, namely the efficiency of the AR camera, which requires an extended focus on specific devices.

VI. CONCLUSIONS

From the overall test results of the application, it can be concluded that this application is fascinating for children to learn based on the increasing Smileyometer results after using the application and the Again-Again Table, which most users want to use again. Meanwhile, the usability factor-based question method results show that the average is more than 4 (above happy).

In terms of user experience, the Kupuku application scores 76 for the SUS, which can be categorized as good. The UEQ results obtain an average score of excellent. Therefore, from these results, it can be concluded that users can use the Kupuku application properly. The development of this application can still be carried out for things like fixing bugs, adding AR variations and quizzes, and developing on other platforms.

ACKNOWLEDGMENT

The authors would like to acknowledge funding from DIKTI through the PDUPT program in 2021, Universitas Gadjah Mada.

REFERENCES

- [1] A. Kaya, R. Ozturk, and C. Altin Gumussoy, "Usability Measurement of Mobile Applications with System Usability Scale (SUS)," in *Industrial Engineering in the Big Data Era*, F. Calisir, E. Cevikcan, and H.C. Akdag, Eds., New York, USA: Springer International Publishing, 2019, pp. 389-400.
- [2] R. Alturki and V. Gay, "Usability Testing of Fitness Mobile Application: Methodology and Quantitative Results," *7th International Conference on Computer Science, Engineering & Applications*, 2017, pp. 97-114.
- [3] J.C. Read, "Validating the Fun Toolkit: An Instrument for Measuring Children's Opinions of Technology," *Cognition, Technology & Work*, Vol. 10, No. 2, pp. 119-128, Apr. 2008.
- [4] Z. Sharfina and H.B. Santoso, "An Indonesian Adaptation of the System Usability Scale (SUS)," *2016 International Conference on Advanced Computer Science and Information Systems*, 2016, pp. 145-148.
- [5] B. Laugwitz, T. Held, and M. Schrepp, "Construction and Evaluation of a User Experience Questionnaire," in *HCI and Usability for Education and Work. USAB 2008. Lecture Notes in Computer Science*, A. Holzinger, Ed., Berlin, Germany: Springer, 2008, pp. 63-76.
- [6] S.B. Zaibon and N. Shiratuddin, "Heuristics Evaluation Strategy for Mobile Game-Based Learning," *The 6th IEEE International Conference on Wireless, Mobile, and Ubiquitous Technologies in Education*, 2010, pp. 127-131.
- [7] M.D. Corry, T.W. Frick, and L. Hansen, "User-Centered Design and Usability Testing of a Web Site: An Illustrative Case Study," *Educational Technology Research and Development*, Vol. 45, No. 4, pp. 65-76, Dec. 1997.
- [8] J.R. Lewis and B. Raton, "Usability Testing," in *Handbook of Human Factors and Ergonomics*, G. Salvendy, Ed., 3rd ed., New Jersey, USA: John Wiley & Sons, Inc., 2006, pp. 1275-1316.
- [9] A. Valerian, H.B. Santoso, M. Schrepp, and G. Guarddin, "Usability Evaluation and Development of a University Staff Website," *2018 Third International Conference on Informatics and Computing*, 2018, pp. 1-6.
- [10] G. Sim and M. Horton, "Investigating Children's Opinions of Games: Fun Toolkit vs. This or That," *Proceedings of the 11th International Conference on Interaction Design and Children*, 2012, pp. 70-77.
- [11] A.A.I.I. Paramitha, G.R. Dantes, and G. Indrawan, "The Evaluation of Web Based Academic Progress Information System Using Heuristic Evaluation and User Experience Questionnaire (UEQ)," *2018 Third International Conference on Informatics and Computing*, 2018, pp. 1-6.
- [12] S. Khodambashi and Ø. Nytrø, "Usability Evaluation of Published Clinical Guidelines on the Web: A Case Study," *2016 IEEE 29th International Symposium on Computer-Based Medical Systems*, 2016, pp. 140-145.
- [13] Kharis, P.I. Santosa, and W.W. Winarno, "Evaluasi Usability pada Sistem Informasi Pasar Kerja Menggunakan System Usability Scale (SUS)," *Prosiding Seminar Nasional Sains dan Teknologi*, Vol. 1, No. 1, 2019, pp. 240-245.
- [14] Y.N. Ortega-Gijon and C. Mezura-Godoy, "Usability Evaluation of Brain Computer Interfaces: Analysis of Methods and Tools," *2019 IEEE International Conference on Engineering Veracruz*, 2019, pp. 1-8.
- [15] F. Muhammad, N. Faradilla, E. Muslim, and D.N. Adimia, "User Experience Evaluation on the Usage of Commuter Line Train Ticket Vending Machine," *2017 6th International Conference on Industrial Technology and Management*, 2017, pp. 164-170.
- [16] J.C. Read and S. Macfarlane, "Using the Fun Toolkit and Other Survey Methods to Gather Opinions in Child Computer Interaction," *Proceedings of the 2006 Conference on Interaction Design and Children*, 2016, pp. 81-88.
- [17] J. Brooke, "SUS: A Retrospective," *Journal of Usability Studies*, Vol. 8, No. 2, pp. 29-40, Feb. 2013.
- [18] A. Bangor, P. Kortum, and J. Miller, "Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale," *Journal of Usability Studies*, Vol. 4, No. 3, pp. 114-123, May 2009.
- [19] M. Schrepp. *User Experience Questionnaire Handbook Version 8*. (2019). Access date: Feb. 2, 2020. [Online], <https://www.ueq-online.org/Material/Handbook.pdf>.
- [20] H.B. Santoso, M. Schrepp, R.Y.K. Isal, Y. Utomo, and B. Priyogi, "Measuring User Experience of the Student-Centered E-Learning Environment," *Journal of Educators Online*, Vol. 13, No. 1, pp. 58-79, Jan. 2016.