Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th June 2024. Vol. 3. No. 3; e-ISSN: 2808-4519

Integrating Heuristic Evaluation and Think-Aloud Protocols By Applying Nielsen's Metrics on Indonesian e-Pangan Application

Rahma Fitria[1]*, Cut Meurah Nurul 'Akla[2], Rini Meiyanti[3], Nova Kartika[4], Rifqa[5]

[1,2,3,4,5]*Universitas Malikussaleh*
*rahmafitria@unimal.ac.id**

Abstract

Usability testing has emerged as one of the most important aspects of software testing to provide high-quality software to users. The purpose of this study is to evaluate the usability of an Indonesian mobile government application known as e-pangan. Because of the growing number of mobile phone users, the government has deployed mobile applications to make it easier for people to obtain information. Unfortunately, mobile government is not user-friendly and still has a lot of issues, as evidenced by user reviews in the Google Play store. The study employed a hybrid usability methodology that included heuristic evaluation and think-aloud protocols. The mobile government application Panel Harga Pangan will be tested to represent the study of this testing. Furthermore, the heuristic evaluation is performed by a usability expert, while the think-aloud protocols are performed by ten users. Nielsen's metrics are applied to measure the learnability, efficiency, memorability, error rate, and satisfaction of the application. The results of the research show that the application has an average score of 66%, indicating that while some users might discover it challenging others might discover it to be straightforward. This research is expected to be beneficial for usability practitioners or developers to develop a user-friendly e-government application.

Keywords: *Usability Testing, Heuristic Evaluation, Think-aloud Protocols, Nielsen Metric, Mobile Evaluation*

1. Introduction

Nowadays, the growth of information technology and the internet has revolutionized how data is transmitted. The government also takes advantage of this breakthrough technology to disseminate information in real-time via an application. This is a shift in data transformation from the conventional approach to the e-government paradigm so as for the government to provide good community services. Agriculture food security is one of the government's top concerns in achieving national development since agricultural production is a basic human necessity and sustaining Indonesia's economic stability. Government agencies require a well-integrated system to better administrative systems and community service delivery. As a result, consumers are in need of an application that serves as a regional agriculture food price information center, allowing economic actors to obtain the most up-to-date agriculture food pricing information on a daily basis and increase economic efficiency.

In addition, fluctuations in agricultural production prices in the market caused by a lack of up- to-date agriculture food price information impair agricultural production price stability, resulting in extremely large differences in the pricing of basic commodities in each region of Indonesia. Unfortunately, an e-Pangan Indonesian government application still has many defects in terms of UI/UX design, as well as many bugs that contain inconsistent information, making users confused in finding the desired information and deciding to leave the application. According to preliminary observations made by researchers of various application visitors, there were many issues noticed by users. User dissatisfaction will affect access to the Indonesia e-Pangan application. User satisfaction is a key aspect in the execution of usability testing. Usability is derived from the word usable, which signifies that it is useful or functions properly when tested with end-users [1], [2]. Usability is a scientific field with numerous evaluation methods produced by previous scholars. Usability testing measures, tests, and evaluates a system in terms of the user's ease of use of the system interface to fulfil its goals, which could lead to user satisfaction values when using it [3]. To be useable, an application must successfully provide users with the opportunity to complete the actions carried out to achieve their aims or objectives in the application as well as possible [4]. Usability testing is inextricably linked to the scientific subject of Human-Computer Interaction (HCI), in which system users (people) interact with the system. Since the early 1970s, the scientific area of HCI has been evolving. This science emphasizes the importance of design, user interface, and visual appearance to make it simple and easy for users to engage with the system.

The goal of the research is to collect feedback from usability experts and randomly selected end- users. Usability can be measured by user feedback regarding user impressions when using apps, or it can be reinforced and strengthened by expert opinions or experts with experience in the subject of usability. To evaluate the usability of the e-Pangan Indonesian government application, various evaluation methods can be used. One such method is the think-aloud protocol approach, which involves users verbalizing their thoughts and actions while interacting with the application. This approach can provide insights into users' thought processes and help identify usability issues. Additionally, the viewpoint of usability specialists can be incorporated by using the heuristic evaluation method, which involves experts evaluating the application based on a set of established usability guidelines. By combining these two methods, a comprehensive evaluation

of the e-Pangan application can be conducted. In 2012, Nielsen claimed that the think-aloud protocols method allows researchers to explicitly hear users talk about how they use the application in general [3]. Meanwhile, Nielsen defines heuristic evaluation as an expert-based strategy for identifying usability issues in user interface design [5]. It is predicted that combining the two methods will yield more accurate findings.

Usability is a key aspect of the software development cycle, and it is defined by five components: learnability, efficiency, memorability, errors, and satisfaction. Learnability refers to how easy it is for users to complete basic tasks the first time they encounter the design. Efficiency refers to how quickly users can perform tasks after learning the design. Memorability refers to when users return to the design after a period of not using it, how easily they can rebuild proficiency. Errors refer to how many mistakes users make, how severe these mistakes are, and how easily they can be corrected. By evaluating the e-Pangan application based on these five components, we can identify its strengths and weaknesses and make improvements to enhance its usability and user satisfaction [3].

Previous research on three Indonesian online news portals used a combination of two methodologies, namely heuristic evaluation and think-aloud. Another previous study found minor severity during HE and 85 recommendations for improvement when using the think-aloud approach [6]. Furthermore, the study in usability testing on the website of Syarief Hidayatullah State Islamic University Jakarta generates quite satisfactory results, except for an error section in several functions on the website and out-of-date news [7]. The Nielsen metrics (learnability, efficiency, memorability, error, and satisfaction) are used in this study.

Usability testing of the Janabadra website used questionnaire methods and the Nielsen model to measure metrics. The Janabadra University website's test results are easy to use. Some people find it difficult to remember the options and information offered on this website [8]. Another study has investigated the integration approach between Heuristic Evaluation (HE) and Thinking Aloud (TA) on game-based learning materials by evaluating in terms of expert review and user review [9]. Several defects were discovered throughout the test following heuristic principles, as well as through observations and interviews conducted during the implementation of the thinking-aloud approach. Although experts and users provided positive feedback on the effectiveness of this learning material website, the user experience still needs to be improved. Furthermore, a hybrid method combining heuristic evaluation and user experience has been tested on several web domains, with the results indicating that the issues discovered during user experience testing were also discovered during heuristic evaluation testing [10].

2. Research Methodology

In general, at the data collecting stage, this research study employs a combination of methods technique (quantitative and qualitative). The present research is an experimental study that combines two approaches of usability evaluation. During this test, task scenarios were implemented, observations were made, and interviews were conducted. The data is analysed to assess the Indonesian e-Pangan application's level of usability.

The following are the stages of study on the execution of usability testing on the e-Pangan application utilizing the heuristic evaluation method and think-aloud with the Nielsen metric measurement, as shown in Figure 1.

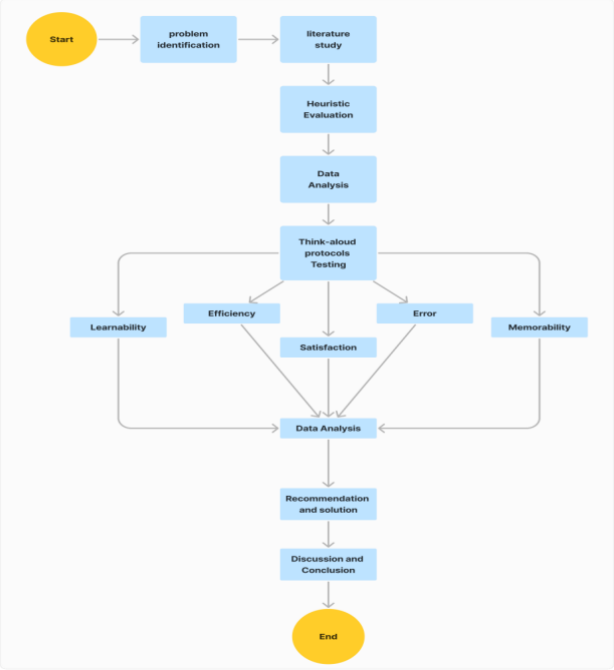

Fig. 1: Research design methodology

The problems that occurred as a result of the prior survey were highlighted in the first step. Then, to determine which approaches are appropriate for this research, a literature review was undertaken. Following that, heuristic evaluation (HE) was performed to obtain input from usability experts. This evaluation should focus on the touchscreen-based mobile device interface of the application and assess its usability based on the established principles of heuristic evaluation. By conducting this evaluation, we can identify usability issues and provide recommendations to improve the application's design and functionality, ultimately leading to a better user experience. All issues discovered by experts will be combined and examined to provide a heuristic evaluation report [11]. Furthermore, the third phase is followed by a usability test, which is typically completed by 10 users every test round [12]. The test scenario is designed prior to executing the usability test based on the concerns identified in the heuristic evaluation report. In addition, before the exam begins, users will be requested to complete a demographic questionnaire. During the testing process, the user will be prompted to complete the test scenario. Evaluators,

on the other hand, will assess the user's performance by observing and analyzing what the user performs. There are five usability metrics will be measured. The metrics of learnability, efficiency, memorability, and mistakes are recorded while users take the test. Meanwhile, satisfaction is measured by distributing a satisfaction questionnaire. The questions and answers are organized on a 5-point Likert scale ranging from 1 to 5.

All of the questions are aimed to elicit feedback from users on how they feel about the e- Pangan (Panel Harga Pagan) application, which has been chosen as an agriculture food price panel application. It is a vital application that might represent the usability testing investigation. More than one thousand people have downloaded the application. The Likert scale is used to express how strongly users or participants agree or disagree with using this mobile application, with "strongly disagree" equaling "1" and "strongly agree" equaling "5". The application's usability score can be determined in % by using five usability metrics: learnability, efficiency, memorability, error and satisfaction. The fifth phase of this study will provide recommendations and solutions based on the observations made throughout the test. During this phase, the issues discovered will be thoroughly examined before making recommendations to improve the application. Finally, the sixth step of this study will explain and finish all of the research findings.

The current research's data collection methods include observation, questionnaires, and interview. The data gathered during the test observation is examined to determine usability metrics of learnability, efficiency, memorability, and errors. The fourth metrics are measured by assessing the success rate of users in completing criteria breakdowns from test scenarios [13]. Participants in the usability testing has been chosen from among novice users who have no prior experience with the application or who have familiarity with another mobile application. The sample of users will be chosen at random, with a minimum of five to ten users, although increasing the number of users has little effect on the outcome [14]. Meanwhile, testers for heuristic evaluation are recruited from persons who have worked as usability engineers or have skills in this area.

The data gathered to determine the percentages of learnability, efficiency, memorability, error rate and satisfaction is entered into the equation. The data will be analysed and presented in a bar chart using Microsoft Excel. The conclusions of this mobile application assessment are based on observations made while users complete test scenarios during usability testing and a questionnaire administered after the test. The data collected is analysed and the learnability, efficiency, memorability, errors and satisfaction are calculated in percentages. The average of these five values is used to compute the application's usability score, which is expressed as a percentage between 1 and 100. As a result, according to the issues discovered during the observation, recommendations and solutions are presented.

3. Result and Analysis

3.1. This Heuristic Evaluation Result

An individual heuristic evaluation is a process in which a usability expert reviews an application to identify any potential usability issues. The resulting heuristic evaluation report includes all of the issues discovered by the expert, along with suggested solutions for improvement. This can help ensure that the application meets the needs and expectations of its users, and ultimately improve user satisfaction and engagement. The table below highlights the most major usability concerns identified by usability expert, as well as solutions for improving them.

Table 2: Heuristic Evaluation Result

No	Issues	Heuristics	Severity	Recommendation
1	The location selected in application returning incomplete and inconsistent data every several times	Consistency and standards	Major	The location selected in application returning incomplete and inconsistent data every several times
2	The Filter function for location selection is not working	Functionality	Major	The Filter function for location selection is not working
3	The filter function provides inconsistent data	Consistency	Major	The filter function provides inconsistent data
4	The sub menu “bantuan aplikasi” in profile menu is not working	Error preventation	Major	The sub menu “bantuan aplikasi” in profile menu is not working
5	The button “kirim” in bantuan aplikasi sub menu does not give any notification	Visibility of system status	Major	The button “kirim” in bantuan aplikasi sub menu does not give any notification
6	Arrangement of 2 different types of filter functions confused users	Match between system and real word	Minor	Arrangement of 2 different types of filter functions confused users
7	The indicator color on the price up and down is not consistent	Error prevention	Minor	Fix the indicator color that represent the price up and down correctly
8	The menu provided in application is confusing at first time using the application	Recognition rather than recall	Minor	Develop a simple menu to use
9	Edit profile is not working	Flexibility and efficiency of use	Minor	Better if user is able to edit profil
10	The FAQ menu provides no information	Help and documentation	Minor	The FAQ menu provides no information

The expert has identified five major and five minor issues based on the preceding table. Ten Nielsen's heuristic concepts were followed in the evaluation procedure [18]. The severity ranking scale (SRS) developed by Nielsen is used to determine the issues' level of severity [19]. Jakob Nielsen offered four severity grades, one of which is usability catastrophe, meaning that the problem must be fixed before the product is released. The second scale, a major usability issue, is deemed urgently needing attention. In the interim, addressing minor usability issues should be a low priority. Finally, if additional time is available for the project, the cosmetic issue must be resolved.

3.2. Think-Aloud Protocol Test Scenario Design

To test the application, a test scenario is created before test execution. After identifying every feature, a data matrix was constructed.

Table 2: Test scenario design.

Scenario	Task Description
Task 1	Choose your location: Province: Aceh Aceh District/City: Banda Aceh
Task 2	Check the price of red chili pepper in Banda Aceh Task 3
Task 3	View monthly history of tuna price in Banda Aceh
Task 4	View agriculture production commodity prices in Banda Aceh
Task 5	Find the national producer-level price of shallots.
Task 6	View the table showing onion prices across Indonesia
Task 7	Submit a food price enquiry via “Bantuan Aplikasi” submenu

3.3. The Learnability Nielsen’s Metrics Measurement

Users are observed during the test execution process in order to determine the learnability metric. Task criteria included "user understands the flow of how to do the task," "user does not have difficulty taking the right steps to do the task," "user can easily complete the task," and "user understands the action taken is right for wrong." These criteria were observed in every task in the test scenario. For every criteria that is met, the task that fails will be marked "no" for no credit, "partial" for 50% credit, and "yes" for 100% credit.
Based on the observation result, user find difficulty in conducting task 3 where 60% of user can not perform the task successfully. The task is asking the user to find a specific price of the product. Based on the data analysis, the effectiveness of each user can be depicted as the table below:

Table 3: Learnability metrics success rate

	U 1	U2	U3	U 4	U5	U6	U 7	U 8	U 9	U10	Subtotal
Yes	20	16	12	16	13	12	20	30	16	24	179
Partial	9	8	7	8	14	5	4	2	6	0	63
No	3	8	13	8	5	15	8	0	10	8	78
Total											320

The above table lists 32 task criteria, each requiring 10 attempts, for a total of 320 attempts. There were 63 partially successful efforts and 179 successful ones. A total of 78 tasks have not been completed; these will be disregarded because 78 x 0% = 0. As a result, we apply the following equation to determine the overall learnability rating for the following set of tasks:

$$\text{Learnability (\%)} = \text{Yes} + (\text{Partial} \times 0.5)) / \text{Total} \times 100\% \quad (1)$$
$$= (179 + (63 \times 0.5)) / 320 \times 100\%$$
$$= 65.78\%$$

From the above equation, the rating of learnability metrics that has been conducted usability testing with 10 users has showed the learnability success rate is approximately 65.78%.

3.4. The Efficiency Nielsen’s Metrics Measurement

The metric of efficiency is determined by observing users throughout test execution. The task criteria that were observed in every task in the test scenario, such as " User easily performs the task for the first time when looking at the interface," " Users can complete tasks quickly," " Users easily resolve menu selection errors," and " Users make very minimal errors." Every successful task requirement will be marked "yes" for 100% credit, "partial" for 50% credit, and "no" for 0% credit for the failure task.
Based on the observation result, user find difficulty in conducting task 3 and task 4 where 60% of user can not perform the task successfully. The task is asking the user to find a specific price of the product and monthly history product. Based on the data analysis, the effectiveness of each user can be depicted as the table below:

Table 4. Efficiency metrics success rate.

	U 1	U2	U3	U 4	U5	U6	U 7	U 8	U 9	U10	Subtotal
Yes	15	17	13	14	19	13	19	30	16	24	180
Partial	9	3	5	6	4	3	3	2	4	1	40
No	8	12	14	12	9	16	10	0	12	7	100
Total											320

Table above shows 32 task criteria with 10 attempts per task, totaling 320 attempts. 180 attempts were successful and 40 were partially successful. There are total of E unsuccessful tasks which will be ignored as 100 x 0% = 0. Therefore, to achieve the overall efficiency rating for this set of tasks we use the following equation:

$$\text{Efficiency (\%)} = \text{Yes} + (\text{Partial} \times 0.5)) / \text{Total} \times 100\% \quad (2)$$
$$= (180 + (40 \times 0.5)) / 320 \times 100\%$$
$$= 62.5\%$$

From the above equation, the rating of efficiency metrics that has been conducted usability testing with 10 users has showed the success rate is approximately 62.5%.

3.5. The Memorability Nielsen's Metrics Measurement

Measuring memorability is done by observing users while the test is being run. Every task in the test scenario was subject to the following task criteria: "User selects the right menu on the first try," and "User can complete the task on the next attempt." For every criteria that is met, the task that fails will be marked "no" for no credit, "partial" for 50% credit, and "yes" for 100% credit.

Based on the observation result, user find difficulty in conducting task 3 and 4 where about 60% of user can not perform the task successfully. The task is asking the user to find a specific price of the product ad monthly history of product. Based on the data analysis, the memorability of each user can be depicted as the table below:

Table 5: Memorability metrics success rate.

	U 1	U2	U3	U 4	U5	U6	U 7	U 8	U 9	U10	Subtotal
Yes	11	12	8	11	12	9	12	16	7	13	111
Partial	1	0	0	0	2	0	0	0	1	1	5
No	4	4	8	5	2	7	4	0	8	2	44
						Total					160

The above table lists 16 task criteria, each requiring 10 attempts, for a total of 320 attempts. Five of the 111 efforts were just partially successful. There are 44 failed tasks in all, and since 44 x 0% = 0, they will be disregarded. As a result, we employ the following equation to determine the overall memorability rating for the set of tasks:

$$\text{Memorability (\%)} = \text{Yes} + (\text{Partial} \times 0.5)) / \text{Total} \times 100\% \quad (3)$$
$$= (111 + (5 \times 0.5)) / 160 \times 100\%$$
$$= 70.93\%$$

From the above equation, the rating of memorability metrics that has been conducted usability testing with 10 users has showed the success rate is approximately 70.93%.

3.6. The Errors Nielsen's Metrics Measurement

Throughout the test, users are observed to identify the error metric. The task requirements, which included elements like "Users select the correct menu on the first try," "Users can easily fix errors and mistakes made," "Users make few errors and mistakes," and "Users successfully complete the task," were all observed in every task in the test scenario. For every criteria that is met, the task that fails will be marked "no" for no credit, "partial" for 50% credit, and "yes" for 100% credit.

Based on the observation result, user find difficulty in conducting task 3 and task 4 where more 50% of user cannot perform the task successfully. The task is asking the user to find a specific price of the product and monthly history product. Based on the data analysis, the errors of each user can be depicted as the table below:

Table 6: Errors Metrics Success Rate

U 1	U2	U3	U 4		U5	U6	U 7	U 8	U 9		U10	Subtotal
Yes	23	21	14	20	21	17	21	30	17	24		208
Partial	1	2	3	3	4	2	2	1	3	2		23
No	8	9	15	9	7	13	9	1	12	6		89
						Total						320

Table above shows 32 task criteria with 10 attempts per task, totaling 320 attempts. 208 attempts were successful and 23 were partially successful. There are total of 89 unsuccessful tasks which will be ignored as 89 x 0% = 0. Therefore, to achieve the overall errors rating for this set of tasks we use the following equation:

$$\text{Errors (\%)} = \text{Yes} + (\text{Partial} \times 0.5)) / \text{Total} \times 100\% \quad (4)$$
$$= (208 + (23 \times 0.5)) / 320 \times 100\%$$
$$= 68.59\%$$

From the above equation, the rating of errors metrics that has been conducted usability testing with 10 users has showed the success rate is approximately 68.59%.

3.7. The Satisfaction Nielsen's Metrics Measurement

The computer vision techniques were investigated in order to classify nutmeg based on the Indonesian national standard. System Usability Scale (SUS) post-questionnaires were used to test user satisfaction after they had utilized the system. Practitioners use this questionnaire the most frequently [6]. Likert scales with five points are used to organize the questions and responses. The System Usability Scale (SUS) and this likert scale pair well [15]. The scale had a range of 1 to 5.

Table 7: SUS Questionnaire

No	Questionnaire
1	I'm sure I would like to engage in frequent use of this system.
2	I thought the system was overly complicated.
3	I found the system is user-friendly.
4	I believe that in order to use this system, I would require the assistance of a technical expert.
5	I discovered that this system's numerous features were well incorporated.
6	I felt that this system offered numerous inconsistent features.
7	The majority of people should have no trouble picking up how to use this method fairly quickly.
8	I found it somewhat difficult to use the system.
9	I was really relaxed while use the system.
10	Before I could start using this application, I had a lot to learn.

To achieve the satisfaction rating for the Panel Harga Pangan application, we use the following equation:

$$\begin{aligned} \text{Satisfaction}(\%) &= \text{Answer Poin/ Total Point x } 100\% \\ &= 320 / 500 \text{ x } 100\% \\ &= 64\% \end{aligned} \quad (5)$$

3.8. Usability Score

With each usability parameter now represented as a percentage, the metrics of learnability, efficiency, memorability, errors and satisfaction. The index between 1 and 100 is employed to represent the usability of the e-Pangan application (Panel Harga Pangan) by averaging these five scores. The following formula can be used to determine the panel harga pangan application's usability score for all users:

$$\begin{aligned} \text{Usability (\%)} &= \text{(Learnability+ Efficiency + Memorability + Errors + Satisfaction) /5} \\ &= (65.78+62.5+70.93+68.59+64)/\ 5 \\ &= 66.36\% \end{aligned} \quad (6)$$

4. Discussion and Conclusion

The Indonesian e-Pangan application was evaluated by an expert and 10 users for usability evaluation in the present research. The main findings of this study revealed that the users' perception is the complex matters in conducting think aloud method by expressing their feelings and thinking. The issues found in heuristic evaluation are also found in usability think-aloud method. These results substantiate to Ayuni's research in 2023 where the issues found by users in think-aloud approach revealed more detailed issues than expert review [16]. In addition, the issues found by the combination method of heuristics and think-aloud produce more reliable and effective results [17]. The usability problem that was found in this research include (a) the inconsistent data in location selection, which affected the efficiency of use of the application; (b) the submenu (filter menu) is sometimes not working; (c) 60% of users' assessment is hard to understand; (d) some button in edit profile is not working; (e) there is no indicator of error message.These issues will be suggested with the recommendation for future development of other similar applications. The five metrics of learnability, efficiency, memorability, error rate and satisfaction are strongly matched in this usability measurement. In other words, the observation and satisfaction questionnaire results are strongly interrelated. However, the think-aloud qualitative result indicates the user's comments toward the application show differently with quantitative results, 60% of users answered the interview session after the testing said "The application is hard to understand". In contrast, all users answered "neutral, agree and strongly agree" to the satisfaction questionnaire question "I found the system is user- friendly".
Overall, the usability score of national e-Pangan application (Panel Harga Pangan) has shown some problems that user faces during the evaluation which is average (66%). It indicates, some users found the application straightforward to use, others found it challenging to finish the assigned task scenario. More testing needed to be conducting with more expert review to evaluate using Nielsen's heuristics principles to validate these results. Thus, closed interview in think-aloud method may present the decent outcome. Nevertheless, the results of this assessment called for additional research, which would involve enlisting well-known individuals who would be able to frankly communicate their feelings. Obtaining the user's help in sketching the desired prototype as anticipated is highly encouraged. More tools, such as eye tracking, produce more dependable result.

Acknowledgments

This research is funded by Advanced Knowledge and Skills for Sustainable Growth Project 2023- Asian Development Bank-Universitas Malikussaleh.

References

[1] Hussain A and Fitria R 2018 Mobile flight and hotel booking application: a heuristic and UX Test Journal of Telecommunication, Electronic and Computer Engineering pp 93-101
[2] Fitria R 2022 User experience test of frog virtual learning environment website Journal Informatika Kaputama (Medan) pp 26–47
[3] Nielsen J 2012 Usability 101: introduction to usability (https://www.nngroup.com/articles/usability- 101-introduction-to-usability)
[4] Alshammari T, Alhadreti O, and Mayhew, P J 2015 When to ask participants to think aloud: a comparative study of concurrent and restrospective think-aloud methods International Journal of Human Computer Interaction (IJHCI)
[5] Nielsen J 1994 Enhancing the explanatory power of usability heuristics Proc. of the SIGCHI Conference on Human Factors in Computing Systems (CHI'94) pp 152-158
[6] Fajarini P T, Wirdiani N K A and I.P. A. Dharmadi I P A 2020 Evaluasi portal berita online pada aspek usability menggunakan heurtistic evaluation dan think aloud Jurnal Teknologi Informasi Dan Ilmu Komputer
[7] Larasati I 2020 evaluasi penggunaan website universitas islam negeri syarif hidayatullah jakarta dengan menggunakan metode usability testing Journal of Computer Science and Information Systems pp 68-77
[8] Yumarlin M Z 2016 QUEST e-learning portal: applying heuristic evaluation, usability testing and eye tracking Universal Access in the Information Society pp 531–543
[9] Chang, Chang C C and Johnson T 2021 Integrating heuristics and think-aloud approach to evaluate the usability of game-based learning material Journal of Computers in Education pp 137- 157
[10] Paz F, Paz F A, Villanueva D, and Pow-Sang J A 2015 Heuristic evaluation as a complement to usability testing: a case study in WEB Domain Proc. 12th International Conference on Information Technology: New Generations (ITNG) pp 546–551
[11] Nielsen J 2005 Usability engineering methods for software developers (ACM) pp 71-74
[12] Kaikkonen A, Kekalainen A, Cankar M, Kallio T and Kankainen A 2005 Usability testing of mobile application: a comparison between laboratory and field testing J. Usability Stud pp 4-16.
[13] Nielsen J 2001 First Rule of Usability? Don't Listen to Users (https://www.nngroup.com/articles/first-rule-of-usability-dont-listen-to-users/)
[14] Gunduz F and Pathan A S K 2013 On the key factors of usability in small-sized mobile touch- screen application Int. J. Multimed. Ubiquitous Eng pp 115-138
[15] Moller A, Roalter L, Diewald S, Scherr J, Kranz M, Hammerla N, Olivier P, and Plotz T 2012 GymSkill: a personal trainer for physical exercises, Int. Conf. Pervasive Comput. Commun. PerCom (IEEE) pp 213–220
[16] Ayuni R and Ahsyar T K 2023 Evaluasi usability pada sistem informasi beasiswa menggunakan metode heuristic evaluation dan think aloud KLIK: Kajian Ilmiah Informatika dan Komputer vol 3 no 6 pp 737–745
[17] Chang C C and Johnson T 2020 Integrating heuristics and think-aloud approach to evaluate the usability of game-based learning material Journal of Computers in Education vol 8 no 1 pp 137–157
[18] Nielsen J 1990 World leaders in research-based user experience: 10 usability heuristics for user interface design (Nielsen Norman Group, https://www.nngroup.com/articles/ten-usability- heuristics/)
[19] Nielsen J 1994 World leaders in research-based user experience: severity ratings for usability problems (Nielsen Norman Group, https://www.nngroup.com/articles/how-to- rate-the-severity- of-usability-problems)