(RESEARCH ARTICLE)

Usability evaluation of online transportation using Nielsen model

Nia Budi Puspitasari *, Yussy Aulia and Zainal Fanani Rosyada

Department of Industrial Engineering, Diponegoro University, Prof. Soedarto, SH Street, Tembalang, Semarang, 50275, Indonesia.

Abstract

Maxim is one of the online transportation services in Indonesia. Based on the feedback provided by users on the App Store and Google Play store, there are still shortcomings, such as the application often having errors when used, GPS accuracy is not precise, and the system feeling difficult when canceling orders. Based on these problems, usability testing is needed. The method used in this research is usability testing based on the standard Nielsen Model conducted by five respondents. In this study, the Usability System Scale questionnaire was also filled in to get the satisfaction value carried out by 100 respondents using Maxim. The usability test results show that the respondents complained that the payment methods were not varied and there was no customer service chat. This usability test gives results on each indicator: the learnability value of 92%, the efficiency value of 0.42 goals/sec, the error value of 0.1, and the satisfaction value of 57.2.

Keywords: Usability; Nielsen Model; Usability System Scale; Online transportation

1. Introduction

A practical and user-friendly user interface is essential, especially for non-expert users in the field applied to the system. The quality of an application consists of various aspects, including an application must-have features that function correctly according to its purpose, can be used easily by users, and can be accessed anywhere and anytime.

Usability measures the quality of the user experience when interacting with products or systems, software applications, mobile technology, or other equipment operated by users [1]. Usability testing or measurement is a series of activities to measure efficiency, effectiveness, and user satisfaction with software [2]. Usability testing aims to determine whether the related design has problems or not for its users and collects qualitative and quantitative data, which can later be used to determine the level of user satisfaction with the product.

In recent years, online transportation or what is generally called online taxis has grown rapidly, especially in Indonesia, because of its use, which makes daily activities easier. One of the online transportation service companies growing in Indonesia is Maxim, which can be downloaded for free on Google Playstore and App Store. Maxim was initially founded in 2003 in Chadrinsk, Ural Mountains, Russia. In Indonesia itself, Maxim opened its first branch in Jakarta in 2018.

Based on a preliminary survey conducted on 30 respondents using the Maxim application, the respondents complained about the inaccurate map accuracy, which caused the pick-up point to be incorrect. The respondents also said it was difficult to distinguish the red and blue circle logos as a pick-up location and destination marker. The payment methods were less varied, and the application's appearance was less pleasing to the eye. This problem is part of the usability problem wherein the usability guideline must meet user expectations regarding the navigation and content of a system [3]. Based on the problems above, research is needed to analyze the usability evaluation of the Maxim application using

* Corresponding author: Nia Budi Puspitasari

the Usability Testing method based on the standard Nielsen Model to determine the usability of the application. Researchers use the Nielsen model because the Nielsen Model is an effective method to measure how the performance of a system assists users in operating it to make users feel easy, satisfied, and comfortable in using it [4]. If you use users as participants, researchers will find real problems and direct users about the issues encountered when using the application because the user is already using the system.

A previous study by Permana (2018), "Usability Evaluation of the Grab Application Using the Usability Testing Method" tested four aspects, namely learnability, efficiency, error, and satisfaction. The study used five respondents Grab users with the criteria of respondents aged between 15 and 64 years and live in Surabaya. The research method used was the usability testing method [5]. Then, in a previous study by Farouqi (2018), "Usability Evaluation of the Go-Jek Application Using the Usability Testing Method", four aspects were tested, namely learnability, efficiency, error, and satisfaction. The study used five respondents who were new users of the Go-Jek application and lived in Surabaya. The research method used to solve the existing problems was the usability testing method [6]. In this study, we will examine four aspects as in previous studies. The satisfaction aspect will be measured using a System Usability Score (SUS) questionnaire. The difference between previous research and this research lies in the online motorcycle taxi application used.

The existence of research with Maxim as the object will provide an evaluation that can be given regarding the level of usability and satisfaction of Maxim users in Indonesia because, of course, there are adjustments between Maxim and new users in Indonesia. This research is expected to be able to provide an overview of user satisfaction, what users need, and what things Maxim must improve from user complaints as well as to measure how easy it is for users to carry out tasks, how quickly users can get the information needed to complete tasks, and error rates occur when the user uses the Maxim application.

2. Literature Review

2.1. Usability Testing

Usability testing or measurement is a series of activities to measure or evaluate efficiency, effectiveness, and user satisfaction [2]. In addition, usability testing can also be interpreted as one of the categories of methods by observing the users of a design. Then the data is collected for analysis. The purpose of the test is to determine whether the design has problems for its users and collect quantitative and qualitative data to assess user satisfaction with the product [4].

In the usability evaluation process, the user experience when using the product will be tested. From this process, you will find any errors in the product to be an assessment to develop the product to be better [7]. There are several aspects of usability evaluation as follows [8].

- It can operate according to user needs.
- Easy to use by users.
- It can be used efficiently.
- Ease of remembering the use of the product.
- User satisfaction in using the product.
- User's comprehension in using the product.

2.2. Nielsen Model

Nielsen defines usability as a measure of quality that examines and measures how easy an interface is to use by users [9]. According to Nielsen, there are five main categories of usability, including:

- Learnability
Learnability is an indicator that describes how easily users can understand how to use it and the functions contained on the website

- Efficiency
Efficiency is a benchmark that measures the speed and accuracy of users in accessing a system.

- Memorability

Memorability is related to the user's memory in running a system, namely how far the user remembers the location of the features and the system's appearance.

- Errors

Error is an indicator that measures how many user errors while using the system.

- Satisfaction

Satisfaction is an indicator that measures how satisfied users are with the system they use, including how users feel when accessing the system and whether the users feel comfortable accessing the system without feeling burdened.

2.3. System Usability Scale (SUS)

System Usability Scale (SUS) is a questionnaire that can be used to measure the system's usability from the user's subjective point of view. This method was developed by John Broke in 1986. This method has ten questions with a Likert scale response of 1 (strongly disagree) to 5 (strongly agree) [10;11]. Odd-numbered questions have a positive meaning, and even-numbered questions have a negative meaning. The assessment of positive questions is calculated by the formula (x-1). Negative question scores are calculated by the formula (5-x). Once all scores have been determined, the total score is multiplied by 2.5 to get an SUS score between 0 and 100.

After getting the final results of the respondent's assessment, the next step is to determine the grade of the assessment results. Two ways can be used to determine the grade of research results [12]. The first method is seen from the level of user acceptance which consists of three categories: not acceptable, marginal, and acceptable. Meanwhile, in terms of grade level, there are five scales, namely A, B, C, D, and F. Furthermore, the adjective ratings consist of worst imaginable, poor, ok, good, excellent, and best imaginable, as can be seen in Figure 1.
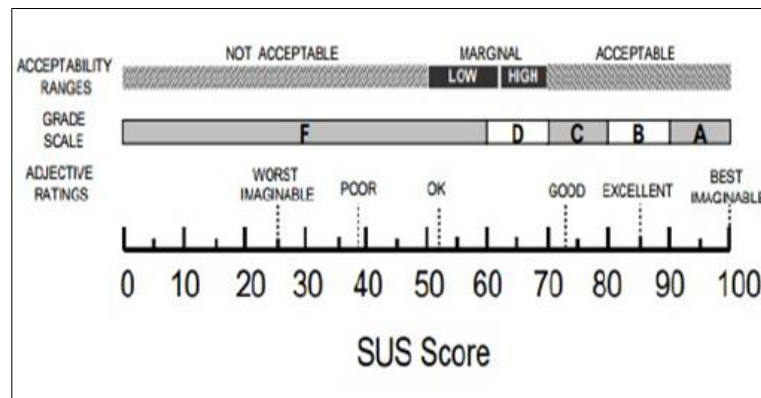

Figure 1 System Usability Scale Score

3. Material and method

The research method is a research stage that must be determined before research to be carried out clearly and systematically to achieve research objectives.

3.1. The Initiation of Necessity

This stage is the initial stage in research, where the researcher will collect the needs needed to conduct research. At this stage, identification of problems in the Maxim application will be carried out, literature studies, determining the respondents, developing task scenarios, determining performance measurement indicators, and preparing SUS questionnaires.

3.2. User Testing

The next stage is the user testing stage, where at this stage, the researcher will conduct usability testing with a predetermined method to obtain data. Usability testing was carried out by five respondents as follows.

Table 1 Usability Testing Respondents

Respondents' name	Respondents' Code
Respondent 1	R1
Respondent 2	R2
Respondent 3	R3
Respondent 4	R4
Respondent 5	R5

Five respondents carried out usability testing by performing 5 task scenarios given by the researcher to measure the value of learnability, efficiency, and error indicators. The five respondents were asked to work on 5 task scenarios as follows.

Table 2 Usability Testing Task Scenarios

Task 1 (T1)	Order an online taxi from the pick-up point of the E-walk Mall Balikpapan to the drop-off point of the Mall Plaza Balikpapan.
Task 2 (T2)	Adding a tip for drivers of IDR 5,000
Task 3 (T3)	Make payments in cash.
Task 4 (T4)	Place an order immediately.
Task 5 (T5)	Choose a motorbike taxi service option to travel.

After usability testing was carried out, respondents were interviewed to find out the problems experienced by respondents during the use of Maxim. The next step is filling out the System Usability Scale questionnaire by 100 respondents online. All data that has been obtained is then processed with predetermined equations to get results from each indicator.

3.3. Post-User Testing

At this stage, an analysis will be carried out on the results of testing the Maxim application, the results of interviews, and the results of filling out questionnaires that respondents have carried out. After the analysis is carried out, based on these results, recommendations for improvements to the Maxim application will be compiled.

4. Result and Discussion

The following can be seen the calculation results of each Nielsen Model indicator and its analysis.

4.1. The Analysis of Learnability Indicator

In measuring the level of learnability, it required data on the respondents' success in working on task scenarios [15]. Data on respondents' success in completing task scenarios can be seen in Table 3. Assessment of the data collection, namely success (s), partial success (p), and fail (f).

Table 3 Data of Respondents' Success

Respondent	T1	T2	T3	T4	T5
R1	S	S	S	S	S
R2	S	S	S	S	S
R3	S	S	S	S	F
R4	S	S	F	S	S
R5	S	S	S	S	S

The learnability indicator can be calculated using the success rate equation as follows.

$$success\ rate = \frac{success+(partial\ success \times 0,5)}{total\ task}\ x\ 100\%.... (1)$$

$$= \frac{23 + (0 \times 0,5)}{25} \times 100\%$$

$$= 92\%$$

The results of measuring the learnability indicator using the success rate equation obtained results of 92%. According to Sauro, based on 115 usability tests, the average task completion rate is 78% [13]. So, it can be concluded that the level of learnability of the Maxim application with 92% results is above the average.

4.2. The Analysis of Efficiency Indicator

To calculate the efficiency indicator, time-based efficiency will be used to analyze the time required by the user to complete the task scenario. In Table 4, it can be seen that the respondents' success in completing the task scenario is written with the number 1 if the task scenario is carried out well without any errors and the number 0 if the respondent cannot do the task scenario well.

Table 4 Respondents' Completion of Each Task

Respondent	T1	T2	T3	T4	T5
R1	1	1	1	1	1
R2	1	1	1	1	1
R3	1	1	1	1	0
R4	1	1	0	1	1
R5	1	1	1	1	1

Next, Table 5 shows the time spent by respondents working on each task scenario.

Table 5 Time Spent on Each Task

Respondent	Time (s)				
	T1	T2	T3	T4	T5
R1	25	5	3	2	2
R2	32	5	3	4	2
R3	21	3	3	3	0
R4	31	3	0	2	2
R5	28	6	1	2	1

The efficiency indicator can be calculated using the equation as follows.

$$Time\ Based\ Efficiency = \frac{\sum_{j=1}^{R} \sum_{i=1}^{N} \frac{n_{ij}}{t_{ij}}}{NR}... (2)$$

$$= \frac{\frac{1}{25} + \frac{1}{5} + \frac{1}{3} + \cdots + \frac{1}{2}}{5x5}$$

$$= \frac{10{,}504}{25}$$

$$= 0{,}42 \ goals/sec$$

Based on the results of the time-based efficiency calculations that have been carried out, the results of the efficiency evaluation of 5 respondents in working on 5 task scenarios are 0.42 goals/second. This shows that users can complete tasks as much as 0.42 per second based on the total time spent by all users to complete task scenarios, which is 189 seconds.

4.3. The Analysis of Error Indicator

To measure the error indicator, it is necessary to obtain data based on the number of errors made by the respondent at each step in working on the task scenario. The number of errors made by respondents can be seen in Table 6.

Table 6 Number of Errors Made by Respondents

Respondent	T1		T2		T3		T4		T5	
	F	K	F	K	F	K	F	K	F	K
R1	0	5	0	4	0	2	0	2	0	2
R2	0	5	0	4	0	2	0	2	0	2
R3	0	5	0	4	0	2	0	2	2	2
R4	2	5	0	4	2	2	0	2	0	2
R5	0	5	1	4	0	2	0	2	0	2

Description:
K = Number of Opportunities
F = Number of Mistakes

The error indicator can be calculated using the equation as follows.

$$defective \ rate = \frac{total \ defects}{total \ opportunities} \dots (3)$$

$$= \frac{7}{15x5}$$

$$= 0{,}093$$

$$\approx 0{,}1 = 10\% \ defect$$

Based on the calculation of the defect rate that has been carried out, the evaluation of errors from 5 respondents in working on 5 task scenarios is 0.1 or 10% defects. This shows that all respondents made seven errors out of the possible 75 errors that could occur. From these results, the quality level is 90%. According to Sauro, the average error value per task is 0.7, i.e., errors are made by 2 out of 3 users [13]. This shows that the error rate in the Maxim application is above average.

4.4. The Analysis of Satisfaction Indicator

In calculating the final SUS score, the SUS score of each respondent will be calculated. Then all the SUS scores of each respondent are added up to find the average to get the final SUS score. The average SUS final score can be calculated by equation four as follows.

$$SUS \ score = \bar{X} = \frac{\sum((\sum odd \ score - \sum even \ score)x \ 2{,}5)}{n} \dots\dots (4)$$

$$\bar{X} = \frac{5720}{100}$$

6

$$\bar{X} = 57,2$$

Based on the results of the calculation of the SUS questionnaire, the average overall SUS score was 57.2. According to Sauro, the average SUS score of the 500 studies he has conducted is 68 [12]. This shows that the average SUS score of the Maxim application is still below the average. But apart from that, in determining the results of the SUS score, there are three points of view, namely [16]:

Acceptability, the score is included in the marginal level. Acceptability is used to see the level of user acceptance of the software [12].

Grade scale, the score is included in grade D with a score range of >= 51 and <68. A grade scale is used to see the level of software [14].

Adjective rating, the score is included in the OK level. Adjective rating is used to see the rating of the software [12].

4.5. Analysis of Interview Results

After the respondents completed the usability test, interviews were conducted with the five respondents to find out the problems found when the respondents used the Maxim application. There are eight problem responses obtained from interviews with the five respondents as can be seen in Table 7.

Table 7 Interview Results

No.	Respondent	Problem Found	Problem Code
1.	R1	Not many payment method options	M1
2.	R2	The reroute option is not available	M2
		Not many payment method options	M1
		Unattractive UI	M3
		Maps are not up-to-date	M4
3.	R3	No promo features	M5
		Not many payment method options	M1
		There is an additional fee if the driver has waited for more than 5 minutes	M6
		There is no FAQ menu, and customer service chat	M7
4.	R4	Not many payment method options	M1
		There is a confirmation pop up that makes user confused	M8
5.	R5	Not many payment method options	M1

It can be seen in Table 7 that there are eight problems found during interviews with respondents. Each respondent mentioned the M1 problem five times, namely, not many payment method options.

4.6. Alternative Recommendations

Based on the results of the analysis of the five indicators and the findings of the problems obtained from interviews with respondents, we can draw up recommendations for improvement for the Maxim application. Suggestions for design improvements for the Maxim application using the 8 Golden Rules guidelines are as follows.

Eliminate unnecessary confirmation pop-ups on motorbike taxi service options to avoid confusion for new unfamiliar users when using the Maxim application.

Added frequently asked questions (FAQ) menus and customer service chat to help users find problems during application use.

In the delivery service, it is necessary to add an information column regarding the name of the recipient and the telephone number of the recipient of the delivery item so that the user no longer needs to send a separate message regarding the recipient's name and telephone number to the driver.

Added changing travel destinations when the user has already traveled. The intended destination change feature is, for example, if the user has traveled intending to go to the Prof. Soedarto Auditorium, but in the middle of the trip, the user wants to change the destination to the Imam Barjo Auditorium. Then, if the change in the destination of the trip becomes closer than the initial destination, it is hoped that the set fare will be cheaper, not the other way around. The fare conditions are expected to be recalculated from the pick-up point to the new drop-off point, not calculated from the pick-up point to the drop-off point, and added to the fare from the pick-up point to the new point (doubled fare).

Referring to the M1 problem, it is necessary to have an e-payment feature as a payment option other than using cash payments in the current cashless era. The e-payment feature needs to have an easy top-up option as the competing applications, namely top-up balances through driver-partners, ATMs, and internet banking.

It is necessary to integrate Maxim's maps into Google Maps to make the maps more accurate.

5. Conclusion

Based on the research that has been done, the conclusions that can be drawn are that the learnability indicator results are above the average. The efficiency indicator results show that users can complete tasks as much as 0.42 per second based on the overall time spent by all users in completing task scenarios. The results of the error indicator show that the error rate for the Maxim application is above the average. The SUS score for the satisfaction indicator is marginal, not acceptable, with a D grade and an OK rating. From the usability testing that has been done, the researchers also found a total of eight problems shared by the five respondents, such as not many payment method options, no option to reroute in the middle of the trip, the UI is not attractive, maps are not updated, there are no promo features, there are additional costs if the driver has waited more than 5 minutes, there is no FAQ menu and customer service chat, and there is a confirmation pop-up making the user confused. Based on the results of these tests, recommendations were made for Maxim's improvement, such as adding an e-payment feature, changing travel destination in the middle of the trip features, adding a FAQ menu and customer service chat, adding an information column regarding the name and number of recipients in the delivery service, and integrate the maps into google maps to make it more accurate.

Suggestion

With very little research available on the usability evaluation of Maxim applications, there is still plenty of room to conduct similar research. So, the suggestion that can be submitted regarding this research is to conduct usability testing with a more significant number of respondents and different methods such as ISO 9241-11, QUIM model, Heuristic Evaluation, Cognitive Walkthrough, and Domain Specific Inspection so that in future research, the limitations of Nielsen's method used to test the usability of the Maxim application in this study can be seen.

Compliance with ethical standards

Acknowledgments

Disclosure of conflict of interest

No conflict of interest is to be disclosed.

References

[1]   J. Nielsen, Usability 101: Introduction to usability, (2012).

[2]   B. Albert, and T. Tullis, Measuring the User Experience. MA: Burlington, (2008).

[3]   M. O. Leavitt, Research-based web design & usability guidelines, vol. 2009, no. July 12. Washington: U.S. Government Printing Office, (2006).

[4] J. Rubin, and D. Chisnell, Handbook of Usability: How to Plan, Design, and Conduct Effective Test. Wiley Publishing, Inc, (2008).

[5] N. R. Permana, I. Aknuranda, and R. I. Rokhmawati, Evaluasi Usability pada Aplikasi Grab dengan Menggunakan Metode Pengujian Usability, Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, Vol. 2, No. 10, (2018).

[6] M. I. Farouqi, I. Aknuranda, and A. D. Herlambang, Evaluasi Usability pada Aplikasi Go-Jek dengan Menggunakan Pengujian Usability, Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, Vol. 2, No. 9, (2018).

[7] H. Belson, and J. Ho, A Fresh Graduate's Guide to Software Development Tools and Technologies, (2012).

[8] S. Lauesen, User Interface Design A Software Engineering Perspective, (2005).

[9] J. Nielsen, Usability Engineering. Academic Press INC ed, (1994).

[10] J. Brooke, SUS - A quick and dirty usability scale. Usability evaluation in industry, vol. 189, no. 194, pp. 4–7, (1996), doi: 10.4236/9781618961020_0002.

[11] U. Ependi, F. Panjaitan, and H. Hutrianto, System Usability Scale Antarmuka Palembang Guide Sebagai Media Pendukung Asian Games XVIII, J. Inf. Syst. Eng. Bus. Intell., vol. 3, no. 2, p. 80, (2017), doi: 10.20473/jisebi.3.2.80-86.

[12] J. Brooke, SUS : A Retrospective, J. usability Stud., vol. 8, no. 2, pp. 29–40, (2013).

[13] J. Sauro, Measuring Usability With The System Usability Scale (SUS), (2011), retrieved from https://measuringu.com/sus/

[14] J. R. Lewis and J. Sauro, Item Benchmarks for the System Usability Scale, J. Usability Stud., vol. 13, no. 3, pp. 158–167, (2018).

[15] J. Sauro, How to Measure Learnability, (2013), retrieved from https://measuringu.com/measure- learnability/>

[16] A. Bangor, P. Kortum, and J. Miller, Determining what individual SUS scores mean: Adding and adjective rating scale, Journal of usability studies, vol. 4, pp. 114-123, (2009).