



Assessing User Experience and Usability in the OVO Application: Utilizing the User Experience Questionnaire and System Usability Scale for Evaluation

Ali Ibrahim¹, Onkky Alexander², Ken Ditha Tania³, Pacu Putra⁴, Allsela Meiriza⁵

^{1,2,3,4,5}Department of Information Systems, Faculty of Computer Science, Universitas Sriwijaya, Palembang, Indonesia

³Engineering and Technology, Universiti Teknologi Malaysia, Johor Bahru, Malaysia

^{1,2,3,4,5}Management of Information Systems and Business (MISaB) Research Group, Faculty of Computer Science Universitas Sriwijaya, Palembang, Indonesia

¹aliiibrahim@unsri.ac.id, ²onkkya26@gmail.com, ³ken.tania@gmail.com, ⁴pacuputra@unsri.ac.id, ⁵allsela@unsri.ac.id

Abstract

The OVO application, despite having a large user base in Indonesia, has received low ratings compared to other digital wallet apps on the Google Play Store and App Store. Users frequently complain about the user experience, which greatly affects their overall satisfaction. This study evaluates the user experience and usability of the OVO application using the User Experience Questionnaire (UEQ) and System Usability Scale (SUS). The UEQ results show that efficiency is excellent (1.55), while attractiveness, perspicuity, dependability, and stimulation are above average (1.56, 1.67, 1.33, and 1.16, respectively). However, the novelty aspect falls below average (0.64), indicating a need for improvement. The SUS score is 77.53, classifying the app as "Acceptable" with a "C" grade and an overall "Good" rating. Addressing the identified shortcomings can enhance the user experience and usability, ultimately improving user satisfaction. This study contributes valuable empirical data to the field, offering insights for researchers and practitioners in assessing the user experience and usability of mobile applications.

Keywords: digital wallet; user experience; usability; user experience questionnaire; system usability scale

1. Introduction

Today's modern transactions continue to shift from cash-based transactions to electronic-based transactions. Equal connectedness through Information and Communication Technology (ICT) has contributed significantly to the market transformation of their financial and operational businesses. The trend towards digitization and internet use has brought about significant changes in how the global economy operates. The emergence of various financial technology (FinTech) applications is enabling consumers to go beyond conventional cash-based payment systems. Digital payments are becoming the norm in people's daily lives. This rapid development in the financial sector led to the invention of many digital payment technologies, where payers and payees use digital applications to send and receive money. As such, payment systems are rapidly changing from coin and paper-based cash to convenient, fast and cost-effective forms of digital payments [1]. The development of non-cash transactions is expected to increase yearly globally. In 2022, the growth of non-cash transactions

was estimated to reach 1,045.5 billion USD, with the highest growth in developing countries in Asia and the Middle East [2]. Digital wallets are now necessary for people to carry out their activities and meet their needs [3], [4]. This positive trend must be followed by good user experience and application usability [5]. E-Wallet is an electronic service that functions to store data and as a payment instrument. In principle, E-Wallet is similar to mobile banking or Internet banking services, but the depositor does not use a bank but a digital wallet. E-wallet applications in Indonesia include OVO, Dana, GoPay, Shopeepay, Jenius, LinkAja, and others [6].

OVO is an electronic wallet application in Indonesia that users have used since 2016. OVO offers easy payments for phone credit, data packages and insurance. Nevertheless, OVO got some negative reviews on Google Play and App Store. One of the negative reviews that users feel about the OVO application is a user experience problem which causes the application's user experience to work better than the user expects. Some users complained that the OVO application response process was slow and that the

payment process using OVO took too long for them. Negative reviews on OVO have significant implications for its overall performance and user engagement. They can harm OVO's reputation, deter potential users, decrease user engagement and retention, impact competitiveness, and damage the brand's image. Addressing these negative reviews is crucial for OVO's success in the digital payment industry. The application's usability includes the user experience issues experienced by OVO users. Usability refers to how quickly and easily application users can complete tasks [7].

User Experience (UX) focuses on the overall experience users have when interacting with a product. Usability measures how easy and effective it is to use. Both are evaluated in the field through methods such as user experience questionnaires, system usability scales, usability testing, heuristic evaluation, user surveys, analytics, A/B testing, expert reviews, etc. These methods provide insights into user satisfaction, behavior, and areas for improvement.

By incorporating UEQ and SUS, researchers aim to gain valuable insights into the user experience and usability of the OVO application, thereby enriching their findings and contributing to a more comprehensive understanding of the topic. To support the researcher's assertions, the researcher has included relevant examples and case studies that illustrate the successful application of UEQ and SUS in similar research studies or within similar application contexts. These examples serve as tangible evidence of the efficacy and relevance of these methods in evaluating user experience and usability.

Research using the user experience questionnaire method and system usability scale was conducted by Guntur Eka Saputra, Rakhmi Khalida, and Ratu Nurmalika from Gunadarma University entitled "Evaluation of User Experience TLX Training Gate for Competitive Programming Learning using User Experience Questionnaire and System Usability Scale". In this study, the measurement results were obtained on 6 UEQ scales, namely the attractiveness scale (1.27), perspicuity (0.85), efficiency (1.12), dependability (1.13), stimulation (1.35) and novelty (0.81). All scales get positive impressions; the SUS score is 75 [8].

Furthermore, research was conducted by Nina Setiyawati and Dwi Hosanna Bangkalang entitled "The Comparison of Evaluation on User Experience and Usability of Mobile Banking Applications Using User Experience Questionnaire and System Usability Scale". In this study, the 6 UEQ scale measurements on four mobile banking applications received a positive impression on each scale except for BNI Mobile (Efficiency and Novelty) and Livin (Novelty), which received a neutral impression. SUS scores were obtained for the four mobile banking applications,

namely BCA Mobile (72.76), Octo Mobile (71.47), BNI Mobile (71.49), and Livin (72.4) [5].

This study aims to evaluate the user experience and measure the usability of the OVO application. The user experience in the OVO application is evaluated using a user experience questionnaire by analyzing six scales or aspects, namely attractiveness, perspicuity, dependability, efficiency, stimulation, and novelty [9][10]. Meanwhile, to measure usability in the OVO application, the system usability scale is used by analyzing three categories: acceptability ranges, grade scales, and adjective ratings [11], [12].

2. Research Methods

This study aims to assess and quantify the user experience and usability of the OVO application by employing the User Experience Questionnaire (UEQ) and System Usability Scale (SUS). The research methodology and process are illustrated in Figure 1.

2.1. Research Design

The research design in this article is evaluative and descriptive. It aims to measure and explain the success of a specific product, program, or activity, allowing conclusions to be drawn about its feasibility, relevance, effectiveness, and efficiency. This design provides a framework for assessing and analyzing the subject of the study in order to gain insights into its various aspects and evaluate its overall performance. By employing an evaluative and descriptive research design, the researchers can gather data, analyze it, and draw meaningful conclusions about the topic under investigation.

2.2. Research Process

The research framework used as a reference in the research to be carried out is shown in Figure 1.

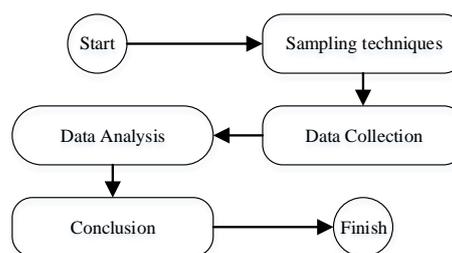


Figure 1. Research Process

Figure 1 explains that the research process begins with the sampling technique and ends with drawing conclusions. Detailed information about each research activity will be presented in the subsequent discussion subsection.

2.3. Sampling Technique

This stage aims to determine the sample and sampling technique used during the study. The population that is

the focus of this research are those who use the OVO application. This study uses the Lemeshow formula to determine the number of samples with an unknown population [16]. Through the Lemeshow formula, the number of samples to be taken is at least 100 respondents. The sampling technique used in this study is Non-Probability Sampling, namely Purposive Sampling, a sampling technique selected based on specific criteria that the researcher wants. The criteria used in this study are : OVO application users; Located di Indonesia; Minimum age of 15 years.

The purposeful Sampling method was chosen to ensure a targeted selection of participants who meet the specific criteria essential for this study. By utilizing this method, the researcher aimed to include OVO application users from various backgrounds, including diverse age groups, regions within Indonesia, etc. This approach allows researchers to gather a wide range of perspectives and experiences, contributing to a more comprehensive assessment of the user experience and usability of the OVO application.

Although the non-probability nature of purposeful Sampling does not guarantee a representative sample of the entire OVO user base, researchers made efforts to ensure diversity within the selected sample. The researcher's intention was to include participants from different demographics to increase the generalizability and relevance of our findings. The researcher reached out to potential participants through various channels, including social media platforms, online communities, and direct invitations to OVO users who matched the researcher's criteria. By employing this approach, researchers aimed to capture a broad spectrum of users and mitigate potential biases that may arise from a more limited sample.

It is important to note that while the researcher's sample may not represent the entire population of OVO application users, the focus of this study is to assess user experience and usability rather than provide statistically representative data. Nonetheless, the insights gained from this diverse sample will contribute valuable findings and recommendations for enhancing the user experience and usability of the OVO application.

2.4. Data Collection Figure

At this stage, data collection was carried out from predetermined respondents. In this study, the instrument used was a questionnaire. The questionnaire included questions about the respondent's identity, the general use of the OVO application, 26 user experience questionnaire statements, and ten system usability scale statements. To minimize potential biases, researchers implemented several measures in the data collection process.

Firstly, to address the response bias common in online surveys, researchers employed a diverse recruitment

strategy to ensure a representative sample. The researcher reached out to OVO application users through various channels, including social media platforms such as WhatsApp, Telegram, Twitter, Instagram, etc. By utilizing multiple platforms, researchers aimed to reduce the risk of excluding certain user groups that may have different usage patterns or experiences. Additionally, the researcher encouraged participants to share the survey link with their acquaintances who were OVO application users, which helped the researcher reach a wider audience.

Secondly, to mitigate non-response bias, researchers made efforts to maximize the response rate and minimize missing data. Extended the survey duration from February 9 to March 16, 2023, allowing participants ample time to complete the questionnaire at their convenience. The researcher also sent out reminders at regular intervals to encourage respondents to participate. Moreover, to handle incomplete or inappropriate responses, researchers implemented validation checks within the online survey platform (Google Forms) to ensure that all required questions were answered and responses within a reasonable range were recorded. In the case of incomplete or inappropriate responses, the researcher excluded them from the final analysis to maintain the validity of the study.

The samples obtained during the deployment were 166 respondents, but 11 were not users of the OVO application, resulting in a remaining sample size of 155 respondents. The User Experience Questionnaire is a questionnaire that provides an overview of the level of user satisfaction based on user experience. The user experience questionnaire has been tested in several cases to provide an overview of user satisfaction. It usually takes 3-5 minutes to read and complete the user experience questionnaire. One of the other advantages of the user experience questionnaire is its free availability, which is available in the Indonesian language version. User experience questionnaire data analysis was carried out using the UEQ Data Analysis Tool, which compared the value of each aspect with existing product data [13].

The user experience questionnaire consisted of six scales divided into 26 indicator questions, as shown in Table 1. The user experience questionnaire used a 7-point semantic differential scale. Respondents were asked to rate from 1 to 7 on 26 UEQ indicator items according to their subjective assessment. The User Experience Questionnaire (UEQ) was used to measure user experience, consisting of 26 question components covering various aspects [17]:

Attractiveness: The product should look attractive, enjoyable, friendly, and pleasant; Efficiency: I should perform my tasks with the product fast, efficient, and in a pragmatic way; Perspicuity: The product should be

easy to understand, clear, simple, and easy to learn; Dependability: The interaction with the product should be predictable, secure, and meets my expectations; Stimulation: Using the product should be interesting, exiting, and motivating; Novelty: The product should be innovative, inventive, and creatively designed.

The components of the UEQ questions based on the aspects assessed are shown in Table 1.

Table 1. UEQ Testing Instruments

Scale	Indicator		Item
Attractiveness	Inconvenient	Enjoyable	ATT1
	Good	Bad	ATT2
	Dislike	Gratifying	ATT3
	Uncomfortable	Comfortable	ATT4
	Attractive	Unattractive	ATT5
	User-friendly	User-unfriendly	ATT6
Perspicuity	Not understood	Understandable	PER1
	Easily understood	Hardly understood	PER2
	Complicated	Simple	PER3
	Clear	Confusing	PER4
Efficiency	Fast	Slow	EFF1
	Inefficient	Efficient	EFF2
	Impractical	Practical	EFF3
	Organized	Disorganized	EFF4
Dependability	Unpredictable	Predictable	DEP1
	Obstruct	Supportive	DEP2
	Safe	Unsafe	DEP3
	Meet expectations	Not meeting expectations	DEP4
Stimulation	Beneficial	Less beneficial	STI1
	Tedious	Engaging	STI2
	Unappealing	Interesting	STI3
	Motivational	Unmotivating	STI4
Novelty	Creative	Monotonous	NOV1
	Innovative	Conventional	NOV2
	Commonplace	Leading-edge	NOV3
	Conservative	Innovative	NOV4

John Brooke created the SUS questionnaire at the Digital Equipment Corporation in England in 1986 [14]. This questionnaire measures three crucial aspects. The first aspect is the effectiveness of using this technology to achieve user goals. The second aspect is efficiency, namely how much user effort and resources are expended in achieving these goals. The third aspect is satisfaction, or how satisfying is the user experience? [15]. The system usability scale questionnaire consists of 10 statements, as shown in Table 2. The system usability scale questionnaire uses a 5-point Likert scale. Respondents were asked to provide an assessment of "Strongly Disagree", "Disagree", "Neutral", "Agree", and "Strongly Agree" on the 10 SUS statements according to their subjective assessment. The System Usability Scale (SUS) measures the usability attributes of the OVO application, namely aspects of effectiveness, efficiency, satisfaction, satisfaction, easy to learn, ease to remember and few errors. SUS gives an overall score between 0 and 100. The SUS half section (odd statements, i.e. 1, 3, 5, 7, and 9) describes a

positive evaluation (items with positive polarity). The other half of the sections (even statements, i.e. 2, 4, 6, 8, and 10) depict negative evaluations (items with negative polarity). For items with positive polarity, answers were coded as 0 to 4 from disagreement to agreement. Whereas for items with negative polarity, the answers are coded from 4 to 0 [18]. The list of SUS statements is shown in Table 2.

Table 2. SUS Testing Instruments [14]

Questions	Item
I think that i would like to use this system frequently	SUS1
I found the system unnecessarily complex	SUS2
I thought the system was easy to use	SUS3
I think that I would need the support of a technical person to be able to use this system	SUS4
I found the various functions in this system were well integrated	SUS5
I thought there was too much inconsistency in this system	SUS6
I would imagine that most people would learn to use this system very quickly	SUS7
I found the system very cumbersome to use	SUS8
I felt very confident using the system	SUS9
I needed to learn a lot of things before I could get going with this system	SUS10

2.5. Data Analysis

At this stage, data inconsistencies analysis, quantitative data analysis, demographic analysis, and descriptive statistical analysis of the data that has been obtained are carried out. The data obtained will be processed using IBM SPSS Statistics 25, UEQ Data Analysis Tool, and Microsoft Excel. Additionally, measures were taken to mitigate potential errors or biases in the analysis and handle outliers or missing data.

To address potential errors or biases in the data analysis process, several steps were taken. Firstly, data inconsistencies were analyzed using the UEQ Data Analysis Tool Version 12. This analysis involved assessing the seriousness of respondents answers to the questionnaire and identifying any suspicious data. Specifically, a critical value greater than 2 and a critical length value exceeding 15 were used as criteria to detect errors in questionnaire completion. In cases where such errors were identified, the respective data points were removed from the analysis.

In the analysis of quantitative data, a validity test and a reliability test will be carried out on the data that has been obtained. The validity test was carried out by looking at the Pearson correlation value of each indicator for each variable. In contrast, the reliability test was carried out by looking at Cronbach's alpha (α) value of each research variable.

Regarding demographic analysis, respondent data was categorized based on gender, age, duration of use, and frequency of use. This categorization allowed for a better understanding of potential variations in user experience and usability based on these demographic

factors. The results of the demographic analysis were presented in the form of charts or graphs to facilitate comprehension and interpretation.

In the descriptive statistical analysis using UEQ Data Analysis Tool Version 12 and Microsoft Excel. The data presented in this descriptive statistical test shows data that can be seen from the mean, which is the average value of each measured scale; the maximum, which is the highest value of each measured scale; the minimum, which is the lowest value of each measured scale, and the standard deviation used to determine the distribution of data from the sample and used to describe each research variable.

To handle outliers or missing data, specific procedures were implemented. Outliers, which are data points that deviate significantly from the overall pattern, were identified and assessed for their impact on the analysis results. Depending on the nature and extent of the outliers, options such as excluding them from the analysis or conducting sensitivity analyses were considered. Additionally, missing data points were identified, and appropriate strategies, such as imputation techniques or the exclusion of incomplete cases, were employed to ensure a comprehensive analysis.

By implementing these measures, researchers aimed to mitigate potential errors or biases in the data analysis process and address outliers or missing data effectively. These steps enhance the robustness and reliability of the findings, providing a more comprehensive assessment of the user experience and usability of the OVO application.

Several rules must be considered when transforming scores on questionnaire data using UEQ: Each answer in the UEQ questionnaire is rated on a scale of 1 to 7, indicating the level of user acceptance from "negative" to "positive"; These items have a scale from -3 to +3. Thus, -3 represents the most negative answer, 0 is a neutral answer, and +3 is the most positive answer [19].

After the data transformation, only the average or mean assessment can be carried out for each scale or question item on the UEQ from each respondent's answer. The following are the rules for the average or mean rating scale in UEQ which can be seen in Table 3.

Table 3. UEQ Mean Rating Scale

Mean Value Range	Explanation
> 0.8	Positive Evaluation
-0.8 – 0.8	Neutral Evaluation
< -0.8	Negative Evaluation

If the mean value of an item is more significant than 0.8, then the item will enter into the positive evaluation category and, in the diagram, is in the green area. If the mean value of an item is between -0.8 to 0.8, then the item will fall into the normal or neutral evaluation category and, in the diagram, is in the yellow area.

Meanwhile, if the mean value of an item is less than -0.8, then the item will enter into the negative evaluation category and in the diagram, it is in the red area.

Then several rules must be considered when calculating scores on questionnaire data using SUS:

For every odd-numbered question (1, 3, 5, 7, 9), the score obtained from user responses will be reduced by 1 can be seen in Equation 1.

$$\text{odd weight} = xi - 1 \quad (1)$$

Equation 2 shows that each even-numbered question (2, 4, 6, 8, 10) will have its final score calculated by subtracting the user's score from 5

$$\text{even weight} = 5 - xi \quad (2)$$

The SUS score is obtained by summing up the scores of each question and then multiplying it by 2.5 as seen in Equation 3.

$$\text{SUS Score} = (\text{odd weight} + \text{even weight}) \times 2,5 \quad (3)$$

The scoring rules mentioned in Equation 3 apply to one respondent. In Equation 4 for multiple respondents, the SUS scores of each respondent are summed up and then divided by the number of respondents to calculate the average SUS score.

$$\bar{x} = \frac{\sum x}{n} \quad (4)$$

Where \bar{x} represents the average SUS score, $\sum x$ denotes the sum of the SUS scores, and n indicates the number of respondents.

3. Results and Discussions

3.1. Analysis of Data Inconsistencies

To ensure the reliability of the responses, the study employed UEQ Data Analysis Tools Version 12 to analyze data inconsistencies, the result can be seen in Table 4. This approach adds value to the study by filtering out potentially careless or insincere responses, thereby enhancing the credibility of the results. By setting critical value parameters, the seriousness of respondents responses to the questionnaire was assessed, and any haphazard or insincere responses were identified, along with suspicious data. To identify errors in completing the questionnaire, a critical value greater than two and a critical length exceeding 15 were considered, indicating inconsistencies. Based on these criteria, it is recommended to remove such data from the analysis. This rigorous step strengthens the methodological aspects of the study and underscores the significance of accurate and thoughtful input from respondents in deriving reliable and insightful conclusions.

After the update by removing data, 155 respondent data was reduced to only 148.

Table 4. Inconsistencies Data

No	Scales with inconsistent answers							Critical length
	Attractiveness	Perspicuity	Efficiency	Dependability	Stimulation	Novelty	Critical?	Same answer for
32		1			1	1	3	12
44							0	26
112	1	1		1			3	14
126		1	1		1	1	4	9
132	1	1	1	1			6	26
136		1	1		1	1	4	18
153	1	1		1	1	1	5	11

3.2. Quantitative Data Analysis

In this stage, a testing or pilot test is conducted before the actual research to assess the suitability of a questionnaire. Before data collection, the author distributed the questionnaire to 30 respondents to test its validity and reliability. Valid questionnaires with reliable indicators were distributed to respondents who met the criteria and matched the predetermined sample size.

The validity test is conducted on 30 respondents to assess the validity of the questionnaire. An item is considered valid if the calculated r exceeds the tabled r. The significance level is 0.05 or 5% for 30 respondents, resulting in a tabled r of 0.361. Table 5 until 11 are the validity test results for each variable item.

Table 5. Validity test of the attractiveness scale

Item	The calculated r value	The tabled r value	Description
ATT1	0.646	0.361	Valid
ATT2	0.751	0.361	Valid
ATT3	0.877	0.361	Valid
ATT4	0.582	0.361	Valid
ATT5	0.713	0.361	Valid
ATT6	0.538	0.361	Valid

Table 5 demonstrates that all items in the attractiveness scale are deemed valid as the calculated r values are more significant than the tabled r value.

Table 6. Validity test of the perspicuity scale

Item	The calculated r value	The tabled r value	Description
PER1	0.373	0.361	Valid
PER2	0.523	0.361	Valid
PER3	0.464	0.361	Valid
PER4	0.712	0.361	Valid

Table 6 shows that all items in the perspicuity scale are considered valid as the calculated r values are more significant than the tabled r value.

Table 7. Validity test of the efficiency scale

Item	The calculated r value	The tabled r value	Description
EFF1	0.645	0.361	Valid
EFF2	0.713	0.361	Valid
EFF3	0.678	0.361	Valid
EFF4	0.660	0.361	Valid

Table 7 shows that all items in the efficiency scale are deemed valid as the calculated r values are more significant than the tabled r value.

Table 8. Validity test of the dependability scale

Item	The calculated r value	The tabled r value	Description
DEP1	0.720	0.361	Valid
DEP2	0.444	0.361	Valid
DEP3	0.504	0.361	Valid
DEP4	0.788	0.361	Valid

Table 8 shows that all items in the dependability scale are considered valid as the calculated r values are more significant than the tabled r value.

Table 9. Validity test of the stimulation scale

Item	The calculated r value	The tabled r value	Description
STI1	0.666	0.361	Valid
STI2	0.776	0.361	Valid
STI3	0.578	0.361	Valid
STI4	0.741	0.361	Valid

Table 9 displays that all items in the stimulation scale are deemed valid as the calculated r values are more significant than the tabled r value.

Table 10. Validity test of the novelty scale

Item	The calculated r value	The tabled r value	Description
NOV1	0.754	0.361	Valid
NOV2	0.469	0.361	Valid
NOV3	0.615	0.361	Valid
NOV4	0.544	0.361	Valid

Table 10 shows that all items in the novelty scale are considered valid as the calculated r values are more significant than the tabled r value.

Table 11. Validity test of SUS

Item	The calculated r value	The tabled r value	Description
SUS1	0.475	0.361	Valid
SUS2	0.822	0.361	Valid
SUS3	0.423	0.361	Valid
SUS4	0.687	0.361	Valid
SUS5	0.387	0.361	Valid
SUS6	0.782	0.361	Valid
SUS7	0.433	0.361	Valid
SUS8	0.785	0.361	Valid
SUS9	0.406	0.361	Valid
SUS10	0.671	0.361	Valid

Table 11 shows that all items on the SUS are deemed valid because the computed r-value is greater than the critical r-value.

Reliability Test results can be seen in Table 12 until 14.

Table 12. Reliability Test of 26 UEQ Items

Reliability Statistics	
Cronbach's Alpha	N of Items
0.993	26

Table 12 shows that all indicator items in the UEQ are deemed reliable because the Cronbach's alpha values are greater than 0.60.

Table 13. Reliability Test of the 6 UEQ Scales

Scale	Cronbach's Alpha
Attractiveness	0.90
Perspicuity	0.90
Efficiency	0.87
Dependability	0.80
Stimulation	0.91
Novelty	0.83

Table 13 displays that all scales in the UEQ are considered reliable as the Cronbach's alpha values are greater than 0.60.

Table 14. Reliability Test of SUS

Reliability Statistics	
Cronbach's Alpha	N of Items
0.795	10

Table 14 shows that all items in the SUS questionnaire are considered reliable, as the Cronbach's alpha values are greater than 0.60.

3.3. Analysis of Demographic Data

Respondents' characteristics can be grouped based on gender, age, residence, highest education level, duration of app usage, and intensity of app usage to provide an overview of the respondents' conditions. Table 15 until 20 are the results of the demographic analysis of the respondents:

Table 15. Characteristics Based on Gender

Gender	Total	Percentage
Male	52	35.1%
Female	196	64.9%

According to Table 15, there are 52 male respondents, accounting for 35.1% of the total, and 96 female respondents, accounting for 64.9%. Therefore, it can be concluded that the majority of OVO users in this study are female.

Table 16. Characteristics Based on Age

Age	Total	Percentage
15 – 23 years	112	75.7%
24 – 32 years	14	9.5%
33 – 41 years	10	6.8%
42 – 50 years	10	6.8%
> 50 years	2	1.4%

According to Table 16, it can be seen that there are 112 respondents aged 15-23 years old, accounting for 75.7%, 14 respondents aged 24-32 years old, accounting for 9.5%, 10 respondents aged 33-41 years old, accounting for 6.8%, 10 respondents aged 42-50 years old, accounting for 6.8%, and 2 respondents aged over 50 years old, accounting for 1.4%. Therefore, it can be concluded that the majority of OVO users in this study are aged between 15 and 23 years old.

Table 17. Characteristics Based on Residence

Residence	Total	Percentage
Banten	3	2%
Bengkulu	1	0.7%
DKI Jakarta	37	25%
West Java	25	16.9%
Central Java	1	0.7%
East Java	4	2.7%
Bangka Belitung	6	4.1%
Riau	1	0.7%
Lampung	1	0.7%
South Sulawesi	1	0.7%
South Sumatra	66	44.6%
North Sumatra	2	1.4%

According to Table 17, it can be seen that there are respondents from various provinces in Indonesia. There are three respondents (2%) from Banten Province, 1 respondent (0.7%) from Bengkulu Province, 37 respondents (25%) from DKI Jakarta Province, 25 respondents (16.9%) from West Java Province, 1 respondent (0.7%) from Central Java Province, four respondents (2.7%) from East Java Province, six respondents (4.1%) from Bangka Belitung Islands Province, 1 respondent (0.7%) from Riau Islands Province, 1 respondent (0.7%) from Lampung Province, one respondent (0.7%) from South Sulawesi Province, 66 respondents (44.6%) from South Sumatra Province, and two respondents (1.4%) from North Sumatra Province.

From the data, most OVO users involved in this study are from South Sumatra Province.

Table 18. Characteristics based on highest education level

Highest education level	Total	Percentage
Elementary School	1	0.7%
Junior High School	5	3.4%
Senior High School	97	65.5%
Diploma	2	1.4%
Bachelor's Degree	40	27%
Master's Degree	1	0.7%
Other	2	1.4%

According to Table 18, can be seen that there is one respondent with an elementary school education, accounting for 0.7%, five respondents with a junior high school education, accounting for 3.4%, 97 respondents with a senior high school education, accounting for 65.5%, two respondents with a diploma (D3) education, accounting for 1.4%, 40 respondents with a bachelor's degree (S1) education, accounting for 27%, one respondent with a master's degree (S2)

education, accounting for 0.7%, and two respondents with other last education, accounting for 1.4%. Therefore, most OVO users in this study have a high school education.

Table 19. Characteristics based on usage duration

Usage duration	Total	Percentage
< 1 year	33	22.3%
1 – 3 years	57	38.5%
> 3 years	58	39.2%

Based on Table 19, it can be seen that 33 respondents have been using the OVO application for less than one year, accounting for 22.3%, 57 respondents who have been using the OVO application for 1 to 3 years, accounting for 38.5%, and 58 respondents who have been using the OVO application for more than three years, accounting for 39.2%. Therefore, most OVO users in this study have used the OVO application for over three years.

Table 20. Characteristics based on usage intensities

Usage intensities	Total	Percentage
Rarely	29	19.6%
Sometimes	52	35.1%
Frequently	60	40.5%
Very frequently	7	4.7%

According to Table 20, it can be seen that 29 respondents rarely use the OVO application, accounting for 19.6% of the total, 52 respondents sometimes use the OVO application, accounting for 35.1%, 60 respondents who frequently use the OVO application, accounting for 40.5%, and seven respondents who use the OVO application very frequently, accounting for 4.7%. Most OVO users in this study frequently use the OVO application.

3.4. Descriptive Statistical Analysis

Descriptive statistical analysis of the User Experience Questionnaire (UEQ) method is conducted by calculating the mean scores for each UEQ scale and each question item. However, data transformation needs to be performed before conducting the descriptive statistical analysis. Table 21 and Figure 2 show the mean scores for pragmatic and hedonic quality.

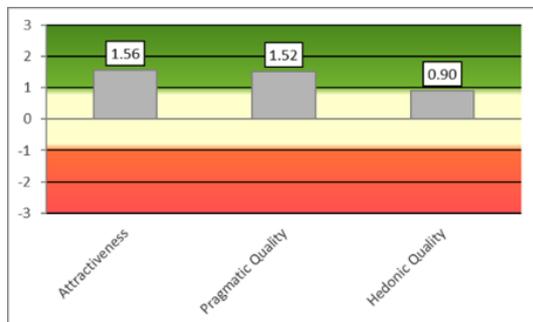


Figure 2. Graph of Mean Scores for Pragmatic and Hedonic Quality

Based on the data presented in Table 21 and Figure 2, it can be observed that attractiveness, pragmatic quality, and hedonic quality of the OVO application are in the green area, indicating positive evaluation scores. Table 22 and 23, also Figure 2 and 3 show the mean scores for the 6 UEQ scales.

Table 21. Mean Score of Pragmatic and Hedonic Quality

Pragmatic and Hedonic Quality	
Attractiveness	1.56
Pragmatic Quality	1.52
Hedonic Quality	0.90

Based on the data presented in Table 22 and Figure 3, it can be observed that the attractiveness scale, perspicuity scale, efficiency scale, dependability scale, and stimulation scale of the OVO application are in the green area, indicating positive evaluation scores. On the other hand, the novelty scale is in the yellow area, indicating a neutral evaluation score.

Table 22. Mean Scores of the 6 UEQ Scales

UEQ Scales	
Attractiveness	1.563
Perspicuity	1.671
Efficiency	1.546
Dependability	1.331
Stimulation	1.162
Novelty	0.644

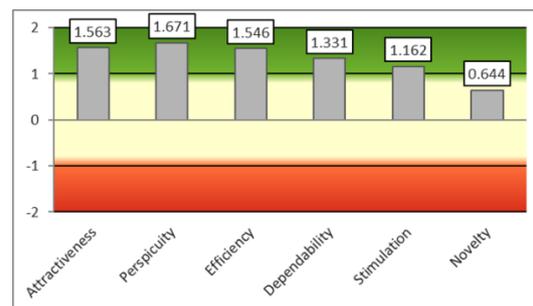


Figure 3. Graph of Mean Scores for UEQ Scales

Based on Table 23, it can be observed that all UEQ items received positive evaluations except for item STI2 (boring/engaging), item NOV2 (conventional/innovative), item NOV3 (ordinary/leading), and item NOV4 (conservative/innovative) which received neutral evaluations.

Table 23. The mean scores for each UEQ item

Item	Mean	Left	Right
ATT1	1.5	troublesome	pleasant
PER1	1.8	incomprehensible	understandable
NOV1	1.0	creative	monotone
PER2	1.6	easy to learn	hard to learn
STI1	1.5	beneficial	less useful
STI2	0.7	boring	exciting
STI3	1.2	not attractive	interesting
DEP1	0.9	unpredictable	predictable
EFF1	1.4	fast	slow
NOV2	0.3	inventive	conventional
DEP2	1.6	obstruct	support
ATT2	1.8	Good	bad

Item	Mean	Left	Right
PER3	1.5	complicated	simple
ATT3	1.4	dislike	exhilarating
NOV3	0.7	common	front
ATT4	1.6	uncomfortable	comfortable
DEP3	1.5	safe	not safe
STI4	1.2	motivating	not motivating
DEP4	1.3	meet expectations	expectations
EFF2	1.5	not efficient	efficient
PER4	1.8	clear	confusing
EFF3	1.8	impractical	practical
EFF4	1.5	organized	untidy
ATT5	1.4	attractive	not attractive
ATT6	1.8	user friendly	not user friendly
NOV4	0.7	conservative	innovative

Based on Table 24 and Figure 4, the benchmark results indicate that the efficiency scale is rated "Good". The attractiveness, perspicuity, dependability, and stimulation scales are rated as "Above Average". However, the novelty aspect is still rated as "Below Average".

Table 24. Benchmark Results

Scale	Mean	Comparison to benchmark
Attractiveness	1.56	Above Average
Perspicuity	1.67	Above Average
Efficiency	1.55	Good
Dependability	1.33	Above Average
Stimulation	1.16	Above Average
Novelty	0.64	Below Average

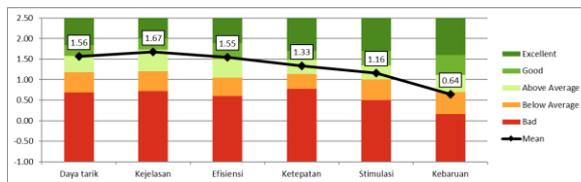


Figure 4. Benchmark Results Graph

Descriptive statistical analysis using the System Usability Scale (SUS) method in this study involved calculating the average or mean SUS scores. However, before conducting the descriptive statistical analysis, data transformation is necessary using the rules of the SUS method. After the data transformation, the average or mean SUS scores can be assessed. From the final scores, it can be determined whether the system is rated as good or not. Table 25 are the results of calculating the average or mean SUS scores using the rules of the SUS method.

Table 25. Summary of SUS Score Calculation Results

R	Calculated Score		R	Calculated Score	
	Total	Score		Total	Skor
1	29	72.5	75	38	95
2	33	82.5	76	20	50
3	40	100	77	40	100
4	40	100	78	22	55
5	30	75	79	32	80
6	30	75	80	28	70
7	26	65	81	37	92.5
8	29	72.5	82	37	92.5
9	29	72.5	83	30	75
10	28	70	84	36	90

R	Calculated Score		R	Calculated Score	
	Total	Score		Total	Skor
11	30	75	85	34	85
12	33	82.5	86	20	50
13	35	87.5	87	37	92.5
14	34	85	88	23	57.5
15	28	70	89	28	70
16	36	90	90	34	85
17	34	85	91	29	72.5
18	34	85	92	24	60
19	28	70	93	40	100
20	39	97.5	94	35	87.5
21	40	100	95	35	87.5
22	40	100	96	35	87.5
23	38	95	97	36	90
24	34	85	98	33	82.5
25	32	80	99	40	100
26	18	45	100	26	65
27	40	100	101	25	62.5
28	26	65	102	30	75
29	25	62.5	103	33	82.5
30	28	70	104	30	75
31	31	77.5	105	19	47.5
32	29	72.5	106	32	80
33	23	57.5	107	28	70
34	36	90	108	36	90
35	24	60	109	24	60
36	39	97.5	110	29	72.5
37	34	85	111	21	52.5
38	26	65	112	35	87.5
39	26	65	113	36	90
40	33	82.5	114	26	65
41	36	90	115	37	92.5
42	29	72.5	116	40	100
43	26	65	117	28	70
44	30	75	118	40	100
45	32	80	119	32	80
46	35	87.5	120	25	62.5
47	34	85	121	40	100
48	35	87.5	122	37	92.5
49	29	72.5	123	32	80
50	33	82.5	124	37	92.5
51	35	87.5	125	35	87.5
52	31	77.5	126	31	77.5
53	36	90	127	25	62.5
54	38	95	128	20	50
55	40	100	129	18	45
56	29	72.5	130	31	77.5
57	34	85	131	20	50
58	25	62.5	132	30	75
59	36	90	133	31	77.5
60	24	60	134	20	50
61	26	65	135	16	40
62	27	67.5	136	20	50
63	26	65	137	36	90
64	33	82.5	138	35	87.5
65	34	85	139	28	70
66	33	82.5	140	20	50
67	34	85	141	39	97.5
68	38	95	142	33	82.5
69	26	65	143	34	85
70	28	70	144	34	85
71	30	75	145	29	72.5
72	40	100	146	33	82.5
73	25	62.5	147	26	65
74	29	72.5	148	22	55
Average SUS score					
77.53					

Table 25 and Figure 5 summarize SUS score calculations from the questionnaires distributed to 148 respondents, resulting in an average or mean score of

77.53 according to the System Usability Scale (SUS) method.

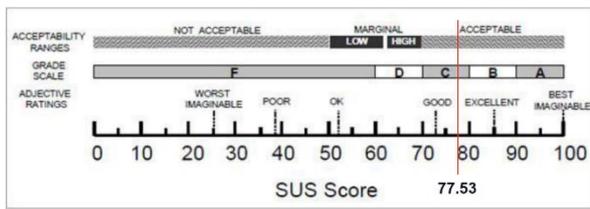


Figure 5. SUS Score Value [20]

After obtaining the SUS score, the next step is interpreting the results. There are three perspectives to determine the interpretation of the SUS score calculations:

Acceptability ranges as seen in Figure 6 consist of three levels: not acceptable, marginal (low and high), and acceptable. Acceptability is used to assess the level of user acceptance of the application.

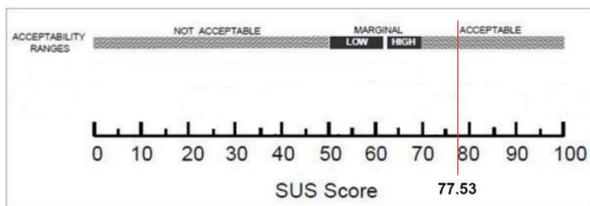


Figure 6. Acceptability Ranges

In the calculation of the SUS score, the previously obtained score was 77.53, indicating that the user acceptance level of the OVO application is categorized as "ACCEPTABLE".

The grade scale consists of A, B, C, D, and F, which are used to determine the grade level of the application can be seen in Figure 7.

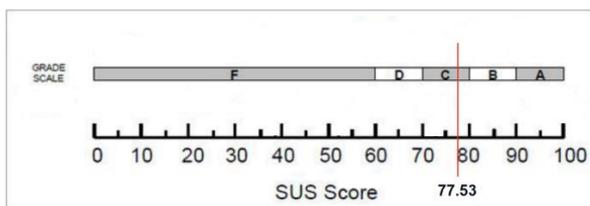


Figure 7. Grade Scales

In the calculation of the SUS score, the previously obtained score was 77.53, indicating that the OVO application falls under the "C" grade.

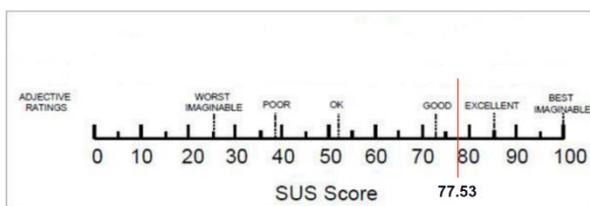


Figure 8. Adjective Ratings

Adjective ratings consist of the categories worst imaginable, poor, ok, good, and best imaginable. Adjective ratings are used to determine the rating of the application can be seen in Figure 8.

In the calculation of the SUS score, the previously obtained score was 77.53, indicating that the OVO application falls under the "GOOD" category.

3.5. Improvement Recommendations

Based on the evaluation of the user experience conducted using the User Experience Questionnaire (UEQ) method and processed with UEQ Data Analysis Tool Version 12, the recommendations can be made for enhancing the OVO application: Consider incorporating more innovative, cutting-edge, and creative services or features into the OVO application. This could involve adopting new and innovative features that align with current trends in the industry. These additions will help improve novelty, ensuring the application stays up-to-date and provides a unique user experience; Explore the inclusion of more engaging services or features within the OVO application. For instance, consider incorporating gamification elements that allow users to earn OVO Points or integrating captivating animations and enjoyable sound effects. Such enhancements will enhance the stimulation aspect of the application, making the overall user experience more enjoyable and interactive.

By implementing these recommendations, it is anticipated that the OVO application can address the identified issues and provide a more satisfactory user experience.

4. Conclusion

The user experience and usability evaluation using the User Experience Questionnaire (UEQ) and System Usability Scale (SUS) has been successfully conducted, involving 148 competent respondents who assessed the application based on gender, age, duration of usage, and frequency of usage. The benchmark results for the six aspects of UEQ show that one aspect, efficiency, falls into the "good" category with a mean value of 1.55. Additionally, four aspects, namely attractiveness (mean: 1.56), perspicuity (mean: 1.67), dependability (mean: 1.33), and stimulation (mean: 1.16), are classified as "above average" categories. However, one aspect, novelty, falls into the "below average" category with a mean value of 0.64. Regarding the measurement of OVO application usability using the System Usability Scale (SUS) method, the obtained score is 77.53. This score falls within the "Acceptable" range in the Acceptability Ranges category, a "C" grade in the Grade Scale category, and is rated as "Good" in the Adjective Ratings category.

References

- [1] K. Khando, M. S. Islam, and S. Gao, "The Emerging Technologies of Digital Payments and Associated Challenges: A Systematic Literature Review," *Futur. Internet*, vol. 15, no. 1, 2023, doi: 10.3390/fi15010021.
- [2] A. F. Fainusa, R. Nurcahyo, and M. Dachyar, "Conceptual Framework for Digital Wallet User Satisfaction," *ICETAS 2019 - 2019 6th IEEE Int. Conf. Eng. Technol. Appl. Sci.*, pp. 1–4, 2019, doi: 10.1109/ICETAS48360.2019.9117285.
- [3] W. Rachbini, T. Evi, I. S. Pramesworo, and E. Rahmawati, "Effect of E-Wallet Experience on E-Satisfaction and E-Word of Mouth (Study on E-Wallet Applications in Indonesia)," *Cent. Eur. Manag. J.*, vol. 30, pp. 302–316, 2022, doi: 10.57030/23364890.cemj.30.4.25.
- [4] S. N. Shetu, M. M. Islam, and S. I. Promi, "An Empirical Investigation of the Continued Usage Intention of Digital Wallets: The Moderating Role of Perceived Technological Innovativeness," *Futur. Bus. J.*, vol. 8, no. 1, p. 43, 2022, doi: 10.1186/s43093-022-00158-0.
- [5] N. Setiyawati and D. H. Bangkalang, "The Comparison of Evaluation on User Experience and Usability of Mobile Banking Applications Using User Experience Questionnaire and System Usability Scale," *Proceedings*, vol. 82, p. 87, 2022, doi: 10.3390/proceedings2022082087.
- [6] B. Foster, R. Hurriyati, and M. D. Johansyah, "The Effect of Product Knowledge, Perceived Benefits, and Perceptions of Risk on Indonesian Student Decisions to Use E-Wallets for Warunk Upnormal," *Sustain.*, vol. 14, no. 11, 2022, doi: 10.3390/su14116475.
- [7] H. C. Prakoso, M. Iqbal, and A. Kurniawati, "Usability Evaluation OVO Electronic Wallet Application Using Usability Testing and Usefulness, Satisfaction, Ease Of Use Questionnaire," vol. 8, no. 10, pp. 1–6, 2020, doi: <https://doi.org/10.26438/ijcse/v8i10.16>.
- [8] G. E. Saputra, Rakhmi Khalida, and Ratu Nurmalika, "Evaluation of User Experience TLX Training Gate for Competitive Programming Learning Using User Experience Questionnaire and System Usability Scale," *Int. J. Sci. Technol.*, vol. 1, no. 2, pp. 30–37, 2022, doi: 10.56127/ijst.v1i2.142.
- [9] B. Laugwitz, T. Held, and M. Schrepp, "Construction and Evaluation of a User Experience Questionnaire," *HCI Usability Educ. Work*, vol. 5298, pp. 63–76, 2008, doi: 10.1007/978-3-540-89350-9_6.
- [10] H. B. Santoso, M. Schrepp, R. Yugo Kartono Isal, A. Y. Utomo, and B. Priyogi, "Measuring user experience of the student-centered E-learning environment," *J. Educ. Online*, vol. 13, no. 1, pp. 1–79, 2016.
- [11] J. Brooke, "SUS: A 'Quick and Dirty' Usability Scale," *Usability Eval. Ind.*, no. November 1995, pp. 207–212, 2020, doi: 10.1201/9781498710411-35.
- [12] Z. Sharfina and H. B. Santoso, "An Indonesian Adaptation of the System Usability Scale (SUS)," *2016 Int. Conf. Adv. Comput. Sci. Inf. Syst. ICACISIS 2016*, pp. 145–148, 2017, doi: 10.1109/ICACISIS.2016.7872776.
- [13] I. M. A. O. Gunawan, G. Indrawan, and Sariyasa, "User experience evaluation of academic progress information systems using retrospective think aloud and user experience questionnaire," *J. Phys. Conf. Ser.*, vol. 1810, no. 1, 2021, doi: 10.1088/1742-6596/1810/1/012015.
- [14] J. Brooke, "SUS: A 'Quick and Dirty' Usability Scale," *Usability Eval. Ind.*, pp. 207–212, 2020, doi: 10.1201/9781498710411-35.
- [15] W. H. Cheah, N. Mat Jusoh, M. M. T. Aung, A. Ab Ghani, and H. Mohd Amin Rebutan, "Mobile Technology in Medicine: Development and Validation of an Adapted System Usability Scale (SUS) Questionnaire and Modified Technology Acceptance Model (TAM) to Evaluate User Experience and Acceptability of a Mobile Application in MRI Safety Screening," *Indian J. Radiol. Imaging*, vol. 33, no. 1, pp. 36–45, 2023, doi: 10.1055/s-0042-1758198.
- [16] R. T. Andi and M. M. Ali, "Analysis of the Influence of Lifestyle on Purchasing Decision for Samsung Smartphone Products in Jakarta," *Eur. J. Bus. Manag. Res.*, vol. 4, no. 4, 2019, doi: 10.24018/ejbmr.2019.4.4.72.
- [17] A. Hinderks, M. Schrepp, F. J. Domínguez Mayo, M. J. Escalona, and J. Thomaschewski, "Developing a UX KPI based on the user experience questionnaire," *Comput. Stand. Interfaces*, vol. 65, no. July, pp. 38–44, 2019, doi: 10.1016/j.csi.2019.01.007.
- [18] M. Schrepp, "A Comparison of SUS, UMUX-LITE, and UEQ-S," *J. User Exp.*, vol. 18, no. 2, pp. 86–104, 2023.
- [19] M. Schrepp, "User Experience Questionnaire Handbook Version 8," *URL* https://www.Res.net/publication/303880829_User_Experience_Questionnaire_Handbook_Version_2 (Accessed 02.02. 2017), pp. 1–15, 2019, [Online]. Available: www.ueq-online.org.
- [20] John Brooke Independent, "SUS: A Retrospective," *Physiologist*, vol. 8, no. 2, pp. 29–40, 2013.