Assessing User Experience of ChatGPT Website Employing the User Experience Questionnaire (UEQ)

[1]Dila Okta Dwi Putri, [2]Dinda Lestarini[*]
[1,2]Sistem Informasi, Fakultas Ilmu Komputer, Universitas Sriwijaya
[1,2]Jl. Raya Palembang - Prabumulih No.KM. 32, Indralaya
*e-mail: *dinda.lestarini@gmail.com*

(***received*:** 3 September 2024, ***revised*:** 30 October 2024, ***accepted*:** 11 November 2024)

Abstract

One form of information technology advancement is the use of AI in website development, one of which is ChatGPT website. ChatGPT has a poor user experience in various aspects, so it is necessary to evaluate the user experience on the ChatGPT website using User Experience Questionnaire (UEQ). The attractiveness variable obtained a positive evaluation value of 1.503. The perspicuity variable obtained a positive evaluation value of 1.661. The efficiency variable obtained a positive evaluation value of 1.615. The dependability variable obtained a positive evaluation value with an overall average value of 1.286. The stimulation variable obtained a positive evaluation value of 1.182. The novelty variable obtained a positive evaluation value of 0.942. The ChatGPT website has shown good quality because it has a positive evaluation value from user assessments. However, in the attractiveness, dependability, and novelty variables, there are still several items that get neutral ratings. So product improvements are still needed to increase user satisfaction.

Keywords : AI, user experience, user experience questionnaire, website

1 Introduction

The era of globalization makes the progress and development of information technology grow rapidly, which can affect all activities, giving rise to new business opportunities through technology [1]. The development of technology aims to facilitate and improve the efficiency of activities, especially in terms of accessing and obtaining information [2]. One form of advancement in information technology is the use of artificial intelligence (AI) in website development. AI denotes the capability of automatons or computational algorithms to execute functions that usually require human cognitive abilities, including speech recognition, natural language understanding, decision-making, and problem-solving [3]. ChatGPT is an AI that uses AI-based language models to generate conversational text that resembles human interaction. ChatGPT is chatbot that utilizes artificial intelligence that is able to interact and assist in performing various tasks or activities [4]. ChatGPT OpenAI is a machine learning technology that applies deep learning models and uses artificial neural networks [5].

Based on the analysis conducted, ChatGPT has the disadvantage of not being able to display the reference requested by the prompt formulated by the user [4]. “ChatGPT has poor user experience in various aspects. These aspects or categories include ChatGPT providing irrelevant or off-topic information, users having difficulty in getting the desired output from ChatGPT, connections or systems experiencing frequent problems due to service overload or technical issues, ChatGPT having limited information in its database, ChatGPT failing to produce output that supports creative activities, ChatGPT repeating output or presenting similar information, and ChatGPT communicating in a poor manner” [6].

The ChatGPT website requires an evaluation of the UX to find out and assess the user experience when interacting with the ChatGPT website. Assessing UX of the ChatGPT website is executed utilizing the User Experience Questionnaire (UEQ) method. UEQ has advantages over other methods because it

 ISSN:2302-8149
e-ISSN:2540-9719

provides a comprehensive description of UX, ranging from classical usability to aspects of UX and is equipped with analytical tools to interpret evaluation results appropriately [7].

2 Literature Review

Prior research seek to acquire comparative data and reference benchmarks. In addition, previous research is used to avoid the assumption of similarity with this research. The first research conducted as a literature review in this study is research focuses on identifying user experience of ChatGPT website using an early user questionnaire study [6]. In this prior research reveals that ChatGPT user experience is influenced by both pragmatic and hedonic attributes, where users value practical outcomes and enjoyable interactions. The User Experience Questionnaire (UEQ) is a robust tool for assessing ChatGPT's UX, covering dimensions such as attractiveness, reliability, efficiency, and emotional aspects of interaction. Utilizing UEQ allows researchers to pinpoint ChatGPT's strengths and weaknesses in delivering a positive user experience. Prior researches indicate that users often experience dissatisfaction when ChatGPT fails to provide expected responses on the first attempt. This highlights the need for a systematic approach to UX evaluation, including thorough analysis of user interactions and the development of support for more effective prompt crafting. Connecting UX research on ChatGPT with prior findings can offer comprehensive insights into user interactions and opportunities for enhancing the experience.

The second research focuses on analyzing user experience of Jogjakita application as a medium for Yogyakarta residents to be able to express and inspire each other [8]. This research employs the User Experience Questionnaire (UEQ) to assess the UX of the JogjaKita, utilizing data from the user base to derive a measurement of user requirements. It aims to evaluate both positive and negative perceptions of the application. Researchers conducted the study to assess the quality of the application from the user's perspective. The User Experience Questionnaire (UEQ) furnishes an all-encompassing evaluation of UX, concentrating on facets of usability and experiential interaction. UEQ is proficient in assessing UX by evaluating dimensions such as functionality, appeal, and effectiveness. UEQ delivers a thorough evaluation through a survey that addresses various elements. In the context of the ChatGPT application, utilizing UEQ facilitates an understanding of user engagement and highlights areas needing refinement. Applying UEQ to ChatGPT enables the assessment of how well the application satisfies user needs and identifies its strengths and weaknesses. Overall, employing UEQ for UX evaluation provides crucial insights for further refinement and future improvements.

The third research focuses on analyzing user experience of Placeplus application as a platform that helps users find, explore, and get information about various locations or places [9]. Assessing User Experience (UX) is essential in developing interactive systems, such as AI applications like ChatGPT. The User-Centered Design (UCD) approach is pertinent to this research because it prioritizes user's needs and expectations throughout the design and evaluation phases. UCD employs iterative processes—analysis, design, evaluation, and implementation—to optimize UX. Ongoing UX evaluation facilitates the collection of user feedback, which is crucial for design enhancements. Based on this third prior research, it has been demonstrated that employing the UEQ method is more proficient in assessing a range of applications and systems, offering profound insights into user experience.

The fourth research is a study that analyzes user experience on chatbot reminders [10]. This research states that a comparison between chatbot reminder and ChatGPT shows that both have their own advantages and challenges in terms of user experience. Chatbot reminder, with its user-centered design, offers a simple interface and specific functionality for reminder management, which improves user satisfaction in specific contexts. Meanwhile, ChatGPT, as an AI-based language model, provides more complex and flexible dialog capabilities, enabling more in-depth interactions. However, users may need time to understand how to optimally utilize its features.

Overall, based on the literature review, a constellation of insights emerges concerning the strengths and limitations of preceding UX research. Prior studies have adeptly employed the User Experience Questionnaire (UEQ) to gauge user satisfaction and interaction quality, particularly in aspects of usability

 ISSN:2302-8149
e-ISSN:2540-9719

and emotive allure. The strength of these inquiries lies in their capacity to illuminate both pragmatic and hedonic dimensions of UX, thus providing a nuanced equilibrium between functionality and user contentment. Nonetheless, conspicuous limitations are evident in more specialized realms, such as prompt optimization and the facilitation of user guidance within complex AI features, as exemplified by ChatGPT.

This study introduces novelty by not only undertaking a holistic UX evaluation of ChatGPT through the UEQ but also pinpointing latent gaps where UX could be substantively augmented, particularly in guiding users toward enhancing prompt efficacy. Furthermore, it addresses prospective avenues for developing support systems aimed at ameliorating dissatisfaction when initial responses fall short of user anticipations. Hence, this research expands upon previous findings by offering actionable insights to refine ChatGPT's user experience, especially in optimizing user assistance and support.

3 Research Method

Researchers evaluate the user experience of the ChatGPT website using the UEQ as a research method. To facilitate the research, a research process or flow is used which is described in the research sequence as presented in Figure 1.

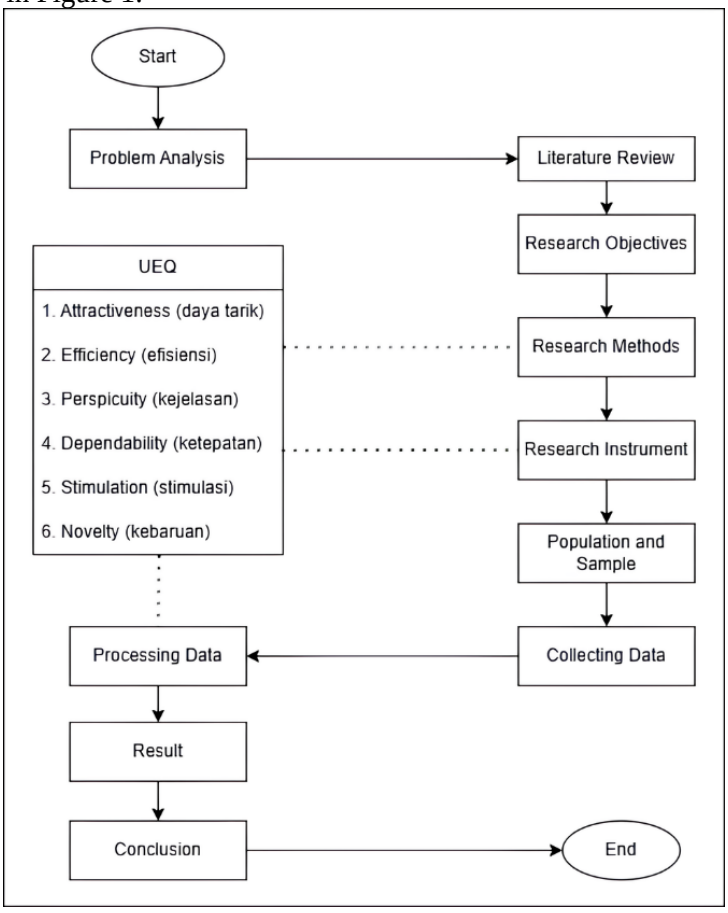

Figure 1. Research sequence

Based on Figure 1, the research sequence consist of problem analysis, literature review, research objectives, research methods, research instruments, population and sample, collecting data, processing data, result, and conclusion.

1) Problem Analysis
This stage is carried out by analyzing and formulating the problems that will be carried out in the research, as for the problems analyzed, namely regarding the level of website user satisfaction based on users on the ChatGPT website.
2) Literature Review
Literature review is carried out by searching for information from various sources such as journals, books, articles, trusted links, and previous research related to the author's research.
3) Research Objectives
The stage of determining research objectives is important in research, so that it can harmonize the results in accordance with the desired objectives. This research seeks to assess user experience through the ChatGPT website using the User Experience Questionnaire (UEQ) method.
4) Research Methods
The stage of determining the research method must be done before starting the other stages. The research method used in this research is the User Experience Questionnaire (UEQ) method.
5) Research Instruments
This research instrument was prepared employing the User Experience Questionnaire (UEQ) method. UEQ is an instrument designed to evaluate user experience in interacting with a product, system, or service [11]. The instrument is determined based on aspects of Attractiveness, Pragmatic Quality, and Hedonic Quality, which are then divided into 6 scales. UEQ questions are categorized into 6 scales, resulting in 26 questions representing aspects of user experience that can be assessed by respondents [12].
6) Population and Sample
Population is an area that includes subjects / objects that possess specific attributes and features to be generalized [13]. The population in this research includes all individuals utilizing of the ChatGPT website. The sample is minor portion taken from population members based on predetermined procedures to enable it to stand in for the population [14]. The sample in this study will be determined with the Lemeshow's formula. The Lemeshow's formula is used because the population in this study is very large and always changes [15]. Then the sample in this research can be determined using the Lemeshow's formula (1).

$$n = \frac{z^2 p\,(1-p)}{d^2} \tag{1}$$

With n denotes the quantity of samples, z is the standard value of 1.96, p is the maximum estimate of 50% or equal to 0.5, and d is alpha (0.10) or sampling error of 10.

$$n = \frac{1{,}96^2 \times\ 0{,}5\,(1-0{,}5)}{0{,}1^2} = 96{,}02 = 96 \tag{1}$$

Deriving from the aforementioned calculations, it can be deduced that the sample size in this study comprised 96 respondents.
7) Data Collection
Data collection is conducted as material in processing research data in the form of primary data such as questionnaires distributed to respondents who are users of the ChatGPT website so that responses or responses are given regarding the indicators of each variable. In this research, collecting data was conducted by disseminating questionnaires through Google Forms from July 3, 2024, to July 30, 2024. From the distributed questionnaires, the total of 176 respondents were acquired for this research.

 ISSN:2302-8149
e-ISSN:2540-9719

8) Data Processing
Data processing of all data that has been collected needs to be used to facilitate data or information that can be interpreted and analyzed further. Data processing was conducted using the UEQ Data Analysis Tool Version 12.
9) Result
After processing the data, results can be obtained regarding the evaluation value of the ChatGPT website user experience as a reference used in drawing conclusions.
10) Conclusion
Based on the data analyzed and processed with reference to the hypothesis according to the research method used, conclusions can be determined which are used as references for further research references.

4 Result and Analysis

Assessing user experience (UX) on the ChatGPT website employing the UEQ method was conducted by disseminating questionnaires using Google Forms to all individuals utilizing of the ChatGPT website. Of all the questionnaires distributed, 176 participants were engaged in this study. Based on the 176 respondents collected, they were categorized according to various factors, including age, gender, occupation, and usage frequency. The frequency and percentage of the respondent profiles are detailed in Table 1.

Table 1. Respondents profile

Summary	Frequency	Percentage
Age :		
≤ 20	58	20 %
21 – 30	79	44.9 %
31 – 40	27	15.3 %
41 – 50	7	4 %
≥ 50	5	2.8 %
Gender :		
Female	112	36.4 %
Male	64	63.6 %
Occupation :		
Student	34	19.3 %
College Student	95	54 %
State Civil Apparatus	9	5.1 %
Private Employees	18	10.2 %
Entrepreneur	16	9.1 %
And others	4	2.3 %
Usage Frequently :		
< 1 month	25	14.2 %
1 – 3 months	39	22.2 %
4 – 6 months	41	23.3 %
7 – 12 month	26	14.8 %
> 1 year	45	25.6 %

 ISSN:2302-8149
e-ISSN:2540-9719

The respondent data obtained based on Table 1 will be processed using the UEQ Data Analyst Tool, thus enabling a more structured and data-driven approach in improving the design and function of the ChatGPT website to better meet user needs and expectations. The UEQ methodology comprises 26 inquiries designed to assess user experience and identify areas that are insufficient and require enhancement. The findings related to the mean value, variance, standard deviation result in this study are depicted in Table 2.

Table 2. Mean, variance, and standard deviation results

No	Mean	Variance	Std. Dev.	Left	Right	Scale
1	1.8	1.2	1.1	annoying	enjoyable	Attractiveness
2	2.1	1.2	1.1	not understandable	understandable	Perspicuity
3	0.6	2.8	1.7	creative	dull	Novelty
4	1.3	3.2	1.8	easy to learn	difficult to learn	Perspicuity
5	1.4	4.0	2.0	valuable	inferior	Stimulation
6	0.9	2.1	1.4	boring	exciting	Stimulation
7	1.5	1.6	1.3	not interesting	interesting	Stimulation
8	0.8	2.6	1.4	unpredictable	predictable	Dependability
9	1.5	2.5	1.6	fast	slow	Efficiency
10	0.8	1.5	1.6	inventive	conventional	Novelty
11	1.7	2.0	1.2	obstructive	supportive	Dependability
12	1.5	1.9	1.4	good	bad	Attractiveness
13	1.7	1.3	1.4	complicated	easy	Perspicuity
14	1.6	2.0	1.1	unlikable	pleasing	Attractiveness
15	1.3	1.9	1.4	usual	leading edge	Novelty
16	1.8	1.2	1.1	unpleasant	pleasant	Attractiveness
17	1.2	2.8	1.7	secure	not secure	Dependability
18	1.0	2.1	1.5	motivating	demotivating	Stimulation
19	1.4	1.9	1.4	meets expectations	don’t meet expectation	Dependability
20	1.7	1.8	1.4	inefficient	efficient	Efficiency
21	1.6	1.6	1.3	clear	confusing	Perspicuity
22	1.9	1.4	1.2	impractical	practical	Efficiency
23	1.4	2.3	1.5	organized	cluttered	Efficiency
24	0.8	2.5	1.6	attractive	unattractive	Attractiveness
25	1.5	2.2	1.5	friendly	unfriendly	Attractiveness
26	1.2	2.6	1.6	conservative	innovative	Novelty

According to Table 2, the mean, variance, and standard deviation results for the 26 UEQ items are presented. UEQ encompasses six aspects, namely attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty [16]. The range of rating scale values is divided into three categories, namely values below -0.8 are considered negative, values between -0.8 to 0.8 are considered neutral, and values above 0.8 are considered positive [17]. Each item's value exceeds 0.8, indicating a favorable evaluation for all question items. The data illustrated in the figure serves as a basis for calculating the six scales. Subsequently, the largest and smallest scales of the UEQ will be examined and summarized. Each scale comprises several questions from the 26 items, and the average value is computed accordingly [18]. The summary of the descriptive statistical evaluation based on variables are presented in Table 3.

 ISSN:2302-8149
e-ISSN:2540-9719

Table 3. Summary of the descriptive statistical evaluation based on variables

Variable	Mean	Assessment
Attractiveness	1,512	Positive
Perspicuity	1,652	Positive
Efficiency	1,620	Positive
Dependability	1,285	Positive
Stimulation	1,184	Positive
Novelty	0,951	Positive

Based on table 3, it can be concluded that the first variable, which is attractiveness variable obtained a positive evaluation score showing an overall average value of 1.512. Then the perspicuity variable obtained a positive evaluation score, showing an overall average value of 1.652. The efficiency variable obtained a positive evaluation, reflecting an average score of 1.620. Then the dependability variable obtained a positive evaluation, reflecting an average score of 1.284. The stimulation variable obtained a positive evaluation, reflecting an average score of 1.184. Lastly, the novelty variable obtained a positive evaluation, reflecting an average score of 0.951.

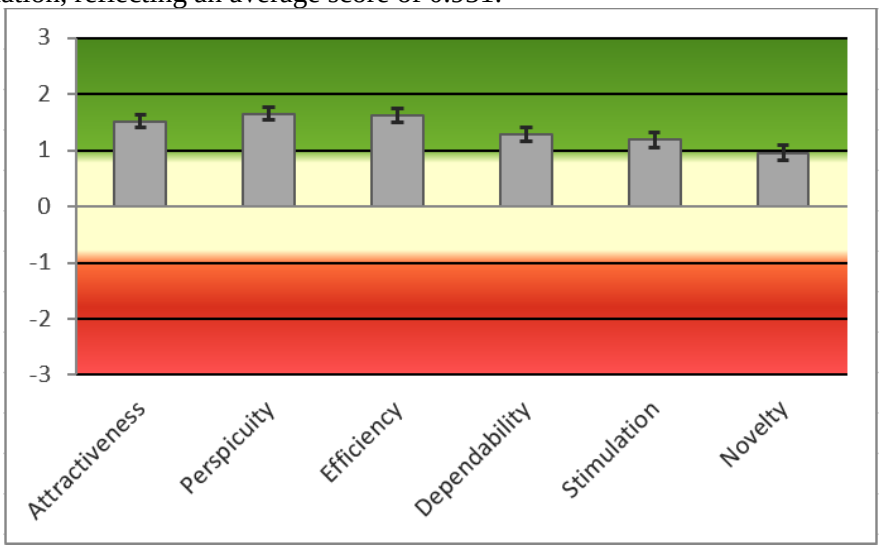

Figure 2. UEQ scale value of chatgpt website

Figure 2 presents the findings of the user experience assessment variables employing the UEQ method on the ChatGPT website involving 176 respondents. All aspects of UX measurement in this study, which are attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty, managed to get an average value (mean) above 0.8 or at a positive evaluation level indicated by the green area on the diagram. Of all aspects, the novelty aspect has a lower average value compared to other aspects. This is due to the users' perception that the ChatGPT website is a monotonous and generic system. This result is in line with the factor test which shows that the monotonous and general aspects have a mediocre impression, so efforts are needed to increase the value by creating a more creative and innovative design.

Each average value derived from the UEQ calculations is juxtaposed with the benchmark values outlined in Table 2 to elucidate the significance of each score. The comparative results of the UEQ scale measurements against the benchmark values are illustrated in Figure 3 and table 4 below.

 ISSN:2302-8149
e-ISSN:2540-9719

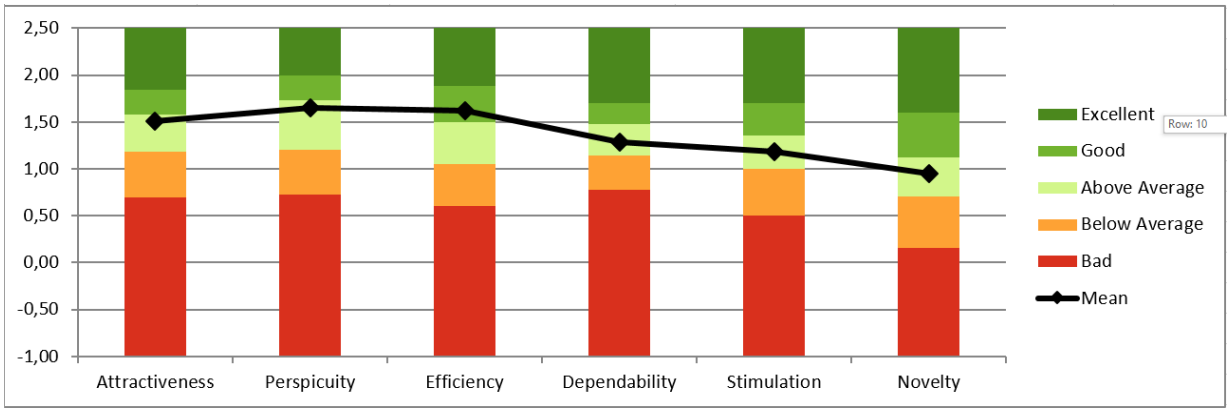

Figure 3. chatgpt website UEQ scale value benchmark results

Table 4 Interpretation of Benchmark Comparison Value

Scale	Mean	Benchmark	Interpretation
Attractiveness	1.51	Above Average	25% of the result were greater, 50% were lesser
Perspicuity	1.65	Above Average	25% of the result were greater, 50% were lesser
Efficiency	1.62	Good	10% of the result were greater, 75% were lesser
Dependability	1.29	Above Average	25% of the result were greater, 50% were lesser
Stimulation	1.18	Above Average	25% of the result were greater, 50% were lesser
Novelty	0.95	Above Average	25% of the result were greater, 50% were lesser

Based on Figure 3 and table 4 presented, it can be concluded that when compared with other products, the ChatGPT website gets a score above average on the variables of attractiveness, perspicuity, dependability, stimulation, and novelty. Whereas the efficiency variable gets a good score. At the scale value level, the good value is a higher level than the above average value. So it can be said that the efficiency variable has the highest value at the benchmark scale value level compared to the other five variables. Conversely, other variables garnered evaluation scores that fell below the average threshold. Consequently, it can be asserted that, in comparison to alternative products or systems, the ChatGPT website exhibits deficiencies that necessitate prompt rectification or enhancement, ensuring that ChatGPT users are afforded a superior user experience.

5 Conclusion

Research employing the User Experience Questionnaire (UEQ) method to measure or evaluate user experience on the ChatGPT website obtained 176 respondents. The ChatGPT website gets a positive evaluation value in all aspects. The variable of attractiveness obtained a positive assessment value of 1.503. The variable of perspicuity obtained a positive assessment value of 1.661. The efficiency variable obtained a positive assessment value of 1.615. The variable of dependability obtained a positive assessment value with an overall average value of 1.286. The variable of stimulation obtained a positive assessment value of 1.182. The novelty variable obtained a positive assessment value of 0.942. The ChatGPT website has shown good quality because it has a positive assessment value from user assessments. However, in the attractiveness, dependability, and novelty variables there are still several items that get a neutral assessment. So that product improvement is still needed in increasing user satisfaction.

Reference

[1] Resa, Kiky R.N.W., "User Interface Dan User Experience Website Bpkad Provinsi Sumatera Selatan Menggunakan Metode Heuristic Evaluation," *Jurnal Sistem Informasi,* vol. 4, no. 2, pp 90-95, 2022

[2] Jagad, A Dewantara., "Pengaruh Teknologi dalam Sikap Moralitas dan Tanggung Jawab," *Jurnal Kewarganegaraan,* vol. 7, no. 1, pp. 52-53, 2023.

[3] M. Sobron, Yamin Lubis,, "Implementasi Artificial Intelligence Pada Sistem Manufaktur Terpadu," *SEMNASTEK UISU,* vol. 4, no. 1, pp. 1-7, 2021.

[4] A. Faiz, K. Imas, "Tantangan Penggunaan ChatGPT dalam Pendidikan Ditinjau dari Sudut Pandang Moral," *Jurnal Ilmu Pendidikan,* vol. 5, no. 1, 2023

[5] Adi, Setiawan, & Ulfah, K, "Penggunaan ChatGPT Untuk Pendidikan di Era Education 4.0: Usulan Inovasi Meningkatkan Keterampilan Menulis", *Jurnal PETISI,* vol. 4, no. 1, pp. 49-54, 2023.

[6] Bjaaland, Marita S., Petter Bae, & Asbjorn Folstad, "The User Experience of ChatGPT : FIndings from a Questionnaire Study of Early Users", *CUI '23: ACM conference on Conversational User Interfaces,* vol. 23, no. 2, 2023

[7] Alghifari, Hartzani, "Evaluasi User Experience pada Dompet Digital Ovo Menggunakan User Experience Questionnaire (UEQ)," *Jurusan Sistem Informasi Universitas Islam Negeri Syarif Hidayatullah : Skripsi Diterbitkan,* 2021.

[8] A. Prayoga, et al., "Analysis of Jogjakita's User Experience Using the User Experience Questionnaire (UEQ)," *TEKNIMEDIA,* vol. 4, no.1, pp. 53-59, 2023

[9] Yordan, Patra, et al., "Analisis *User Experience* pada Pendekatan User Centered dalam Rancangan Aplikasi Placeplus," *Jurnal AUTOMATA,* vol.1, vol.2, 2020.

[10] Silvana, Rasio H., Rika Perdana S, "Evaluasi User Experience Sistem Informasi Akedemik Mahasiswa pada Perguruan TInggi Menggunakan User Experience Questionnaire", *Jurnal Politeknik Caltex Riau,* vol.6, no.1, pp.69-78, 2020

[11] Rao, R.R., & Setyadi, "Analisis UX Pada Aplikasi SISMIOP Bapenda Kab.Pemalang Menggunakan Metode User Experience Questionnaire," *Kajian Ilmiah Informatika dan Komputer,* vol. 3, no. 6, pp. 1265, 2023

[12] Martin, Schrepp, "User Experience Questionnaire (UEQ)" [Online]. Available: https://www.ueq-online.org/. [Accessed: Jun. 13, 2024].

[13] Jasmalinda, "Pengaruh Citra Merek dan Kualitas Produk terhadap Keputusan Pembelian Konsumen Motor Yamaha di Kabupaten Padang Pariaman," *Jurnal Inovasi Penelitian,* vol. 1, no. 10, pp. 2200, 2023

[14] Ambar, Lukitaningsih, & Firti Lestari, "Pengaruh Brand Image, Brand Trust dan Brand Ambassador terhadap Keputusan Pembelian Produk Smarthphone," *Jurnal Ekonomi,*

Manajemen dan Akutansi, vol. 25, no. 1, pp. 92-93, 2023

[15] Sirti, Nurwahida, "Penerapan metode Dematel dan Aras dalam Pengambilan Keputusan Multi Kriteria : Studi Kasus Pemilihan Aplikasi EWallet di Kalangan Mahasiswa Universitas Sumatera Utara", *JIKEM*, vol.3, no.2, pp 2060-2078, 2023

[16] Achmad, Nugroho A.R., & Erwin Apriliyanto, "Analysis Usability of User Experience of The Srawung with The User Experience Questionnaire (UEQ) Method," *Jurnal Ilmu Komputer An Nuur*, vol. 3, no. 1, pp. 1-2, 2023

[17] Sabam, Y.R.M, Nia Nuraeni, "Evaluasi User Experience Website E-Learning My-Elnusa Menggunakan User Experience Questionnaire (UEQ)," *Jurnal SWABUMI*, vol.11, no.1, pp. 78-84, 2023.

[18] Umi, Fariha, et.al., "Evaluasi User Experience terhadap Aplikasi OVOmenggunakanMetodeUserExperienceQuestionnaire(UEQ)", *Jurnal SISTEMASI : Sistem Informasi*, vol.13, no.5, pp.2227, 2024.