Elinvo (Electronics, Informatics, and Vocational Education)
Vol 11, No. 1, 2026, pp. 55-67
ISSN 2580-6424 (printed) | ISSN 2477-2399 (online)

Comparing User Perceptions of User Interface and User Experience in Learning Management Systems: A UEQ-Based Study Across Two Universities

Tri Setya Darmawan[1], Agung Fatwanto[2*]

[1,2]State Islamic University Sunan Kalijaga, Yogyakarta, Indonesia
[1]tsdarmawan@gmail.com; [2]agung.fatwanto@uin-suka.ac.id;
* corresponding author

Article Info

Article history:
Received May 31, 2025
Revised May 1, 2026
Accepted May 14, 2026
Available Online May 30, 2026

Keywords:
Learning Management System, Equivalence Test, Paired t-test, Wilcoxon, UEQ, Statistical Testing, Statistical Power

Abstract

User perceptions toward software application interface and experience are important factors for the successful implementation of software systems. Therefore, this study was aiming to investigate these success factors by comparing user perceptions towards software application interface and experience when using software applications. This research used a inferential quantitative approach. The instrument used in this study was the UEQ (User Experience Questionnaire) formulated by the UEQ Team. The object of this study was two e-learning systems deployed at two state universities in Yogyakarta. Data on user perceptions toward software application interface and experience were collected twice. The collected data were then statistically analysed to obtain results for equivalence test, paired t-tests, Wilcoxon tests, statistical power, and Cohen's D values. Based on equivalence testing on the data gathered during the first and second collection stages, it was found that user perceptions toward software application interface and experience were equivalent and not different for most criteria and sub-criteria, except for the enjoyable, good, pleasing, and pleasant sub-criteria at University A and enjoyable, good, and motivating at University B. Meanwhile, based on the results of paired t-tests and Wilcoxon tests, it was found that user perceptions toward software application interface and experience were also equivalent and not different for most criteria and sub-criteria, except for the pleasing and good sub-criteria at University A, and the enjoyable and pleasing sub-criteria at University B. Therefore, it is concluded that there was generally no difference in user perceptions toward software application interface and experience when using those e-learning systems.

INTRODUCTION

The domain of human-computer interaction necessitates the incorporation of constituent elements to facilitate effective communication between the human user and the computer system. Among these fundamental components, user interface and user experience emerge as pivotal facilitators of interaction. These components are intricately linked in functionality, whereby human users engage with interfaces, and their resultant experiences dictate the efficacy of interactive products [1]. The user interface represents the locus of interaction between the human user and the system, mediating exchanges through commands such as content utilization and data input [2]. Furthermore, the user interface assumes a foundational role within computer-based systems [3]. An information system characterized by an interface that is difficult to understand or navigate can significantly impede institutional performance [4]. In contrast, user experience refers to the multidimensional interaction that encompasses users'

perceptions, responses, behaviors, cognitive processes, and emotional states during system use [2], [5], [6]. ISO 9241-210 views user experience as a combination of users' perceptions and responses that are shaped by both their interaction with a system and their expectations before the interaction occurs [7], [8]. Therefore, user experience reflects users' subjective evaluation of the product or system they interact with [6], [9].

Recent studies continue to highlight the importance of examining the relationship between UI and UX. Previous studies have examined UI and UX from various perspectives, including interface design, usability evaluation, software development, and user experience assessment [10]. Overall, these studies consistently report that UI quality influences users' perceptions and interactions, while UX plays an important role in determining user satisfaction and the overall success of interactive systems. In particular, recent research has increasingly applied standardized evaluation instruments to assess UI and UX in educational software and Learning Management Systems (LMSs) [11]. However, most existing studies focus on evaluating a single platform or a single institutional implementation, making it difficult to compare user perceptions across different LMS environments using the same evaluation framework.

The present study addresses this gap by comparing users' perceptions of UI and UX between two different LMS implementations: a MOODLE (Modular Object-Oriented Dynamic Learning)-based LMS at University A and a non-MOODLE LMS at University B. Rather than evaluating only one platform, this study applies the same evaluation approach to both LMS implementations, enabling a more direct comparison of users' perceptions across different institutional contexts.

The instrument employed in this study is the User Experience Questionnaire (UEQ). UEQ is commonly utilized in research pertaining to user experience. There exist various research methodologies concerning user experience. Among these methodologies are frameworks such as SUPR-Q (Standardized User Experience Percentile Rank Questionnaire), QUIS (Questionnaire for User Interaction Satisfaction), SUMI (Software Usability Measurement Inventory), and SUS (System Usability Scale) [12], [13]. Compared to these four performance frameworks, UEQ (User Experience Questionnaire) offers remarkable advantages due to its provision of a comprehensive user experience impression, encompassing both classic usability aspects and user experience facets [12]. In this study, the utilization of UEQ extends beyond user experience components alone and is also applied to user interface components. This decision is motivated by the importance of designing the overall user experience rather than just the interface [1].

Accordingly, this study has two primary objectives: (1) to evaluate users' perceptions of UI and UX within each LMS implementation using the UEQ framework, and (2) to compare these perceptions between the MOODLE-based LMS at University A and the non-MOODLE LMS at University B. The contribution of this study is to provide an empirical comparison of two different LMS implementations using a common evaluation framework, thereby offering practical insights for higher education institutions and LMS developers in improving interface design and user experience.

METHODS

This study employs an inferential quantitative design, in which numerical data are collected and statistically analyzed to identify patterns, examine relationships among variables, generate predictions, and draw conclusions that can be generalized to the target population [14]–[16]. In this study, the inferential aspect refers to the use of statistical techniques that allow conclusions drawn from the sample to be extended to the target population [15], [17]. The accuracy of inferential statistics is influenced by the selection of an appropriate sample that can adequately represent the population [18], [19]. Appropriate statistical procedures are applied to quantitative variables to test research hypotheses, with results interpreted based on probability values, significance levels, and effect size measures [15], [17].

The processes involved in conducting this research include questionnaire development, data collection, and data analysis.

Research Instrument

This questionnaire used in this study was developed based on the User Experience Questionnaire (UEQ), a widely adopted instrument for evaluating the user experience of interactive systems. UEQ consists of 26 items grouped into six dimensions: attractiveness, efficiency, perspicuity, dependability, stimulation, and novelty, as presented in Table 1 [20], [21]. Attractiveness has aspects with pragmatic and hedonic qualities. The pragmatic quality aspect includes efficiency, perspicuity, and dependability. The hedonic quality aspect includes stimulation and novelty. Both pragmatic and hedonic quality aspects influence the attractiveness and preferences of a product [6], [22]. Pragmatic and hedonic qualities represent two categories that summarize various quality aspects. Pragmatic quality primarily reflects how effectively a product supports users in accomplishing their intended tasks, whereas hedonic quality emphasizes experiential attributes that extend beyond task completion, such as interface aesthetics and design originality [22], [23].

Table 1. Questionnaire Criteria Item List

Criteria	Number of Items	Sub-Criteria
Attractiveness	6	Enjoyable, Good, Pleasing, Pleasant, Attractive, Friendly
Efficiency	4	Fast, Efficient, Practical, Organized
Perspicuity	4	Understandable, Easy to Learn, Easy, Clear
Dependability	4	Predictable, Supportive, Secure, Meets Expectations
Stimulation	4	Valuable, Exciting, Interesting, Motivating
Novelty	4	Creative, Inventive, Leading edge, Innovative

This questionnaire is developed based on the User Experience Questionnaire (UEQ), which has a broad scope of evaluation [6]. The UEQ instrument is formulated by the UEQ Team. UEQ is an efficient and user-friendly tool or questionnaire used to measure user experience [24]–[26]. UEQ is used primarily to measure the user experience of interactive products quickly and promptly [24], [26]. The UEQ has been widely employed to assess user experience across a broad range of applications, including business systems, software development environments, websites, online services, and social networking platforms [24], [25]. The UEQ approach gathers feedback from e-learning system users by means of questionnaire completion. According to [24], there are 6 rating scales in UEQ, including attractiveness, efficiency, perspicuity, dependability, stimulation, and novelty. Each scale has subsections as mentioned in [24] stating that the assumed scale structure of UEQ includes attractiveness consisting of annoying or enjoyable, unlikable or pleasing, bad or good, unpleasant or pleasant, unattractive or attractive, and unfriendly or friendly; efficiency consisting of slow or fast, inefficient or efficient, impractical or practical, and cluttered or organized; perspicuity consisting of not understandable or understandable, difficult to learn or easy to learn, complicated or easy, and confusing or clear; dependability consisting of unpredictable or predictable, obstructive or supportive, not secure or secure, and does not meet expectations or meets expectations; stimulation consisting of inferior or valuable, boring or exciting, not interesting or interesting, and demotivating or motivating; and novelty consisting of dull or creative, conventional or inventive, usual or leading edge, and conservative or innovative. In an official website of the UEQ Team, it is stated that the UEQ scale encompasses an overall assessment of the user experience**.** There are two aspects measured within it, namely the classic usability aspect comprising efficiency, perspicuity, and dependability, and the user experience aspect comprising originality and stimulation. In this study, the original UEQ was adapted to evaluate both user interface and user experience perceptions. Therefore, the questionnaire should be regarded as a UEQ-based adapted questionnaire rather than the original UEQ instrument.

Before data collection, the adapted questionnaire underwent validity and reliability testing to confirm its suitability for the study. Construct validity was examined using the Pearson product–moment correlation, with questionnaire items considered valid when their correlation coefficients exceeded the critical value at the 0.05 significance level. Reliability was evaluated by calculating Cronbach's alpha to assess the internal consistency of the instrument. Alpha coefficients above 0.70 were interpreted as indicating acceptable reliability, whereas values greater than 0.90 reflected excellent internal consistency. Detailed findings from both the validity and reliability analyses are reported in the Results section.

After the questionnaire was developed, a four-point Likert scale was adopted to measure respondents' perceptions. The questionnaire consisted of paired statements, with user experience (UX) statements presented in odd-numbered items and user interface (UI) statements presented in even-numbered items. Responses were rated using four categories: Strongly Disagree (1), Disagree (2), Agree (3), and Strongly Agree (4). A neutral midpoint was intentionally omitted to encourage respondents to express either agreement or disagreement, thereby reducing the tendency to select ambiguous middle responses [27], [28]. The exclusion of a neutral midpoint is supported by previous studies, which indicate that respondents often interpret the midpoint inconsistently, as example "no opinion", "unsure", or "don't care", potentially reducing the interpretability of responses [27], [29]. In addition, the midpoint may function as an escape option for respondents who are uncertain or reluctant to express a clear opinion [27], [30]. Because the original UEQ employs a 7-point bipolar semantic differential scale, the instrument used in this study should be regarded as an adapted UEQ-based questionnaire rather than the original UEQ instrument [21], [28].

Data Collection

In data collection, it is essential to determine the required sample size initially. The purpose of determining the sample size is to ensure that the data collected are neither excessive nor insufficient to represent the entire population in a study [31], [32]. There are various methods for determining sample size in a population, and one commonly used approach in survey research is the Slovin/Yamane formula such as (1) [32].

$$n = \frac{N}{1+Ne^2} \quad \text{..... (1)}$$

Regarding these parameters, n specifies the necessary sample size, N corresponds to the overall population, and e reflects the permissible margin of error. Using a 10% margin of error [32], the minimum required sample sizes were calculated for each university, as presented in Table 2. As shown in Table 2, all datasets exceeded the minimum sample size requirement.

Table 2. Minimum Required Sample Size

University	Population (N)	Minimum Sample Size	Total Responses (1)	Total Responses (2)
University A	24.000	100	176	372
University B	21.000	100	387	375

After obtaining the sample size, researchers distribute it to respondents. Sampling techniques are generally classified into two broad categories: probability sampling and non-probability sampling. However, this study utilizes probability sampling, specifically random sampling. A study [33] states that the use of probability sampling methods better represents the target population. Random sampling involves selecting samples from a population by giving an equal chance to each population member to become a sample [34].

Two samples are the objects of this study. The first sample is derived from respondents of University A, and the second sample is from respondents of University B. Each respondent uses a

different object. The application study used by respondents of University A is MOODLE, while the application study used by respondents of University B is non-MOODLE. Moodle is an open-source learning management system that enables educators, administrators, and students to manage teaching and learning activities within a secure and integrated online environment. MOODLE is also a cloud-based medium accessible via smartphones or computers connected to the internet, minimizing student smartphone misuse during learning and fostering ICT (Information and Communication Technology) learning habits to prepare for the digital era [35]. Meanwhile, the non-MOODLE used by University B is a learning system management created independently by the institution, thus it is likely tailored to specific needs only. The difference in the basis of learning system management is intended to observe differences in user perceptions from the perspective of user interface and user experience. The two LMS implementations were selected because they represent different development approaches. University A employs a MOODLE-based LMS, whereas University B uses an institutionally developed non-MOODLE LMS. Comparing these two implementations enables the study to examine users' perceptions of UI and UX across different LMS platforms. Both systems provide similar educational functions but differ in their underlying platform architecture and implementation.

Data were collected in two independent waves at each university. At University A, the first wave was conducted from July to October 2020, whereas the second wave was conducted from August 2021 to January 2022. At University B, the first wave was conducted from May to October 2020 and the second wave from August to September 2021. The two-wave design was intended to replicate the same analysis using independent respondent groups, allowing the consistency of the findings to be examined across different data collection periods before comparing the two LMS implementations. A total of 176 and 372 valid responses were obtained from University A during Wave 1 and Wave 2, respectively, while 387 and 375 valid responses were obtained from University B. Responses with incomplete questionnaire data were excluded prior to statistical analysis.

Data Analysis

The collected data were analyzed using three statistical procedures: the equivalence test, the paired *t*-test, and the Wilcoxon signed-rank test. All statistical analyses were performed using IBM SPSS Statistics. The equivalence test is a comparison between two samples or groups within certain boundaries [36], [37]. The equivalence test follows the Neyman–Pearson hypothesis-testing framework, enabling researchers to evaluate equivalence while maintaining control error rates and achieving the desired statistical power [36], [37]. This testing method provides more specific predictions about effect sizes deemed valuable by researchers for investigation [38]. Subsequently, inferential statistical tests were performed to examine whether significant differences existed between each pair of responses.

In conducting the equivalence test, several values need to be determined beforehand. These include the mean values (2), the mean difference between user interface and user experience, standard deviation (3), error (σ) (4) (5), and MoE (error (t) $p < 0.05$) (6). These formulations are derived from a book [53].

$$\bar{x} = \sum_{i=1}^{r} f_i x_i \,/\, \sum_{i=1}^{r} f_i \quad \text{..... (2)}$$

$$S_r = \sqrt{\sum (x_i - \bar{x})^2 \,/n} \quad \text{..... (3)}$$

$$E = t_{hitung} . s_r / \sqrt{\bar{n}} \quad \text{..... (4)}$$

$$t_{hitung} = (\bar{x} - \mu_0)/(s/\sqrt{n}) \quad \text{..... (5)}$$

$$E = 1{,}96 . s_r / \sqrt{\bar{n}} \quad \text{..... (6)}$$

ISSN 2580-6424 (printed) | ISSN 2477-2399 (online)

In this formula, x_i represents the center value of the i-th class interval, while n denotes the total sample size, and $\mu 0$ signifies the assumed population mean. To determine whether the mean difference between paired UI and UX scores indicates practical equivalence, the Margin of Error (MoE) is utilized to establish the permissible threshold [36], [38]. If the confidence interval of the paired mean difference falls entirely within the equivalence margin, the two measures are regarded as equivalent [37], [38]. Otherwise, they are considered not equivalent [37].

These values are used to perform the equivalence test. There are four possible conclusions drawn from the test results: equivalent and not different, equivalent and different, not equivalent and not different, and not equivalent and different [37]. These conclusions can be determined by comparing the mean values of paired items and presenting them in a graph like Figure 1, where A represents equivalent and different, B represents not equivalent and not different, C represents not equivalent and different, and D represents equivalent and not different [39].

Following the equivalence analysis, additional inferential statistical tests were conducted to determine whether paired observations differed significantly. The selection of either a parametric or a non-parametric procedure was based on the distributional characteristics of the paired differences.

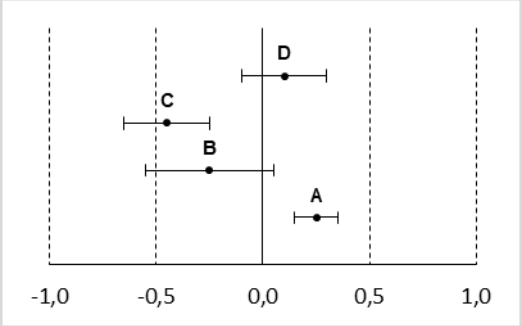

Figure 1. Graph of Comparison Mean

Prior to selecting the inferential test, the normality of the paired differences was evaluated using the Shapiro–Wilk test [39], [40]. The normality assessment indicated that the paired differences were not normally distributed. Consequently, the Wilcoxon signed-rank test was considered the primary inferential test, whereas the paired *t*-test was also performed to provide a parametric comparison of the results [39]. The paired t-test compares two paired data by examining parametric differences and testing the significance level of the independent variable's influence on the dependent variable partially [41]. These two data mutually influence each other as they originate from the same group [41], [42]. The formula for the paired t-test is as follows (7).

$$t = \frac{Mx_1 - Mx_2}{\sqrt{\{(\sum x_1^2 + \sum x_2^2)/(n_1+n_2-2)\}\{1/n_1+2/n_2\}}} \quad \text{..... (7)}$$

Where M_x is the mean calculated mean, x is the sum of the deviation scores, and n is the number of subjects [43], [44]. The subsequent step after the t-test is testing the significance value of t. This significance test is conducted by comparing the t-value to the critical t-value at the predetermined significance level [44]. If the t-value > critical t-value, there is a significant difference, and if the t-value < critical t-value, there is no significant difference [44]. Another criterion of the paired t-test is if the Sig (2-tailed) α value is above 0.05, there is no difference between the two compared values, while if the Sig (2-tailed) α value is below 0.05, there is a difference between the two compared values [43]–[45]. In addition to the significance value ($p < 0.05$), effect size (Cohen's *d*) and statistical power were reported to provide information regarding the magnitude of the observed differences and the adequacy of the statistical analysis [43]–[45].

As a non-parametric statistical approach, the Wilcoxon test evaluates the variance between paired samples by accounting for both the direction and size of their differences to draw population inferences [41], [46]–[48]. The formula for the Wilcoxon test is as follows (8).

$$W = \sum_{i=1}^{N} \left[sgn\left(x_{2,1} - x_{1,i}\right) \cdot R_i \right] \quad (8)$$

Within this equation, each paired measurement i out of N is represented by the vector $x_i = (x_{1,i}, x_{2,i})$, with R_i signifying its assigned rank. Here, R_i reflects the specific placement of the observation when the absolute discrepancies $|x_{2,i} - x_{1,i}|$ are arranged in an ordered sequence. In drawing conclusions, pairs with larger absolute differences have higher ranks R_i, making them determinants in W. Meanwhile, pairs with smaller absolute differences have low ranks R_i because of their minimal influence. This is due to the test's reliance more on ranks than on measurements themselves. Therefore, the Wilcoxon test can be considered as a test of shifting median values between two groups [46], [47]. For the Wilcoxon signed-rank test, statistical significance was determined using the *Asymp. Sig.* (*p*-value). A *p*-value below 0.05 indicated a statistically significant difference between the paired observations, whereas a *p*-value of 0.05 or higher indicated that no statistically significant difference was detected [47], [48].

RESULT AND DISCUSSION

Based on the criteria mentioned in Table 1, the questionnaire was divided into six parts, namely: first, regarding the attractiveness criteria consisting of 12 statements with details of 6 statements about user interface and user experience when using the application with different sub-criteria for each statement, including enjoyable, good, pleasing, pleasant, attractive, and friendly; second, regarding the efficiency criteria consisting of 8 statements with details of 4 statements about user interface and user experience when using the application with different sub-criteria for each statement, including fast, efficient, practical, and organized; third, regarding the clarity criteria consisting of 8 questions with details of 4 statements about user interface and user experience when using the application with different sub-criteria for each statement, including understandable, easy to learn, easy, and clear; fourth, regarding the dependence criteria consisting of 8 statements with details of 4 statements about user interface and user experience when using the application with different sub-criteria for each statement, including predictable, supportive, secure, and meets expectations; fifth, regarding the stimulation criteria consisting of 8 statements with details of 4 statements about user interface and user experience when using the application with different sub-criteria for each statement, including valuable, exciting, interesting, and motivating; and sixth, regarding the novelty criteria consisting of 8 statements with details of 4 statements about user interface and user experience when using the application with different sub-criteria for each statement, including creative, inventive, leading edge, and innovative. Then, validity tests were conducted as shown in Table 3 and Table 4 to assess the validity of each questionnaire instrument. The validity assessment confirmed that all questionnaire items satisfied the established validity criteria and were therefore considered suitable for subsequent analysis. Reliability testing further showed excellent internal consistency across all six UEQ dimensions, with Cronbach's alpha coefficients exceeding 0.98. These findings demonstrate that the adapted UEQ instrument produced highly consistent measurements for both UI and UX evaluations.

After obtaining data from the respondents, the next step was to conduct equivalence testing. The results of the equivalence test are displayed in Table 3 for the learning management system at University A and Table 4 for the learning management system at University B. To improve table readability, the equivalence outcomes are represented using the abbreviations E is Equivalent & Not Different (Equivalent – Conclusive) and U is Not Equivalent & Not Different (Undetermined – Inconclusive).

ISSN 2580-6424 (printed) | ISSN 2477-2399 (online)

Table 3. Results of Equivalence Test between UX and UI in University A

Criteria	Sub-criteria	Equivalence test		Conclusion
		(1)	(2)	
Attractiveness	Enjoyable	U	E	U
	Good	U	E	U
	Pleasing	U	U	U
	Pleasant	U	E	U
	Attractive	E	E	E
	Friendly	E	E	E
Efficiency	Fast	E	E	E
	Efficient	E	E	E
	Practical	E	E	E
	Organized	E	E	E
Perspicuity	Understandable	E	E	E
	Easy to learn	E	E	E
	Easy	E	E	E
	Clear	E	E	E
Dependability	Predictable	E	E	E
	Supportive	E	E	E
	Secure	E	E	E
	Meets expectation	E	E	E
Stimulation	Valuable	E	E	E
	Exciting	E	E	E
	Interesting	E	E	E
	Motivating	E	E	E
Novelty	Creative	E	E	E
	Inventive	E	E	E
	Leading edge	E	E	E
	Innovative	E	E	E

Table 4. Results of Equivalence Test between UX and UI in University B

Criteria	Sub-criteria	Equivalence test		Conclusion
		(1)	(2)	
Attractiveness	Enjoyable	U	U	U
	Good	E	U	U
	Pleasing	E	E	E
	Pleasant	E	E	E
	Attractive	E	E	E
	Friendly	E	E	E
Efficiency	Fast	E	E	E
	Efficient	E	E	E
	Practical	E	E	E
	Organized	E	E	E
Perspicuity	Understandable	E	E	E
	Easy to learn	E	E	E
	Easy	E	E	E
	Clear	E	E	E
Dependability	Predictable	E	E	E
	Supportive	E	E	E
	Secure	E	E	E
	Meets expectation	E	E	E
Stimulation	Valuable	E	E	E
	Exciting	E	E	E
	Interesting	E	E	E
	Motivating	E	U	U
Novelty	Creative	E	E	E
	Inventive	E	E	E
	Leading edge			

Criteria	Sub-criteria	Equivalence test		Conclusion
		(1)	(2)	
	Innovative	E	E	E
		E	E	E

Based on the results of the equivalence test, it can be concluded that most pairs of user interface and user experience sub-criteria when using the application showed user perceptions that are equivalent and not different except for the enjoyable, good, pleasing, and pleasant sub-criteria at University A and the enjoyable, good, and motivating sub-criteria at University B. This is because the pairs of user interface and user experience sub-criteria when using the application showed user perceptions that are not equivalent and not different.

Table 5. Result Test between UX and UI in University A

Criteria	Sub-criteria	Paired T test		Wilcoxon test		Statistical Power		Cohen's D	
		(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Attractiveness	Enjoyable	0,00	0,06	0,00	0,07	0,9	0,15	0,33	0,07
	Good	0,02	0,02	0,03	0,02	0,33	0,15	0,16	0,07
	Pleasing	0,03	0,00	0,03	0,00	0,28	0,31	0,14	0,11
	Pleasant	0,00	0,32	0,00	0,32	0,57	0,07	0,23	0,03
	Attractive	0,72	0,84	0,76	0,76	0,04	0,04	0,02	0,01
	Friendly	0,53	0,29	0,53	0,29	0,06	0,07	0,04	0,04
Efficiency	Fast	0,22	0,20	0,22	0,20	0,05	0,09	0,04	0,05
	Efficient	0,32	0,62	0,32	0,66	0,05	0,04	0,04	0,01
	Practical	0,39	0,08	0,38	0,09	0,06	0,09	0,04	0,05
	Organized	0,32	0,35	0,36	0,34	0,07	0,05	0,05	0,02
Perspicuity	Understandable	0,64	0,02	0,63	0,02	0,04	0,12	0,03	0,06
	Easy to learn	0,27	0,52	0,27	0,52	0,07	0,04	0,05	0,01
	Easy	0,75	0,26	0,62	0,26	0,03	0,05	0,01	0,02
	Clear	0,41	0,28	0,46	0,29	0,06	0,05	0,04	0,02
Dependability	Predictable	0,06	0,73	0,06	0,73	0,12	0,04	0,09	0,01
	Supportive	0,25	0,11	0,25	0,12	0,1	0,09	0,07	0,05
	Secure	0,76	0,29	0,74	0,29	0,03	0,05	0,01	0,02
	Meets expectation	0,43	1,00	0,43	1,00	0,04	0,02	0,03	0,00
Stimulation	Valuable	0,02	0,13	0,02	0,13	0,18	0,07	0,12	0,04
	Exciting	0,02	0,27	0,03	0,27	0,16	0,05	0,11	0,02
	Interesting	0,30	0,72	0,30	0,72	0,07	0,04	0,05	0,01
	Motivating	0,02	0,48	0,02	0,48	0,16	0,05	0,10	0,02
Novelty	Creative	0,66	0,18	0,67	0,20	0,02	0,07	0,00	0,04
	Inventive	0,12	0,82	0,12	0,86	0,1	0,04	0,07	0,01
	Leading edge	0,47	0,43	0,46	0,43	0,05	0,05	0,04	0,02
	Innovative	0,13	0,70	0,13	0,70	0,09	0,02	0,07	0,00

Although the paired t-test was performed for comparison purposes, the normality assumption for paired differences was not satisfied. Therefore, the Wilcoxon signed-rank test was used as the primary inferential analysis, while the paired t-test results are presented only as complementary information. The results obtained from the paired t-test indicated that the majority of pairs of user interface sub-criteria and experience had significance values or Sig (2-tailed) greater than 0.5 α, indicating that users' perceptions did not differ except for the sub-criteria of good and pleasing in University A as shown in Table 5, and the sub-criteria of enjoyable and good in University B as shown in Table 6. This is because pairs of user interface sub-criteria and experience had significance values or Sig (2-tailed) below 0.5 α, namely the good sub-criteria valued at 0.02 in both the first and second data collections, and the pleasing sub-criteria valued at 0.03 in the first data collection and 0 in the second data collection in University A, as well as the enjoyable sub-criteria valued at 0 in both the first and second data collections, and the good sub-criteria valued at 0.04 in the first data collection and 0 in the second data collection in

University B. These results are presented for comparison with the non-parametric analysis and are not used as the primary basis for interpretation because the assumption of normality was violated.

Following that, a non-parametric test was conducted using the Wilcoxon test. The results obtained from the Wilcoxon test indicated that the majority of pairs of user interface sub-criteria and experience had significance values (Asymp.Sig) greater than α, indicating that users' perceptions did not differ except for the sub-criteria of good and pleasing in University A as shown in Table 5, and the sub-criteria of enjoyable and good in University B as shown in Table 6. This is because these pairs of sub-criteria had significance values (Asymp.Sig) less than α, namely the good sub-criteria valued at 0.03 in the first data collection and 0.02 in the second data collection, and the pleasing sub-criteria valued at 0.03 in the first data collection and 0 in the second data collection in University A, as well as the enjoyable sub-criteria valued at 0 in both the first and second data collections, and the good sub-criteria valued at 0.03 in the first data collection and 0 in the second data collection in University B. These findings are consistent with the equivalence test, where most UI and UX pairs demonstrated similar user perceptions across both measurement periods. Only a limited number of sub-criteria showed statistically significant differences.

The reported effect sizes (Cohen's d) were generally below 0.20 across both universities, indicating that the observed differences between paired UI and UX responses were negligible in practical terms even when statistical significance was detected. Similarly, statistical power varied across individual sub-criteria, with higher power observed only for those showing statistically significant differences. Overall, these findings suggest that the practical differences between UI and UX perceptions were limited.

Table 6. Result Test between UX and UI in University B

Criteria	Sub-criteria	Paired T test		Wilcoxon test		Statistical Power		Cohen's D	
		(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Attractiveness	Enjoyable	0,00	0,00	0,00	0,00	0,97	0,98	0,29	0,26
	Good	0,04	0,00	0,03	0,00	0,18	0,69	0,08	0,16
	Pleasing	1,00	0,02	0,93	0,02	0,02	0,26	0,00	0,09
	Pleasant	0,12	0,14	0,12	0,14	0,08	0,16	0,04	0,07
	Attractive	0,47	0,10	0,40	0,11	0,05	0,13	0,03	0,06
	Friendly	0,73	0,07	0,80	0,08	0,04	0,16	0,01	0,07
Efficiency	Fast	0,66	0,56	0,65	0,56	0,04	0,05	0,01	0,02
	Efficient	0,29	0,76	0,37	0,68	0,05	0,04	0,03	0,01
	Practical	0,29	0,01	0,37	0,01	0,05	0,21	0,03	0,08
	Organized	0,35	0,09	0,35	0,09	0,04	0,12	0,01	0,06
Perspicuity	Understandable	0,25	0,35	0,25	0,45	0,07	0,06	0,04	0,03
	Easy to learn	0,03	0,07	0,03	0,07	0,13	0,13	0,06	0,06
	Easy	0,63	0,29	0,53	0,27	0,04	0,06	0,01	0,03
	Clear	0,90	0,36	0,89	0,35	0,02	0,06	0,00	0,03
Dependability	Predictable	0,91	0,29	0,89	0,29	0,02	0,05	0,00	0,03
	Supportive	0,10	0,47	0,13	0,51	0,14	0,06	0,07	0,03
	Secure	0,82	0,83	0,74	0,82	0,04	0,04	0,01	0,01
	Meets expectation	0,91	1,00	0,90	0,96	0,02	0,02	0,00	0,00
Stimulation	Valuable	0,04	0,05	0,04	0,05	0,14	0,19	0,07	0,08
	Exciting	0,31	0,00	0,30	0,01	0,08	0,21	0,04	0,08
	Interesting	0,52	0,06	0,57	0,07	0,04	0,19	0,01	0,08
	Motivating	0,55	0,00	0,56	0,00	0,04	0,34	0,01	0,11
Novelty	Creative	0,70	0,00	0,79	0,00	0,04	0,21	0,01	0,08
	Inventive	0,30	0,16	0,24	0,16	0,07	0,07	0,04	0,03
	Leading edge	1,00	0,01	0,99	0,01	0,02	0,11	0,00	0,05
	Innovative	0,52	0,25	0,52	0,25	0,04	0,07	0,01	0,04

Based on the combined results of the equivalence test and the Wilcoxon signed-rank test, most UI and UX sub-criteria produced consistent findings across both data collection periods. Nevertheless,

several sub-criteria, particularly enjoyable, good, pleasing, pleasant (University A), and enjoyable, good, and motivating (University B), remained inconclusive because they did not satisfy the equivalence criteria despite showing no statistically significant differences. The present study did not investigate the reasons underlying these inconclusive findings; therefore, further research is required to determine the factors contributing to these results.

CONCLUSION

The findings indicate that users' perceptions of LMS interface quality and user experience remained largely consistent across the two data collection phases. Most UEQ dimensions and sub-criteria demonstrated equivalent results, with only a small number of items, primarily within the Attractiveness dimension, showing inconsistent perceptions. Overall, these results suggest that users tend to perceive interface quality and user experience consistently across repeated evaluations, regardless of whether the LMS is based on Moodle or a non-Moodle platform.

From both theoretical and practical perspectives, the findings indicate that UI and UX should not always be treated as entirely separate constructs when evaluating Learning Management Systems. For most UEQ dimensions, users perceived interface quality and user experience consistently, suggesting that improvements to interface design are likely to influence the overall user experience simultaneously. Therefore, universities and LMS developers may use the UEQ as a practical instrument for monitoring user perceptions and identifying interface aspects that require further improvement.

This study is limited by the inclusion of only two public universities and a relatively short interval between repeated measurements. Consequently, the findings should be interpreted within similar educational contexts. Future research may extend this work by involving more diverse institutions, different LMS platforms, and longer observation periods to examine whether similar patterns of perceptual consistency are maintained across broader learning environments.

REFERENCES

[1] H. Kivijärvi and K. Pärnänen, “Instrumental Usability and Effective User Experience: Interwoven Drivers and Outcomes of Human-Computer Interaction,” *Int. J. Human–Computer Interact.*, vol. 39, no. 1, pp. 34–51, Jan. 2023, doi: 10.1080/10447318.2021.2016236.

[2] I. F. B. Andoro, “Analisis Tingkat Usability dan User Experience LMS Institut Widya Pratama dengan Pendekatan UEQ (User Experience Questionnaire),” *J. Ilm. Sist. Inf. dan Ilmu Komput.*, 2025, doi: 10.55606/juisik.v5i1.1471.

[3] P. R *et al.*, “Human-Computer Interaction: Enhancing User Experience in Interactive Systems,” in *E3S Web of Conferences*, 2023. doi: 10.1051/e3sconf/202339904037.

[4] F. Fitria, M. Yahya, M. Ali, P. Purnamawati, and A. M. Mappalotteng, “The Impact of System Quality and User Satisfaction: The Mediating Role of Ease of Use and Usefulness in E-Learning Systems,” *Int. J. Environ. Eng. Educ.*, 2024, doi: 10.55151/ijeedu.v6i2.134.

[5] H.-S. Wong and H. Ip, “Human Computer Interaction,” pp. 289–293, 2015, doi: 10.1007/978-3-319-03068-5.

[6] J. Kollmorgen, A. Hinderks, and J. Thomaschewski, “Selecting the Appropriate User Experience Questionnaire and Guidance for Interpretation: the UEQ Family,” *Int. J. Interact. Multim. Artif. Intell.*, vol. 9, pp. 126–139, 2024, doi: 10.9781/ijimai.2024.08.005.

[7] M. Maguire, “Using Human Factors Standards to Support User Experience and Agile Design,” pp. 185–194, 2013, doi: 10.1007/978-3-642-39188-0_20.

[8] R. Juárez-Ramírez, “User-centered design and adaptive systems: toward improving usability and accessibility,” *Univers. Access Inf. Soc.*, vol. 16, pp. 361–363, 2017, doi: 10.1007/s10209-016-0480-1.

[9] P. Vlachogianni and N. Tselios, “Perceived usability evaluation of educational technology using the System Usability Scale (SUS): A systematic review,” *J. Res. Technol. Educ.*, vol. 54, pp. 392–409, 2021, doi: 10.1080/15391523.2020.1867938.

[10] A. Schankin, M. Budde, T. Riedel, and M. Beigl, “Psychometric Properties of the User Experience Questionnaire (UEQ),” in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 2022. doi: 10.1145/3491102.3502098.

[11] A. Saleh, H. Y. Abuaddous, I. Alansari, and O. Enaizan, “The Evaluation of User Experience on Learning

Management Systems Using UEQ," *Int. J. Emerg. Technol. Learn.*, vol. 17, pp. 145–162, 2022, doi: 10.3991/ijet.v17i07.29525.
[12] D. Quiñones and L. Rojas, "Understanding the customer experience in human-computer interaction: a systematic literature review," *PeerJ Comput. Sci.*, vol. 9, 2023, doi: 10.7717/peerj-cs.1219.
[13] R. K. Paredes and A. A. Hernandez, "Measuring the quality of user experience on web services: A case of university in the Philippines," in *2017IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*, 2017, pp. 1–6. doi: 10.1109/HNICEM.2017.8269446.
[14] W. M. Lim, "What Is Quantitative Research? An Overview and Guidelines," *Australas. Mark. J.*, vol. 33, pp. 325–348, 2024, doi: 10.1177/14413582241264622.
[15] G. Kotronoulas *et al.*, "An Overview of the Fundamentals of Data Management, Analysis, and Interpretation in Quantitative Research.," *Semin. Oncol. Nurs.*, p. 151398, 2023, doi: 10.1016/j.soncn.2023.151398.
[16] H. Mohajan, "Quantitative Research: A Successful Investigation in Natural and Social Sciences," *J. Econ. Dev. Environ. People*, 2020, doi: 10.26458/jedep.v9i4.679.
[17] D. Chawla and A. Deorari, "Inferential Statistics: Introduction to Hypothesis Testing," *J. Neonatol.*, vol. 19, pp. 259–264, 2005, doi: 10.1177/0973217920050313.
[18] R. Wilcox, "Inferential Statistics," *Oxford Res. Encycl. Bus. Manag.*, 2021, doi: 10.1093/acrefore/9780190224851.013.280.
[19] A. Kenny, "From Sample to Population," pp. 89–104, 1982, doi: 10.1016/b978-0-08-024281-1.50012-7.
[20] S. R. Henim and R. P. Sari, "User Experience Evaluation of Student Academic Information System of Higher Education Using User Experience Questionnaire," *J. Komput. Terap.*, 2020, doi: 10.35143/jkt.v6i1.3582.
[21] M. Schrepp, A. Hinderks, and J. Thomaschewski, "Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S)," *Int. J. Interact. Multim. Artif. Intell.*, vol. 4, pp. 103–108, 2017, doi: 10.9781/ijimai.2017.09.001.
[22] M. Hassenzahl, "The Effect of Perceived Hedonic Quality on Product Appealingness," *Int. J. Human–Computer Interact.*, vol. 13, pp. 481–499, 2001, doi: 10.1207/s15327590ijhc1304_07.
[23] D. S. S. Sahid, P. I. Santosa, R. Ferdiana, and E. N. Lukito, "Evaluation and measurement of Learning Management System based on user experience," in *2016 6th International Annual Engineering Seminar (InAES)*, 2016, pp. 72–77. doi: 10.1109/INAES.2016.7821910.
[24] M. Schrepp, A. Hinderks, and J. Thomaschewski, "Construction of a Benchmark for the User Experience Questionnaire (UEQ)," *Int. J. Interact. Multim. Artif. Intell.*, vol. 4, pp. 40–44, 2017, doi: 10.9781/ijimai.2017.445.
[25] M. Rauschenberger, M. Schrepp, M. Cota, S. Olschner, and J. Thomaschewski, "Efficient Measurement of the User Experience of Interactive Products. How to use the User Experience Questionnaire (UEQ).Example: Spanish Language Version," *Int. J. Interact. Multim. Artif. Intell.*, vol. 2, pp. 39–45, 2013, doi: 10.9781/ijimai.2013.215.
[26] B. Laugwitz, T. Held, and M. Schrepp, "Construction and Evaluation of a User Experience Questionnaire," pp. 63–76, 2008, doi: 10.1007/978-3-540-89350-9_6.
[27] M. Kankaraš and S. Capecchi, "Neither agree nor disagree: use and misuse of the neutral response category in Likert-type scales," *METRON*, vol. 83, pp. 111–140, 2024, doi: 10.1007/s40300-024-00276-5.
[28] B. Zhang, J. Luo, and J. Li, "Moving beyond Likert and Traditional Forced-Choice Scales: A Comprehensive Investigation of the Graded Forced-Choice Format," *Multivariate Behav. Res.*, vol. 59, pp. 434–460, 2023, doi: 10.1080/00273171.2023.2235682.
[29] J. Nadler, R. Weston, and E. Voyles, "Stuck in the Middle: The Use and Interpretation of Mid-Points in Items on Questionnaires," *J. Gen. Psychol.*, vol. 142, pp. 71–89, 2015, doi: 10.1080/00221309.2014.994590.
[30] K. Kurtulus, C. Basfirinci, and Z. C. Uk, "Beyond the middle: rethinking midpoint and 'no idea' options in Likert scale for consumer sensitive issues," *Humanit. Soc. Sci. Commun.*, 2026, doi: 10.1057/s41599-026-07037-x.
[31] A. Adam, "A Study on Sample Size Determination in Survey Research," pp. 125–134, 2021, doi: 10.9734/bpi/nicst/v4/2125e.
[32] A. Adam, "Sample Size Determination in Survey Research," *J. Sci. Res. Reports*, pp. 90–97, 2020, doi: 10.9734/jsrr/2020/v26i530263.
[33] B. Mukti, "Methods in health research: Probability and non-probability sampling," *Heal. Sci. Int. J.*, 2025, doi: 10.71357/hsij.v3i2.64.
[34] S. Singh, "Simple Random Sampling," pp. 71–136, 2003, doi: 10.1007/978-94-007-0789-4_2.
[35] S. H. P. W. Gamage, J. R. Ayres, and M. B. Behrend, "A systematic review on trends in using Moodle for teaching and learning," *Int. J. STEM Educ.*, vol. 9, no. 1, p. 9, 2022, doi: 10.1186/s40594-021-00323-x.

[36] R. Kelter, "Bayesian Hodges-Lehmann tests for statistical equivalence in the two-sample setting: Power analysis, type I error rates and equivalence boundary selection in biomedical research," *BMC Med. Res. Methodol.*, vol. 21, 2021, doi: 10.1186/s12874-021-01341-7.

[37] D. Lakens, A. Scheel, and P. Isager, "Equivalence Testing for Psychological Research: A Tutorial," *Adv. Methods Pract. Psychol. Sci.*, vol. 1, pp. 259–269, 2018, doi: 10.1177/2515245918770963.

[38] P. Riesthuis, "Simulation-Based Power Analyses for the Smallest Effect Size of Interest: A Confidence-Interval Approach for Minimum-Effect and Equivalence Testing," *Adv. Methods Pract. Psychol. Sci.*, vol. 7, 2024, doi: 10.1177/25152459241240722.

[39] R. Tapio, "The Role of Data Assumptions in Selecting Between Parametric and Nonparametric Tests," *Asian J. Probab. Stat.*, 2025, doi: 10.9734/ajpas/2025/v27i11830.

[40] J. A. Montero, "A Structured Guide to Univariate Test Selection Based on Normality, Variance Homogeneity, and Graphical Data Exploration.," *J. Surg. Res.*, vol. 318, pp. 230–240, 2026, doi: 10.1016/j.jss.2025.10.053.

[41] G. Ghadhban and H. Rasheed, "Robust Tests for the Mean Difference in Paired Data using Jackknife Resampling Technique," *Iraqi J. Sci.*, 2021, doi: 10.24996/ijs.2021.62.9.23.

[42] U. Tonggumnead, "A COMPARATIVE STUDY ON THE EFFICIENCY OF TEST STATISTICS IN TESTING THE DIFFERENCES BETWEEN TWO DEPENDENT DATASETS," *Int. J. Apllied Math.*, 2021, doi: 10.12732/ijam.v34i1.2.

[43] S. Dankel and J. Loenneke, "Effect Sizes for Paired Data Should Use the Change Score Variability Rather Than the Pre-test Variability," *J. Strength Cond. Res.*, vol. 35, pp. 1773–1778, 2018, doi: 10.1519/jsc.0000000000002946.

[44] B. Langenberg, M. Janczyk, V. Koob, R. Kliegl, and A. Mayer, "A tutorial on using the paired t test for power calculations in repeated measures ANOVA with interactions," *Behav. Res. Methods*, vol. 55, pp. 2467–2484, 2022, doi: 10.3758/s13428-022-01902-8.

[45] H. Kang, "Sample size determination and power analysis using the G*Power software," *J. Educ. Eval. Health Prof.*, vol. 18, 2021, doi: 10.3352/jeehp.2021.18.17.

[46] B. Gerald and T. F. Patson, "Parametric and Nonparametric Tests: A Brief Review," *Int. J. Stat. Distrib. Appl.*, 2021, doi: 10.11648/j.ijsd.20210703.12.

[47] Budiono and A. Prasetia, "STUDI PERBANDINGAN HASIL UJI WILCOXON PADA DATA HASIL PENGUKURAN DAN HASIL KATEGORI DATA PENELITIAN KESEHATAN TINGKAT STRESS TEKANAN DARAH DAN MOTORIK HALUS," *J. Ilm. Pamenang*, vol. 4, no. 2, pp. 8–15, 2022, doi: https://doi.org/10.53599/jip.v4i2.94.

[48] F. Fadilatunnisyah, R. F. S, E. A. Fasha, A. K. Putri, and D. A. J. D. Putri, "Penggunaan Uji Wilcoxon Signed Rank Test untuk Menganalisis Pengaruh Tingkat Motivasi Belajar Sebelum dan Sesudah Diterima di Universitas Impian," *Indones. J. Educ. Dev. Res.*, vol. 2, no. 1, 2024, doi: https://doi.org/10.57235/ijedr.v2i1.1887.