



A Statistical Analysis of System Usability Scale (SUS) Evaluations in Online Learning Platform

Ozzi Suria

Information System Department, Universitas Mercu Buana Yogyakarta, Yogyakarta, Indonesia
Email: ozzisuria@mercubuana-yogya.ac.id

Abstract

This study evaluates the usability of a Moodle-based online learning system at a university using the System Usability Scale (SUS). Introduced in 2019, the platform has been instrumental in facilitating access to educational resources, enhancing the distribution of course materials, and supporting academic activities through interactive features. The SUS was employed to gather subjective feedback from 120 student respondents, providing insights into the system's efficiency, effectiveness, and user satisfaction. Statistical analysis was performed using methods including the Shapiro-Wilk test to address the skewed distribution of the original dataset with a mean SUS score of 71.52 and a high standard deviation of 16.18, indicating varied user experiences. Analysis of sample means was also conducted to achieve a more stable estimate of usability, resulting in a mean of 71.51 with a significantly reduced standard deviation of 2.44, suggesting a more consistent assessment across users. The final SUS score placed the system in the 'C+' category according to the Sauro and Lewis grading scale, indicating that while usability is acceptable, substantial improvements are necessary.

Keywords: online learning, System Usability Scale (SUS), Moodle, statistical analysis, usability improvement

1. INTRODUCTION

Online learning systems have become essential in colleges and universities, serving as accessible and flexible tools that enhance the teaching and learning process [1], [2], [3]. These digital platforms facilitate a seamless transition to blended learning environments, where traditional classroom settings are augmented with online resources, enhancing the learning experience [4]. Recent studies have underscored the benefits of online learning, including the flexibility, improved access to educational content anytime and anywhere, increased student engagement, and enhanced collaboration through digital platforms [5], [6], [7], [8], [9]. Furthermore, online learning platforms are not constrained by the limitations associated with physical classroom spaces [10], [11].



At the university, the adoption of a Moodle-based online learning platform has significantly transformed the educational experience for both students and university. Implemented in 2019, this platform was introduced to enhance the distribution and accessibility of course materials, facilitate online lectures, and streamline the scheduling of academic facilities amidst spatial limitations. The move towards this online learning system also aligns with the university's objectives to ensure that educational resources are available to students 24/7. By replacing an outdated system with limited functionality, the new Moodle-based platform has introduced a range of interactive features that support quizzes, assignments, and robust course material sharing. This transition not only reflects the university's commitment to leveraging technology for educational excellence but also its responsiveness to the evolving needs of its academic community.

The application of the System Usability Scale (SUS) in evaluating the usability of the online learning system at the university is grounded in the tool's reliability and simplicity [12], [13]. SUS provides a quick, yet comprehensive, measure of a system's usability, encompassing aspects of efficiency, effectiveness, and user satisfaction [14]. By employing a standardized questionnaire, the SUS allows for the collection of subjective feedback on the online learning platform's usability from a user-centric perspective. This approach is crucial in identifying areas of improvement and ensuring the system meets the users' needs. Furthermore, the SUS's widespread adoption in usability studies across various domains offers a benchmark for comparing the online learning system's performance against established standards.

Statistical analysis plays a pivotal role in supporting the SUS scores obtained from the evaluation of the online learning system. By applying statistical methods to analyze the sample means from SUS score data, we can derive insights into the central tendency of the data [15], [16], [17], providing a clear indication of the overall system usability. This analysis is crucial for supporting the validity of the final SUS score results.

In previous research conducted by [18], a comprehensive systematic review was performed to evaluate the perceived usability of educational technology systems using the System Usability Scale (SUS). The study meticulously analyzed results from 104 research papers, employing statistical analysis to understand usability across various educational technologies. The results indicated a generally good level of usability with average SUS scores around 70.09, demonstrating variations depending on the type of technology and educational context. This statistical approach provided a robust foundation for assessing the effectiveness of the technologies in facilitating educational processes, thus guiding improvements in educational technology design and implementation. The findings serve as a critical reference for educational technology stakeholders aiming to enhance user experience and educational outcomes.

Another research conducted by [19], the paper focuses on evaluating the usability of the Shopee website in Indonesia using the System Usability Scale (SUS). They applied the SUS method to collect and analyze user feedback on the website's usability, emphasizing its practicality and effectiveness in gauging user satisfaction and ease of use. The study involved distributing SUS questionnaires to respondents and statistically analyzing the data to derive usability scores. The results indicated that the Shopee website achieved an average usability score of 67.08, categorizing it within the "OK" range of SUS scores. This outcome suggests that while the website's usability is acceptable, there is potential for further enhancement to improve user experience. The use of statistical analysis and SUS provided a quantitative measure of the website's usability, supporting the development of more user-centric design improvements.

Previous studies by [20], the study explores the quantitative usability assessment of academic websites using experiential and specific task-based System Usability Scale (SUS) ratings. The authors employed the SUS methodology to evaluate the usability of 50 academic websites, comparing the SUS scores derived from general user experience with those obtained from task-specific assessments. Statistical analysis, particularly independent sample t-tests, was used to discern significant differences in mean SUS scores between the two assessment methods. The results indicated a strong positive correlation between the task success rate, an ISO metric of system effectiveness, and the SUS scores, suggesting that successful task completion correlates with higher usability ratings. Overall, the study found that 85.29% of the websites had higher mean SUS scores when evaluated based on specific tasks compared to experiential assessments, highlighting the effectiveness of task-based evaluations in usability studies.

Drawing on prior research, the purpose of employing the System Usability Scale (SUS) to evaluate the online learning system in this study is to objectively assess and enhance the usability of educational platforms. This approach ensures that the system not only meets the technical needs but also enhances user satisfaction and effectiveness in a real-world educational setting. We employ statistical analysis to support the results of the SUS scores, which provide a reliable measure of the system's overall usability. By leveraging these insights, we aim to optimize the online learning experience, making it more intuitive and accessible for all users, ultimately leading to a more effective educational environment.

2. METHODS

The online learning system evaluated in this study was previously developed and has been actively used within the educational framework of our institution. As part of the methodology, we begin by disseminating a questionnaire designed to gather comprehensive feedback from users regarding their interactions with the system.

To provide a clear overview of the research flow and subsequent processes, Figure 1 illustrates the sequential steps from the distribution of the questionnaire to the final analysis of the data.

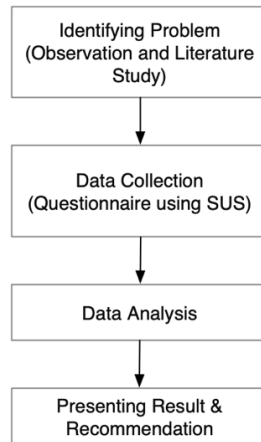


Figure 1. Research steps

3.1. Identifying Problem

At university, an assessment of the existing Moodle-based online learning platform has revealed several areas that require attention. Primarily, students and faculty have expressed concerns about the intuitiveness and responsiveness of the user interface, which can sometimes hinder seamless interaction with course materials. Addressing these problems is essential not only for enhancing the usability and functionality of the online learning system but also for improving overall academic performance and student satisfaction. An extensive literature study also helps contextualize the system within the broader scope of online learning platforms, assessing how similar challenges have been addressed in other academic institutions.

3.2. Data Collection

In this step, a System Usability Scale (SUS) questionnaire was completed by 120 respondents, who are active users of the Moodle-based online learning system per 8 April 2024. The questionnaire consists of 10 statements, allowing users to rate their agreement on a Likert scale from 1 (strongly disagree) to 5 (strongly agree). The questions used for the SUS questionnaire is shown in Table 1.

Table 1. 10 questions of SUS [14]

No	Questions
1	I think that I would like to use this system frequently.
2	I found the system unnecessarily complex.
3	I thought the system was easy to use.
4	I think that I would need the support of a technical person to be able to use this system.
5	I found the various functions in this system were well integrated.
6	I thought there was too much inconsistency in this system.
7	I would imagine that most people would learn to use this system very quickly.
8	I found the system very cumbersome to use.
9	I felt very confident using the system.
10	I needed to learn a lot of things before I could get going with this system.

This standardized tool is chosen for its reliability and ability to gauge the system's effectiveness, efficiency, and user satisfaction. The comprehensive sample size ensures a rich dataset, reflecting the diverse interactions and levels of satisfaction that users have with the system. By targeting a sample of the user population, we ensure a broad representation of experiences and perceptions of the system's usability.

3.3. Data Analysis

During the data analysis phase, two substantial tasks are undertaken to glean insights from the questionnaire data collected using SUS method. Initially, to interpret the data gathered from the SUS questionnaire, we will compute an overarching usability score in accordance with the scoring system established by SUS. This score serves as a quantitative representation of the system's usability from each user's perspective.

We will utilize the following equations to process each individual response: For each question bearing an odd number (Qo), we employ the subsequent equation:

$$Qo = \sum_{i=1,3,5,7,9} x-1 \quad (1)$$

For each question bearing an even number (Qe), we employ the subsequent equation:

$$Qe = \sum_{i=2,4,6,8,10} 5-x \quad (2)$$

x = value from questionnaire,

Following that, the SUS score for each response will be determined using this specific equation:

$$S = (Q_o + Q_e) * 2.5 \quad (3)$$

Lastly, to derive the overall SUS Score from all responses, we implement the equation below:

$$SUS = \frac{\sum_{i=1}^n S_i}{n} \quad (4)$$

n = total number of responses.

To sum up, the equation employed for computing the ultimate SUS score is as follows:

$$SUS = \frac{\sum_{i=1}^n (\sum_{i=1,3,5,7,9}^5 x - 1 + \sum_{i=2,4,6,8,10}^5 5 - x) * 2.5}{n} \quad (5)$$

Subsequently, a comprehensive statistical analysis of the collected SUS scores is to discern the central tendency of the data, which serves as a vital indicator for accurately assessing the system's usability. The Shapiro-Wilk test is conducted to examine potential skewness in the SUS scores distribution. Should the results indicate a non-normal distribution, the central tendency will be determined by analyzing the sample means extracted from the original data. If, however, the distribution is normal, the analysis will proceed with the original dataset. This approach ensures the most representative central value for the SUS scores is ascertained, a key measure of the system's overall usability as experienced by its users. The flow of statistical analysis is shown in figure 2.

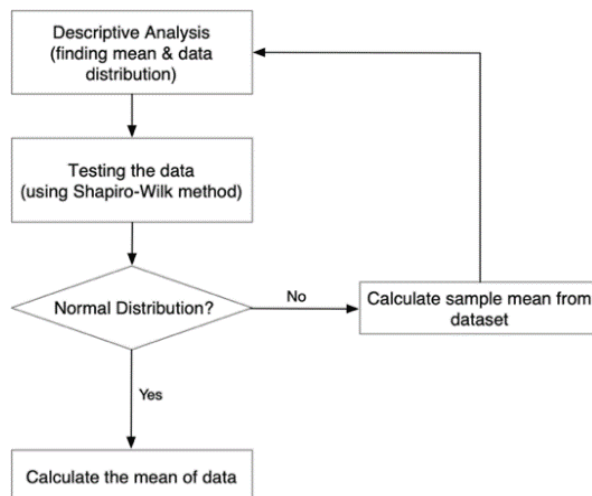


Figure 2. Statistical analysis flow

Following the depiction of the statistical analysis flow in Figure 2, each step of this process will be thoroughly examined in the "Results and Discussion" section. The discussion will follow the sequence outlined in Figure 2, interpreting the statistical outcomes and correlating them with the result of usability assessments of the online learning system.

3. RESULTS AND DISCUSSION

Beginning with a detailed examination of the usability of the university's Moodle-based online learning system, this study involved a total of 120 students as a sample who participated in the System Usability Scale (SUS) questionnaire. These respondents are actively engaged with the platform for a variety of academic activities such as accessing course materials, participating in discussions, and submitting assignments. The online learning system, accessible both on campus via Wi-Fi and off-campus using private data connections, ensures that students have the flexibility to engage with their courses at any time and from any location.

3.1. SUS Analysis Result

Following the analytical procedures outlined in equations (1)-(5) from the previous section, the results of the questionnaire using the SUS method are presented in Table 2. The initial values in the questions column (Q1, Q2, ..., Q10) are derived from a Likert scale ranging from 1 to 5 from the questionnaire and are transformed using equations (1) and (2). The values in the SUS score column are calculated using equation (3), which provides the score for each recorded response. Finally, to determine the final SUS score listed in the bottom row, equation (4) is employed, which calculates the average value of all the SUS scores from each response.

Referring to Table 2, the developed online learning system has attained a final SUS score of 71.52. To categorize this score, the general grading scale for interpreting SUS scores, as detailed by [21], will be utilized. This grading scale is depicted in Table 3.

Table 2. SUS analysis result

R	Questions										SUS Score
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	
1	2	0	1	2	2	2	2	2	2	2	42.5
2	4	4	4	0	4	4	4	0	4	0	70
3	4	4	4	4	4	4	4	4	4	4	100
4	4	4	4	2	4	3	3	4	4	1	82.5
5	3	1	2	1	3	1	2	2	3	2	50
6	4	0	1	4	3	3	1	0	2	3	52.5
7	3	3	3	3	3	2	3	3	2	3	70

8	4	3	3	2	3	3	4	2	3	4	77.5
9	3	2	2	2	3	1	3	3	2	2	57.5
10	4	1	3	2	2	2	2	2	3	3	60
...	...	...	...	...	...	...	...	...	...	...	...
R111	4	4	4	4	4	4	2	3	3	4	90
R112	4	4	4	4	4	4	4	4	4	0	90
R113	4	3	3	3	3	3	3	4	4	2	80
R114	4	4	4	3	4	3	4	4	4	3	92.5
R115	4	3	3	3	3	3	4	4	3	3	82.5
R116	4	3	4	2	4	2	4	4	4	0	77.5
R117	2	3	4	4	2	3	4	4	3	3	80
R118	4	1	3	2	3	1	2	1	3	1	52.5
R119	2	2	3	2	2	1	3	1	3	1	50
R120	3	3	3	3	3	3	3	3	3	3	75
AVERAGE											71.52

Table 3. General grading scale interpretation of SUS scores [21]

SUS Score Range	Grade	Percentile Range
84.1-100	A+	96-100
80.8-84.0	A	90-95
78.9-80.7	A-	85-89
77.2-78.8	B+	80-84
74.1-77.1	B	70-79
72.6-74.0	B-	65-69
71.1-72.5	C+	60-64
65.0-71.0	C	41-59
62.7-64.9	C-	35-40
51.7-62.6	D	15-34
0.0-51.6	F	0-14

The final SUS score obtained is 71.52, which places it within the 'C+' grade category. This classification indicates that the usability of the system is slightly above average but not yet reaching the levels of 'good' or 'excellent.' A 'C+' grade indicates that while the system is deemed functional and generally acceptable, there are still notable areas for improvement to enhance user satisfaction and overall experience. This grade suggests that the system meets basic usability standards but could benefit significantly from targeted refinements. To see that the obtained SUS score is accurate, we will conduct a statistical analysis in the next section.

3.2. Statistical Analysis

3.2.1. Descriptive Analysis

Statistical analysis was conducted using Python, utilizing various libraries including pandas, numpy, seaborn, and pyplot to facilitate the analysis of the questionnaire data shown in Table 2. The execution of the code depicted in Figure 3 yielded

results consistent with those listed in Table 2, including the standard deviation. The calculated mean, which matches the final SUS score reported in Table 2, was found to be 71.52 with a standard deviation of 16.18.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
data = pd.read_csv('/sus_scores.csv', header=0)
descriptive_stats = data['sus_score'].describe()
print("Descriptive Statistics:\n", descriptive_stats)

Descriptive Statistics:
count    120.000000
mean     71.520833
std      16.176858
min       22.500000
25%       60.000000
50%       75.000000
75%       85.000000
max      100.000000
Name: sus_score, dtype: float64
```

Figure 3. Descriptive analysis result from data in table 2

The standard deviation of 16.18 points to the spread of the SUS scores around the mean. This relatively high standard deviation indicates a wide variability in user perceptions of usability. The wide range in SUS scores suggests that different users may have significantly different experiences and opinions about the usability of the system. Some users may find the system quite usable, while others may encounter difficulties.

3.2.2. Testing the Data

To evaluate the distribution of the data, the Shapiro-Wilk test is employed to check for normality within the dataset. The Shapiro-Wilk test is particularly effective for assessing the normality of data sets when the sample sizes do not exceed 2000 [22]. Should the p-value from the Shapiro-Wilk test fall below the predetermined significance threshold of $\alpha = 0.05$, it indicates that the data does not follow a normal distribution. This significance level of 0.05 is selected given the relatively moderate sample size. Upon running the script shown in Figure 4, a histogram representing the data distribution was generated, as depicted in Figure 5.

```
sns.histplot(data['sus_score'], kde=True)
plt.title('SUS Score Distribution')
plt.xlabel('SUS Score')
plt.ylabel('Frequency')
plt.show()

shapiro_test = stats.shapiro(data['sus_score'])
print("Shapiro-Wilk Test:\n W-statistic={:.3f}, p-value={:.3f}".format(shapiro_test[0], shapiro_test[1]))
if shapiro_test[1] > 0.05:
    print("The distribution of SUS scores appears to be normal.")
else:
    print("The distribution of SUS scores does not appear to be normal.")
```

Figure 4. Generate histogram of data and testing the distribution

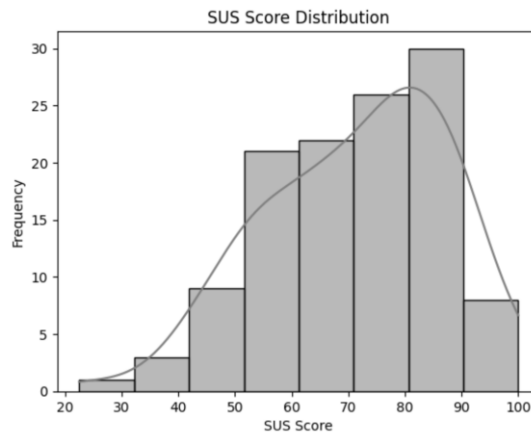


Figure 5. SUS scores data histogram

The results of the Shapiro-Wilk test yielded a W statistic of 0.971 and a p -value of 0.010. The W value, being close to 1, indicates that the data might be normally distributed. However, since the p -value is below the threshold of 0.05, it leads to the conclusion that the data are not normally distributed.

3.2.3. Sample Means Approach

Given that the original dataset from Table 2 does not exhibit a normal distribution, an alternative approach involves using sample means. Using sample means applies the Central Limit Theorem to data with a non-normal distribution, enabling the use of normal distribution principles when dealing with large sample sizes [23], [24]. This method entails drawing multiple samples from the SUS scores data and calculating the mean for each sample to form a distribution of sample means. This technique aims to identify the central tendency of the SUS scores, particularly useful given the large standard deviation of the original scores. By implementing the procedure outlined in Figure 6, sample mean data was derived from the original dataset in Table 2. To calculate these sample means, 1200 samples were used, each consisting of 40 data points, to ensure a robust estimation of the sampling distribution.

```
def cal_sample_means(data, n_smp=1200, smp_size=40):
    means = []
    for _ in range(n_samples):
        sample = np.random.choice(data, size=sample_size, replace=True)
        means.append(np.mean(sample))
    return means
data2 = pd.read_csv('/sus_scores_akuntansi.csv', header=0)
sample_means = cal_sample_means(data2['sus_score'], n_smp=1200, smp_size=40)
```

Figure 6. Get sample means

After obtaining the sample means, the code depicted in Figures 3, 4, and 5 can be utilized to perform a descriptive analysis of these means and their distribution. The outcomes of this analysis are presented in Table 4.

Table 4. Sample means result

No	Sample Mean
1	68.375
2	72.187
3	69.187
4	74
5	72.5
...	...
1196	68.625
1197	67.937
1198	71.562
1199	70.687
1200	72.187

The descriptive analysis reveals a mean SUS score of 71.51 with a standard deviation of 2.44, indicating that the usability of the online learning system is similar to the result of original dataset. The relatively low standard deviation suggests that the SUS scores are consistently clustered around the mean, implying a reliable and uniform user experience among the respondents. The Shapiro-Wilk test on the sample means data yielded a W value of 0.998 and a p-value of 0.338. Given these results and a significance level of $\alpha = 0.05$, it can be concluded that the data is normally distributed. The distribution of the sample means data is depicted in Figure 7.

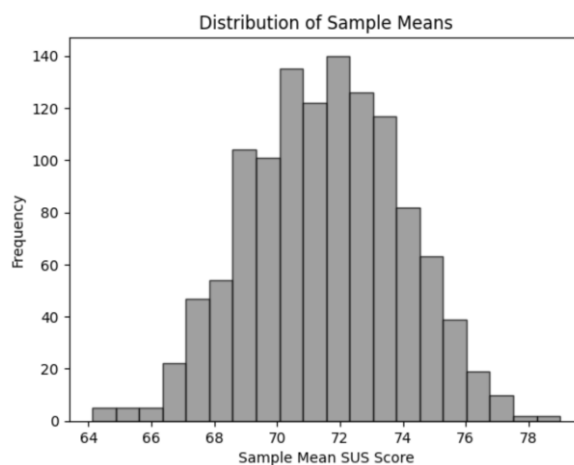


Figure 7. Sample means distribution

The final SUS score from Table 2, along with its standard deviation of 16.18, and the SUS score from sample means result in Table 4, with a standard deviation of 2.44, are both visually displayed in Figure 8.

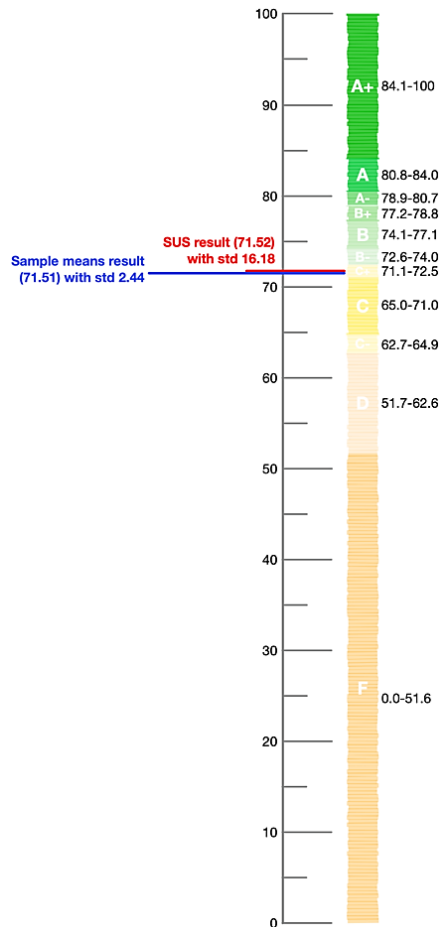


Figure 8. SUS and sample means result based on grading scale interpretation of SUS score (blue line represents sample means result, red line represents SUS result)

3.3. Discussion

In the previous section, the analysis determined that the online learning platform received a System Usability Scale (SUS) score of 71.52, placing it in the 'C+' category as per the Sauro and Lewis grading scale. This indicates that while the platform's usability meets acceptable standards, there remains potential for enhancing user satisfaction. The results align with broader usability research

trends, which typically display significant variation in user experiences, as evidenced by the high initial standard deviation of 16.18. The reduction of this variability to 2.44 in the sample means analysis suggests a more uniform assessment of the platform, averaging across various samples. This emphasizes the need to account for individual differences in user experiences, which can significantly impact overall usability evaluations, and illustrates how the Central Limit Theorem helps achieve a more accurate estimation of central tendencies in skewed distributions.

In practical terms, utilizing the System Usability Scale (SUS) provided a solid framework for measuring usability, which was complemented by statistical analysis. However, the limitation of this approach is evident in the initial non-normal distribution of data, which required adjustment by analyzing sample means. Practical implications from this study suggest targeted enhancements in areas such as user interface design and interactive features, which could elevate the platform from merely acceptable to excellent, enhancing both user satisfaction and educational outcomes. Additionally, the sample means method proved effective in stabilizing variability, offering a more consistent measure of central tendency. This approach underscores the importance of ongoing usability testing and iterative design improvements to better meet user needs.

4. CONCLUSION

Based on the results and analysis from the previous section, the System Usability Scale (SUS) score for the online learning system stands at 71.52, categorized as 'C+' on the Sauro and Lewis grading scale. This indicates that while the system's usability is slightly above average, there is room for significant improvements. The close alignment of the SUS score with the central tendency obtained from the sample means method, which is 71.51 with a much lower standard deviation of 2.44 compared to the original 16.18, highlights the consistency and reliability of these findings. The reduced standard deviation in the sample means method points to a narrower spread of user opinions, suggesting a more uniform assessment of usability across the surveyed users. This convergence in user perception underscores the system's functional adequacy but also indicates clear opportunities for enhancements to elevate the user experience to higher usability grades. This result suggests that by focusing on specific improvements to the online learning system, the university can enhance student engagement and satisfaction, better supporting their academic success.

Future work should focus on addressing the specific areas of usability that have room for improvement within the online learning system. Detailed user feedback sessions can be conducted to identify the precise aspects that users find lacking or cumbersome. Subsequent studies could also explore the impact of these enhancements on user satisfaction by conducting follow-up SUS evaluations after

the improvements are implemented. Additional methods of evaluation can also be used as a comparison. Ultimately, this iterative approach to usability testing and system refinement is crucial for continuously adapting the online learning platform to the dynamic needs of the university's educational environment.

REFERENCES

- [1] U. Mukarromah and W. Wijayanti, "Implementation of the online learning at vocational high school during Covid-19: Between obligations and barriers," *Jurnal Pendidikan Vokasi*, vol. 11, no. 1, Mar. 2021, doi: 10.21831/jpv.v11i1.37110.
- [2] V. Singh and A. Thurman, "How Many Ways Can We Define Online Learning? A Systematic Literature Review of Definitions of Online Learning (1988-2018)," *American Journal of Distance Education*, vol. 33, no. 4, pp. 289–306, Oct. 2019, doi: 10.1080/08923647.2019.1663082.
- [3] E. Enoch, B. Qamaruzzaman, and Q. Y. Zaqiyah, "Pengembangan dan Implementasi E-Learning di Universitas Islam Bandung," *Ta'dib Jurnal Pendidikan Islam*, vol. 12, no. 2, pp. 573–584, Aug. 2023, doi: 10.29313/TJPI.V12I2.12312.
- [4] H. C. Kuong, "Enhancing Online Learning Experience: From Learners' Perspective," *Procedia Soc Behav Sci*, vol. 191, pp. 1002–1005, Jun. 2015, doi: 10.1016/j.sbspro.2015.04.403.
- [5] N. K. A. Suwastini, K. I. R. Dewi, and G. R. Dantes, "Benefits of online learning according to recent studies," *Jurnal Inovasi dan Teknologi Pembelajaran*, vol. 10, no. 1, pp. 32–42, Mar. 2023, doi: 10.17977/UM031V10I12023P032.
- [6] L. Pei and H. Wu, "Does online learning work better than offline learning in undergraduate medical education? A systematic review and meta-analysis," *Med Educ Online*, vol. 24, no. 1, p. 1666538, Jan. 2019, doi: 10.1080/10872981.2019.1666538.
- [7] P. Paudel, "Online Education: Benefits, Challenges and Strategies During and After COVID-19 in Higher Education," *International Journal on Studies in Education*, vol. 3, no. 2, pp. 70–85, Sep. 2021, doi: 10.46328/IJONSE.32.
- [8] M. D. B. Castro and G. M. Tumibay, "A literature review: efficacy of online learning courses for higher education institution using meta-analysis," *Educ Inf Technol (Dordr)*, vol. 26, no. 2, pp. 1367–1385, Mar. 2021, doi: 10.1007/S10639-019-10027-Z/METRICS.
- [9] H. N. Y. Nhi and T. T. Nhan, "Online Education During The Covid-19 Pandemic - Opportunities And Challenges: A Case Study In The South Of Vietnam," *European Journal of Open Education and E-learning Studies*, vol. 7, no. 1, Apr. 2022, doi: 10.46827/ejoe.v7i1.4222.
- [10] A. Kilburn, B. Kilburn, and T. Cates, "Drivers of Student Retention: System Availability, Privacy, Value And Loyalty In Online Higher Education," *Academy of Educational Leadership Journal*, vol. 18, no. 4, 2014.

- [11] W. Bhuasiri, O. Xaymoungkhoun, H. Zo, J. J. Rho, and A. P. Ciganek, "Critical success factors for e-learning in developing countries: A comparative analysis between ICT experts and faculty," *Comput Educ*, vol. 58, no. 2, pp. 843–855, Feb. 2012, doi: 10.1016/J.COMPEDU.2011.10.010.
- [12] J. Brooke, "SUS: a retrospective," *J Usability Stud*, vol. 8, no. 2, pp. 29–40, 2013.
- [13] B. Klug, "An Overview of the System Usability Scale in Library Website and System Usability Testing," *Weave: Journal of Library User Experience*, vol. 1, no. 6, Apr. 2017, doi: 10.3998/weave.12535642.0001.602.
- [14] J. Brooke, "SUS-a quick and dirty usability scale," in *Usability Evaluation In Industry*, P. W. Jordan, B. Thomas, and I. Lyall, Eds., London, Bristol: Taylor & Francis, 1996, pp. 189–194.
- [15] S. Manikandan, "Measures of central tendency: The mean," *J Pharmacol Pharmacother*, vol. 2, no. 2, pp. 140–142, Jun. 2011, doi: 10.4103/0976-500X.81920.
- [16] M. A. Barri, "A Simulation Showing the Role of Central Limit Theorem in Handling Non-Normal Distributions," *Am J Educ Res*, vol. 7, no. 8, pp. 591–598, Aug. 2019, doi: 10.12691/education-7-8-8.
- [17] N. Clark, M. Dabkowski, P. J. Driscoll, D. Kennedy, I. Kloo, and H. Shi, "Empirical Decision Rules for Improving the Uncertainty Reporting of Small Sample System Usability Scale Scores," *Int J Hum Comput Interact*, vol. 37, no. 13, pp. 1191–1206, Aug. 2021, doi: 10.1080/10447318.2020.1870831.
- [18] P. Vlachogianni and N. Tselios, "Perceived usability evaluation of educational technology using the System Usability Scale (SUS): A systematic review," *Journal of Research on Technology in Education*, vol. 54, no. 3, pp. 392–409, May 2022, doi: 10.1080/15391523.2020.1867938.
- [19] F. G. Sembodo, G. F. Fitriana, and N. A. Prasetyo, "Evaluasi Usability Website Shopee Menggunakan System Usability Scale (SUS)," 2021.
- [20] K. Sagar and A. Saha, "Quantitative usability assessment relying on experiential and specific task-based SUS ratings," *Journal of Statistics and Management Systems*, vol. 23, no. 2, pp. 333–348, Feb. 2020, doi: 10.1080/09720510.2020.1736316.
- [21] J. Sauro and J. R. Lewis, *Quantifying the User Experience: Practical Statistics for User Research*, 2nd Edition. Cambridge: Elsevier, 2016. doi: 10.1016/C2010-0-65192-3.
- [22] J. P. Royston, "An Extension of Shapiro and Wilk's W Test for Normality to Large Samples," *J R Stat Soc Ser C Appl Stat*, vol. 31, no. 2, pp. 115–124, Jun. 1982, doi: 10.2307/2347973.
- [23] S. G. Kwak and J. H. Kim, "Introduction Basic Concepts of Central Limit Theorem Central limit theorem: the cornerstone of modern statistics KJA," *Korean J Anesthesiol*, vol. 70, no. 2, pp. 144–156, 2017, doi: 10.4097/kjae.2017.70.2.144.

- [24] R. Bono, M. J. Blanca, J. Arnau, and J. Gómez-Benito, "Non-normal distributions commonly used in health, education, and social sciences: A systematic review," *Front Psychol*, vol. 8, no. SEP, p. 267072, Sep. 2017, doi: 10.3389/FPSYG.2017.01602/BIBTEX.