

Evaluation of STIESIA E-Learning User Experience Using User Experience Questionnaire in the Implementation of Blended Learning

Achmad Fakhri Dzulfikar

Sekolah Tinggi Ilmu Ekonomi Indonesia Surabaya, Jl Menur Pumpungan No.30, Surabaya, 60118

ARTICLE INFO

Article historys:

Received : 27/07/2026

Revised : 15/08/2026

Accepted : 08/09/2026

Keywords:

Blended Learning; E-Learning Platform; User Experience; User Experience Questionnaire

ABSTRACT

This study evaluates user experience (UX) of STIESIA Surabaya's e-learning platform within its blended learning program, with the aim of identifying specific strengths and gaps. Using a quantitative descriptive-evaluative design, researcher distributed User Experience Questionnaire (UEQ) to 90 Accounting and Management students, each with at least one semester of platform use. After screening 96 initial submissions, researcher analyzed 90 verified responses against a UEQ benchmark of more than 20.000 products. All six dimensions scored positively in absolute terms. Perspicuity rated highest (1.33 Above Average), followed by Attractiveness (1.20) and Efficiency (1.13), and Stimulation (1.13), all above average. Dependability (1.00) and Novelty (0.43) fell Below Average against benchmark. An independent t-test found no significant differences between programs. Platform handles core pragmatic tasks well but shows relative weakness in novelty and dependability, which points to a need for user-centered investigation before any institution-wide redesign.



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

Corresponding Author:

Achmad Fakhri Dzulfikar
Sekolah Tinggi Ilmu Ekonomi Indonesia Surabaya, 60118, Surabaya
Email: achmadfakhri@stiesia.ac.id

1. INTRODUCTION

Blended learning is now a common model across Indonesian universities, combining face to face instruction with online components to give students more flexibility in accessing course materials and activities[1]. Some research suggest a mix of 70-80% online and 20-30% in person instruction works well in certain contexts[2], but exact ratio matters less than whether supporting platform is one students can actually use. At STIESIA, blended learning means regular classroom sessions supplemented by mandatory online activities: content delivery, assignment submission, discussion forums, and assessments, all run through a dedicated e-learning platform.

COVID-19 forced most institutions to move fully online within weeks. Blended models that followed were often built with little time to check whether underlying platforms actually worked[3]. STIESIA Surabaya, a private university with roughly 1.500 active students, runs its blended programme on a home-grown e-learning platform. Having a platform in place, however, does not guarantee that students learn well from it. Prior research shows that poor platform design leads students to disengage, get frustrated, and find workarounds that defeat purpose of blended learning in first place[4].

ISO 9241-210 defines user experience as everything a person perceives, feels, and concludes when interacting with a system[5]. In an e-learning context, those responses are not separate from learning,

they shape how much students engage with material and how much stays with them. A study from UAE found that student's overall judgment of whether blended learning was working for them was closely tied to how well platform itself held up[6]. That finding has a practical consequence: assessing platform's UX is not an optional audit. It is the only way to know what students are actually experiencing when they log in.

This study asks four questions: (1) How does STIESIA's e-learning platform score across six UEQ dimensions? (2) How do students perceive blended learning experience platform is meant to support? (3) which dimensions receive highest and lowest ratings? and (4) What patterns of strength and weakness emerge from dimensional scores? Fourth question stays descriptive by design, focused on interpreting score patterns rather than claiming what causes them. Any insight into underlying causes comes from open-ended survey responses, analysed through thematic coding to identify recurring themes.

Study uses User Experience Questionnaire (UEQ), developed by Martin Schrepp[7]. Instrument covers three pragmatic dimensions (Perspicuity, Efficiency, and Dependability) and two hedonic dimensions (Stimulation, and Novelty), plus an overall Attractiveness score [8]. Students complete it in three to five minutes, which made total sampling across full cohort practical. UEQ also comes with an international benchmark database, so STIESIA's results can be compared against data from more than 20.000 commercial and non-commercial products rather than read in isolation[9].

Scores run on a -3 to +3, and official UEQ guidelines use two separate interpretations. First concerns practical significance: scores above +0.8 count as positive, scores between -0.8 and +0.8 as neutral, and scores below -0.8 as negative. Second is benchmark based classification, which places a score into one of five categories, Excellent, Good, Above Average, Below Average, or Bad, using percentile cutoffs specific to each dimension and to the benchmark dataset's version[10]. These two interpretations answer different questions: first tells you whether a score is favourable in absolute terms, second tells you where platform stands relative to a large reference sample. This study reports both, using official UEQ Data Analysis Tool (www.ueq-online.org. version 14 2024) to compute means, confidence intervals, and benchmark comparisons[11]. Data were collected through a quantitative, descriptive-evaluative design using total sampling.

2. RESEARCH METHOD

This study used a quantitative descriptive-evaluative approach to assess user experience (UX) on STIESIA Surabaya's e-learning platform in a blended learning context. Main instrument was the User Experience Questionnaire (UEQ), a validated tool with 26 pairs of opposing adjectives rated on a 7 point scale (-3 to +3), covering six dimensions: Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty. Table 1 lists each dimension with its corresponding indicators and item numbers.

Table 1. Parameter UEQ

Variable	Indicator	Number Indicator
Attractiveness	Annoying / Enjoyable	1
	Bad / Good	12
	Unlikeable / Pleasing	14
	Unpleasant / Pleasant	16
	Unattractive / Attractive	24
	Unfriendly / Friendly	25
Perspicuity	Not understandable / Understandable	2
	Difficult to learn / Easy to learn	4
	Complicated / Easy	13
	Confusing / Clear	21
Efficiency	Slow / Fast	9
	Inefficient / Efficient	20
	Impractical / Practical	22
	Cluttered / Organized	23

Variable	Indicator	Number Indicator
Dependability	Unpredictable / Predictable	8
	Obstructive / Supportive	11
	Not secure / Secure	17
	Does not meet expectations / Meet Expectations	19
Stimulation	Inferior / Valuable	5
	Boring / Exciting	6
	Not interesting / Interesting	7
	Demotivating / Motivating	18
Novelty	Dull / Creative	3
	Conventional / Inventive	10
	Usual / Leading edge	15
	Conservative / Innovative	26

2.1. Population and Sample

Institutional population comprises all active students at STIESIA Surabaya (approximately 1.500). accessible population, however, was defined more narrowly: all students enrolled in two courses taught by researcher during odd semester of 2025/2026 academic year, one in the Accounting programme and one in the Management programme, who had used e-learning platform for at least one prior semester. This accessible frame totalled 90 students across two course sections.

Because this number was relatively small, a total sampling (census) approach was applied: all 90 eligible students were invited to participate ($n=N=90$). Intent was to maximise data coverage within this specific cohort, not to claim representativeness of entire institutional student body[12]. Sampling frame was chosen for practical reasons: manageable data access, researcher's first-hand knowledge of platform use in these course, and a semester long research timeline. Including both Accounting and Management sections also allowed for exploratory cross programme comparisons.

Drawing sample from researcher's own classes introduces potential convenience selection bias and may limit generalizability to other courses, instructors, or semesters. To reduce authority or coercion effects, participation was entirely voluntary, and questionnaire was administered anonymously through an external platform (Google Forms) with no students identifiers linked to responses. Student were told, both invitation on the first page of the form, that participating or withdrawing would have no effect on their course grades or academic standing.

2.2. Instrument and Data Collection

Data were collected using UEQ questionnaire, distributed online via Google Forms. Form had three sections: respondent demographics, 26 UEQ items, and open-ended questions on contextual factors in blended learning experience.

Before full distribution, a pilot test was run with 10 respondents drawn from same accessible population but not enrolled in two target course sections. Pilot checked item clarity, readability, and estimated completion time. No major wording changes were needed, minor formatting adjustments improved display on mobile devices. These 10 pilot respondents were excluded from final analysis to avoid data contamination. Main data collection ran for four weeks, with reminder prompts sent in second and third weeks to sustain response rates.

Institutional permission was obtained before data collection began. All participants gave informed consent via a checkbox on the first page of the online form. Respondents were assured of anonymity, data confidentiality, and right to withdraw at any point without consequence. Only aggregated, non identifiable data are reported.

	1	2	3	4	5	6	7		
annoying	☐	☐	☐	☐	☐	☐	☐	enjoyable	1
not understandable	☐	☐	☐	☐	☐	☐	☐	understandable	2
creative	☐	☐	☐	☐	☐	☐	☐	dull	3
easy to learn	☐	☐	☐	☐	☐	☐	☐	difficult to learn	4
valuable	☐	☐	☐	☐	☐	☐	☐	inferior	5
boring	☐	☐	☐	☐	☐	☐	☐	exciting	6
not interesting	☐	☐	☐	☐	☐	☐	☐	interesting	7
unpredictable	☐	☐	☐	☐	☐	☐	☐	predictable	8
fast	☐	☐	☐	☐	☐	☐	☐	slow	9
inventive	☐	☐	☐	☐	☐	☐	☐	conventional	10
obstructive	☐	☐	☐	☐	☐	☐	☐	supportive	11
good	☐	☐	☐	☐	☐	☐	☐	bad	12
complicated	☐	☐	☐	☐	☐	☐	☐	easy	13
unlikable	☐	☐	☐	☐	☐	☐	☐	pleasing	14
usual	☐	☐	☐	☐	☐	☐	☐	leading edge	15
unpleasant	☐	☐	☐	☐	☐	☐	☐	pleasant	16
secure	☐	☐	☐	☐	☐	☐	☐	not secure	17
motivating	☐	☐	☐	☐	☐	☐	☐	demotivating	18
meets expectations	☐	☐	☐	☐	☐	☐	☐	does not meet expectations	19
inefficient	☐	☐	☐	☐	☐	☐	☐	efficient	20
clear	☐	☐	☐	☐	☐	☐	☐	confusing	21
impractical	☐	☐	☐	☐	☐	☐	☐	practical	22
organized	☐	☐	☐	☐	☐	☐	☐	cluttered	23
attractive	☐	☐	☐	☐	☐	☐	☐	unattractive	24
friendly	☐	☐	☐	☐	☐	☐	☐	unfriendly	25
conservative	☐	☐	☐	☐	☐	☐	☐	innovative	26

Figure 1. Indicator and Scale UEQ

2.3. Data Analysis

Responses were cleaned for outliers and internal consistency using UEQ Data Analysis Tool[10], and each score was converted from 7-point scale to the -3 to +3 range. Dimensional averages were then computed with 95% confidence intervals and benchmarked against UEQ reference database (over 20.000 products), which classifies performance into five tiers: Excellent (>1.75), Good (0.80-1.75), Above Average (0.52-0.80), Below Average (0.00-0.52), and Bad (<0.00)[9]. An independent samples t-test was run to compare UX perceptions between Accounting and Management groups, and open-ended responses on blended learning context were analyzed descriptively[13].

3. RESULTS AND DISCUSSION

Data were collected via an online Google Form over four weeks, with reminder prompts sent in the second and third weeks to encourage participation. A total of 96 initial submissions were received. After screening for completeness and potential duplicates, six responses were excluded, resulting in a final verified sample of 90 respondents fully matching defined accessible population. Final sample comprised 50 Accounting students (55,6%) and 40 Management students (44,4%), as shown in Table 2. This verified sample serves as basis for all subsequent analyses.

Table 2. Respondent Profile

Study Program	Frequency (n)	Percentage (%)
S1 Akuntansi	50	55,6%
S1 Manajemen	40	44,4%
Total	90	100%

3.1. Evaluation of E-Learning User Experience

A total of 90 verified responses were processed using UEQ Data Analysis Tool (Version 14, obtained from official website in July 2026) at www.ueq-online.org. Tool converts raw 1-7 scale values to a -3 to +3 transformed scale and computes item level and dimension level descriptive statistics.

Item	Mean	Variance	Std. Dev.	No.	Left	Right	Scale
1	1.2	1.8	1.3	90	annoying	enjoyable	Attractiveness
2	1.5	1.4	1.2	90	not understandable	understandable	Perspicuity
3	0.7	2.7	1.7	90	creative	dull	Novelty
4	1.1	2.6	1.6	90	easy to learn	difficult to learn	Perspicuity
5	1.6	2.3	1.5	90	valuable	inferior	Stimulation
6	0.7	1.5	1.2	90	boring	exciting	Stimulation
7	0.9	1.7	1.3	90	not interesting	interesting	Stimulation
8	0.1	2.0	1.4	90	unpredictable	predictable	Dependability
9	0.3	2.5	1.6	90	fast	slow	Efficiency
10	0.0	2.3	1.5	90	inventive	conventional	Novelty
11	1.5	1.5	1.2	90	obstructive	supportive	Dependability
12	1.4	1.8	1.3	90	good	bad	Attractiveness
13	1.3	1.3	1.1	90	complicated	easy	Perspicuity
14	0.9	1.5	1.2	90	unlikable	pleasing	Attractiveness
15	0.4	1.9	1.4	90	usual	leading edge	Novelty
16	1.4	1.4	1.2	90	unpleasant	pleasant	Attractiveness
17	1.4	2.2	1.5	90	secure	not secure	Dependability
18	1.3	1.8	1.3	90	motivating	demotivating	Stimulation
19	1.0	1.7	1.3	90	meets expectations	does not meet expectations	Dependability
20	1.5	1.2	1.1	90	inefficient	efficient	Efficiency
21	1.4	1.9	1.4	90	clear	confusing	Perspicuity
22	1.3	1.9	1.4	90	impractical	practical	Efficiency
23	1.4	2.0	1.4	90	organized	cluttered	Efficiency
24	0.9	1.6	1.3	90	attractive	unattractive	Attractiveness
25	1.4	1.5	1.2	90	friendly	unfriendly	Attractiveness
26	0.6	1.9	1.4	90	conservative	innovative	Novelty

Figure 2. Item Average

Three lowest scoring items were item 10 Inventive vs Conventional (0.0, Novelty dimension), item 8 Unpredictable vs Predictable (0.1, Dependability dimension), and item 9 Fast vs Slow (0.3, Efficiency dimension)

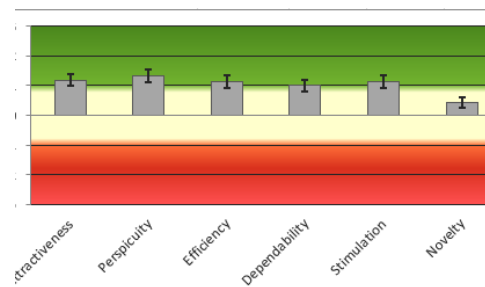


Figure 3. 6 UEQ

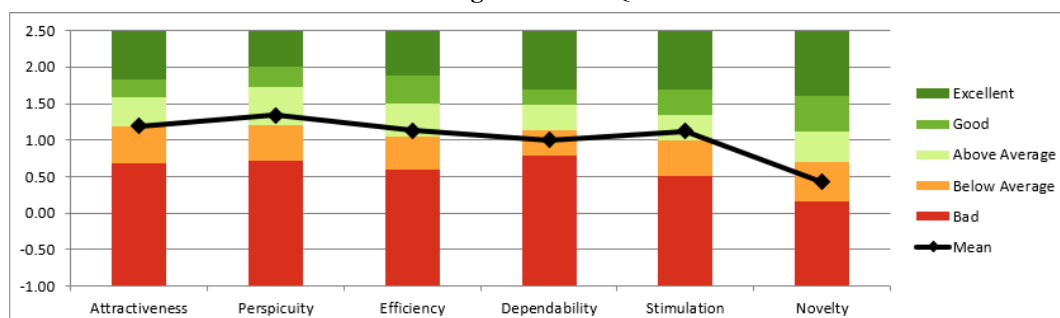


Figure 4. Benchmark UEQ

Under standard UEQ absolute interpretation, scores between -0.8 and +0.8 count as neutral in practical terms. These three item scores fall in that range, so they indicate neither strong dissatisfaction nor positive endorsement. Students do not perceive platform as notably conventional, unpredictable, or slow, but they do not rate these aspects as strong points either.

All six dimension means were positive. Two separate interpretive lenses apply to these results. First is absolute interpretation, or practical significance, scores above +0.8 count as positive, scores between -0.8 and +0.8 as neutral, and scores below -0.8 as negative. Second is benchmark classification, which places each score against a reference database of more than 20.000 product evaluations built into Version

14 of the tool. In this classification, Above Average means 25% of products in database score higher and 50% score lower, while Below Average means 50% score higher and 25% score lower.

Table 3 reports dimension means, standard deviations, 95% confidence intervals, benchmark classifications, and percentile interpretations.

Table 3 UEQ Benchmark Classification

UEQ Dimension	Mean	Std. Dev	95% CI	Classification	Interpretation
Attractiveness	1.20	1.24	[0.94, 1.46]	Above Average	25% score higher, 50% score lower
Perspicuity	1.33	1.23	[1.07, 1.59]	Above Average	25% score higher, 50% score lower
Efficiency	1.13	1.25	[0.87, 1.39]	Above Average	25% score higher, 50% score lower
Dependability	1.00	1.28	[0.73, 1.27]	Below Average	50% score higher, 25% score lower
Stimulation	1.13	1.28	[0.86, 1.40]	Above Average	25% score higher, 50% score lower
Novelty	0.43	1.42	[0.13, 0.73]	Below Average	50% score higher, 25% score lower

In absolute terms, Perspicuity (1.33), Attractiveness (1.20), Efficiency (1.13), Stimulation (1.13) and Dependability (1.00) all exceed +0.8 threshold, which confirms a practically positive perception on each. Students find platform easy to learn, visually acceptable, efficient for routine tasks, reasonably motivating, and sufficiently reliable. Against benchmark, first four dimensions rank Above Average, platform outperforms median product in reference database on these measures.

Dependability (1.00) is positive in absolute terms but falls Below Average against benchmark. Low score on predictability (Item 8 = 0.1) suggests that occasional inconsistencies in system feedback or navigation responses may mildly undermine user confidence. This does not point to a critical failure, but it does mark an area worth refining.

Novelty (0.43) is lowest scoring dimension. It sits inside neutral range in absolute terms (between -0.8 and +0.8) and Below Average against benchmark. Students do not actively dislike platform's creativity or innovation. They simply do not experience it as modern or cutting-edge. This matches Santoso, who found that educational platforms often prioritize functional pragmatism over hedonic qualities such as visual originality and contemporary interaction design[14]. Score points to room for interface improvement rather than to user dissatisfaction or a barrier to learning.

3.2. Comparative Analysis

To check whether UX perceptions differed between Accounting and Management students, an independent samples t-test was run for each of six UEQ dimensions. Shapiro-wilk test and Q-Q plots assessed normality, Levene's test checked homogeneity of variances, and assumptions held reasonably well across all scales. Table 4 reports group descriptive statistics, mean differences, 95% confidence intervals, test statistics, and effect sizes (Cohen's d).

Table 4. Comparative Analysis

UEQ Dimension	Accounting Mean (SD)	Managemen t Mean (SD)	Mean Diff	95% CI of Diff	t-value	df	p-value	Cohen's d
Attractiveness	1.21 (1.24)	1.19 (1.25)	0.02	[-0.50, 0.54]	0.08	88	0.938	0.02
Perspicuity	1.35 (1.23)	1.31 (1.24)	0.04	[-0.48, 0.56]	0.15	88	0.881	0.03
Efficiency	1.14 (1.25)	1.12 (1.26)	0.02	[-0.50, 0.54]	0.08	88	0.938	0.02
Dependability	1.02 (1.28)	0.98 (1.29)	0.04	[-0.49, 0.57]	0.15	88	0.881	0.03
Stimulation	1.12 (1.27)	1.14 (1.30)	-0.02	[-0.55, 0.51]	-0.08	88	0.938	-0.02
Novelty	0.44 (1.42)	0.42 (1.43)	0.02	[-0.57, 0.61]	0.07	88	0.948	0.01

Across all six dimensions, p-values exceed 0.05, and 95% confidence intervals for mean differences cross zero. Cohen's d effect sizes are negligible, ranging from 0.01 to 0.03.

These results need a cautious reading. Failing to reject null hypothesis does not prove two groups are identical or practically indistinguishable. It only means that, within this specific cohort and sample, data do not support a program based distinction.

Sampling frame itself limits how far this reading can go, it comprised two courses taught by same researcher. Shared course content, similar assignment structures, or instructor specific effects could produce this uniformity and mask real program level differences that exist elsewhere. At most, findings show that within this semester and these two courses sections, students from both programs reported comparable UX perceptions. This offers preliminary evidence that UX issues such as low novelty relative to benchmark are unlikely to be tied strongly to academic discipline. It does not establish that a single institution wide intervention will benefit all students equally across every course, instructor, or academic year[15].

4. CONCLUSION

STIESIA Surabaya's e-learning platform shows positive pragmatic usability across most UEQ dimensions: Perspicuity (1.33), Attractiveness (1.20), Efficiency (1.13), and Stimulation (1.13) all score above +0.8 and rank Above Average against benchmark, while Dependability (1.00) and Novelty (0.43) are positive to neutral in absolute terms but Below Average against reference database. Comparative analysis found no statistically significant differences between Accounting and Management students in this sample, but that absence of significance should not be read as proof of program wide uniformity, given limited sampling frame (two courses, one instructor, one semester) and potential confounding factors involved. These results offer a diagnostic baseline rather than a prescriptive roadmap: UEQ identifies where platform falls short but does not reveal which specific interface features or functional issues drive Novelty and Dependability scores. Rather than move straight to a broad visual redesign, institution should first run task-based usability testing and semi structured interviews with a more diverse student sample, expand survey to additional courses and instructors, and use those qualitative findings to guide targeted, pilot tested improvements before any institution-wide rollout.

REFERENCES

- [1] D. Masveta and M. E. Manyangara, "The UX-UI continuum: exploring the interplay between user experience and user interface in e-learning platforms," *Cogent Educ.*, vol. 12, no. 1, Dec. 2025, doi: 10.1080/2331186X.2025.2536531.
- [2] Bokolo Anthon Jnr, "Examining Blended Learning Adoption Towards Improving Learning Performance in Institutions of Higher Education," *Technol. Knowl. Learn.* 2024 293, vol. 29, no. 3, pp. 1401–1435, Jan. 2024, doi: 10.1007/S10758-023-09712-3.
- [3] E. Aboagye, J. A. Yawson, and K. N. Appiah, "COVID-19 and E-Learning: The Challenges of Students in Tertiary Institutions," *Soc. Educ. Res.*, pp. 1–8, Jun. 2021, doi: 10.37256/SER.212021422.
- [4] D. Norman, *Living with Complexity*. Cambridge: The MIT Press, 2016.
- [5] *ISO 9241-210:2019 Ergonomics of human-system interaction Part 210: Human-centred design for interactive systems*, 2nd ed. ISO (International Organization for Standardization), 2019.
- [6] B. A. Zardari, Z. Hussain, A. A. Arain, W. H. Rizvi, and M. S. Vighio, "Development and Validation of User Experience-Based E-Learning Acceptance Model for Sustainable Higher Education," *Sustain.* 2021, Vol. 13, Page 6201, vol. 13, no. 11, p. 6201, May 2021, doi: 10.3390/SU13116201.
- [7] J. Kollmorgen, A. Hinderks, and J. Thomaschewski, "Selecting the Appropriate User Experience Questionnaire and Guidance for Interpretation: the UEQ Family.," *Int. J. Interact. Multimed. Artif. Intell.*, vol. 9, no. 4, pp. 126–139, Aug. 2025, doi: 10.9781/IJIMAI.2024.08.005.

- [8] M. Schrepp *et al.*, “On the Importance of UX Quality Aspects for Different Product Categories,” *Int. J. Interact. Multimed. Artif. Intell.*, vol. 8, no. 2, pp. 232–246, 2023, doi: 10.9781/IJIMAI.2023.03.001.
- [9] M. Schrepp, A. Hinderks, and J. Thomaschewski, “Construction of a Benchmark for the User Experience Questionnaire (UEQ),” *Int. J. Interact. Multimed. Artif. Intell.*, vol. 4, no. 4, pp. 40–44, Jun. 2021, doi: 10.9781/IJIMAI.2017.445.
- [10] M. Schrepp, “User Experience Questionnaire Handbook Version 10th,” 2023, *hlm*.
- [11] Georgios Melissourgos and Christos Kastanos, “UEQ-GR and UEQ-S-GR: Towards a Greek Adaptation of the User Experience Questionnaire and its Short Version,” *CHIGreece*, Sep. 2025, doi: 10.1145/3749012.3749059.
- [12] I. Dewi, H. Khotimah, R. Nasution, T. Hariyati, and A. Rusli, “TEORI POPULASI DAN PENGAMBILAN SAMPEL,” Jun. 2023, doi: 10.47783/literasiologi.v9i4.
- [13] B. Laugwitz, T. Held, and M. Schrepp, “Construction and evaluation of a user experience questionnaire,” in *Symposium of the Austrian HCI and usability engineering group*, Springer, 2020, pp. 63–76.
- [14] H. B. Santoso, M. Schrepp, L. M. Hasani, R. Fitriansyah, and A. Setyanto, “The use of User Experience Questionnaire Plus (UEQ+) for cross-cultural UX research: evaluating Zoom and Learn Quran Tajwid as online learning tools,” *Heliyon*, vol. 8, no. 11, p. e11748, Nov. 2022, doi: 10.1016/J.HELIYON.2022.E11748.
- [15] L. I. Burhan, “Efektivitas Blended Learning dalam Meningkatkan Motivasi Belajar Mahasiswa Pendidikan di Masa Pascapandemi,” *CENDEKIA J. Pendidik. Terintegrasi*, vol. 1, no. 1, pp. 25–37, Jul. 2025, doi: 10.63982/P8PSPZ63.