Jurnal Penelitian dan Pengkajian Ilmu Pendidikan: e-Saintika
https://journal-center.litpam.com/index.php/e-Saintika/index

July 2026 Vol. 10, No. 2
e-ISSN: 2615-6881
pp. 700-723

Comparing Perceived Usability Across Online Coding Learning Platforms Using the System Usability Scale

[1] Muhammad Isfa' Lana, *[1] Nuur Wachid Abdul Majid, [2] Mohamed Nor Azhari Azman

[1] Information Systems & Technology Education, Universitas Pendidikan Indonesia, Indonesia
[2] Faculty of Technical and Vocational, Universiti Pendidikan Sultan Idris, Malaysia

*Corresponding Author e-mail: nuurwachid@upi.edu

Received: January 2026; Revised: June 2026; Published: July 2026

Abstract

This study evaluated and compared the perceived usability of three widely used online coding learning platforms—W3Schools, Coursera, and Udemy—among university students in Purwakarta, Indonesia. A quantitative, non-randomized, cross-sectional between-subjects design was employed using the System Usability Scale (SUS), with each participant evaluating one platform that they had previously used. The final sample comprised 120 students, with 40 participants in each platform group. Descriptive statistics, reliability analysis, normality testing, the Kruskal–Wallis test, epsilon-squared effect size, and supplementary Dunn pairwise comparisons with Bonferroni adjustment were applied. W3Schools obtained the highest mean SUS score (85.4) and was classified as "Best Imaginable," whereas Udemy (82.4) and Coursera (80.9) were both classified as "Excellent." Despite these descriptive differences, the Kruskal–Wallis test showed no statistically significant overall difference in perceived usability among the three platform-user groups ($H = 2.472$, $p = .291$). The corresponding epsilon-squared effect size was small ($\varepsilon^2 = .021$), and none of the Bonferroni-adjusted pairwise comparisons reached statistical significance. These findings indicate that all three platforms were perceived as highly usable by their respective user groups, while the observed numerical differences did not establish statistically significant platform superiority. The study contributes standardized comparative evidence on coding-platform usability within an Indonesian university context. Future research should incorporate larger and more diverse samples, explicitly measure prior platform experience, and combine subjective usability ratings with objective task-based measures and qualitative evidence.

Keywords: Comparative Analysis; e-Learning; Online Coding Learning Platforms; Perceived Usability; System Usability Scale

How to Cite: Lana, M. I., Majid, N. W. A., & Azman, M. N. A. (2026). Comparing Perceived Usability Across Online Coding Learning Platforms Using the System Usability Scale. *Jurnal Penelitian Dan Pengkajian Ilmu Pendidikan: E-Saintika, 10*(2), 700-723. https://doi.org/10.36312/e-saintika.v10i2.4432

10.36312/e-saintika.v10i2.4432

INTRODUCTION

The rapid development of digital technology has substantially expanded the role of online environments in programming education and higher education more broadly (Liu et al., 2020). Web-based tutorials, video-based courses, and self-paced learning environments allow students to access learning materials flexibly and repeatedly according to their individual needs (Bond et al., 2021; Martin & Bolliger, 2018). For students developing programming competencies, such environments provide supplementary opportunities to learn, practice, and revisit technical concepts beyond formal classroom instruction (Lin et al., 2022; Shi et al., 2021). Research on digital learning has further demonstrated the relevance of student engagement, independent learning, learning-system quality, and appropriately designed programming activities, including gamified programming environments (Paiva et al., 2022; Chuenyindee et al., 2022; Tao et al., 2025; Bond et al., 2021; Al-Fraihat et al., 2020). Consequently, evaluating how students experience these digital environments has become an important component of educational-technology research.

Usability constitutes one important dimension of this evaluation. Previous studies have examined usability at different levels, including specific course content in virtual learning environments (Obonya & Kadlečík, 2020), e-learning systems and learning management systems (Alqurni, 2023; Kremser et al., 2021; Setiawan et al., 2020; Sharip et al., 2023), academic information services (Anrahvi et al., 2024), and digital learning materials (Mertasari & Candiasa, 2022). The technological contexts being evaluated are also becoming increasingly diverse, encompassing educational chatbots and artificial-intelligence-supported learning environments (Ait Baha et al., 2024; Zawacki-Richter et al., 2019), as well as virtual and immersive environments (Alnuaimi & Awad, 2025; Bolpagni & Gabrielli, 2025). Recent methodological work has also emphasized the need for more rigorous and transparent usability-evaluation practices in learning-environment design (Prajapati & Das, 2025). This breadth of research demonstrates that usability cannot be assumed from technological functionality alone; it must be examined in relation to particular users and contexts of use.

The relevance of programming-related digital learning is particularly timely in Indonesia. Regulation of the Minister of Primary and Secondary Education Number 13 of 2025, which amended Regulation Number 12 of 2024 concerning the national curriculum, introduced Coding and Artificial Intelligence as elective subjects to be implemented gradually at primary and secondary education levels beginning in the 2025–2026 academic year (Kementerian Pendidikan Dasar dan Menengah, 2025). This policy reflects a broader national emphasis on digital competencies, while research involving younger learners also demonstrates the increasing presence of coding environments within educational practice (Dilmen et al., 2023). Although the regulation directly concerns primary and secondary education rather than universities, this broader educational emphasis reinforces the continuing relevance of programming competence across educational levels. For university students, online coding learning platforms consequently represent one accessible means of supporting continuous and self-directed development of programming knowledge and skills.

As students gain access to a growing range of platforms, perceived usability becomes particularly relevant because technically similar learning goals can be supported through substantially different interfaces and learning structures. Usability has been associated with positive interaction experiences and continued engagement with digital systems (van Doorn et al., 2023; Martins et al., 2015), while educational-technology research emphasizes relationships among usability, system quality, information quality, satisfaction, and continued technology use (Al-Fraihat et al., 2020; Almaiah et al., 2020; Chen et al., 2022). Similar concerns arise in programming-learning environments, where students' and lecturers' perceptions of interface usability can reveal barriers that are not evident from system functionality alone (Ramos et al., 2021). User-centered interface organization and clear interaction structures are therefore relevant considerations in educational systems (Nugroho et al., 2022; Martin & Bolliger, 2018). Importantly, however, perceived usability should be distinguished from instructional quality, learning effectiveness, or educational achievement. The present study evaluates how usable the platforms are perceived to be; it does not test whether one platform teaches programming more effectively than another.

A substantial body of research has evaluated usability in educational technology, but the evidence remains heterogeneous in terms of systems, users, and evaluation approaches. Systematic evidence indicates that educational-technology usability research has often focused on particular systems or implementation contexts rather than directly comparable sets of alternative platforms (Lu et al., 2022). Other studies have applied broader platform-selection frameworks, including multi-attribute group decision-making methods (Su et al., 2022) and hybrid multi-criteria decision-making approaches incorporating pedagogical criteria (Al-Gerafi et al., 2024). Comparative usability research also exists in other digital domains, including wearable technologies (Alshamari & Althobaiti, 2024) and mobile banking applications (Setiyawati & Bangkalang, 2022), illustrating the usefulness of comparative designs while also demonstrating that conclusions are highly dependent on the technologies and users being studied. These studies provide methodological precedents for comparison but should not be treated as direct evidence regarding the usability of coding-learning platforms.

Within programming education itself, comparative evidence is more limited but not absent. Ratnasari and Wibawa (2020), for example, compared the UI/UX quality of Codecademy, W3Schools, and Udemy as online coding-course platforms using an A/B-testing approach in an Indonesian vocational-education context. This prior work is particularly relevant because it demonstrates that differences among coding platforms can be examined comparatively. However, it did not provide a standardized SUS-based comparison of W3Schools, Coursera, and Udemy among university students. More broadly, programming-education research has documented the importance of learning environments and tools for supporting introductory programming (Luxton-Reilly et al., 2018), while studies comparing different modes of educational delivery demonstrate the broader value of evaluating user experiences across learning environments rather than examining a single setting in isolation (Veeraiyan et al., 2022). Accordingly, a more specific gap remains concerning whether university students who use different online coding learning platforms report distinguishable levels of perceived usability when evaluated through the same standardized instrument.

The present study addresses this gap by examining three platforms that support coding-related self-directed learning but differ in their dominant organization and content-delivery models. W3Schools emphasizes tutorial-oriented resources, examples, and interactive coding practice, whereas Coursera represents a more structured online-course model associated with courses and programs developed with universities and other institutional partners; Udemy operates through an online educational marketplace in which independently developed courses are offered across a broad range of subjects. The characterization of these platforms is intended to describe their broad organizational differences rather than to treat instructional architecture as an experimentally manipulated variable. Research on MOOCs has documented structured online-course models and differences in course organization (Hew & Cheung, 2014; Kovanović et al., 2018), while Udemy has been examined explicitly as an online educational marketplace (Mozahem, 2021). An Indonesian comparison has also previously identified W3Schools and Udemy as distinct online coding-course environments (Ratnasari & Wibawa, 2020). Despite these differences, all three platforms can provide resources for independent development of programming-related knowledge and skills (Luxton-Reilly et al., 2018).

This distinction is important for the interpretation of the present study. Because participants were not randomly assigned to platforms and instructional characteristics were not manipulated experimentally, the study does not test whether platform architecture causes differences in usability. Instead, it compares perceived usability across naturally occurring groups of users who had prior experience with the respective platforms. Such contextualization is particularly important because usability and user-experience assessments can vary across populations and settings, as illustrated in studies conducted in different digital and educational contexts (de Dios López et al., 2024; Binyamin et al., 2019). Thus, the scientific contribution lies not in claiming an isolated causal effect of instructional architecture but in providing a standardized cross-platform comparison within a university-student population that remains relatively underrepresented in this specific research context.

The System Usability Scale (SUS) was selected to provide this standardized comparison. Originally introduced as a concise global measure of perceived usability (Brooke, 1996), SUS has subsequently become one of the most widely used standardized subjective usability instruments (Lewis, 2018; Sauro, 2016). Its applicability in educational technology is supported by systematic evidence (Vlachogianni & Tselios, 2022) and by its use in online-learning evaluations, including large-scale shifts toward digital learning during the COVID-19 period (Pal & Vanijja, 2020). The wider literature also demonstrates the adaptability of standardized usability assessment across diverse settings, including immersive systems (Tieri et al., 2025), health-related applications and instruments (de Souza et al., 2024; Maramba et al., 2019), and public-service applications (Dyayu et al., 2023). SUS uses 10 items rated on a five-point response scale to produce a global perceived-usability score (Yulius et al., 2022). It should therefore be understood in this study as a standardized subjective measure of perceived usability, rather than as a direct measurement of objective task effectiveness, efficiency, instructional quality, or learning outcomes. Related systematic work employing other standardized usability instruments, such as PSSUQ/CSUQ, further demonstrates the continuing importance of perceived-usability evaluation in educational technology (Vlachogianni & Tselios, 2023).

Recent research has also increasingly emphasized the importance of evaluating usability and user experience in relation to particular populations and actual-use contexts. Studies addressing accessibility for users with visual impairments (Adnan et al., 2025) and the development of user-experience measures for actual system use in online courses (Wang et al., 2025) illustrate this movement toward more context-sensitive evaluation. Against this background, the present study contributes a SUS-based comparison of W3Schools, Coursera, and Udemy among university students in Purwakarta, Indonesia. Its contribution is defined by the combination of a common standardized usability measure, three differently organized coding-learning platforms, naturally occurring user groups, and a specific Indonesian higher-education context—not by an assumption that one platform architecture is intrinsically superior to another.

Accordingly, the primary objective of this study was to evaluate and compare the perceived usability of W3Schools, Coursera, and Udemy among university students in Purwakarta, Indonesia. A quantitative, non-randomized between-subjects design was used in which each participant evaluated one platform with which they had prior experience. The analysis first characterized platform-level SUS scores and classifications and then examined whether the observed differences were statistically supported. Based on this objective, the study addressed the following research questions:

1. What are the SUS-based perceived usability levels and classifications of W3Schools, Coursera, and Udemy among the sampled university students?
2. Is there a statistically significant overall difference in SUS scores among users of W3Schools, Coursera, and Udemy?
3. Do supplementary pairwise comparisons identify statistically significant differences in SUS scores between individual platform groups after adjustment for multiple comparisons?

For the overall inferential comparison, the null hypothesis (H_0) stated that the distribution of SUS scores does not differ among the three platform-user groups, whereas the alternative hypothesis (H_1) stated that at least one platform-user group differs.

METHOD

Research Design and Platform Selection

This study employed a quantitative approach using a non-randomized, cross-sectional comparative survey design to examine differences in perceived usability across online coding learning platforms. A between-subjects design was applied, in which each participant evaluated only one platform that they had previously used. Participants were therefore grouped according to their prior platform use rather than randomly assigned to experimental conditions. Consequently, the study was designed to compare perceived usability across naturally occurring user groups and does not support causal inference regarding the effects of specific platform architectures on usability.

Before the main study, a preliminary survey involving 82 students was conducted to identify the coding learning platforms most frequently used in the Purwakarta context. As summarized in Table 1, W3Schools was reported by 51 students (62.2%), Coursera by 20 students (24.4%), and Udemy by 8 students (9.8%), while 3 students (3.6%) reported using other platforms. Because usage of the remaining platforms was sparse and distributed across several alternatives, the main study focused on W3Schools, Coursera, and Udemy. This selection ensured that the comparative analysis was restricted to platforms with an identifiable user base in the preliminary survey rather than attempting to compare platforms represented by only a few respondents.

The overall research sequence—from the preliminary survey and platform selection through participant screening, group balancing, SUS administration, and statistical analysis—is illustrated in Figure 1.

Table 1. Preliminary survey results on coding learning platforms used by students

Platform	Number of Users (n)	Usage Percentage (%)
W3Schools	51	62.2
Coursera	20	24.4
Udemy	8	9.8
Other	3	3.6
Total	82	100.0

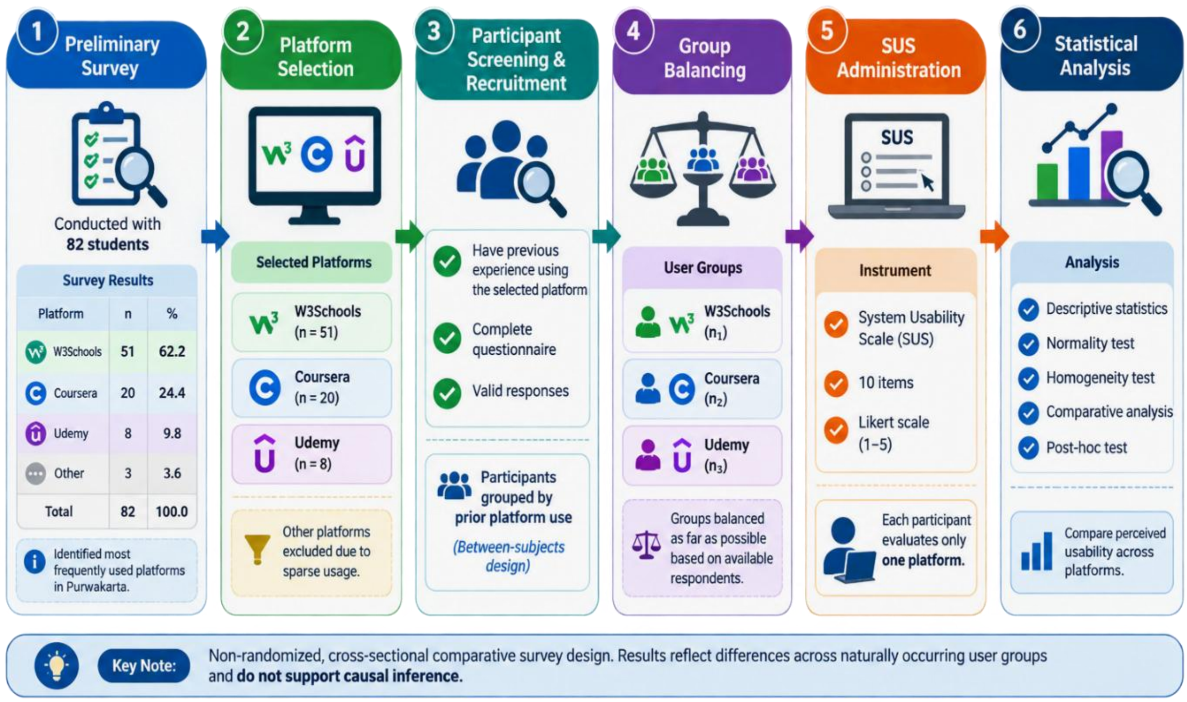

Figure 1. Research Flow Diagram

Participants and Sampling

The main study included 120 university students in Purwakarta who had prior experience using one of the three selected platforms. A purposive sampling technique was employed to ensure that participants had sufficient prior exposure to the platform they evaluated. Eligibility required prior use of the respective platform for coding learning, operationalized as having actively followed or completed at least one course or tutorial module on that platform.

To facilitate balanced comparisons, respondents were intentionally recruited until each platform group contained 40 eligible participants, despite the unequal usage distribution observed in the preliminary survey. Because Udemy had the lowest preliminary usage prevalence (9.8%), additional targeted screening was conducted through student academic groups and technical communities associated with Information Systems and Technology education programs. This recruitment procedure identified eligible Udemy users with genuine prior experience rather than allocating participants to a platform they had not used.

The final sample therefore comprised 40 W3Schools users, 40 Coursera users, and 40 Udemy users. The equal group sizes facilitated direct statistical comparison among the three user groups; however, this intentional balancing does not reproduce the natural platform-use distribution identified in the preliminary survey and may therefore limit population-level representativeness.

Participant characteristics are summarized in Table 2. Participants were 18–24 years old and represented several technical study programs. Most participants were in Semester 7 (64.1%). HTML was the most frequently reported prior coding experience, followed by Java and Python. All 120 participants were classified as active/regular users and had completed at least one relevant course or tutorial module, indicating that the SUS ratings were obtained from participants with substantive prior exposure to the platform being evaluated rather than from first-time users.

Table 2. Participants' Characteristics

Characteristic	W3Schools (n=40)	Coursera (n=40)	Udemy (n=40)	Total (N=120)
Gender				
Male	26	27	22	75
Female	14	13	18	45
Age range	18–24 years	18–24 years	18–24 years	18–24 years
Study Program				
Pendidikan Sistem dan Teknologi Informasi	16	27	23	66
Sistem Telekomunikasi	4	4	5	13
Mekatronika dan Kecerdasan Buatan	7	3	3	13
Teknik Informatika	6	2	4	12
Elektro	2	0	0	2
Teknologi Rekayasa Mekatronika	5	4	5	14
Semester				
Semester 3	4 (10.0%)	6 (15.0%)	10 (25.0%)	20 (16.7%)
Semester 5	7 (17.5%)	9 (22.5%)	7 (17.5%)	23 (19.2%)
Semester 7	29 (72.5%)	25 (62.5%)	23 (57.5%)	77 (64.1%)
Prior Coding Experience				
HTML	32	30	40	102
Python	18	20	7	45
Java	20	23	20	63
Usage Frequency				
Active/Regular	40	40	40	120
Daily Usage Duration				
< 3 hours/day	32	12	13	57
> 3 hours/day	8	28	27	63
Completed Course/Tutorial Module				
Yes	40	40	40	120

Instrument and SUS Scoring

The instrument used in this study was the System Usability Scale (SUS), a standardized questionnaire designed to provide a global measure of perceived system usability (Brooke, 1996). The SUS consists of 10 items with alternating positively and negatively worded statements, rated on a five-point Likert scale ranging from "strongly disagree" to "strongly agree."

In this study, the SUS was administered in Indonesian to facilitate comprehension among respondents. Rather than conducting a new translation, the study used the previously validated Indonesian adaptation developed by Sharfina and Santoso (2016). Using this established adaptation maintained consistency with a version for which linguistic and measurement properties had already been evaluated in the Indonesian context.

SUS scores were calculated according to the standard scoring procedure (Brooke, 1996). For odd-numbered items, the score contribution was calculated by subtracting 1 from the participant's response; for even-numbered items, the response value was subtracted from 5. The resulting item contributions were summed and multiplied by

2.5, producing a composite SUS score ranging from 0 to 100. The SUS score was treated as a global indicator of perceived usability rather than as a direct task-based measure of objective effectiveness or efficiency.

Reliability Assessment

Before internal consistency was evaluated, all negatively worded SUS items were reverse-coded so that higher coded values consistently represented more favorable usability perceptions. Cronbach's alpha was then calculated separately for each platform group. A coefficient above 0.70 was considered an acceptable level of internal consistency (Lewis & Sauro, 2021).

As reported in Table 3, Cronbach's alpha was 0.767 for W3Schools, 0.857 for Coursera, and 0.714 for Udemy. All three values exceeded the adopted 0.70 criterion, indicating acceptable internal consistency of the 10-item SUS within each platform group.

Table 3. Cronbach's Alpha Reliability Results

Platform	Cronbach's Alpha	N of Items
W3Schools	0.767	10
Coursera	0.857	10
Udemy	0.714	10

Data Collection Procedure and Response Quality Control

Data were collected through a confidential online questionnaire administered using Google Forms and distributed through university academic communication channels and student technical communities. At the beginning of the questionnaire, participants were presented with an informed-consent statement and were required to provide explicit electronic consent before proceeding.

Respondents first completed a screening stage to confirm prior experience with one of the selected coding learning platforms. Only respondents who had actively used or completed a course or tutorial module on W3Schools, Coursera, or Udemy were eligible for inclusion. Each eligible respondent evaluated only one platform on the basis of prior experience and completed the SUS with reference to that platform. Because these assessments relied on participants' recollection of previous platform interactions rather than a standardized usability task conducted immediately before SUS administration, the evaluation was retrospective and remained potentially susceptible to recall bias.

Data collection was conducted over one month in December 2025. Google Forms' "limit to one response" function was activated, requiring participants to sign in using an institutional account to reduce duplicate submissions. Because institutional sign-in was required, the study is more accurately described as confidential rather than fully anonymous; identifying information was not reported in the research outputs.

A total of 137 responses were received. Before statistical analysis, responses were screened for completeness and response-quality problems. Patterns considered indicative of insufficient or inconsistent responding included straight-lining—providing the same response category across all 10 SUS items—and internally contradictory patterns across the alternating positively and negatively worded SUS statements. Such response-pattern screening is consistent with established recommendations for identifying potentially careless responses in survey data (Meade

& Craig, 2012). Seventeen responses were excluded based on these criteria, leaving 120 valid responses for the final analysis.

Statistical Analysis

Statistical analyses were conducted using IBM SPSS Statistics version 23, with the significance level set at $\alpha = .05$. The analysis proceeded sequentially. First, descriptive statistics were calculated to summarize SUS scores for each platform. Second, Cronbach's alpha was calculated to assess internal consistency. Third, the distribution of SUS scores within each platform group was evaluated using the Kolmogorov–Smirnov and Shapiro–Wilk tests.

Because at least one platform group deviated from normality, the primary between-group comparison was conducted using the rank-based Kruskal–Wallis test (Kruskal & Wallis, 1952) rather than a parametric one-way analysis of variance. The test examined whether the distributions of SUS scores differed across the three independent platform-user groups.

Statistical significance was supplemented with an effect-size estimate to distinguish the magnitude of between-group variation from the p-value alone. Epsilon-squared was calculated as $\varepsilon^2 = H/(N - 1)$, where H represents the Kruskal–Wallis test statistic and N represents the total number of observations. This expression is algebraically equivalent to the epsilon-squared formulation discussed by Tomczak and Tomczak (2014). For descriptive interpretation, 0.01, 0.06, and 0.14 were used as conventional reference points for small, medium, and large effects, respectively (Cohen, 1988), while the effect-size estimate itself was interpreted in conjunction with the inferential result rather than as a separate test of significance.

For completeness and to characterize specific pairwise patterns among the three platforms, exploratory pairwise comparisons were conducted using Dunn's rank-sum procedure (Dunn, 1964). Bonferroni-adjusted p-values were used to control the familywise Type I error rate across the multiple comparisons. Because the omnibus Kruskal–Wallis test constituted the primary inferential test, the pairwise results were interpreted as supplementary comparisons and not as evidence overriding the omnibus result.

Ethical Considerations

This study was conducted in accordance with established principles of voluntary participation, informed consent, confidentiality, and protection of participant privacy. Explicit electronic informed consent was obtained from all participants before questionnaire completion, and participation was voluntary.

All collected data were handled confidentially, and respondent identities were not included in the research report. Formal ethics approval was waived by the institutional review board of Universitas Pendidikan Indonesia because the study involved a non-invasive survey procedure and was considered minimal risk. The institutional-account sign-in used to restrict duplicate questionnaire submissions should therefore not be described as anonymous data collection; confidentiality and de-identification in reporting more accurately characterize the procedure.

RESULTS AND DISCUSSION

Descriptive SUS Scores and Usability Classification

The descriptive analysis showed high SUS scores across all three online coding learning platforms. To interpret the platform-level scores consistently, the SUS grade, percentile, adjective, acceptability, and Net Promoter Score (NPS) classification framework adopted in this study is presented in Table 4 (Lana & Majid, 2024).

Table 4. Interpretation of SUS Score

Grade	SUS	Percentile Range	Adjective	Acceptability	NPS
A+	84.1–100	96–100	Best Imaginable	Acceptable	Promoter
A	80.8–84.0	90–95	Excellent	Acceptable	Promoter
A−	78.9–80.7	85–89	Good	Acceptable	Promoter
B+	77.2–78.8	80–84		Acceptable	Passive
B	74.1–77.1	70–79		Acceptable	Passive
B−	72.6–74.0	65–69		Acceptable	Passive
C+	71.1–72.5	60–64		Acceptable	Passive
C	65.0–71.0	41–59	OK	Marginal	Passive
C−	62.7–64.9	35–40		Marginal	Passive
D	51.7–62.6	15–34		Marginal	Detractor

The descriptive SUS results for each platform are summarized in Table 5. W3Schools obtained the highest mean SUS score at 85.4, based on a total score of 3415, with scores ranging from 62.5 to 100 and a mode of 87.5. According to the classification shown in Table 4, this mean corresponded to grade A+ and the "Best Imaginable" category. Udemy obtained a mean SUS score of 82.4 from a total score of 3297.5, with scores ranging from 52.5 to 100 and a mode of 75.0, corresponding to grade A and the "Excellent" category. Coursera obtained a mean score of 80.9 from a total SUS score of 3237.5, with scores ranging from 47.5 to 100 and a mode of 85.0; this score was also classified as grade A and "Excellent."

Table 5. SUS Score Results

Platform	N	Total Score	Mean	Min	Max	Mode	Grade	Category
W3Schools	40	3415.0	85.4	62.5	100	87.5	A+	Best Imaginable
Coursera	40	3237.5	80.9	47.5	100	85.0	A	Excellent
Udemy	40	3297.5	82.4	52.5	100	75.0	A	Excellent

Descriptively, W3Schools had the highest mean SUS score, followed by Udemy and Coursera. However, these numerical differences represent descriptive patterns only; their statistical significance was examined separately using the between-group analyses reported further.

Assessment of Score Distribution

Before conducting the between-platform comparison, the distribution of SUS scores within each platform group was evaluated using the Kolmogorov–Smirnov and Shapiro–Wilk tests. The complete normality-test results are presented in Table 6. Because each platform group contained 40 participants, the Shapiro–Wilk results were used as the primary basis for evaluating normality.

As shown in Table 6, the Shapiro–Wilk test did not indicate statistically significant departures from normality for W3Schools, $W = .958$, $p = .143$, or Udemy, $W = .956$, $p = .126$. In contrast, Coursera showed a statistically significant departure from normality, $W = .939$, $p = .033$. Therefore, the assumption of normality was not satisfied across all three groups, supporting the use of the nonparametric between-group analysis specified in the Method section.

Table 6. Normality Test Results

Measure	Platform	K–S Statistic	df	Sig.	S–W Statistic	df	Sig.
Score	W3Schools	.139	40	.050	.958	40	.143
	Coursera	.099	40	.200*	.939	40	.033
	Udemy	.107	40	.200*	.956	40	.126

** This is a lower bound of the true significance. a. Lilliefors Significance Correction.*

Overall Comparison of Perceived Usability

Differences in SUS scores among the three independent platform groups were examined using the Kruskal–Wallis test. The omnibus test statistics are reported in Table 7, while the corresponding group mean ranks are presented in Table 8.

As shown in Table 7, the Kruskal–Wallis test produced a test statistic of $H = 2.472$, with $df = 2$ and $p = .291$. Because the p-value exceeded the predefined significance level of .05, the analysis did not detect a statistically significant overall difference in SUS-based perceived usability among W3Schools, Coursera, and Udemy.

Table 7. Kruskal–Wallis Test Statistics Results

Test Statistic	Score
Chi-Square (H)	2.472
df	2
Asymp. Sig.	.291

a. Kruskal–Wallis Test. b. Grouping Variable: Platform.

The rank distribution underlying the Kruskal–Wallis analysis is shown in Table 8. W3Schools had the highest mean rank (67.50), followed by Udemy (57.63) and Coursera (56.38). Although the mean ranks differed numerically, the omnibus result reported in Table 7 showed that these differences were not statistically significant.

Table 8. Kruskal–Wallis Test Ranks Results

Platform	N	Mean Rank
W3Schools	40	67.50
Coursera	40	56.38
Udemy	40	57.63
Total	120	

Effect Size

To quantify the magnitude of the observed between-group variation, epsilon-squared (ε^2) was calculated from the Kruskal–Wallis statistic. The resulting effect-size estimate was $\varepsilon^2 = 0.021$. Based on the interpretive criteria specified in the Method section, this value falls within the small-effect range. The effect-size estimate therefore indicates that the magnitude of between-group differences in this sample was small.

Pairwise Comparisons

Supplementary pairwise comparisons were conducted using Dunn's test with Bonferroni-adjusted p-values. The complete pairwise results are presented in Table 9.

As shown in Table 9, none of the three pairwise comparisons reached statistical significance after Bonferroni adjustment. The adjusted p-value was .495 for W3Schools versus Coursera, .654 for W3Schools versus Udemy, and 1.000 for Coursera versus Udemy. Accordingly, no statistically significant pairwise differences in SUS scores were detected between any of the platform groups.

Table 9. Pairwise Comparisons Using Dunn's Post-Hoc Test

Comparison Pair	Mean Rank Difference	p-value	Adjusted p-value*	Decision
W3Schools vs Coursera	11.12	.165	.495	Not significant
W3Schools vs Udemy	9.87	.218	.654	Not significant
Coursera vs Udemy	−1.25	.875	1.000	Not significant

** Bonferroni-adjusted p-value.*

Summary of Statistical Findings

The descriptive results showed that W3Schools obtained the highest mean SUS score and mean rank, followed descriptively by Udemy and Coursera. Nevertheless, the Kruskal–Wallis test did not detect a statistically significant overall difference among the three groups ($H = 2.472$, $p = .291$). The estimated effect size was small ($\varepsilon^2 = 0.021$), and the supplementary Dunn–Bonferroni comparisons likewise identified no statistically significant pairwise differences. Thus, the observed numerical differences in SUS scores were descriptive and were not supported by statistically significant between-platform differences in this sample.

Discussion

Overall Perceived Usability of the Coding Learning Platforms

The findings indicate consistently favorable perceived usability across the three online coding learning platforms. As reported in Table 5, W3Schools achieved a mean SUS score of 85.4 and was classified as "Best Imaginable," whereas Udemy (82.4) and Coursera (80.9) were both classified as "Excellent." These scores indicate that participants generally perceived all three platforms as usable within their respective coding-learning experiences. Importantly, these results should be interpreted as measures of perceived usability rather than direct evidence of instructional effectiveness, programming-skill improvement, or learning achievement, because those outcomes were not measured in the present study.

The generally favorable scores are consistent with the broader educational-technology literature showing that usability constitutes an important dimension of students' interactions with digital learning environments (Lin et al., 2022; Chuenyindee et al., 2022; Alqurni, 2023). A systematic review of SUS applications in educational technology likewise found that the instrument has been extensively used to evaluate a wide range of learning technologies and that perceived usability varies according to the type of system, population, and educational context (Vlachogianni & Tselios, 2022). Similarly, Lu et al. (2022) demonstrated that usability research in educational technology encompasses diverse technologies, methods, and outcome measures, with e-learning applications in higher education representing a major area of investigation.

The present findings also extend this line of research into the more specific context of programming education. Ramos et al. (2021), for example, demonstrated that usability problems in a virtual programming laboratory can affect how students interact with programming-learning tools, emphasizing the importance of evaluating the interface rather than assuming that a technically functional learning environment is necessarily easy to use. Thus, the high SUS scores observed in the present study provide useful evidence that the three platforms were perceived positively by their respective user groups, while remaining distinct from evidence about their pedagogical effectiveness.

Between-Platform Differences and Their Practical Meaning

Although W3Schools, Udemy, and Coursera differed numerically in their SUS scores, the inferential results do not provide evidence that the three user groups differed significantly in perceived usability. As shown in Table 7, the Kruskal–Wallis test yielded $H = 2.472$ and $p = .291$. The corresponding mean ranks presented in Table 8 were 67.50 for W3Schools, 57.63 for Udemy, and 56.38 for Coursera. Therefore, W3Schools occupied the highest descriptive position, but the observed rank differences were insufficient to establish a statistically significant overall difference among the groups.

This distinction between descriptive ranking and inferential evidence is important. A higher sample mean does not by itself establish that one platform has superior usability at the population level. The small epsilon-squared effect size ($\varepsilon^2 = .021$) further indicates that the magnitude of the observed between-group variation was limited in this sample. Consistently, the supplementary Dunn–Bonferroni comparisons reported in Table 9 identified no statistically significant difference for W3Schools versus Coursera, W3Schools versus Udemy, or Coursera versus Udemy. The appropriate interpretation is therefore that the study did not detect statistically significant differences among the three platform groups—not that the platforms are proven to have identical usability.

Comparative usability studies in other digital-system domains reinforce the value of combining descriptive ratings with broader evidence rather than interpreting numerical rankings alone. Alshamari and Althobaiti (2024), for example, combined SUS with customized heuristic evaluation in their assessment of wearable systems, while Setiyawati and Bangkalang (2022) combined SUS with the User Experience Questionnaire when comparing mobile banking applications. Although these studies concern different technological domains and should not be treated as direct evidence about coding platforms, they illustrate why comparative usability assessment benefits from considering multiple aspects of user interaction rather than relying exclusively on differences in aggregate SUS scores.

A similar point emerges within educational technology. Sharip et al. (2023) compared Google Classroom, Microsoft Teams, and Padlet and showed that comparative platform evaluation can reveal differences that extend beyond a single overall usability score. This distinction is particularly relevant to the present findings: perceived usability represents one dimension of platform selection, whereas students may also consider course availability, content organization, instructional style, certification, cost, or established learning habits. Consequently, the absence of a statistically significant SUS difference should not be interpreted as evidence that the three platforms are interchangeable in all educational respects.

Descriptive Advantage of W3Schools and the Role of Prior Experience

The higher descriptive score obtained by W3Schools requires particularly cautious interpretation. Table 5 shows that W3Schools had the highest mean SUS score (85.4), and Table 8 shows that it also had the highest mean rank (67.50). However, neither the omnibus nor the pairwise statistical analyses established a significant superiority of W3Schools. Therefore, the finding supports a descriptive pattern rather than a statistically demonstrated platform advantage.

The preliminary survey provides relevant contextual information because W3Schools was substantially more prevalent among the initially surveyed students, accounting for 62.2% of reported platform use. This prevalence suggests that W3Schools may have been more familiar within the local student population. Nevertheless, the preliminary usage proportion is a group-level characteristic and cannot establish that participants in the W3Schools main-study group were individually more experienced than participants evaluating Coursera or Udemy. Table 2 further shows that all participants were classified as active or regular users and had completed at least one course or tutorial module, while daily usage duration varied across platform groups. Because individual familiarity was not modeled as a predictor of SUS scores, the present study cannot determine whether familiarity explains W3Schools' higher descriptive score.

There is, however, a methodological basis for treating previous experience as a plausible factor requiring control in future studies. McLellan et al. (2012) found that users with more extensive prior experience could provide more favorable SUS ratings than users with limited or no experience. Kortum and Johnson (2013) subsequently examined the relationship between experimentally controlled experience and perceived usability and found that the effect of increasing experience on SUS ratings may occur but is not necessarily uniform across systems. These findings strengthen the rationale for considering prior platform experience when interpreting subjective usability ratings, while also showing why familiarity should not be assumed to explain the current results without direct statistical testing.

The original interpretation linking the W3Schools result directly to familiarity, navigation efficiency, or reduced cognitive load should therefore be avoided. User-centered research emphasizes the importance of understanding users' previous experiences and usage contexts (van Doorn et al., 2023), but the present study did not measure cognitive load or experimentally manipulate familiarity. In this manuscript, prior experience is therefore treated as a plausible contextual factor rather than a demonstrated causal mechanism.

Platform Characteristics and the Limits of Whole-System Comparison

The three platforms differ substantially in how coding-learning experiences are organized. W3Schools primarily offers tutorial-oriented and hands-on materials, Coursera generally presents structured courses, and Udemy operates through a marketplace-based course model. These distinctions provide the conceptual basis for comparing the platforms, but they also constrain interpretation because the present design did not experimentally isolate instructional architecture, interface structure, or individual design features.

Previous research demonstrates that interface organization can influence usability when design variables are directly manipulated. Shi et al. (2021), for example, experimentally compared different navigation structures in an online learning environment and found differences in usability, task efficiency, and cognitive load according to interface organization. Research on MOOCs likewise shows that course organization and learning structure are meaningful features of online learning environments (Kovanović et al., 2018; Hew & Cheung, 2014). Such evidence demonstrates that structural characteristics can matter, but it also highlights an important distinction from the present study: the current participants evaluated naturally occurring commercial platforms rather than experimentally controlled versions of the same system. Consequently, the present data cannot identify whether the observed descriptive SUS differences arose from navigation design, content organization, course structure, familiarity, learner goals, or other platform-specific characteristics.

This distinction is particularly relevant in programming education. Ramos et al. (2021) showed that specific interface features in a virtual programming environment can create usability problems even when the platform serves a valid instructional purpose. Thus, a global SUS score can indicate how users perceive the system overall, but it does not identify which individual interface components produced that evaluation. The absence of statistically significant differences in the current study therefore should not be interpreted as evidence that the three platforms contain equivalent interface designs or pedagogical structures.

The same measurement boundary applies more broadly to user experience. The SUS provides a concise global measure of perceived usability, whereas broader constructs such as user experience, satisfaction with content, motivation, or emotional response may require complementary instruments. Vlachogianni and Tselios (2023), for example, reviewed educational-technology studies employing the PSSUQ/CSUQ family of questionnaires, while de Dios López et al. (2024) demonstrated in another digital domain that usability and broader user-experience characteristics can be examined through multidimensional measurement. These studies should not be interpreted as direct evidence about coding platforms; rather, they support the methodological point that perceived usability represents only one component of a broader user experience.

Methodological Limitations

Several methodological limitations should be considered when interpreting the findings. First, purposive recruitment was used to obtain equal groups of 40 participants for each platform. This balancing facilitated statistical comparison, but it did not reproduce the naturally observed platform-use distribution. The preliminary survey was strongly dominated by W3Schools users, whereas the main study intentionally gave W3Schools, Coursera, and Udemy equal representation. Consequently, the main sample is appropriate for balanced group comparison but should not be interpreted as an estimate of the prevalence or population-level distribution of platform perceptions among all university students in Purwakarta.

Second, the three groups were naturally occurring user groups rather than randomly assigned experimental conditions. Participant characteristics, previous coding experience, platform familiarity, learning objectives, and patterns of platform use could therefore differ among groups. Although Table 2 provides descriptive information on several of these characteristics, the study did not statistically adjust for them. This limitation prevents causal attribution of the observed SUS differences to platform architecture itself.

Third, the evaluation was retrospective and entirely self-reported. Participants completed the SUS based on their previous interactions with the platform rather than after performing a standardized set of equivalent coding tasks. Consequently, the findings capture remembered perceptions of usability but do not provide direct evidence regarding task-completion time, error rates, navigation efficiency, or observed interaction difficulties. Pal and Vanijja (2020) demonstrated the usefulness of SUS for evaluating perceived usability in online learning, but subjective measurement remains conceptually different from objective performance-based usability testing.

Fourth, the three platforms do not provide identical educational environments. They differ in instructional organization, course content, pricing models, certification opportunities, instructor involvement, and likely learner intentions. These dimensions were not measured as independent variables in the present study and may contribute to the overall experience from which participants formed their SUS judgments. Direct comparisons should therefore be interpreted as comparisons of perceived usability among users of three real-world platforms rather than controlled estimates of the isolated effect of interface or pedagogical architecture.

Scientific and Practical Implications

The principal scientific contribution of the study is not the identification of a statistically superior coding platform. Rather, it provides comparative evidence showing that three platforms with different instructional structures can receive similarly favorable perceived-usability evaluations within a specific group of Indonesian university students. This finding complements the broader systematic evidence that usability evaluation in educational technology is highly context-dependent and benefits from clearly specifying the system, users, tasks, and evaluation method (Lu et al., 2022; Vlachogianni & Tselios, 2022).

From a practical perspective, the findings indicate that W3Schools, Coursera, and Udemy were all perceived as highly usable by the sampled students. Because the differences were not statistically significant and the effect size was small, educators and students should not interpret the descriptive ranking as sufficient evidence for selecting W3Schools over the other platforms. Platform selection should instead consider usability together with learning objectives, available programming content, instructional structure, cost, certification requirements, and learner preferences. Alqurni (2023) similarly emphasizes that usability in e-learning involves several interface and interaction dimensions, while comparative LMS research indicates that user evaluation can extend beyond the usability score itself (Sharip et al., 2023).

Directions for Future Research

Future studies should address the current limitations by recruiting larger samples that either reflect the natural distribution of platform use or explicitly account for unequal platform prevalence. Prior experience should also be measured more precisely—for example, through duration of platform membership, cumulative hours of use, number of completed courses or modules, and frequency of recent use—and incorporated into the analysis as a potential explanatory or control variable. This would allow the possible experience effect suggested by previous SUS research (McLellan et al., 2012; Kortum & Johnson, 2013) to be tested directly rather than inferred from platform prevalence.

A mixed-method design would provide a more complete account of why users assign particular SUS ratings. Qualitative interviews, open-ended responses, think-aloud protocols, or observation could identify specific navigation, interface, and learning-workflow characteristics that cannot be isolated from a global SUS score. User-centered mixed-method usability studies such as van Doorn et al. (2023) illustrate the value of integrating subjective ratings with richer accounts of users' experiences. Likewise, the methodological diversity identified in educational-technology usability research supports combining standardized questionnaires with observational or task-based evidence where feasible (Lu et al., 2022).

Finally, future research should incorporate objective usability indicators such as task-completion time, task-success rate, error frequency, and navigation behavior and examine their relationship with perceived usability. Controlled comparisons using equivalent coding tasks would provide stronger evidence about whether specific interface structures account for differences among platforms. Subsequent studies could also examine whether perceived usability is associated with educational outcomes such as course completion, continued platform use, coding-task performance, or programming-skill development. Such analyses would extend the present study from a comparison of perceived usability toward a more comprehensive evaluation of the relationship between usability and learning outcomes.

CONCLUSIONS

This study evaluated and compared the perceived usability of three widely used online coding learning platforms—W3Schools, Coursera, and Udemy—using the System Usability Scale (SUS). All three platforms received favorable usability evaluations among the sampled university students. Descriptively, W3Schools obtained the highest mean SUS score (85.4) and was classified as "Best Imaginable," whereas Udemy (82.4) and Coursera (80.9) were both classified as "Excellent." However, these descriptive differences were not supported by statistically significant between-platform differences. The Kruskal–Wallis analysis showed no significant overall difference in SUS scores among the three platform-user groups, and the small effect size ($\varepsilon^2 = 0.021$) indicated limited between-group variation. Supplementary Dunn–Bonferroni comparisons likewise identified no statistically significant pairwise differences.

Accordingly, the findings do not establish that one platform is statistically superior to the others in perceived usability. Rather, they indicate that W3Schools, Coursera, and Udemy were all perceived as highly usable by their respective user groups in this sample, despite their different organizational and instructional characteristics. Because participants belonged to naturally occurring platform-user groups and were not randomly assigned, the observed descriptive differences cannot be attributed causally to particular interface or instructional-design features.

The contribution of this study lies in providing a standardized, SUS-based cross-platform comparison within the specific context of university students in Purwakarta, Indonesia. The findings extend the evidence on educational-technology usability by showing that differently organized coding-learning platforms can receive similarly favorable perceived-usability evaluations within this population. From a practical perspective, the results suggest that usability alone does not provide sufficient evidence for selecting one of these platforms over the others. Students and educational stakeholders should therefore consider perceived usability together with other factors not examined in this study, such as learning objectives, content structure, course availability, cost, certification, and learner preferences.

These conclusions are limited specifically to perceived usability. The study did not measure instructional quality, objective task performance, programming-skill development, or learning effectiveness and therefore does not establish the broader educational superiority or equivalence of the three platforms.

RECOMMENDATIONS

Future research should extend the present findings through larger and more diverse samples. Studies may either recruit participants in proportions that better reflect the natural distribution of platform use or retain balanced comparison groups while explicitly measuring and statistically controlling for differences in prior platform experience, coding background, and usage patterns. In particular, familiarity should be operationalized using measurable indicators such as duration of platform use, frequency of recent use, cumulative hours of interaction, and number of completed courses or tutorial modules rather than inferred from platform prevalence alone.

A mixed-method approach is also recommended to explain why participants assign particular usability ratings. Standardized SUS assessment could be complemented with interviews, open-ended responses, think-aloud protocols, or direct behavioral observation to identify specific interface, navigation, and learning-workflow characteristics underlying users' evaluations.

Future studies should additionally incorporate objective usability measures, including task-completion time, task-success rate, error frequency, and navigation behavior. Where feasible, participants could perform equivalent standardized coding tasks across platforms, allowing researchers to distinguish subjective perceived usability from observable interaction performance and to examine whether particular interface characteristics account for differences in user experience.

Finally, subsequent research should examine the relationship between perceived usability and educational outcomes, including course or module completion, continued platform use, coding-task performance, and programming-skill development. Such analyses would clarify whether favorable usability perceptions are associated with meaningful learning outcomes and would provide a more comprehensive basis for evaluating the educational value of online coding learning platforms beyond usability alone.

Acknowledgment
The authors acknowledge the institutional support provided by the Director and Deputy Directors of Universitas Pendidikan Indonesia, Purwakarta Campus, as well as the students who assisted with the data collection process and provided technical support. The authors also acknowledge the use of ChatGPT as a tool for language paraphrasing and structural refinement to improve the linguistic presentation of the manuscript.

Funding Information
This research received no external funding.

Conflict of Interest Statement
The authors declare no conflict of interest.

Informed Consent
Participation was voluntary, and participants were informed about the purpose and procedures of the study.

Ethical Approval
Formal ethics approval was waived by the institutional review board of Universitas Pendidikan Indonesia because the study involved a non-invasive survey procedure and was considered minimal risk.

Data Availability
The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Author Contributions Statement

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
M Isfa' Lana	✓	✓			✓	✓		✓	✓		✓			
NW Abdul Majid	✓	✓		✓						✓		✓	✓	
M Nor Azhari Azman				✓						✓		✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

REFERENCES

Adnan, F., Putra, J. A., Agustiningsih, M. D., Oktaviana, E., & Robi'atul Adawiyah, N. A. (2025). Exploring the usability of platforms for individuals with visual impairments: A systematic literature review. *Frontiers in Computer Science, 7*, Article 1601621. https://doi.org/10.3389/fcomp.2025.1601621

Ait Baha, T., El Hajji, M., Es-Saady, Y., & Fadili, H. (2024). The impact of educational chatbot on student learning experience. *Education and Information Technologies, 29*(8), 10153–10176. https://doi.org/10.1007/s10639-023-12166-w

Al-Fraihat, D., Joy, M., Masa'deh, R., & Sinclair, J. (2020). Evaluating e-learning systems success: An empirical study. *Computers in Human Behavior, 102*, 67–86. https://doi.org/10.1016/j.chb.2019.08.004

Al-Gerafi, M. A., Goswami, S. S., Khan, M. A., Naveed, Q. N., Lasisi, A., AlMohimeed, A., & Elaraby, A. (2024). Designing of an effective e-learning website using inter-valued fuzzy hybrid MCDM concept: A pedagogical approach. *Alexandria Engineering Journal, 97*, 61–87. https://doi.org/10.1016/j.aej.2024.04.012

Almaiah, M. A., Al-Khasawneh, A., & Althunibat, A. (2020). Exploring the critical challenges and factors influencing the e-learning system usage during COVID-19 pandemic. *Education and Information Technologies, 25*, 5261–5280. https://doi.org/10.1007/s10639-020-10219-y

Alnuaimi, M. A., & Awad, M. (2025). VR environment of digital design laboratory: A usability study. *Frontiers in Virtual Reality, 6*, Article 1566680. https://doi.org/10.3389/frvir.2025.1566680

Alqurni, J. (2023). Assessing the usability of e-learning software among university students. *International Journal of Information Technology and Web Engineering, 18*(1), 1–26. https://doi.org/10.4018/IJITWE.329198

Alshamari, M. A., & Althobaiti, M. M. (2024). Usability evaluation of wearable smartwatches using customized heuristics and System Usability Scale score. *Future Internet, 16*(6), 204. https://doi.org/10.3390/fi16060204

Anrahvi, R., Pratama, N., & Stevani, S. (2024). Penerapan metode System Usability Scale (SUS) dalam mengukur kepuasan mahasiswa terhadap website direktori akademik. *Indonesian Journal of Business Economics and Management, 3*(2), 74–80. https://doi.org/10.57152/ijbem.v3i2.2020

Binyamin, S. S., Rutter, M., & Smith, S. (2019). Extending the technology acceptance model to understand students' use of learning management systems in Saudi higher education. *International Journal of Emerging Technologies in Learning, 14*(3), 4–21. https://doi.org/10.3991/ijet.v14i03.9732

Bolpagni, M., & Gabrielli, S. (2025). Development of a comprehensive evaluation scale for LLM-powered counseling chatbots (CES-LCC) using the eDelphi method. *Informatics, 12*(1), 33. https://doi.org/10.3390/informatics12010033

Bond, M., Bedenlier, S., Marín, V. I., & Händel, M. (2021). Emergency remote teaching in higher education: Mapping the first global online semester. *International Journal of Educational Technology in Higher Education, 18*(1), 50. https://doi.org/10.1186/s41239-021-00282-x

Brooke, J. (1996). SUS: A “quick and dirty” usability scale. In P. W. Jordan, B. Thomas, I. L. McClelland, & B. Weerdmeester (Eds.), *Usability evaluation in industry* (pp. 189–194). Taylor & Francis. https://doi.org/10.1201/9781498710411

Chen, G., Chen, P., Huang, W., & Zhai, J. (2022). Continuance intention mechanism of middle school student users on online learning platform based on qualitative comparative analysis method. *Mathematical Problems in Engineering, 2022*, Article 3215337. https://doi.org/10.1155/2022/3215337

Chuenyindee, T., Montenegro, L. D., Ong, A. K. S., Prasetyo, Y. T., Nadlifatin, R., Ayuwati, I. D., Sittiwatethanasiri, T., & Robas, K. P. E. (2022). The perceived usability of the learning management system during the COVID-19 pandemic: Integrating System Usability Scale, technology acceptance model, and task-technology fit. *Work, 73*(1), 41–58. https://doi.org/10.3233/WOR-220015

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

de Dios López, A., Real, J., Meza, C., Borras-Santos, A., Collado-Borrell, R., Escudero-Vilaplana, V., & Gomis-Pastor, M. (2024). Validation of a questionnaire to assess the usability of and user experience with mobile health applications. *Healthcare, 12*(23), 2328. https://doi.org/10.3390/healthcare12232328

de Souza, J. S., Kliemann, N., Vieira, F. G. K., Al Nahas, A., Reitz, L. K., Aglago, E. K., Copetti, C. L. K., Vieira, L. C., Huybrechts, I., de Pinho, N. B., & Di Pietro, P. F. (2024). Development, content validity and usability of a self-assessment instrument for the lifestyle of breast cancer survivors in Brazil. *Nutrients, 16*(21), 3707. https://doi.org/10.3390/nu16213707

Dilmen, K., Kert, S. B., & Uğraş, T. (2023). Children’s coding experiences in a block-based coding environment: A usability study on Code.org. *Education and Information Technologies, 28*(9), 10839–10864. https://doi.org/10.1007/s10639-023-11625-8

Dunn, O. J. (1964). Multiple comparisons using rank sums. *Technometrics, 6*(3), 241–252. https://doi.org/10.1080/00401706.1964.10490181

Dyayu, A. L., Beny, B., & Yani, H. (2023). Evaluasi usability aplikasi PeduliLindungi menggunakan metode usability testing dan System Usability Scale (SUS). *Jurnal Manajemen Teknologi dan Sistem Informasi (JMS), 3*(1), 395–404. https://doi.org/10.33998/jms.2023.3.1.720

Hew, K. F., & Cheung, W. S. (2014). Students’ and instructors’ use of massive open online courses (MOOCs): Motivations and challenges. *Educational Research Review, 12*, 45–58. https://doi.org/10.1016/j.edurev.2014.05.001

Kementerian Pendidikan Dasar dan Menengah. (2025). *Peraturan Menteri Pendidikan Dasar dan Menengah Nomor 13 Tahun 2025 tentang perubahan atas Peraturan Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi Nomor 12 Tahun 2024 tentang* *kurikulum pada pendidikan anak usia dini, jenjang pendidikan dasar, dan jenjang pendidikan menengah*. JDIH Kemendikdasmen.

Kortum, P., & Johnson, M. (2013). The relationship between levels of user experience with a product and perceived system usability. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 57*(1), 197–201. https://doi.org/10.1177/1541931213571044

Kovanović, V., Joksimović, S., Poquet, O., Hennis, T., Čukić, I., de Vries, P., Hatala, M., Dawson, S., Siemens, G., & Gašević, D. (2018). Exploring communities of inquiry in massive open online courses. *Computers & Education, 119*, 44–58. https://doi.org/10.1016/j.compedu.2017.11.010

Kremser, W., Kranzinger, S., & Bernhart, S. (2021). Design and implementation of a gesture-aided e-learning platform. *Sensors, 21*(23), 8042. https://doi.org/10.3390/s21238042

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association, 47*(260), 583–621. https://doi.org/10.1080/01621459.1952.10483441

Lana, M. I., & Majid, N. W. A. (2024). Analisis kualitas Microsoft Teams menggunakan SUS sebagai platform pembelajaran berbasis e-learning. *SemanTIK: Teknik Informasi, 10*(2). https://doi.org/10.55679/semantik.v10i2.82

Lewis, J. R. (2018). Measuring perceived usability: The CSUQ, SUS, and UMUX. *International Journal of Human–Computer Interaction, 34*(12), 1148–1156. https://doi.org/10.1080/10447318.2017.1418805

Lewis, J. R., & Sauro, J. (2021). Usability and user experience: Design and evaluation. In G. Salvendy & W. Karwowski (Eds.), *Handbook of human factors and ergonomics* (pp. 972–1015). John Wiley & Sons. https://doi.org/10.1002/9781119636113.ch38

Lin, H.-C. K., Liao, Y.-C., & Wang, H.-T. (2022). Eye movement analysis and usability assessment on affective computing combined with intelligent tutoring system. *Sustainability, 14*(24), 16680. https://doi.org/10.3390/su142416680

Liu, Z.-Y., Lomovtseva, N., & Korobeynikova, E. (2020). Online learning platforms: Reconstructing modern higher education. *International Journal of Emerging Technologies in Learning, 15*(13), 4–21. https://doi.org/10.3991/ijet.v15i13.14645

Lu, J., Schmidt, M., & Lee, M. (2022). Usability research in educational technology: A state-of-the-art systematic review. *Educational Technology Research and Development, 70*, 1951–1992. https://doi.org/10.1007/s11423-022-10152-6

Luxton-Reilly, A., Simon, Albluwi, I., Becker, B. A., Giannakos, M., Kumar, A. N., Ott, L., Paterson, J., Scott, M. J., Sheard, J., & Szabo, C. (2018). Introductory programming: A systematic literature review. In *ITiCSE 2018 Companion: Proceedings Companion of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education* (pp. 55–106). Association for Computing Machinery. https://doi.org/10.1145/3293881.3295779

Maramba, I., Chatterjee, A., & Newman, C. (2019). Methods of usability testing in the development of eHealth applications: A scoping review. *International Journal of Medical Informatics, 126*, 95–104. https://doi.org/10.1016/j.ijmedinf.2019.03.018

Martin, F., & Bolliger, D. U. (2018). Engagement matters: Student perceptions on the importance of engagement strategies in the online learning environment. *Online Learning, 22*(1), 205–222. https://doi.org/10.24059/olj.v22i1.1092

Martins, A. I., Rosa, A. F., Queirós, A., Silva, A., & Rocha, N. P. (2015). European Portuguese validation of the System Usability Scale (SUS). *Procedia Computer Science, 67*, 293–300. https://doi.org/10.1016/j.procs.2015.09.273

McLellan, S., Muddimer, A., & Peres, S. C. (2012). The effect of experience on System Usability Scale ratings. *Journal of Usability Studies, 7*(2), 56–67.

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*(3), 437–455. https://doi.org/10.1037/a0028085

Mertasari, N. M. S., & Candiasa, I. M. (2022). Formative evaluation of digital learning materials. *Journal of Education Technology, 6*(3), 507–514. https://doi.org/10.23887/jet.v6i3.44165

Mozahem, N. A. (2021). The online marketplace for business education: An exploratory study. *The International Journal of Management Education, 19*(3), 100544. https://doi.org/10.1016/j.ijme.2021.100544

Nugroho, K. T., Julianto, B., & Nur, D. F. M. S. (2022). Usability testing pada sistem informasi manajemen AKN Pacitan menggunakan metode System Usability Scale. *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI), 11*(1), 74. https://doi.org/10.23887/janapati.v11i1.43209

Obonya, J., & Kadlečík, M. (2020). Assessing the intensity of the usability of the course content within the virtual learning environment. *International Journal of Emerging Technologies in Learning, 15*(17), 165–180. https://doi.org/10.3991/ijet.v15i17.13715

Paiva, J. C., Queirós, R., Leal, J. P., Swacha, J., & Miernik, F. (2022). Managing gamified programming courses with the FGPE platform. *Information, 13*(2), 45. https://doi.org/10.3390/info13020045

Pal, D., & Vanijja, V. (2020). Perceived usability evaluation of Microsoft Teams as an online learning platform during COVID-19 using System Usability Scale and technology acceptance model in India. *Children and Youth Services Review, 119*, Article 105535. https://doi.org/10.1016/j.childyouth.2020.105535

Prajapati, S. P., & Das, S. (2025). Is "quick and dirty" good enough? An analysis of the usability evaluation practices for learning environment design. In *Proceedings of the 2025 ACM International Conference on Interactive Media Experiences* (pp. 313–321). Association for Computing Machinery. https://doi.org/10.1145/3706370.3727857

Ramos, V. F. C., Cechinel, C., Magé, L., & Lemos, R. (2021). Student and lecturer perceptions of usability of the virtual programming lab module for Moodle. *Informatics in Education, 20*(2), 1–19. https://doi.org/10.15388/infedu.2021.14

Ratnasari, N., & Wibawa, A. P. (2020). Analisis perbandingan kualitas UI/UX platform online coding course pada pembelajaran daring pemrograman komputer dengan metode A/B testing. *JEPIN (Jurnal Edukasi dan Penelitian Informatika), 6*(2). https://doi.org/10.26418/jp.v6i2.40771

Sauro, J. (2016). Standardized usability questionnaires. In *Quantifying the user experience: Practical statistics for user research* (2nd ed., pp. 185–248). Morgan Kaufmann.

Setiawan, D., Wicaksono, S. L., & Rafianto, N. (2020). Evaluasi usability e-learning Moodle dan Google Classroom menggunakan SUS questionnaire. *JAMI: Jurnal Ahli Muda Indonesia, 1*(1), 55–64. https://doi.org/10.46510/jami.v1i1.13

Setiyawati, N., & Bangkalang, D. H. (2022). The comparison of evaluation on user experience and usability of mobile banking applications using User Experience

Questionnaire and System Usability Scale. *Proceedings, 82*(1), 87. https://doi.org/10.3390/proceedings2022082087

Sharfina, Z., & Santoso, H. B. (2016). An Indonesian adaptation of the System Usability Scale (SUS). In *2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS)* (pp. 145–148). IEEE. https://doi.org/10.1109/ICACSIS.2016.7872776

Sharip, A. A., Md Zain, S. M., Abdul Rahid, N. N., & Md Saidi, R. (2023). Usability evaluation of learning management systems: Google Classroom, Microsoft Teams and Padlet. *International Journal of Academic Research in Progressive Education and Development, 12*(2). https://doi.org/10.6007/IJARPED/v12-i2/17224

Shi, A., Huo, F., & Han, D. (2021). Role of interface design: A comparison of different online learning system designs. *Frontiers in Psychology, 12*, Article 681756. https://doi.org/10.3389/fpsyg.2021.681756

Su, W., Luo, D., Zhang, C., & Zeng, S. (2022). Evaluation of online learning platforms based on probabilistic linguistic term sets with self-confidence multiple attribute group decision making method. *Expert Systems with Applications, 208*, Article 118153. https://doi.org/10.1016/j.eswa.2022.118153

Tao, L., Cukurova, M., & Song, Y. (2025). Learning analytics in immersive virtual learning environments: A systematic literature review. *Smart Learning Environments, 12*(1), 43. https://doi.org/10.1186/s40561-025-00381-6

Tieri, G., Costa, A., Zabberoni, S., Tenaglia, E., & De Simone, M. S. (2025). User experience, system usability, and feasibility of two novel immersive virtual reality memory tasks for cognitive training: A pilot study. *Brain Sciences, 15*(12), 1289. https://doi.org/10.3390/brainsci15121289

Tomczak, M., & Tomczak, E. (2014). The need to report effect size estimates revisited: An overview of some recommended measures of effect size. *Trends in Sport Sciences, 21*(1), 19–25.

van Doorn, M., Monsanto, A., Boeschoten, C. M., van Amelsvoort, T., Popma, A., Öry, F. G., Alvarez-Jimenez, M., Gleeson, J., Jaspers, M. W. M., & Nieman, D. H. (2023). Moderated digital social therapy for young people with emerging mental health problems: A user-centered mixed-method design and usability study. *Frontiers in Digital Health, 4*, Article 1020753. https://doi.org/10.3389/fdgth.2022.1020753

Veeraiyan, D. N., Varghese, S. S., Rajasekar, A., Karobari, M. I., Thangavelu, L., Marya, A., Messina, P., & Scardina, G. A. (2022). Comparison of interactive teaching in online and offline platforms among dental undergraduates. *International Journal of Environmental Research and Public Health, 19*(6), 3170. https://doi.org/10.3390/ijerph19063170

Vlachogianni, P., & Tselios, N. (2022). Perceived usability evaluation of educational technology using the System Usability Scale (SUS): A systematic review. *Journal of Research on Technology in Education, 54*(3), 392–409. https://doi.org/10.1080/15391523.2020.1867938

Vlachogianni, P., & Tselios, N. (2023). Perceived usability evaluation of educational technology using the Post-Study System Usability Questionnaire (PSSUQ): A systematic review. *Sustainability, 15*(17), 12954. https://doi.org/10.3390/su151712954

Wang, M., Ramasamy, S. S., Dawod, A. Y., & Yu, X. (2025). Development and TAM-based validation of a user experience scale for actual system use in online courses. *Education Sciences, 15*(7), 855. https://doi.org/10.3390/educsci15070855

Yulius, R., Nasrullah, M. F. A., & Thohari, A. H. (2022). Analisis usability pada aplikasi Amboo Mothercare menggunakan System Usability Scale. *Jurnal Cakrawala Ilmiah, 1*(10), 2349–2358. https://doi.org/10.53625/jcijurnalcakrawalailmiah.v1i10.2564

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators? *International Journal of Educational Technology in Higher Education, 16*(1), 39. https://doi.org/10.1186/s41239-019-0171-0