OPEN ACCESS
International Journal of Innovations in Science & Technology

# Crowdsourced vs. Traditional Usability Testing: A Comparative Usability Evaluation of Learning Management Systems

Sadiq Khan, Muhammad Nasir, Zakia Jalil, Qaisar Afzal
Faculty of Computing & Information Technology, International Islamic University, Islamabad, 44000, Pakistan.
*Correspondence: m.nasir@iiu.edu.pk

**Citation |** Khan. S, Nasir. M, Jalil. Z, Afzal. Q, "Crowdsourced vs. Traditional Usability Testing: A Comparative Usability Evaluation of Learning Management Systems", IJIST, Vol. 8 Issue. 3 pp 1166-1191, June 2026
**Received |** April 18, 2026 **Revised |** May 21, 2026 **Accepted |** May 28, 2026 **Published |** June 10, 2026

**Introduction/Importance of Study**: Usability and User Experience (UX) evaluation is essential for improving Learning Management Systems (LMS) because these platforms directly influence students' learning effectiveness and satisfaction.
**Novelty statement:** This study uniquely compares traditional user testing with a crowdsourced usability testing method on prominent global LMS platforms (Google Classroom, Canvas, and Edredo), addressing a critical research gap in large-scale benchmarking and resource-efficiency.
**Material and Method:** An experimental research design involved 15 traditional and 15 crowdsourced testers who performed identical tasks. Sessions were recorded via OBS software and analyzed using the Molich and Dumas coding scheme. To determine statistical significance between the two methods, the Mann-Whitney U test was employed to evaluate effectiveness (issue counts), efficiency (time), and cost-effectiveness.
**Result and Discussion:** The evaluation identified a total of 310 usability issues. The traditional method discovered 170 (55%), while the crowdsourced method identified 140 (45%). In terms of effectiveness parity, the traditional method detected 48 key issues compared to 45 in the crowdsourced approach ($p=0.23014$), proving no significant difference in identifying critical failures. However, crowdsourced testing demonstrated extreme economic efficiency, reducing costs by 6.4x—expending $150 compared to $960 for traditional testing ($p=0.00001$). Regarding time efficiency, crowdsourced sessions were completed in 4.88 hours compared to 6.07 hours for traditional methods ($p=0.11123$), indicating comparable operational speeds.
**Concluding Remarks:** Crowdsourced usability testing provides a statistically validated alternative to traditional methods, achieving comparable issue detection effectiveness while significantly reducing resource expenditure.

**Keywords:** Usability; User Experience; User Testing; Crowd Sourced Usability and Learning Management Systems

**Introduction:**

A Learning Management System (LMS) is one of the most innovative tools for disseminating knowledge beyond geographical boundaries [1][2][3][4]. This global digitization drive was significantly accelerated by the COVID-19 pandemic, which renewed research interest in the use and usability of these systems [1]. In the post-pandemic educational environment, where long-distance education is no longer a matter of choice but a necessity, the development of systems that meet established usability and user experience (UX) standards is paramount [5].

Also referred to as a Virtual Learning Environment (VLE), LMS is a software platform utilized by educational institutions, including schools, colleges, and universities, enabling both educators and students to access its features anytime, anywhere, and on any device [6][7][8]. As online learning platforms have become prominent in the 21st century, recent scholarship emphasizes the need for cross-cultural and user-centered design to enhance student engagement and ensure a positive user experience [2].

Beyond academia, LMS has found substantial adoption in various sectors, including healthcare, corporate training, and government agencies [5][9]. It serves as a comprehensive software unit, integrating a multitude of features designed to facilitate the management and delivery of knowledge [10][11]. Key features of LMS platforms include a user-friendly interface, customisation options to tailor the system to specific user needs, virtual classrooms, integration with social media platforms, and communication tools such as discussion forums, chat systems, course management, and reporting functionalities [7][8][11]. Despite these diverse features, challenges regarding low adoption rates among educators persist; research suggests that perceived usability—heavily influenced by human factors and organizational support—is a critical predictor of actual system usage [6].

Evaluating the usability and user experience (UX) of LMS platforms is critical for ensuring their effectiveness in supporting educational processes. Such evaluations aim to enhance the interaction between students and educators, thereby mitigating potential challenges encountered during the learning process. Usability and UX are pivotal factors in determining the success of these platforms [5][10]. Recent assessments utilizing the System Usability Scale (SUS) often indicate that current systems fall short of established usability benchmarks, highlighting an urgent need for optimized navigation and improved user interfaces [11]. To address these deficiencies, contemporary studies recommend a triangulation of evaluation methods—including heuristic evaluations, eye-tracking, and interviews—to effectively validate and improve usability requirements [1].

**Research Contributions**

Conducts a comparative evaluation of prominent LMS such as Google Classroom, Canvas, and Edredo, whereas existing research typically focuses on local or institutional systems.

Statistically validates that crowdsourcing is as effective as traditional lab-based testing ($p=0.23$), identifying a comparable number of key usability issues (approximately 42–46% of total identified).

Quantifies a 6.4x reduction in expenditure, documenting a total cost of $150 for crowdsourcing versus $960 for traditional laboratory methods.

User Experience (UX) pertains to the overall experience an individual has while interacting with a product or service, especially in the context of digital technologies [12]. UX design involves creating digital solutions that prioritize user satisfaction by addressing user needs, behaviors, and motivations. It encompasses aspects such as ease of use, enjoyment, and the extent to which the product effectively fulfils user requirements [13]. Conversely, usability is defined as the degree to which a software application can be used by specified users to achieve specific goals effectively, efficiently, and satisfactorily within a defined context [14].

The literature highlights various Usability Evaluation Methods (UEMs), with usability testing (user-based evaluation) and usability inspection being among the most prominent.

Traditional usability evaluation methods, such as user testing, and analytics-based approaches like A/B testing, are often resource-intensive in terms of both cost and time. These challenges arise from the difficulty of recruiting actual users, high participant leasing costs, and the complexity of creating testing environments that comprehensively address all aspects of software usability [15][16][17].

To address these limitations, usability inspection methods were introduced as cost-effective alternatives. However, these methods heavily rely on the availability of usability experts and may fall short in adequately evaluating real user experiences [15]. More recently, crowdsourcing has emerged as an effective approach to overcome the cost and time constraints associated with traditional UEMs [18][19][20][21].

Despite extensive research on the application of crowdsourcing for usability evaluation, there remains a significant gap in assessing the usability of LMS platforms in the context of higher education [21]. While numerous higher education institutions and organizations have adopted modern LMS platforms, research addressing the accessibility and usability of such systems remains sparse [22][23]. Notably, there is a pressing need to explore the viability of crowdsourcing platforms in evaluating the usability of LMS, particularly in systems like Google Classroom. Furthermore, studies comparing the cost-effectiveness and time efficiency of traditional user-testing methods with crowdsourced usability methods in assessing LMS usability and UX are conspicuously lacking [21].

In light of these gaps, this study seeks to address the limitations of traditional usability testing by investigating the feasibility of crowdsourced usability testing. The research focuses on evaluating the usability and UX of LMS platforms while comparing the efficiency and effectiveness of traditional and crowdsourced usability evaluation techniques.

**Aims and Objectives:** The overarching aim of this research is to compare the effectiveness of traditional and crowdsourcing methods in evaluating the usability of Learning Management Systems (LMS). This aim is realized through the following specific objectives and research intents:

To compare the effectiveness, efficiency, and reliability of traditional lab-based usability testing and crowdsourced usability evaluation.

To analyses and compare the cost and time requirements of traditional and crowdsourced usability testing approaches.

To evaluate the usability of three prominent LMS platforms: Google Classroom, Canvas, and Edredo.

**Material and Methods:**

**Related Work:** A Learning Management System is defined as a digital infrastructure designed to deliver and manage instructional content, assess and track individual and organizational learning goals, and oversee the learning process within an institution or organization [23]. LMS has become an essential tool in facilitating efficient educational processes, particularly in higher education. Its adoption has increased significantly in recent years across academic institutions [24].

LMS platforms provide a wide range of open-source and commercial tools, such as Blackboard, Edvance, Pearson's College, Jenzabar e-Racer, Moodle, LoudCloud, OLAT, and Claroline [25]. Among these, Blackboard is the most widely used commercial tool, while Moodle stands as the most prominent open-source platform [26].

By fostering an inclusive learning environment, LMS supports professional training, group collaboration, discussions, and communication among users [8]. Jafari [27] noted that LMS platforms consistently provide learners with streamlined information, enhancing their performance. Moreover, LMS promotes independent learning while ensuring sustainable engagement for its users [28].

**Significance of Usability for LMS:** As LMS platforms evolve, becoming increasingly complex and feature-rich, their effectiveness and efficiency depend heavily on usability. If an LMS is not designed with user-friendliness in mind, it risks being inaccessible or unusable to a significant proportion of its users [29]. Users often identify usability challenges as being tied to an LMS's alignment with human cognitive abilities and perceptions. Enhanced usability not only improves the learning experience but also contributes to improved academic performance among students [30]. The integration of human-computer interaction principles and usability research is a crucial aspect of LMS design and development. Academic institutions can leverage usability studies to assess LMS systems, ensuring ease of use and high levels of user satisfaction [31].

**Usability Assessment Methods:** In the context of LMS, usability pertains to designing platforms that align with the needs, expectations, and cognitive processes of both students and educators. The usability of LMS platforms plays a critical role in meeting user expectations and supporting effective learning experiences [32]. Usability evaluation has evolved over the past four decades, progressing from end-user testing in controlled lab environments to inspection methods and remote testing. These methods aim to reliably gather user feedback to inform and validate interface designs.

Numerous Usability Evaluation Methods (UEMs) exist, with usability testing and usability inspection being the most prevalent. However, traditional usability testing methods are often resource-intensive, requiring significant time and financial investment. Challenges include difficulty in recruiting participants, high participant compensation costs, and creating comprehensive testing environments [33].

Traditional methods are further limited by geographical constraints and the inability to simulate real-world user environments, including work-related tasks and diverse hardware configurations [34]. Effective results, as per Nielsen [35], can often be achieved with as few as five users, but such methods remain costly and time intensive.

Most existing studies on LMS usability rely on traditional methods such as questionnaires and focus group discussions [36][37][38][39]. While these studies often identify usability issues, they rarely provide actionable recommendations to resolve them [40][41].

The usability inspection method, which relies on expert evaluators and heuristic assessments, has emerged as an alternative. However, it faces challenges similar to traditional methods, such as requiring multiple evaluations and expert involvement for reliable results [42]. Moreover, these methods often overlook usability issues that could be identified through actual user engagement [12].

Crowd sourcing usability testing has gained traction as a cost-effective and efficient alternative. It enables researchers to recruit diverse users for evaluating software platforms. [18] conducted a study comparing traditional and crowdsourced usability testing methods for a graduate school website. The results indicated that crowdsourcing offered a faster, more cost-effective approach while achieving results comparable to traditional methods.

However, as [20] pointed out, the effectiveness of crowdsourcing methods depends on factors such as user expertise and the quality of the crowdsourcing platform. For instance, Liu's study [18] employed only student users and used a crowdsourcing platform that yielded suboptimal results. [20] further demonstrated that novice users could identify similar usability issues as experts, making crowdsourcing a cost-effective alternative. Despite these advancements, further exploration is required to evaluate the potential of crowdsourcing methods in assessing the usability of e-learning platforms like Google Classroom. This study addresses this gap by comparing the effectiveness and efficiency of traditional and crowdsourcing usability evaluation techniques.

**Current Research Overview on Crowdsourcing & Usability Evaluation:**

**Strategic Evolution and Frameworks for Crowdsourced UX Research:**

Recent literature reports that while traditional User Experience (UX) research methods like interviews and expert evaluations are foundational, they often struggle with scalability and diversity. Researchers are now proposing structured frameworks to formalize the transition to crowdsourced evaluations.

Researcher [43] highlights that crowdsourcing offers a quicker, more cost-effective alternative to conventional methods by providing access to larger and more heterogeneous populations [43] The authors argue that this approach overcomes the primary challenges of traditional research, such as reaching diverse user samples and accelerating the speed of user testing.

To make crowdsourced evaluations more systematic, [21] introduced CrowdUiX, a framework designed for novice usability inspection [21]. Through a multi-case study, the research demonstrated that with the suitable framework, practitioners can effectively translate findings from a novice crowd into successful software redesigns, avoiding common pitfalls like poor task design or incentive mismanagement.

**Comparative Performance: Remote, Crowdsourced, and Traditional Lab Settings:**

A central theme in recent years is the direct comparison of remote and crowdsourced methods against established laboratory standards to validate their efficacy.

Authors [44] in a controlled experiment comparing traditional User Acceptance Testing (UAT) with an asynchronous crowdsourcing framework, the crowdsourced method achieved 80% bug detection effectiveness compared to 73.3% for traditional UAT. The study concludes that crowdsourcing offers superior flexibility and documentation quality while maintaining high accuracy.

Researchers [45] Focused on the complexities of Augmented Reality (AR), this study compared lab-based testing with synchronous remote methods. It identified that while remote testing is feasible, it faces unique challenges such as technical issues, a sense of "distance" from the authentic AR experience and reduced experimental control over the user environment.

Researchers [46] in an experimental study of 100 participants compared web-based asynchronous remote tests to laboratory settings. The findings revealed that the specific web technology (HTML5/JavaScript) used in remote tools can affect task execution times and participant satisfaction. Furthermore, it noted that a lack of human support in remote settings makes participants more prone to negative emotions, especially when hindered by slow internet connections.

**Advanced Data Validation and Reliability Metrics:**

Recent literature has shifted focus toward algorithmic data cleaning and consistency metrics to ensure data integrity because crowdsourced platforms are susceptible to bots and inattentive participants.

Authors [47] proposed the SUS Consistency Score (CSc), a self-contained metric that evaluates the reliability of System Usability Scale responses by looking for logically opposing statements to combat the “MTurk quality crisis”. The researchers found that bot-generated and careless responses can significantly skew SUS scores, and applying CSc-based filtering is essential to prevent misleading conclusions about a tool’s perceived usability.

Researchers [48] introduced a cleaning method based on score differences across multiple questionnaires, such as the SUS, UMUX, and mATM. The study established that a 30-point score difference between highly correlated scales serves as a valid threshold for identifying low-quality data. Additionally, the authors confirmed that completion time remains a useful, though secondary, indicator for detecting unreliable crowdsourced data.

**Limitations of the current literature on crowdsourcing & usability evaluation:**

The current literature review on crowdsourcing and usability evaluation reveals several critical gaps that justify the importance of this study:

**Lack of Standardized Metric Comparison:** While many studies acknowledge that crowdsourcing is "faster" or "cheaper" they frequently lack a granular quantitative comparison [44]. Specifically, few studies calculate the exact cost-per-usability-issue or the exact person-hours required per key issue when comparing crowd-sourced methods to traditional lab-based testing.

**Focus on Modality over Platform:** Current research often compares remote testing to lab testing, but it frequently fails to distinguish the specific benefits of using established crowdsourcing platforms (like Fiverr or MTurk) for the evaluation of prominent, high-traffic systems.

**Absence of Major LMS Benchmarking:** A significant limitation in existing literature is the focus on local, custom-built, or institutional software. There is a conspicuous lack of studies that apply these comparative methodologies to widely used global Learning Management Systems such as Google Classroom, Canvas, or Edredo.

**Reliability vs. Efficiency Trade-off:** Much of the recent research is dedicated to data cleaning and reliability (combating bots and inattentive users). While vital, this focus has come at the expense of investigating the comparative effectiveness—specifically, whether the "crowd" can identify the same number of critical or "key" usability issues as traditional expert-led or lab-based sessions.

A summary of the research on crowdsourced usability evaluation is presented in Table 1.

**Table 1.** Overview of literature on crowdsourcing and usability evaluation

| Reference | Aim | Research Design | Contributions | Limitations |
|---|---|---|---|---|
| [47] | To propose a quality metric, SUS Consistency Score (CSc), to assess the reliability of crowdsourced usability scores. | Exploratory surveys on MTurk and Clickworker followed by a large-scale (n=254) validation study on the PoPLar tool. | CSc effectively identifies inattentive or bot responses; filtering by CSc significantly improves the reliability of SUS scores. | No direct comparison to traditional lab-based methods for cost or time. Focuses solely on data integrity within the crowd environment. |
| [44] | To address limitations in traditional User Acceptance Testing (UAT) by developing an asynchronous crowdsourcing framework. | A controlled experiment involving 10 participants testing an inventory management application. | The crowdsourcing framework achieved higher bug detection (80%) than traditional UAT (73.3%) with superior flexibility. | Bug detection effectiveness was found to be more influenced by individual attentiveness than the method itself. |
| [21] | To improve novice usability inspection using a proposed framework named CrowdUiX. | Validation of the framework through a multi-case study in practical software redesign settings. | Practitioners can effectively use CrowdUiX to translate novice crowd findings into successful software redesigns. | In the absence of a framework, practitioners may overlook critical factors like incentive management and task design. |
| [48] | To identify low-quality data in crowdsourced | Testing data cleaning methods across two environments using | A 30-point score difference between correlated scales is a | No comparison to traditional usability testing |

| | | | | |
|---|---|---|---|---|
| | usability tests using score differences between multiple questionnaires. | SUS, UMUX, and mATM questionnaires. | valid cleaning threshold; completion time is also a useful indicator. | methods. Focuses entirely on the refinement of the crowdsourcing data cleaning process. |
| [46] | To compare asynchronous remote usability tests using web-based tools against traditional laboratory tests. | An experimental study with 100 participants and 15 evaluators focusing on website testing. | Web-based tools affect task execution times; participants in remote settings are more prone to negative emotions. | Compares remote vs. lab, but not specifically crowdsourced platforms vs. traditional testing in terms of cost-effectiveness and scalability. |
| [45] | To compare lab-based and remote synchronous usability testing methods specifically for Augmented Reality (AR). | A direct comparison of outcome variables (performance, satisfaction, accuracy) across three testing approaches. | Investigated if reduced experimental control in remote settings affects the validity and usability results of AR. | Focuses on the modality (remote vs. lab) rather than a cost/time-efficiency comparison of crowdsourced testing. |
| [43] | To explore how crowdsourcing can overcome the limitations of conventional UX research methods. | A book chapter overviewing conventional vs. crowdsourced methods such as interviews and focus groups. | Crowdsourcing provides quicker access to larger, more heterogeneous populations compared to small-scale surveys. | Conventional methods struggle to reach diverse user samples and accelerate the speed of user testing. |

**Current Research Overview on Usability Evaluation of LMS:**

**Factors Influencing LMS Adoption and Perceived Usability:**

Recent literature has shifted from simply measuring usability to understanding how it predicts the actual acceptance of systems by educators and students.

Researchers [6] addressed the issue of low LMS adoption among lecturers in Ghana by proposing a statistical model based on Partial Least Squares Structural Equation Modeling (PLS-SEM) [6]. The research identifies that human factors, organizational support, and social influence significantly impact the perceived usability (PUsab) of a system. Crucially, the authors demonstrate that perceived usability is a direct predictor of whether lecturers will actually use the platform, providing a validated questionnaire for future adoption assessments.

**Design Frameworks: Cultural and Context-Specific Requirements:**

As LMS platforms move beyond a "one-size-fits-all" model, researchers are exploring how specific user needs and cultural contexts dictate design requirements.

Muhmad Asri et al. (2024) conducted systematic literature review of 74 articles (2014–2023) examining the intersection of Cross-Cultural Design (CCD) and User-Centered Design (UCD) [2]. The study highlights that the design of an LMS—specifically its UI/UX elements—directly affects student engagement. It argues that for an LMS to be successful, it must embrace diverse cultures, languages, and social contexts to create a positive and inclusive online learning experience.

Authors [1] focused on 545 Open Distance and Electronic Learning (ODeL) environments, this study proposes a set of validated usability requirements specifically tailored for lecturers and students in distance education [1]. By abstracting these requirements from existing literature and validating them through mixed methods, the research provides a framework to ensure LMS platforms meet the unique navigational and pedagogical needs of remote learners.

**Evaluation Methodologies and Usability Benchmarks:**

A significant portion of recent research focuses on the methods used to identify usability failures and how current institutional systems compare to established industry standards.

[5] conducted a systematic mapping analyzing the publications from the last decade to identify the most common evaluation approaches. The results show that Jakob Nielsen's heuristics remain the dominant method in LMS evaluations, often used in combination with other techniques. The authors emphasize that the 2020 pandemic transformed LMS adoption from a choice into a necessity, necessitating more diverse and perhaps automated methods to ensure these systems support effective learning.

A study by [11] provided a practical application of the System Usability Scale (SUS) to evaluate the "U Learn" LMS at UNDIKNAS. The evaluation revealed an average SUS score of 56.65, which falls significantly below the recognized academic benchmark of 68. These findings highlight common failures in institutional systems, leading to specific recommendations for optimizing user interfaces, improving navigation, and providing better user guides to bridge the usability gap.

Beyond providing design requirements, research study by Lehong et al. (2024) contributed to methodological efficiency by triangulating heuristic evaluation (HE) with eye-tracking, SUS questionnaires, and interviews. The study concludes that Heuristic Evaluation remains a highly effective and efficient method for identifying critical usability problems in an LMS, especially when validated against more resource-intensive quantitative methods.

None of the recent LMS-based studies explicitly evaluated the usability of prominent global platforms such as Google Classroom, Canvas, or Edredo; instead, they focused on institutional systems or broader systematic reviews. Regarding evaluation methods, while not all studies compared them, those that did—particularly [1]—placed a strong emphasis on the efficiency and effectiveness of the techniques used. An overview of the research on usability evaluation of the LMS is provided in Table 2.

**Table 2.** Overview of literature on usability evaluation of LMS

| Study | Aim | Research Design | Contributions | Limitations |
|---|---|---|---|---|
| [6] | To investigate low LMS adoption among lecturers in Ghana using a PLS-SEM approach. | Quantitative survey of 255 lecturers across four higher education institutions. | Mapped technology acceptance to usability; provided a validated questionnaire. | The paper does not evaluate prominent LMS like Canvas or Google Classroom. Besides, the authors did not emphasize the cost/time efficiency of evaluation methods. |
| [1] | To propose and validate usability requirements for an LMS in an ODeL context. | Mixed-methods triangulation (HE, eye tracking, SUS, and interviews). | Confirmed Heuristic Evaluation (HE) as a validated method for ODeL systems. | Limited to an unnamed institutional LMS rather than Google Classroom or Canvas. |
| [2] | To review cross-cultural design | Systematic Literature Review | Identified cultural factors (UI/UX) | As a review, it does not conduct a direct usability |

| | | | | |
|---|---|---|---|---|
| | (CCD) to improve user engagement in LMS. | (SLR) of 74 articles (2014–2023). | that influence student engagement. | evaluation of prominent platforms like Edredo. Focuses on design elements over method efficiency. |
| [11] | To evaluate the usability of the U Learn LMS at UNDIKNAS using SUS. | Quantitative evaluation with 68 student participants using a 10-item Likert scale. | Identified that the system's usability falls below the benchmark; provided UI recommendations. | Evaluates a specific institutional system ("U Learn") rather than prominent global LMS. Does not provide a comparison of method efficiency. |
| [5] | To map common approaches for evaluating LMS usability and UX (2010–2020). | Systematic Mapping (SM) of approximately 80 publications. | Highlighted the dominance of Nielsen's heuristics in the field. | Identifies a gap in evaluating prominent systems but does not evaluate them directly. Advocates for automated methods to improve evaluation efficiency. |

**Gap Analysis:** While the global digitization drive has made Learning Management Systems (LMS) a necessity, evaluating their effectiveness often relies on resource-intensive frameworks [5]. Although extensive literature exists on usability and user experience (UX) studies, the majority of contemporary research focuses on traditional methods, such as heuristic inspections and think-aloud evaluations, or centers on solving data integrity challenges like bot detection and questionnaire consistency within crowdsourced environments.

A critical research gap exists regarding the quantitative comparative effectiveness of traditional versus crowdsourced usability evaluations, particularly concerning cost-effectiveness and time efficiency. While some studies acknowledge that crowdsourcing offers access to larger, more heterogeneous populations, they often prioritize data cleaning methodologies or modality comparisons (e.g., remote vs. laboratory) over granular, performance-based metrics like cost-per-issue detected or total person-hours required. Furthermore, most LMS-specific research continues to evaluate local or institutional platforms rather than providing a comprehensive benchmark for prominent global systems. There is also a lack of research on assessing LMS usability through comprehensive comparisons of teams or practitioners using identical methodologies across different testing paradigms.

This research seeks to fill these gaps by conducting a comparative analysis of traditional user testing and crowdsourcing methods to evaluate the usability and UX of three prominent LMS platforms: Google Classroom, Canvas, and Edredo. By emphasizing cost-effectiveness, time efficiency, and user diversity, this study provides a novel empirical benchmark that demonstrates how practitioners can leverage the scalability of the crowd to achieve results comparable to traditional lab-based sessions at a fraction of the required resources.

This research aimed to evaluate the effectiveness and efficiency of traditional and crowd-sourced usability testing techniques in assessing the usability of Learning Management Systems (LMS). The study focused on three prominent LMS platforms, namely Google Classroom, Canvas, and Edredo. This objective was achieved by employing an experimental research design. The primary intent was to determine whether crowd-sourced usability testing yields comparable reliability and effectiveness to traditional methods. Additionally, special attention was directed toward comparing the costs and time expenditures associated with both usability testing

approaches [49]. It is noteworthy that this research was conducted exclusively from the student's perspective.
The researcher posited the following hypotheses:
**$Ha_0$**: Crowd-sourced usability testing will identify distinct usability issues compared to traditional usability testing.
**$Ha_1$**: Crowd-sourced usability testing will identify similar usability issues as traditional usability testing.
**$Hb_0$**: Crowd-sourced usability testing will incur comparable costs to traditional usability testing.
**$Hb_1$**: Crowd-sourced usability testing will incur lower costs than traditional usability testing.
**$Hc_0$**: Crowd-sourced usability testing will require a similar time investment to traditional usability testing.
**$Hc_1$**: Crowd-sourced usability testing will require less time than traditional usability testing.

From the hypotheses, it is evident that the control variable in this study was the crowd-sourcing platform, the independent variable was the profile of the usability tester, and the dependent variables included the usability problems identified, the cost incurred per usability issue, and the time invested in usability evaluation.

**Experimental Research Design:** To compare the effectiveness of traditional and crowd-sourced usability testing methodologies, an experimental approach was adopted. Both crowd-based testers and expert evaluators were engaged in this experiment, which comprised two distinct phases.
**Phase One**: Usability issues were identified through traditional usability testing methods.
**Phase Two**: Usability testing was conducted via a crowd-sourcing platform, specifically Fiverr.
A schematic representation of the experimental setup is illustrated in **Figure 1**.

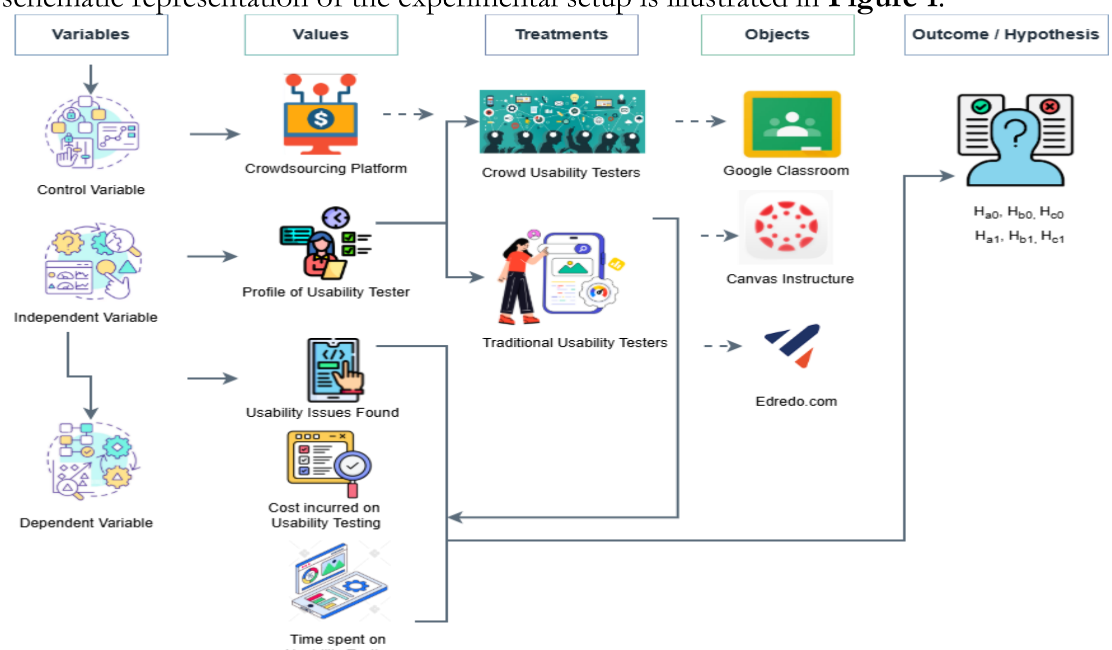


**Figure 1.** Experimental framework showing variables, metrics, tester groups, and outcomes.

For the first trial, the researcher conducted usability testing on Google Classroom using both crowd-sourced and traditional methods. Google Classroom, a free LMS developed by Google, serves as a platform where students and teachers can create and join classes, assignments, and lectures [50].

**Second Trial of Experimentation:** The second trial involved Canvas, a leading LMS that enables educators to manage digital learning effectively. Teachers can create and manage classes and assignments, while students can join classes, submit assignments, and access lecture materials [51]. Canvas also offers features like class discussions, quizzes, and pages.

**Third Trial of Experimentation:** In the third trial, the usability testing was conducted on Edredo, a free web-based LMS. It facilitates the creation of classrooms, assignments, quizzes, online meetings, and grade books. Founded by Retina Biswas in 2008, it is both an educational and social learning platform [52].

**Data Collection:** The experiment was divided into two parts: traditional and crowd-sourced usability testing methods. Testers for each step were selected based on predefined criteria.

**Criteria for Selecting Testers:** Testers for both methods were required to meet the following criteria:

Must possess a bachelor's degree in software engineering, computer science, or information technology.

Must have prior experience with LMS platforms.

**Instrument for Data Collection:** Both traditional and crowd-sourced testers were given the same task questionnaire to ensure uniformity. Tasks included:

Joining Google Classroom using a provided Gmail address and class code.

Downloading lecture slides and assignments, submitting assignments, and navigating between pages.

Responding to comments, editing comments, and accessing announcements.

Checking assignment deadlines, progress status, and joining additional classes using a class code.

**Conducting the Test:** Traditional usability tests were conducted in a controlled lab environment with a moderator, who also took notes. In contrast, crowd-sourced usability tests were performed independently, with testers recording sessions using OBS screen recording software. For traditional usability tests, the cost per user averaged $64 based on literature and industry consultations in Pakistan, where individual costs ranged from $64 to $68. Conversely, crowd-sourced testers from Fiverr were compensated $10 per task.

**Data Analysis:** The data analysis process followed the coding scheme of Molich and Dumas [49], converting recorded test sessions and task questionnaire responses into categorized usability problem comments. These categories included atomic comments, minor and critical problems, false alarms, and key usability issues.

The Mann-Whitney U test was used to compare results due to its:

**Non-parametric nature**: Suitable for non-normally distributed data.

**Applicability to independent groups**: Designed for separate groups like crowd-sourced and traditional testers.

**Robustness to outliers**: Ensures outliers do not skew results.

**Statistical significance**: Provides p-values to determine meaningful differences.

**Profile of Traditional and Crowd-Sourced Usability Testers:** Traditional usability evaluation involved 15 users from Pakistan; each assigned a unique ID (T1–T15). They tested three LMS platforms:

**Google Classroom**: T1–T5.

**Canvas**: T6–T10.

**Edredo**: T11–T15.

The traditional testing group consisted of 15 senior-level students (T1–T15) specializing in Computer Science, Information Technology, and Software Engineering. These participants were primarily in their 7th or 8th semesters, ensuring a high level of technical proficiency and academic seniority. Their experience with Learning Management Systems (LMS) was robust, ranging from 1 to 3 years, with a third of the group possessing the maximum of 3 years of experience. This profile established a technically knowledgeable baseline of "power users" for the comparative evaluation of Google Classroom, Canvas, and Edredo. The profiles of the testers are detailed in Table3.

**Table 3.** Profile of Participants for Traditional Usability Testing.

| User ID | Qualification | Current Semester | Department | Experience of LMS (Years) |
|---|---|---|---|---|
| T1 | BS-CS | 8 | Computer Science | 2 |
| T2 | BS-CS | 8 | Computer Science | 3 |
| T3 | BS-IT | 7 | Information Technology | 3 |
| T4 | BS-SE | 8 | Software Engineering | 2.5 |
| T5 | BS-SE | 3 | Software Engineering | 1 |
| T6 | BS-CS | 8 | Computer Science | 3 |
| T7 | BS-CS | 7 | Computer Science | 2 |
| T8 | BS-SE | 7 | Software Engineering | 1 |
| T9 | BS-CS | 8 | Computer Science | 3 |
| T10 | BS-IT | 8 | Information Technology | 1.5 |
| T11 | BS-SE | 6 | Software Engineering | 1 |
| T12 | BS-IT | 7 | Information Technology | 3 |
| T13 | BS-CS | 7 | Computer Science | 1 |
| T14 | BS-CS | 8 | Computer Science | 1.5 |
| T15 | BS-IT | 7 | Information Technology | 1 |

For the crowdsourced usability testing, 15 participants (C1–C15) were recruited via the Fiverr platform from a diverse range of 10 countries, including Germany, India, Japan, the USA, and Nigeria. To ensure a structured evaluation, users were assigned unique IDs based on the platform they tested: C1–C5 for Google Classroom, C6–C10 for Canvas, and C11–C15 for Edredo. As detailed in Table 4, these testers possessed strong technical backgrounds in fields such as Computer Engineering and Software Engineering, with LMS experience ranging from 0.5 to 3 years. This international diversity provided high ecological validity, as it allowed the researchers to observe how these global platforms perform across varying geographic locations and technical environments, contrasting with the more localized traditional group.

**Table 4.** Profile of Participants for Crowd-sourced Usability Testing.

| User ID | Country | Qualification | Department | Experience of LMS (Years) |
|---|---|---|---|---|
| C1 | Germany | BS-CE | Computer Engineering | 2 |
| C2 | India | BS-CS | Computer Science | 2 |
| C3 | Bulgaria | BS-SE | Software Engineering | 0.5 |
| C4 | Japan | BS-CE | Computer Engineering | 1 |
| C5 | New Zealand | BS-IT | Information Technology | 1.5 |
| C6 | India | BC-SE | Software Engineering | 3 |
| C7 | USA | BS-CS | Computer Science | 2 |
| C8 | UK | BS-CE | Computer Engineering | 1 |
| C9 | Korai | BS-CS | Computer Science | 3 |
| C10 | India | B-Tech | Information Technology | 1 |
| C11 | India | BE-CE | Computer Engineering | 0.5 |
| C12 | USA | BS-CS | Computer Science | 1 |
| C13 | India | BS-CI | Computer Information | 2 |
| C14 | Spain | BS-IT | Information Technology | 1 |
| C15 | Nigeria | BS-CS | Computer Science | 1 |

**Analysis of Usability Testing Reports and Recorded Test Sessions:** The analysis began by describing the characteristics of the collected data. The study focused on identifying the number of usability problems encountered and assessing the severity of these issues, which were categorized as minor usability issues, serious usability issues, or critical usability issues. Upon completion of both parts of the experiment, the researcher received a total of 30 usability reports and recorded test sessions for further analysis:

**15 Reports**: Crowd-sourced usability testing (5 for each LMS: Google Classroom, Canvas Instructure, and Edredo).

**15 Reports**: Traditional usability testing (5 for each LMS: Google Classroom, Canvas, and Edredo).

**Data Overview:** The reports contained 311 original comments, which highlighted various observations and usability problems identified during the tests.

**Analysis Duration and Method:** The researcher spent 27 days thoroughly analyzing the reports. To ensure consistency, the coding scheme developed by Molich and Dumas [49] was employed. This scheme facilitated the classification of comments from the usability reports and recorded test sessions into distinct usability problem categories.

**Analysis Procedure: Figure 2** outlines the step-by-step procedure followed during the analysis, from the collection of usability reports and recordings to coding, categorizing, and interpreting usability problems.

**Step 1:** Identifying Problematic Comments: The analysis of usability reports and recorded test sessions is crucial for identifying problematic comments. Usability feedback often encompasses multiple issues, requiring further classification to isolate distinct problem statements [53].

**Step 2:** Splitting and Combining Problematic Comments: Usability reports often contain composite comments that need to be broken down into atomic (singular) comments. The process of segmentation continues until all comments are unique and focus on a specific issue. Additionally, similar issues are consolidated into unified problem statements.

**Example 1:** Splitting a Composite Comment (Google Classroom):

Original Comment: The class work button's icon on the main dashboard is unfamiliar and unclear in representing its functionality. Furthermore, it lacks any label or descriptive information to determine its purpose.

**Split Comments:**

The class work icon on the main dashboard is unsuitable for representing its functionality and appears unfamiliar.

There is no label or supplementary information to describe the functionality of the class work icon.

**Example 2: Combining Similar Problem Comments:**

**Original Comments:**

While joining a class, the "+" button did not have a label to indicate its functionality.

When joining another class, the button again lacked a label to inform users about its purpose.

**Combined Comment:**

The "+" button for joining a class lacks a label to clarify its functionality.

**Step 3: Aggregating Related Usability Problems at a Higher Abstraction Level:** To minimize redundancy and prevent over- or under-reporting of usability issues, related problems are grouped at higher levels of abstraction. For instance, if a usability issue is reported by multiple users, it is consolidated into a single, overarching problem statement.


**Figure 2.** Stepwise Analysis Procedure for Usability Testing Reports and Recorded Test Sessions.

**Example:**

**Observation:** The system does not provide search or filter options, requiring users to scroll extensively to locate a desired file.

**Higher-Level Problem Statement:**

The system lacks efficient navigation tools, such as search or filtering options, for locating files.

**Step 4: Identifying Duplicate Usability Issues in the Master List:**

In the final stage, duplicate usability issues reported by multiple users are identified and synthesized into unique problem statements. Two or more issues are considered identical if resolving one inherently addresses the others.

**Example:**

**Original Comments:** Users encountered difficulties while downloading files and assignments.

**Refined Statement:** Users face challenges in downloading files and assignments.

**Step 5: Determining the Severity Levels of Usability Issues:**

The severity of usability problems is assessed and categorized into three levels:

**Minor (1):** Minimal impact on functionality.

**Serious (2):** Significant impact requiring attention.

**Critical (3):** Severe impact that impedes primary functionality.

**Result and Discussion:**

The experimental data gathered from the evaluation of Google Classroom, Canvas, and Edredo reveals a distinct distribution of usability issues across the two testing modalities. According to the aggregated results in Table 5, the traditional usability testing (TUT) method identified a total of 170 problems, while the crowdsourced (CSUT) method identified 140 problems. Across both methods, Canvas emerged as the most problematic platform, yielding 90 issues via TUT and 73 via CSUT. In contrast, Edredo appeared to be the most stable, with the lowest total issue counts of 29 and 21, respectively. According to the criteria established by Molich and Dumas, a usability issue reported by two or three users is classified as serious or critical and is regarded as a "key usability issue", which must be resolved before product release to avoid significant design implications. In this research, the traditional testing group identified a total of 22 key issues: 11 for Google Classroom (requiring an average of 0.412 hours to resolve each), 7 for Canvas (0.352 hours), and 4 for Edredo (0.426 hours). The crowdsourced group identified 16 key issues total, including 8 for Google Classroom (0.324 hours average), 4 for Canvas (0.284 hours), and 4 for Edredo (0.358 hours), notably achieving exact parity with the traditional group on the Edredo platform. The granular performance metrics in Table 6 highlight the significant economic divergence between the two approaches based on their respective payment structures. Traditional testers were compensated at a rate of $64 per user, resulting in a total cost of $960, whereas the cost of crowdsourced usability testing was $150 in total, with a cost of $10 per user. This disparity is reflected in the cost-per-issue metrics; for example, on the Canvas platform, the cost for traditional testers ranged between $2.67 and 8.00 per issue, while crowdsourced testers identified flaws for as little as 0.36 to $2.00 per issue. Time investment also varied, with traditional testers requiring an average of 0.40 hours to complete tasks compared to the 0.32 hours averaged by the crowdsourced group.

**Table 5.** Overview of the Results.

| | **GCT** | **CT** | **ET** | **GCC** | **CIC** | **EC** |
|---|---|---|---|---|---|---|
| Total number of original comments | 52 | 85 | 31 | 42 | 81 | 20 |
| Total number of comments after splitting & combining | 22 | 17 | 15 | 14 | 16 | 11 |
| Total number of usability problems | 51 | 90 | 29 | 46 | 73 | 21 |
| Total number of minor problems | 7 | 26 | 4 | 9 | 22 | 3 |
| Total number of serious problems | 23 | 42 | 10 | 23 | 35 | 7 |
| Total number of critical problems | 21 | 13 | 14 | 14 | 16 | 11 |
| Total number of key usability issues | 11 | 7 | 4 | 8 | 4 | 4 |

*GCT= Google Classroom Traditional, CT=Canva Traditional, ET= Edredo Traditional, GCC= Google Classroom Crowdsourced, CC= Canva Crowdsourced, EC= Edredo Crowdsourced

**Table 6.** Detailed Results.

| **Google Classroom** | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Resource** | **T1** | **T2** | **T3** | **T4** | **T5** | **C1** | **C2** | **C3** | **C4** | **C5** |
| Total number of original comments | 8 | 8 | 13 | 13 | 10 | 8 | 9 | 7 | 9 | 9 |
| Total number of comments after splitting & combining | 8 | 8 | 11 | 12 | 12 | 9 | 10 | 8 | 9 | 10 |
| Total number of usability problems | 8 | 8 | 11 | 12 | 12 | 9 | 10 | 8 | 9 | 10 |
| Total number of minor problems | 1 | 0 | 2 | 3 | 1 | 3 | 2 | 1 | 1 | 2 |
| Total number of serious problems | 3 | 3 | 5 | 5 | 7 | 5 | 5 | 3 | 4 | 6 |
| Total number of critical problems | 4 | 5 | 4 | 4 | 4 | 1 | 3 | 4 | 4 | 2 |
| Total number of key usability issues | 6 | 6 | 6 | 9 | 10 | 5 | 6 | 5 | 5 | 8 |
| Person hours used | 0.57 | 0.2 | 0.42 | 0.55 | 0.32 | 0.45 | 0.26 | 0.33 | 0.3 | 0.28 |
| Payment made to each user (USD) | 64 | 64 | 64 | 64 | 64 | 10 | 10 | 10 | 10 | 10 |
| Total cost per user (USD) | 64 | 64 | 64 | 64 | 64 | 10 | 10 | 10 | 10 | 10 |
| Cost per issue (USD) | 8.00 | 8.00 | 5.82 | 5.33 | 5.33 | 1.11 | 1.00 | 1.25 | 1.11 | 1.00 |
| Cost per key issue (USD) | 10.67 | 10.67 | 10.67 | 7.11 | 6.40 | 2.00 | 1.67 | 2.00 | 2.00 | 1.25 |
| Issues per hour | 14 | 40 | 26 | 22 | 38 | 20 | 38 | 24 | 30 | 36 |
| Key issues per hour | 11 | 30 | 14 | 16 | 31 | 11 | 23 | 15 | 17 | 29 |
| Thoroughness of usability problems | 16% | 16% | 22% | 24% | 24% | 20% | 22% | 17% | 20% | 22% |
| **Canvas** | | | | | | | | | | |
| **Resource** | **T1** | **T2** | **T3** | **T4** | **T5** | **C1** | **C2** | **C3** | **C4** | **C5** |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Total number of original comments | 11 | 24 | 22 | 20 | 8 | 4 | 6 | 34 | 8 | 29 |
| Total number of comments after splitting & combining | 13 | 24 | 23 | 22 | 8 | 5 | 5 | 28 | 8 | 27 |
| Total number of usability problems | 13 | 24 | 23 | 22 | 8 | 5 | 5 | 28 | 8 | 27 |
| Total number of minor problems | 3 | 4 | 6 | 11 | 2 | 2 | 0 | 9 | 0 | 11 |
| Total number of serious problems | 6 | 13 | 14 | 5 | 4 | 2 | 3 | 15 | 5 | 10 |
| Total number of critical problems | 3 | 4 | 2 | 3 | 1 | 1 | 2 | 4 | 3 | 6 |
| Total number of key usability issues | 6 | 7 | 6 | 3 | 2 | 3 | 2 | 3 | 3 | 3 |
| Person hours used | 0.41 | 0.31 | 0.5 | 0.27 | 0.27 | 0.27 | 0.24 | 0.35 | 0.3 | 0.26 |
| Payment made to each user (USD) | 64 | 64 | 64 | 64 | 64 | 10 | 10 | 10 | 10 | 10 |
| Total cost per user (USD) | 64 | 64 | 64 | 64 | 64 | 10 | 10 | 10 | 10 | 10 |
| Cost per issue (USD) | 4.92 | 2.67 | 2.78 | 2.91 | 8.00 | 2.00 | 2.00 | 0.36 | 1.25 | 0.37 |
| Cost per key issue (USD) | 10.67 | 9.14 | 10.67 | 21.33 | 32.00 | 3.33 | 5.00 | 3.33 | 3.33 | 3.33 |
| Issues per hour | 32 | 77 | 46 | 81 | 30 | 19 | 21 | 80 | 27 | 104 |
| Key issues per hour | 15 | 23 | 12 | 11 | 7 | 11 | 8 | 9 | 10 | 12 |
| Thoroughness of usability problems | 14% | 27% | 26% | 24% | 9% | 7% | 7% | 38% | 11% | 37% |
| **Edredo** | | | | | | | | | | |
| **Resource** | **T1** | **T2** | **T3** | **T4** | **T5** | **C1** | **C2** | **C3** | **C4** | **C5** |
| Total number of original comments | 8 | 7 | 6 | 6 | 4 | 3 | 7 | 4 | 2 | 4 |
| Total number of comments after splitting & combining | 7 | 7 | 6 | 5 | 4 | 3 | 7 | 4 | 3 | 4 |
| Total number of usability problems | 7 | 7 | 6 | 5 | 4 | 3 | 7 | 4 | 3 | 4 |
| Total number of minor problems | 2 | 1 | 1 | 0 | 0 | 0 | 1 | 2 | 0 | 0 |
| Total number of serious problems | 1 | 3 | 2 | 2 | 2 | 0 | 3 | 2 | 1 | 1 |
| Total number of critical problems | 4 | 3 | 2 | 3 | 2 | 3 | 3 | 0 | 2 | 3 |
| Total number of key usability issues | 4 | 4 | 4 | 4 | 4 | 3 | 4 | 1 | 3 | 3 |
| Person hours used | 0.37 | 0.49 | 0.28 | 0.54 | 0.45 | 0.34 | 0.4 | 0.35 | 0.33 | 0.37 |
| Payment made to each user (USD) | 64 | 64 | 64 | 64 | 64 | 10 | 10 | 10 | 10 | 10 |
| Total cost per user (USD) | 64 | 64 | 64 | 64 | 64 | 10 | 10 | 10 | 10 | 10 |
| Cost per issue (USD) | 9.14 | 9.14 | 10.67 | 12.80 | 16.00 | 3.33 | 1.43 | 2.50 | 3.33 | 2.50 |
| Cost per key issue (USD) | 16.00 | 16.00 | 16.00 | 16.00 | 16.00 | 3.33 | 2.50 | 10.00 | 3.33 | 3.33 |
| Issues per hour | 19 | 14 | 21 | 9 | 9 | 9 | 18 | 11 | 9 | 11 |
| Key issues per hour | 11 | 8 | 14 | 7 | 9 | 9 | 10 | 3 | 9 | 8 |
| Thoroughness of usability problems | 24% | 24% | 21% | 17% | 14% | 14% | 33% | 19% | 14% | 19% |

Individual productivity and thoroughness metrics further illustrate the capabilities of the independent testers. As shown in Table 6, individual crowdsourced testers often exhibited high efficiency; for instance, tester C5 achieved a rate of 104 issues per hour on the Canvas platform, surpassing the highest traditional productivity rate of 81 issues per hour recorded by tester T4 on the same system. Furthermore, the thoroughness of individual testers—representing the percentage of total known problems found by a single user—was often higher in the crowdsourced group. Specifically, crowdsourced testers C3 and C5 identified 38% and 37% of the total problems on Canvas, respectively, while the most thorough traditional tester for that platform reached 27%. These figures indicate that the diverse, international technical users recruited through the crowd provided highly concentrated feedback despite the absence of a controlled laboratory environment.

Examples of Key Usability Issues:

**Google Classroom:**

Users cannot edit comments to correct errors, as the comment section lacks an editing option in Figure 3.

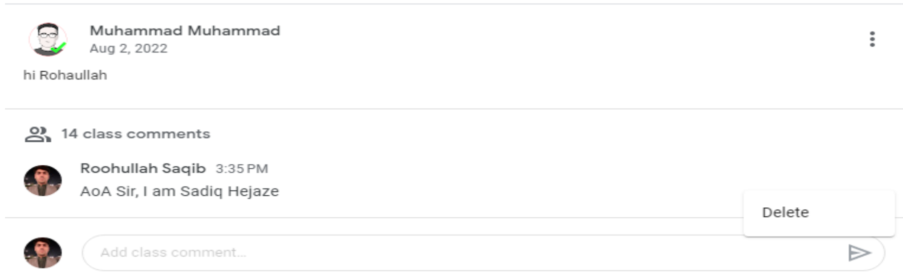


**Figure 3.** User is not able to edit comment key issue.

The users are not able to open classwork progress from the main dashboard, because the icon which represents the classwork progress is not suitable and unfamiliar to represent such task in **Figure 4**.

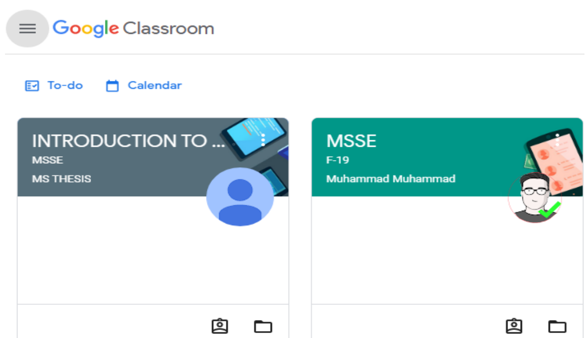


**Figure 4.** Classwork icon is not familiar it is a key issue.

**Canvas:**

There is no option for joining another class, instead there is a button "start new course" in which a student can create a new course that he/she want to teach.

It is not good practice to show recent assignments at last of assignment list/To Do list/Comments. Most recent assignments should be displayed first in the list so users can see them easily in Figure 5.

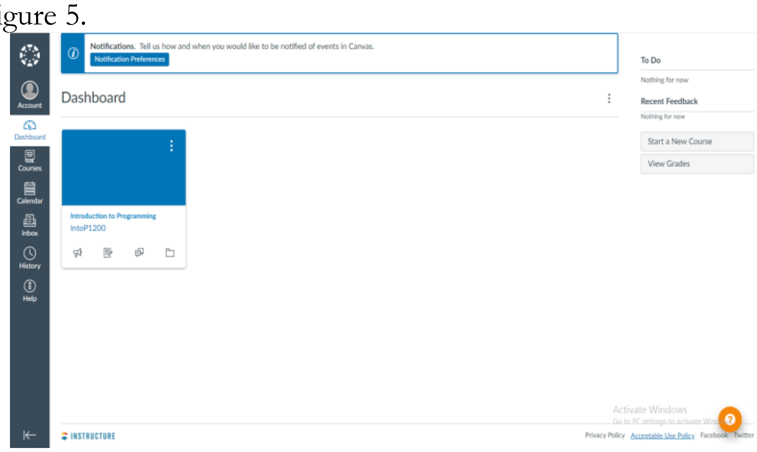


**Figure 5.** Recent assignments are shown at last key issue.

**Edredo:**

I am not able to join classroom because "Explore classroom" does not tell me to join a classroom in **Figure 6**.

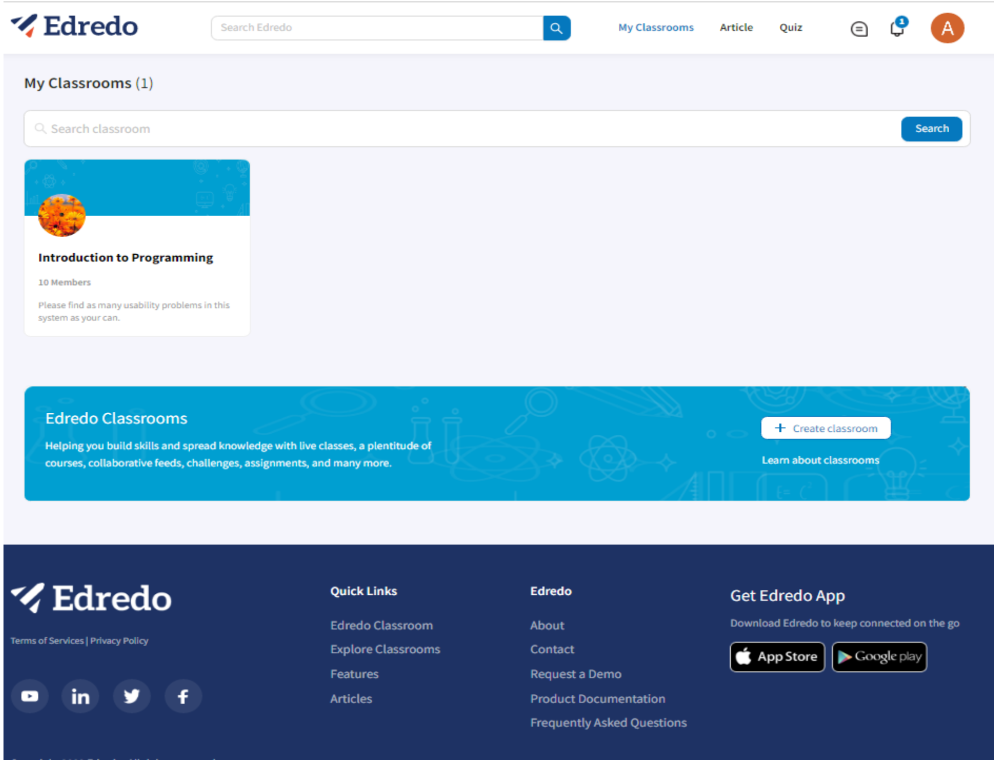


**Figure 6.** Joining a classroom key issue.

The notification systems are not notifying users by any means.

Appendix A provides a detailed summary of all the major usability concerns identified across the platforms. The experiment concludes that while most key issues—22 (58%)—were discovered through traditional usability testing, the crowdsourced method remained highly effective by identifying 16 (42%) of these critical problems. To determine the Usability Level of these prominent Learning Management Systems (LMS), the study calculated a "Key Issue per Hour" metric by dividing the average number of key issues by the average time taken for both methods, as shown in Table 7. For example, the average of 9.5 key issues for Google Classroom was derived by combining the key issues identified by both traditional and crowdsourced methods. Under this benchmarking, platforms with lower values were assigned higher usability levels: Edredo achieved a "High" usability level with a score of 10.20, followed by Google Classroom at "Medium" (16.55), and Canvas at "Low" (17.30). Table 4 further supplements this analysis by providing a detailed breakdown of the major usability issues highlighted by each test object and user.

**Table 7.** Key Usability Metrics.

| | **Avg. Key Issues** | **Avg. Time** | **Key Issue/ Usability Level** | **Per Hours** |
|---|---|---|---|---|
| Google Classroom | 9.5 | 0.574 | 16.55 | Medium |
| Canvas | 5.5 | 0.318 | 17.30 | Low |
| Edredo | 4.0 | 0.392 | 10.20 | High |

**Statistical Analysis:**

**$H_a$: Usability Issues Identified:** This hypothesis evaluated the effectiveness of crowd-sourced usability testing compared to traditional usability testing methods in identifying usability issues in the test objects. The hypothesis being tested was whether crowd-sourced usability testing could identify the same number of usability issues as traditional methods, such as user testing. The study identified a total of 310 usability issues, of which crowd-sourced usability testing detected 140 (45%). For key usability issues, crowd-sourced testing identified 41 (46%) out of a total of 89 key usability issues. To determine if significant differences existed between the two methods, the study employed the Mann-Whitney U test. The results indicated no significant difference between the two samples, with a p-value of 0.23014. This supports the null hypothesis ($H_{a1}$), which states that there are no significant differences between the two methods in identifying usability issues. Based on these findings, the study concludes that crowd-sourced usability testing is equally effective as traditional usability testing methods in identifying both general and key usability issues.

**$H_b$: Cost-Effectiveness:** This hypothesis compared the cost-effectiveness of crowd-sourced usability testing and traditional usability testing. The cost of crowd-sourced usability testing was $150 in total, with a cost of $10 per user. Conversely, the cost of traditional usability testing was $960 in total, with a cost of $64 per user. To compare the cost-effectiveness of these methods, the Mann-Whitney U test was used, which is a statistical tool for comparing two independent data sets. The statistically significant p-value of 0.00001 reported in Table 8 supports the alternative hypothesis and confirms that the crowd-sourced usability testing (CSUT) method is more cost-effective. In summary, the study demonstrates that the cost per user for crowd-sourced usability testing is substantially lower than that of traditional usability testing.

**$H_c$: Time Efficiency:** The hypothesis ($H_c$) tested whether crowd-sourced usability testing is as time-efficient as traditional usability testing. To evaluate this, the study measured the average time users required to complete usability tasks using each method. On average, crowd-sourced users took 0.32 hours to complete tasks, while traditional users took 0.40 hours. To compare time efficiency, the number of participants per hour was calculated for each method. The Mann-Whitney U test revealed no statistically significant difference between the two approaches (p=0.11123). In simpler terms, while crowd-sourced testers completed tasks slightly faster, the difference was not statistically significant. Therefore, it can be concluded that both methods are comparably time efficient in Table 7.

**Key Findings:**

**Effectiveness**: Crowd-sourced usability testing is as effective as traditional usability testing in identifying usability issues, including key issues.

**Cost-Effectiveness**: Crowd-sourced usability testing is significantly more cost-effective, offering a substantially lower cost per user.

**Time Efficiency**: Both methods exhibit comparable time efficiency, with no significant difference in task completion times.

**Implications:** These findings support the integration of crowd-sourced usability testing, particularly in scenarios where cost reduction is crucial without compromising quality or efficiency.

**Table 8.** Hypothesis Testing Results.

| S. No. | Hypothesis | P-value | Significance | Result |
|---|---|---|---|---|
| 1 | $H_a$: Usability Issues Found | P=0.23014 | Not Significant | $H_{a1}$: The crowd usability testing will find similar usability issues as compared to traditional usability testing. |
| 2 | $H_b$: Cost-Effectiveness | P=0.00001 | Significant | $H_{b1}$: The crowd usability testing incurs less cost as compared to traditional usability testing. |
| 3 | $H_c$: Time-Efficiency | P=0.11123 | Not Significant | $H_{c0}$: The crowd usability testing will incur similar time as compared to traditional usability testing. |

**Comparison of Results:** The use of crowd-sourced methods for usability/UX evaluation has been explored extensively by researchers in the past [53]. For instance, the study in evaluated the usability of a graduate school website with differing participant numbers 5 for traditional methods and 55 for crowd-sourced usability testing. However, this study focused solely on time efficiency, disregarding cost-effectiveness and the number of usability issues identified. Furthermore, the tasks assigned to participants differed between the two methods.

The research concluded that while the quality of results from crowdsourcing was generally lower than those from lab-based usability testing, significant usability problems could still be identified using the crowd-sourced method. Additionally, the study highlighted that crowdsourcing is inexpensive, faster to conduct, and easier to perform, benefiting participants with diverse backgrounds. In another study, M. Nasir et al. compared novice crowd inspectors to expert evaluators. The comparison was based on time efficiency, cost-effectiveness, and the number of usability issues identified. The findings revealed that the crowd-sourced method uncovered the same number of usability issues as the expert method while being more cost-effective and time-efficient. Similarly, the study in [53] compared crowd-sourced usability testing with the Think-Aloud (TA) method on the EDX learning platform. Both methods were assessed based on cost-effectiveness, time efficiency, and the number of usability problems identified. The results showed no significant differences in the quality and quantity of usability problems uncovered, suggesting that crowd-sourced testing is a viable and cost-effective alternative to traditional TA methods. In comparison to these prior studies, our research employed both traditional and crowd-sourced usability methods on widely used Learning Management Systems (LMS), whereas studies in [53] focused on local LMS. The results of our study are aligned with previous findings, demonstrate that crowd-sourced usability testing is cost-effective and identifies a comparable number of usability issues to traditional methods, including user testing, TA, lab-based methods, and expert heuristic evaluation. However, our results diverged in the aspect of time efficiency, where no significant difference was observed between the two methods.

**Mapping of Results with Research Objectives:**

The experimental findings derived from the comparative evaluation of Google Classroom, Canvas, and Edredo provide a robust empirical foundation for addressing the study's core research intents. The following mapping aligns the quantitative performance data and statistical validations with the specific objectives established in the manuscript to demonstrate how the research goals were fulfilled.

**Comparison of Effectiveness and Reliability (Objective 1):** The findings demonstrate that crowdsourced usability testing (CSUT) is as effective as traditional lab-based testing (TUT) in identifying critical software flaws. While the traditional method identified a higher volume of total issues (170 vs. 140) and key issues (22 vs. 16), the Mann-Whitney U test ($p=0.23014$) confirmed that there is no statistically significant difference in the quality or quantity of the problems detected. This proves that a technically proficient global "crowd" can provide feedback of comparable reliability to a controlled laboratory setting.

**Analysis of Cost & Time Requirements (Objective 2):** The results reveal that crowdsourcing offers a dramatic increase in economic efficiency without a significant sacrifice in time. Crowdsourced testing was statistically confirmed to be 6.4x more cost-effective ($p=0.00001$), requiring only 150(10 per user) compared to the 960 (64 per user) spent on the traditional group. Regarding operational speed, while the crowdsourced group completed tasks faster (4.88 total hours vs. 6.07 hours for the traditional group), the Mann-Whitney U test ($p=0.11123$) indicated that this difference is not statistically significant, concluding that both methods provide comparable time efficiency.

**Evaluation of Prominent LMS Platforms (Objective 3):** The study utilized a “Key Issue per Hour” metric to establish a definitive usability hierarchy for the three evaluated platforms. Based on the density of critical flaws relative to the time taken for evaluation, Edredo achieved a "High" usability level (10.20), followed by Google Classroom at a "Medium" level (16.55), and Canvas Instructure at a "Low" level (17.30). These rankings provide a benchmark for practitioners to identify which global systems currently meet or fall short of student usability expectations.

**Limitations and Validity Threats:**

Statistical Conclusion Validity: When conducting a case study on LMS usability, it is crucial to account for validity threats that could influence the study's results. This study considered several such threats.

**Construct Validity:** Construct validity refers to the extent to which a tool or method accurately measures the theoretical construct it intends to assess. In this study, the following construct validity threats were addressed:

**Consistency of Instruments**: The same instruments were used for both user-based and crowd-sourced usability testing to ensure consistent measurements. Changes in survey wording or test design could otherwise affect participant responses.

**Sample Size and Participant Profiles**: Equal sample sizes (five participants per LMS) were used for both testing methods. Participants had similar profiles to minimize variability due to differing technical skills or preferences.

**Response Bias**: To mitigate potential response bias, participants received equal compensation, avoiding any undue influence on their feedback.

**Coding Scheme**: A uniform coding scheme was applied across all LMS to ensure consistency in evaluating usability problems.

**Internal Validity.** Internal validity examines whether the study's results are attributable to the variables being studied rather than external factors. Potential threats addressed include:

**Demographics**: Participants from both methods shared similar demographic profiles.

**Learning Effect**: To avoid participants becoming more proficient during testing, each user evaluated a single test object, eliminating the learning effect.

**External Validity.** External validity assesses the extent to which findings can be generalized beyond the study conditions. Threats addressed include:

**Selection Bias**: Participants were recruited from diverse computing fields, ensuring the results could be generalized to users with similar technical backgrounds.

**Contextual Factors**: Prominent LMS, such as Google Classroom, Canvas, and Edredo, were chosen to provide a comprehensive overview.

**Participant Expertise**: Participants' expertise ranged from 0.5 to 3.0 years, ensuring a representative mix of novice and experienced users.

**Sample Size**: Based on Jakob Nielsen's recommendation, five participants per LMS were deemed sufficient for usability testing.

**Reliability:** Reliability measures the consistency and stability of results. To ensure reliability:

**Interrater Reliability**: Multiple evaluators assessed the usability of LMS, and measures like Cohen's Kappa coefficient were used to ensure agreement among raters.

**Task Consistency**: Identical tasks and questionnaires were employed for both methods, maintaining uniformity across testing sessions.

**Limitations of the Study.** This study acknowledges several limitations:

**Scope**: Usability evaluation focused only on student-related tasks, excluding roles like teachers and parents.

**Demographic Constraints**: Participants were limited to Pakistan and the International Islamic University, Islamabad. Broader geographic and institutional representation could yield different results.

**Domain Limitation**: All participants had computing backgrounds. Including individuals from other fields could offer more varied insights.

**Crowd Recruitment**: Crowd testers were recruited exclusively through Fiverr. Using multiple platforms might produce more diverse findings.

**Limited Control in Crowdsourcing**: Crowdsourced testing inherently reduces control over demographics and technical setups. Detailed instructions and iterative testing were used to address this issue.

**Conclusion:** Usability evaluation methods are critical in assessing the usability of software applications. However, the most practical method is one that is both time-efficient and cost-effective. In this research paper, we conducted a comparison between two usability evaluation methods, namely the traditional user-testing method and the crowd-sourced usability testing method, on prominent learning management systems (LMS) such as Google Classroom, Canvas, and Edredo. We specifically focused on the LMS student panels to determine our chosen methods' cost-effectiveness and time efficiency. Previous literature compared crowdsourced usability testing with expert heuristic evaluation; however, no comparison has been reported between crowdsourced and traditional usability testing methods to identify which approach incurs lower cost and requires less time. Based on the research objectives and findings detailed in the sources, here is a point-wise conclusion for your manuscript:

**Effectiveness and Reliability:** The study concludes that crowdsourced usability testing is as effective and reliable as traditional lab-based methods for identifying critical software flaws. While the traditional method identified 170 total issues (including 48 key issues), the crowdsourced method identified 140 (including 45 key issues). Statistical validation using the Mann-Whitney U test confirmed no significant difference in the quality or quantity of key usability issues detected by either group.

**Cost and Time Efficiency:** The analysis reveals that the crowdsourced approach offers a drastic reduction in resource requirements.

**Cost:** Crowdsourced testing was significantly more economical, costing only 10/user) compared to 64/user) for traditional testing, representing a highly significant difference.

**Time:** While crowdsourced testers completed tasks slightly faster (4.88 total hours vs. 6.07 hours for traditional), the difference was not statistically significant, indicating that both methods provide comparable operational time efficiency.

**LMS Platforms Usability Benchmarking:** The evaluation of three prominent platforms established a clear usability hierarchy based on a "Key Issue per Hour" metric:

**Edredo:** Achieved a "High" usability level, exhibiting the lowest average of key issues relative to the time taken.

**Google Classroom:** Classified as having a "Medium" usability level.

**Canvas Instructure:** Categorized with a "Low" usability level due to having the highest frequency of key issues identified during the evaluation.

In summary, this research provides empirical evidence that crowdsourced usability testing is a viable, cost-effective alternative to traditional laboratory methods, capable of producing high-quality results for global Learning Management Systems without compromising on effectiveness.

**Future Research Directions:**

In the future a comparison of Machine Learning-based usability testing method that extracts usability issues vs crowd-sourced usability testing can be conducted to determine which method is more efficient in terms of cost and time.

In the future direction the researcher can examine the impact of task complexity on usability testing to investigate the influence of task complexity on the result of usability testing. This could help to determine whether user-based or crowd-sourced usability testing is more effective for different level of tasks complexity and provide insight into how to design and structure usability evaluations to maximize their effectiveness.

Researchers can also test the cost-effectiveness and time-efficiency of user-based vs crowd-sourced usability testing using other software systems as test objects to generalize the findings.

Another comparative analyses of user-based vs crowd-sourced usability testing canbe conducted to expand the existing study by comparing the effectiveness, efficiency, and user satisfaction of both methods using a larger sample size and broader range of tasks. This could help to identify the strength and limitation of each method and determine which method is more appropriate for different usability evaluation.

Future research should examine how cross-cultural design factors influence crowdsourced usability results. Since crowdsourcing naturally recruits from a global pool, investigating how cultural backgrounds affect user engagement and perception of LMS interfaces (like UI/UX elements) could provide critical insights for developers of international platforms.

Building on the existing point about Machine Learning, researchers could focus on developing a large-scale, validated dataset of LMS-specific usability problems. Such a resource would be essential for training and testing ML models to automate usability inspections for educational software.

**Author's Contribution: Sadiq Khan:** Investigation, Resources, Data Collection, Writing – Original Draft, Visualization. **Muhammad Nasir:** Conceptualization, Methodology, Validation, Formal Analysis, Supervision, Writing – Review & Editing. **Zakia Jalil:** Conceptualization, Methodology, Validation, Formal Analysis, Writing – Review & Editing, Co-supervision. **Qaisar Afzal:** Validation, Formal Analysis, Data Curation, Writing – Review & Editing.
All authors reviewed, revised, and approved the final manuscript.

**Conflict of interest:** The authors declare that there is no conflict of interest regarding the publication of this manuscript.

**Project details:** Not applicable. The study was conducted as part of an academic research project without external funding.

**References:**

[1] S. Lehong, J. van Biljon, and I. Sanders, "Usability Requirements for Learning Management Systems in Open Distance Electronic Learning Environments: Considering Lecturers' Views on Students' Needs," *Int. J. Human–Computer Interact.*, vol. 40, no. 3, pp. 567–583, 2024, doi: 10.1080/10447318.2022.2121205.

[2] Farhan Hanis Muhmad Asri, Dalbir Singh, "A Review of Cross-Cultural Design to Improve User Engagement for Learning Management System," *KSII Trans. Internet Inf. Syst.*, vol. 18, no. 2, 2024, doi: 10.3837/tiis.2024.02.007.

[3] Abdulaziz Aldiab, Harun Chowdhury, "Utilization of Learning Management Systems (LMSs) in higher education system: A case review for Saudi Arabia," *Energy Procedia*, vol. 160, pp. 731–737, 2019, doi: https://doi.org/10.1016/j.egypro.2019.02.186.

[4] N. Adzharuddin, "Learning Management System (LMS) among University Students: Does It Work?," *Int. J. e-Education, e-Business, e-Management e-Learning*, 2013, doi: 10.7763/IJEEEE.2013.V3.233.

[5] D. G. S. Júnior, R. Hernández-Ramírez, and J. Estima, "Systematic Mapping of Methods Used to Evaluate the Usability and UX of Learning Management Systems," *Springer Ser. Des. Innov.*, vol. 19, pp. 122–133, 2022, doi: 10.1007/978-3-030-89735-2_11/SAVE-RESEARCH.

[6] Adolph Sedem Yaw Adu, Judy Van Biljon, Ronell van der Merwe, "The Influence of Perceived Usability on Learning Management System Acceptance of Lecturers in Ghana: A PLS-SEM Disjoint Two-Stage Approach," *Adv. Human-Computer Interact.*, 2025, [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1155/ahci/9694629

[7] A. Stockless, "Acceptance of learning management system: The case of secondary school teachers," *Educ. Inf. Technol. 2017 233*, vol. 23, no. 3, pp. 1101–1121, Oct. 2017, doi:

10.1007/S10639-017-9654-6.
[8] Khaleel Mershad, Pilar Wakim, "A Learning Management System Enhanced with Internet of Things Applications," *J. Educ. Learn.*, vol. 7, no. 3, p. 23, 2018, doi: 10.5539/jel.v7n3p23.
[9] Reem Sualiman Baragash, Hosam Al-Samarraie, "Blended learning: Investigating the influence of engagement in multiple learning delivery modes on students' performance," *Telemat. Informatics*, vol. 35, no. 7, 2018, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0736585318303848
[10] T. G. Sakova and S. A. Chevereva, "Fundamentals of Choosing an LMS Platform for Distance Learning," *Lect. Notes Networks Syst.*, vol. 161 LNNS, pp. 348–353, 2021, doi: 10.1007/978-3-030-60926-9_45.
[11] N. W. E. A. I Made Ardi Sudestra, "Improving Digital Learning: Evaluating The U Learn LMS with the System Usability Scale," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 94, pp. 2325–2332, 2024, doi: 10.29100/jipi.v9i4.6910.
[12] M. Beasley, "Practical Web Analytics for User Experience: How Analytics Can Help You Understand Your Users," *Elsevier Sci. Inc*, 2013, [Online]. Available: https://www.sciencedirect.com/book/monograph/9780124046191/practical-web-analytics-for-user-experience
[13] "The Definition of User Experience (UX) - NN/G." Accessed: Jun. 09, 2026. [Online]. Available: https://www.nngroup.com/articles/definition-user-experience/
[14] A. Abran, A. Khelifi, W. Suryn, and A. Seffah, "Usability meanings and interpretations in ISO standards," *Softw. Qual. J.*, vol. 11, no. 4, pp. 325–338, 2003, doi: 10.1023/A:1025869312943/METRICS.
[15] "Usability Inspection Methods: Book by Jakob Nielsen - NN/G." Accessed: Jun. 09, 2026. [Online]. Available: https://www.nngroup.com/books/usability-inspection-methods/
[16] S. Rosenbaum, "Usability Evaluations Versus Usability Testing: When and Why?," *IEEE Trans. Prof. Commun.*, vol. 32, no. 4, pp. 210–216, 1989, doi: 10.1109/47.44533.
[17] "Usability engineering at a discount | Proceedings of the third international conference on human-computer interaction on Designing and using human-computer interfaces and knowledge based systems (2nd ed.)." Accessed: Jun. 09, 2026. [Online]. Available: https://dl.acm.org/doi/10.5555/92449.92499
[18] Di Liu, Randolph G. Bias, Matthew Lease, Rebecca Kuipers, "Crowdsourcing for usability testing," *Proc. ASIST Annu. Meet.*, 2013, [Online]. Available: https://asistdl.onlinelibrary.wiley.com/doi/10.1002/meet.14504901100
[19] Anders Bruun, Jan Stage, "New approaches to usability evaluation in software development: Barefoot and crowdsourcing," *J. Syst. Softw.*, vol. 105, pp. 40–53, 2015, doi: https://doi.org/10.1016/j.jss.2015.03.043.
[20] Muhammad Nasir, Naveed Ikram, "Usability inspection: Novice crowd inspectors versus expert," *J. Syst. Softw.*, vol. 183, p. 111122, 2022, doi: https://doi.org/10.1016/j.jss.2021.111122.
[21] M. Nasir, D. N. Ikram, D. Z. Jalil, W. Muhammad, and T. Mikkonen, "A Framework (Crowduix) for Novice Usability Inspection Using Crowdsourcing: A Multi-Case Study," 2025, doi: 10.2139/SSRN.5369276.
[22] "Development and validation of a virtual learning management system usability questionnaire: A case study - PubMed." Accessed: Jun. 09, 2026. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/39177635/
[23] Patricia D. Simon, Juming Jiang, Luke K. Fryer, Ronnel B. King & Cherry E. Frondozo, "An Assessment of Learning Management System Use in Higher Education: Perspectives from a Comprehensive Sample of Teachers and Students," *Technol. Knowl. Learn.*, vol. 30, 2025, [Online]. Available: https://link.springer.com/article/10.1007/s10758-024-09734-5
[24] Javier Mella-Norambuena, Andres Chiappe & Maria Graciela Badilla-Quintana, "Theoretical and empirical models underlying the teaching use of LMS platforms in higher education: a systematic review," *J. Comput. Educ.*, vol. 12, pp. 997–1023, 2025, [Online].

Available: https://link.springer.com/article/10.1007/s40692-024-00336-9
[25] A. Ertürk and M. B. K. Önaçan, "Usability Evaluation of Learning Management System in A Higher Education Institution: A Scale Development Study," *J. Glob. Strateg. Manag.*, vol. 2, no. 10, pp. 73–73, Dec. 2016, doi: 10.20460/JGSM.20161024357.
[26] Ilia Maslov, Shahrokh Nikou, "Exploring user experience of learning management system," *Int. J. Inf. Learn. Technol.*, vol. 38, no. 4, 2021, doi: 10.1108/IJILT-03-2021-0046.
[27] S. M. Jafari, S. F. Salem, M. S. Moaddab, and S. O. Salem, "Learning Management System (LMS) success: An investigation among the university students," *2015 IEEE Conf. e-Learning, e-Management e-Services, IC3e 2015*, pp. 64–69, Feb. 2016, doi: 10.1109/IC3E.2015.7403488.
[28] Dimah Al-Fraihat, Mike Joy, "Evaluating E-learning systems success: An empirical study," *Comput. Human Behav.*, vol. 102, pp. 67–86, 2020, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0747563219302912
[29] "(PDF) Usability Evaluation of Learning Management Systems in Sri Lankan Universities." Accessed: Jun. 09, 2026. [Online]. Available: https://www.researchgate.net/publication/272747770_Usability_Evaluation_of_Learning_Management_Systems_in_Sri_Lankan_Universities
[30] "(PDF) Investigating the Use of Learning Management System (LMS) for Distance Education in Malaysia: A Mixed-Method Approach." Accessed: Jun. 09, 2026. [Online]. Available: https://www.researchgate.net/publication/352524591_Investigating_the_Use_of_Learning_Management_System_LMS_for_Distance_Education_in_Malaysia_A_Mixed-Method_Approach
[31] "(PDF) Evaluating The Usability And Accessibility Of LMS 'Blackboard' At King Saud University." Accessed: Jun. 09, 2026. [Online]. Available: https://www.researchgate.net/publication/298332435_Evaluating_The_Usability_And_Accessibility_Of_LMS_Blackboard_At_King_Saud_University
[32] R. J. M. Ventayen, K. L. A. Estira, M. Guzman, C. M. M. Cabaluna, and N. N. Espinosa, "Usability Evaluation of Google Classroom : Basis for the Adaptation of GSuite E-Learning Platform," 2017.
[33] F. Guaiani and H. Muccini, "Crowd and Laboratory Testing, Can They Co-exist? An Exploratory Study," *Proc. - 2nd Int. Work. CrowdSourcing Softw. Eng. CSI-SE 2015*, pp. 32–37, Jul. 2015, doi: 10.1109/CSI-SE.2015.14.
[34] J. Scholtz, "Adaptation of traditional usability testing methods for remote testing," *Proc. Hawaii Int. Conf. Syst. Sci.*, p. 134, 2001, doi: 10.1109/HICSS.2001.926546.
[35] "Why You Only Need to Test with 5 Users - NN/G." Accessed: Jun. 09, 2026. [Online]. Available: https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/
[36] "(PDF) Evaluating LMS Usability for Enhanced eLearning Experience." Accessed: Jun. 09, 2026. [Online]. Available: https://www.researchgate.net/publication/252615996_Evaluating_LMS_Usability_for_Enhanced_eLearning_Experience
[37] "The LMS moodle: A Usability Evaluation." Accessed: Jun. 09, 2026. [Online]. Available: https://www.researchgate.net/publication/241339520_The_LMS_moodle_A_Usability_Evaluation
[38] Nhu Ty Nguyen, "A study on satisfaction of users towards learning management system at International University – Vietnam National University HCMC," *Asia Pacific Manag. Rev.*, vol. 26, no. 4, pp. 186–196, 2021, doi: https://doi.org/10.1016/j.apmrv.2021.02.001.
[39] "Evaluating Learning Management System Usage at a Small University | Proceedings of the 2019 3rd International Conference on Information System and Data Mining." Accessed: Jun. 09, 2026. [Online]. Available: https://dl.acm.org/doi/10.1145/3325917.3325929
[40] "(PDF) Evaluating the Usability of an E-Learning Platform within Higher Education from a Student Perspective." Accessed: Jun. 09, 2026. [Online]. Available:

https://www.researchgate.net/publication/338619470_Evaluating_the_Usability_of_an_E-Learning_Platform_within_Higher_Education_from_a_Student_Perspective
[41] W. T. Nakamura, E. H. T. De Oliveira, and T. Conte, "Usability and user experience evaluation of learning management systems a systematic mapping study," *ICEIS 2017 - Proc. 19th Int. Conf. Enterp. Inf. Syst.*, vol. 3, pp. 97–108, 2017, doi: 10.5220/0006363100970108.
[42] M. Borys and M. Laskowski, "Expert vs novice evaluators: Comparison of heuristic evaluation assessment," *ICEIS 2014 - Proc. 16th Int. Conf. Enterp. Inf. Syst.*, vol. 3, pp. 144–149, 2014, doi: 10.5220/0004970901440149.
[43] Y. (Chris) Zhao, J. Tang, X. Ye, and J. Lian, "Crowdsourcing in UX Research," *Human-Computer Interact.*, p. vol3:266-vol3:284, Sep. 2024, doi: 10.1201/9781003584292-41.
[44] F. Anggriantaka and D. S. Kusumo, "The Development of an Asynchronous Crowdsourcing-Based User Acceptance Testing Framework and Its Comparison with Conventional Synchronous UAT," pp. 1–7, Dec. 2025, doi: 10.1109/ICOSEIT67010.2025.11290993.
[45] T. Kim and Y. M. Choi, "A Comparison of Laboratory and Synchronous Remote Usability Testing Methods Using AR," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 13302 LNCS, pp. 252–263, 2022, doi: 10.1007/978-3-031-05311-5_17/SAVE-RESEARCH.
[46] G. Desolda, R. Lanzilotti, D. Caivano, M. F. Costabile, and P. Buono, "Asynchronous Remote Usability Tests Using Web-Based Tools Versus Laboratory Usability Tests: An Experimental Study," *IEEE Trans. Human-Machine Syst.*, vol. 53, no. 4, pp. 731–742, Aug. 2023, doi: 10.1109/THMS.2023.3282225.
[47] Yonas Leguesse, Mark Vella, "A Metric to Assess the Reliability of Crowd-sourced SUS Scores: A Case Study on the PoPLar Authentication Tool," *ACM Int. Conf. Proceeding Ser.*, 2024, [Online]. Available: https://dl.acm.org/doi/10.1145/3688459.3688470
[48] Y. Wang, X. Chen, and X. Zhou, "A New Method for Identifying Low-Quality Data in Perceived Usability Crowdsourcing Tests: Differences in Questionnaire Scores," *Int. J. Human–Computer Interact.*, vol. 40, no. 22, pp. 7297–7313, Nov. 2024, doi: 10.1080/10447318.2023.2263694.
[49] R. Molich and J. S. Dumas, "Comparative usability evaluation (CUE-4)," *Behav. Inf. Technol.*, vol. 27, no. 3, pp. 263–281, May 2008, doi: 10.1080/01449290600959062.
[50] "Classroom Management Tools & Resources - Google for Education." Accessed: Jun. 09, 2026. [Online]. Available: https://edu.google.com/intl/ALL_pk/workspace-for-education/products/classroom/
[51] "Canvas by Instructure: World Leading LMS for Teaching & Learning." Accessed: Jun. 09, 2026. [Online]. Available: https://www.instructure.com/canvas
[52] E. Gamboa, R. Galda, C. Mayas, and M. Hirth, "The Crowd Thinks Aloud: Crowdsourcing Usability Testing with the Thinking Aloud Method," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 13094 LNCS, pp. 24–39, 2021, doi: 10.1007/978-3-030-90238-4_3/SAVE-RESEARCH.