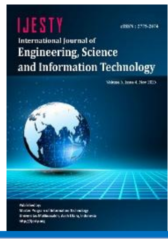
International Journal of Engineering, Science and Information Technology
Volume 5, No. 4 (2025) pp. 597-606
eISSN: 2775-2674
Website: http://ijesty.org/index.php/ijesty
DOI: https://doi.org/10.52088/ijesty.v5i4.1497
Research Paper, Short Communication, Review, Technical Paper



Usability Evaluation of Virtual Reality Metaverse Lab Using Usability Testing With the User-Centered Design Approach

Ketut Sepdyana Kartini[1*], I Made Marthana Yusa[1], I Nyoman Widhi Adnyana[1], I Nyoman Tri Anindia Putra[2]

[1]Institut Bisnis dan Teknologi Indonesia, Bali, Indonesia
[2]Universitas Pendidikan Ganesha, Bali, Indonesia

**Corresponding author Email: sepdyana@instiki.ac.id*

The manuscript was received on 1 April 2025, revised on 16 August 2025, and accepted on 12 November 2025, date of publication 16 December 2025

Abstract

Advances in educational technology have introduced immersive tools such as Virtual Reality (VR) to support complex learning, especially in abstract fields like chemistry. Metaverselab was developed to provide a safe virtual environment for conducting laboratory experiments, yet early implementation revealed substantial usability issues. Students experienced difficulties with the interface, confusion in navigation, and insufficient guidance, which affected the overall learning experience. This study aims to evaluate and improve the usability and user experience of the Metaverselab platform using a User-Centered Design (UCD) approach. The research applied an iterative process involving 40 student respondents and used two standardized instruments: the System Usability Scale (SUS) and the User Experience Questionnaire (UEQ). Both instruments were administered during the initial and final evaluation stages to measure changes in usability and user experience after improvements were introduced. Initial results showed low usability, with an average SUS score of 50.81 and UEQ benchmark values categorized as "Bad" in all dimensions. Based on these findings, specific user requirements were identified and translated into targeted design solutions. Key improvements included an interactive tutorial for first-time users, descriptive information pop-ups for laboratory tools, and undo–redo functions to support error recovery during experiments. Post-implementation testing demonstrated substantial improvement. The average SUS score increased by 44.16% to 73.25, placing the system in the "Good" usability category. UEQ results also improved significantly, with the Efficiency dimension rated "Excellent," while Attractiveness, Perspicuity, and Stimulation were rated "Good." These results confirm that the UCD approach effectively identifies user needs and produces design interventions that enhance the usability, efficiency, and overall learning experience of virtual educational platforms.

Keywords: *Metaverse Lab, System Usability Scale, User-Centered Design, User Experience Questionnaire, Virtual Reality.*

1. Introduction

Advances in educational technology have created new opportunities for more immersive learning experiences, one of which is through the use of virtual reality [1]. In the context of science education, particularly chemistry, laboratory practical activities are an important aspect of reinforcing understanding of abstract concepts that are difficult to grasp based on theory alone [2] [3]. However, conducting real laboratory experiments often faces various obstacles, such as limited laboratory facilities, high operational costs, and safety risks due to the use of hazardous chemicals [4]. Metaverselab is an innovative alternative in chemistry learning through virtual reality (VR) technology. This platform is designed to help students understand concepts and practice laboratory experiments virtually [5] [6]. This technology offers three-dimensional visualization, interactivity, and a learning experience that resembles real conditions in a laboratory. Students can manipulate objects, observe reactions, and receive immediate feedback, thereby increasing student engagement and improving learning outcomes [7] [8]. During implementation, the use of this virtual reality platform did not always run smoothly. Some students still had difficulty adapting to the interface, confusion in navigation, and limitations in interactive features that could reduce learning comfort. This indicates a problem with the usability of the system, which has the potential to cause an unsatisfactory user experience[9]. An unusable system can also hinder student engagement, reduce learning motivation, and even make students reluctant to use the technology in the long term. In other words, without usability evaluation, the development of the metaverse lab risks not providing maximum benefits for students as the primary users [10][9].

The study was conducted by [11], who developed a Virtual Reality (VR)-based serious game for fire evacuation training. The evaluation was carried out using the System Usability Scale (SUS) and User Experience Questionnaire (UEQ) to assess the level of usability and user experience. The results showed that the use of VR was able to provide a more immersive and effective learning experience compared to conventional methods. These findings confirm that VR is not only useful as a simulation medium, but can also improve the quality of learning through user-centered design. [8] then researched the effect of using Augmented Reality (AR) on learning motivation and understanding of acid-base titration concepts in high school students. Using SUS and UEQ, this study evaluated the ease of use and experience of students when using AR applications. The results showed a significant increase in motivation and understanding of chemistry concepts, emphasizing the importance of usability in optimizing the effectiveness of immersive technology-based learning media [12]. A study conducted by [13] analyzed the ticketing and knowledge base application interface using the 8 Golden Rules of Interface Design theory. Evaluating interfaces using design principles successfully identifies critical issues like navigation consistency and feedback efficiency. Building on this, applying User-Centered Design (UCD) is crucial for comprehensive, student-focused evaluation. Consequently, this study evaluates the Virtual Reality Metaverse Lab using UCD to diagnose specific usability problems [14]. The aim is to provide clear insights and development recommendations that effectively support the student learning process.

2. Literature Review

2.1 Virtual Reality in Education

The use of Virtual Reality (VR) in education, especially for sciences such as chemistry, is a solution to overcome the limitations of physical laboratories, such as high costs, safety risks, and limited access [15] [16]. This immersive technology allows students to visualize abstract concepts in three dimensions and conduct experiments repeatedly in a safe environment, thereby supporting deeper understanding and independent learning processes [17].

2.2 The Importance of Usability in Learning Applications

The effectiveness of VR technology as a learning medium is highly dependent on its ease of use. Applications with low usability can cause frustration and hinder the learning process[18]. Therefore, evaluating ease of use is essential to ensure that the application truly supports learning objectives and provides a positive user experience [19] [20].

2.3 User Centered Design

To ensure that technology products have a high level of usability, the User Centered Design (UCD) approach is an essential development philosophy. UCD is an iterative design process that focuses on the needs, goals, and limitations of users in every phase of product development [21]. This approach places users at the center of the design process, from analyzing the context of use, identifying needs, designing solutions, to evaluation [14]. By actively involving users, developers can gain a deep understanding of how the product will be used in real-world scenarios, enabling them to create solutions that are not only functional but also intuitive and satisfying for end users.

2.4 Usability Evaluation

Usability evaluation was conducted using the System Usability Scale (SUS) and User Experience Questionnaire (UEQ) [9] [10]. The SUS is a simple questionnaire containing 10 questions that produce a single score to assess the overall usability of the system, while the UEQ provides a more comprehensive evaluation through six main scales: Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty, enabling it to measure not only classic usability aspects but also emotional and hedonic user experiences [22]. The combination of these two instruments provides a holistic picture, where SUS shows the general level of usability, while UEQ deepens the analysis of the quality of the user experience, as has also been applied in various VR-based application studies[23].

3. Methods

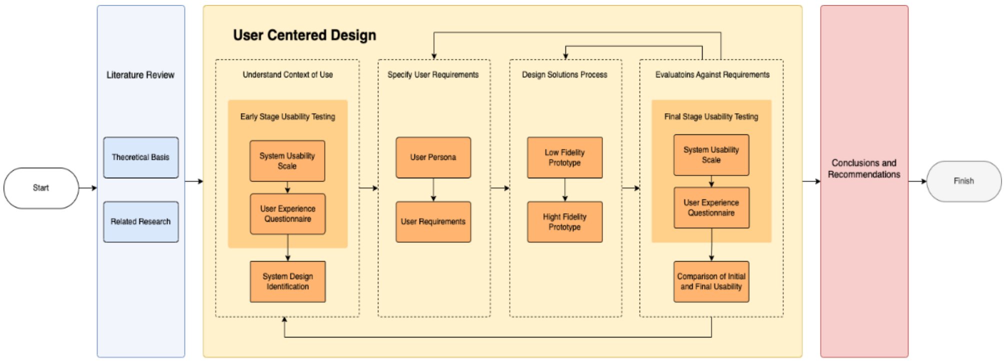


Fig 1. Research Methodology

This study uses the User Centered Design (UCD) approach, a design philosophy that places users at the center of every stage of the development process. This method consists of the stages of understanding the context of use, determining user requirements, designing the solution process, and evaluating against the requirements [24]. This methodology was chosen because of its focus on a deep understanding of user needs, goals, and limitations, which are essential for creating products with high usability. The UCD cycle in this study consists of several main stages. The MetaverseLab system was implemented using the Unity 3D engine and deployed as a standalone virtual reality application. The platform was accessed using Meta Quest head-mounted displays, enabling students to interact with the virtual laboratory without requiring high-performance desktop computers. User interaction was primarily controller-based, allowing object selection, manipulation, and navigation through trigger and ray-casting mechanisms. To ensure user comfort and reduce VR-related discomfort, the system was optimized to maintain a stable frame rate above 72 FPS with lightweight 3D assets and optimized rendering pipelines.

The literature review stage was conducted to build a strong theoretical foundation for the research by reviewing relevant literature and previous studies. The focus of the study included the application of Virtual Reality (VR) technology in education, particularly virtual laboratories; the principles and methodology of User Centered Design (UCD) and its implementation in software development; the use of usability evaluation instruments such as the System Usability Scale (SUS) and User Experience Questionnaire (UEQ), including score interpretation and validity; and usability evaluation case studies on other educational applications to identify best practices and common challenges encountered [23].

3.1. Understand Context of USe

The first stage in the UCD cycle is to gain a deep understanding of how users interact with the product in a real-world context [24]. To achieve this goal, initial usability testing sessions were conducted on the first version of the Metaverse Lab platform. These sessions involved 40 students as respondents, who were the main target users. The purpose of this stage was to collect baseline data on usability and user experience, as well as to identify the main problems encountered by users when performing virtual lab tasks [25]. The instruments used at this stage were the SUS questionnaire for general usability measurement and the UEQ questionnaire for a more detailed analysis of user experience [26]. The participants were undergraduate students enrolled in a chemistry-related program and had prior experience with conventional laboratory procedures. However, prior exposure to virtual reality technology varied among participants, with most respondents having limited or no experience using VR for educational purposes. The participants were aged between 18 and 22 years, with a relatively balanced gender distribution. These characteristics may influence usability perception and are therefore acknowledged as potential limitations of the study.

3.2. Specify User Requirements

Quantitative and qualitative data obtained from the previous stage are thoroughly analyzed to define specific user requirements [24]. Low SUS scores and “Poor” UEQ ratings are further analyzed and translated into a list of concrete functional and non-functional requirements. This process involves identifying the issues most frequently experienced by users, such as navigation difficulties, interface confusion, and lack of system feedback. The result of this stage is a user requirements document that serves as the main guide for the system redesign process [27]. User requirements were derived by clustering recurring usability issues based on their frequency and severity as observed during usability testing sessions. Although formal techniques such as affinity diagramming or card sorting were not explicitly applied, the requirement extraction process followed a systematic interpretation of both quantitative usability scores and qualitative user feedback.

3.3. Design Solutions Process

Based on the list of user requirements that have been specified, the next step is to design and develop solutions to address existing problems [24]. This process is iterative, which may involve creating interface prototypes, redesigning interaction flows (UX), and refining visual elements (UI)[28]. Some concrete solutions implemented in this study include the addition of an interactive tutorial module, simplification of the tool and material selection menu, and improvement of the visual and auditory feedback system for each user action [29]. All of these design changes are specifically designed to meet the requirements identified in the previous stage [30].

3.4. Evaluations Against Requirements

After the design solution was implemented on the Metaverse Lab platform, a second evaluation stage was conducted. This stage aimed to validate whether the changes made had successfully improved usability and user experience, as well as met the established requirements [24]. The reliability of the usability evaluation instruments was supported by prior validation studies of both SUS and UEQ. Internal consistency of the UEQ scales was confirmed using Cronbach’s alpha, which indicated acceptable reliability across all dimensions ($\alpha > 0.7$). To confirm usability improvement, a paired comparison between pre-test and post-test SUS scores was conducted using a Wilcoxon signed-rank test, revealing a statistically significant increase in usability scores ($p < 0.05$). The second usability testing session was conducted with the same respondents and task scenarios as in the initial evaluation. The SUS and UEQ instruments were again used to collect data. The results of this second evaluation were then directly compared with the results of the first evaluation to measure the success rate of the redesign process that had been carried out [31]. This study employed a single major iteration of the UCD cycle, consisting of an initial evaluation, design refinement, and final usability assessment. While multiple iterations may provide deeper insights, the single-cycle approach was sufficient to demonstrate measurable usability improvements within the study scope.

3.5. Conclusions and Recomendations

The final stage of this methodological cycle is to analyze the results of the comparison between the initial and final evaluations. From this analysis, conclusions are drawn regarding the effectiveness of the UCD approach in improving the usability of Metaverse Lab. In addition, at this stage, recommendations are also formulated for future platform development. These recommendations may include suggestions for further improvements, the addition of new features, or other areas that can still be optimized based on user feedback during the evaluation process [21].

4. Result and Discussions

4.1. Understand Context of Use

This stage aims to identify usability issues in the initial version of Metaverse Lab. Testing was conducted on 40 respondents to collect quantitative data on their perceptions and experiences. The results of this evaluation will serve as the basis for system design improvements in the next stage.

4.1.1. System Usability Scale

The initial measurement of system usability was measured with 40 respondents using the SUS method. The weight of the SUS statements ranged from 1 to 5, with the weight of odd-numbered statements converted by removing the value 1 and the weight of even-numbered statements converted by removing the value 5. A summary of the scores in the initial SUS can be seen in Table 1.

Table 1. Summary Scores Results Early Stage

Statistics	Values
Number of Respondents	40
Minimum SUS Score	27.5

Maximum SUS Score	72.5
Average SUS Score	50.81
Usability Category	Poor, Below Acceptable

Based on the results of the initial System Usability Scale (SUS) test on 40 respondents in Table 1, the average score was 50.81. This score is still below the threshold of 70, which is generally used as the standard for a system to be considered as having a good level of usability. Looking at the distribution of respondent scores, there is quite a wide variation, ranging from a low score of 27.5 (R3) to a high score of 72.5 (R39 and R40). This shows that the user experience with the application is still inconsistent: some respondents rated it quite well, but many also still experienced difficulties or obstacles in using it [32]. In general, the initial SUS scores indicate that the application still needs significant improvement, especially in terms of ease of use, navigation consistency, and supporting features such as tutorials and tool information, in order to increase user satisfaction and effectiveness in the virtual learning process.

4.1.2. User Experience Questionnaire

The next usability test was conducted again using the UEQ method with the same procedure and 40 respondents as in the initial stage. The UEQ instrument used consisted of 26 items on a 1-7 Likert scale [33]. The statistical results, such as the mean, variance, and standard deviation from this stage, are presented in Table 2.

Table 2. Mean, Variance, and Standard deviation Early Stage

UEQ Scale	Mean	Variance	Std. Dev.	Description
Attractiveness	0.66	2.33	1.53	Low appeal, users perceived the system as relatively unattractive or unpleasant.
Perspicuity	0.54	2.01	1.42	Users found the system moderately difficult to understand and learn.
Efficiency	0.02	2.22	1.49	Perceived efficiency was very low, indicating slow or effortful interaction.
Dependability	0.07	2.20	1.48	Users felt the system was inconsistent and not fully reliable.
Stimulation	0.13	2.28	1.51	Low motivation and low excitement when using the system.
Novelty	-0.33	2.43	1.56	Users perceived the system as not innovative and lacking creativity.

The initial results of the UEQ with 26 statements summarized in Table 2 show that the user experience is not yet optimal. Although items related to attractiveness and clarity received positive responses, especially “enjoyable” (average 1.6) and “easy to understand” (average 1.8), the high variance indicates inconsistent user perceptions. Conversely, significant negative evaluations emerged regarding uniqueness and reliability. Items such as “creative” (average –1.4) and “motivating” (average –1.8) indicate that the system is considered less innovative and less motivating. Overall, while performance in appeal and clarity is still acceptable, the negative trend in uniqueness and reliability highlights critical weaknesses [34]. These findings emphasize the urgent need for targeted improvements in the next development phase to significantly enhance user engagement and system reliability.

Table 3. UEQ Scales Early Stage

UEQ Scales (Mean and Variance)		
Attractiveness	0,158	0,30
Perspicuity	0,113	0,65
Efficiency	0,050	0,36
Dependability	-0,088	0,78
Stimulation	-0,088	0,54
Novelty	-0,038	0,59

Based on the results of the initial UEQ scale in Table 3, it can be seen that most of the scales are still in the low category. The Attractiveness scale (mean 0.158; variance 0.30), Perspicuity scale (mean 0.113; variance 0.65), and Efficiency scale (mean 0.050; variance 0.36) obtained positive scores, although the values are still very small, indicating that the application is considered somewhat attractive, fairly easy to understand, and slightly efficient. Conversely, Dependability (mean –0.088; variance 0.78), Stimulation (mean –0.088; variance 0.54), and Novelty (mean –0.038; variance 0.59) had negative values, indicating that the application was still unreliable, not sufficiently motivating, and lacked novelty[35]. Overall, these results indicate that in the early stages, the user experience tends to be neutral to negative, with considerable variation in responses (high variance values on several scales). This shows differences in perception among respondents and confirms the need for significant improvements, especially in terms of reliability, stimulation, and innovation [36].

Table 4. Benchmark Results Early Stage

Scale	Mean	Comparisson to benchmark	Interpretation
Attractiveness	0,16	Bad	In the range of the 25% worst results
Perspicuity	0,11	Bad	In the range of the 25% worst results
Efficiency	0,05	Bad	In the range of the 25% worst results
Dependability	-0,09	Bad	In the range of the 25% worst results
Stimulation	-0,09	Bad	In the range of the 25% worst results
Novelty	-0,04	Bad	In the range of the 25% worst results

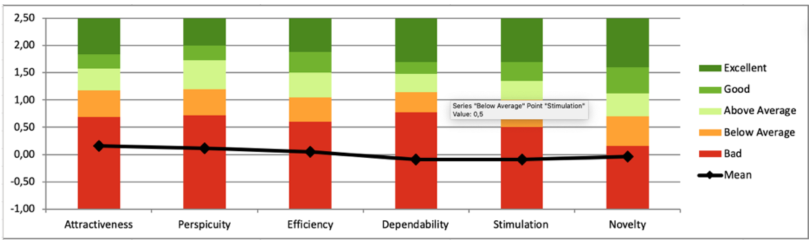


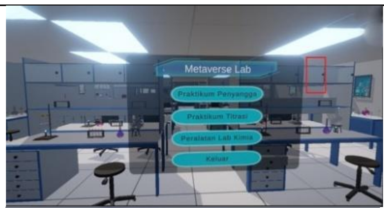
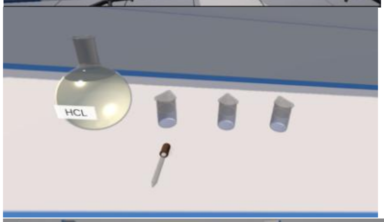
Fig 2. Benchmark Graph Results

Based on the initial UEQ testing in Table 3, average scores across all scales were relatively low. While Attractiveness (0.16), Perspicuity (0.11), and Efficiency (0.05) showed positive but minimal values, Dependability (–0.09), Stimulation (–0.09), and Novelty (–0.04) were negative, indicating the application lacked reliability and motivation. High variance values further highlighted significant inconsistencies in user perception [37]. Crucially, the benchmark analysis in Table 4 and Figure 2 categorized all dimensions as "Bad," placing the application in the bottom 25% of results globally [38]. These findings confirm that the initial usability and user experience were significantly deficient, underscoring the urgent need for comprehensive improvements to reach acceptable standards in the subsequent evaluation.

4.1.3. System Design Identification

The next step is to conduct a system design analysis based on the 8 Golden Rules. This analysis aims to formulate recommendations for improvements to the user interface [13]. The 8 Golden Rules theory can be used as a consideration and reference in assessing the design and performance of the system. Based on this analysis, several design aspects were identified that did not comply with the 8 Golden Rules.

Table 5. System Design Identification

No	Rules	Problems	Figures
1	Reduce Short-Term Memory Load	The virtual reality start menu does not contain instructions on how to use the metaverse lab for chemistry experiments.	
2	Dialog Closure	When the virtual lab does not display information on the virtual chemical tools.	
3	Reversal of Action	There is no Undo feature when an error occurs during a virtual lab session.	

4.2. Specify User Requirements

This stage focuses on identifying the obstacles experienced by users, represented by 40 respondents. The findings from each respondent are abstracted into user personas. Next, the analysis of the obstacles and recommendations for improvement from each persona are synthesized into a user requirements document.

4.2.1. User Persona

Fig 3. User Personas

4.2.2. User Requirements

Table 6. User Requirements

No.	User Requirements	Description
1	Tutorial menu feature	Users need an introductory tutorial before conducting practical work so that they are not confused when using the application.
2	Information on virtual tools	Every chemical tool in the virtual laboratory must display its name, function, and benefits to support student understanding.
3	Undo and redo button	Users need the ability to undo or redo actions so they can learn from their mistakes during practical work.
4	Clear navigations and instructions	The application should provide simpler, clearer, and easier-to-understand instructions and navigation for users.
5	Confirmation at each step of the practicum	The application must clearly display every confirmation provided by the user.

4.3. Design Solution Process

The design solution process stage involves compiling recommendations for design improvements based on problems identified in initial usability testing and feedback provided by respondents. At this stage, the focus is on overcoming obstacles experienced by users in order to improve their interaction experience [39]. One important improvement that was implemented was the addition of a tutorial menu feature before entering the main practicum in MetaverseLab. This was because in the initial version, the absence of a tutorial menu had the potential to cause confusion for students when using the virtual laboratory for the first time. With the tutorial menu, users can understand the flow of application usage, navigation, and practicum steps in stages so that the learning process becomes easier, more focused, and more effective.

Fig 4. Recommendations for Improving the Tutorial Menu at the Beginning of the Practicum

The second improvement addresses the lack of functional information in the virtual chemistry practicum. Previously, laboratory equipment did not display explanations of its purpose, making it difficult for students to use the tools correctly. To solve this, interactive labels or pop-up descriptions were added, allowing students to easily understand each tool's function and improving overall learning effectiveness.

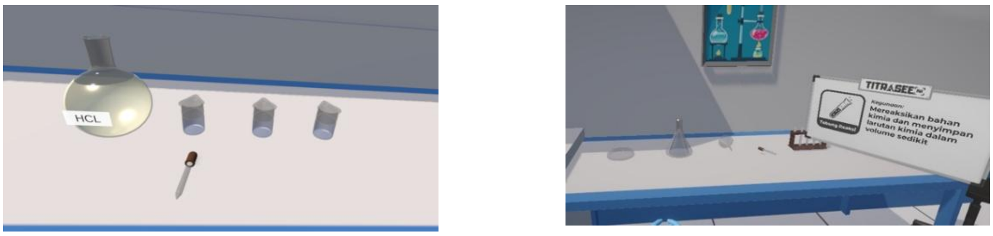


Fig 5. Recommendations for Improving ToolIinformation Usage

The third improvement addresses the lack of undo and redo buttons in the virtual lab. Without these features, students cannot correct mistakes or repeat steps during experiments, which limits the learning process. Adding undo and redo allows students to freely explore, correct errors, and engage in a more interactive and effective learning experience.

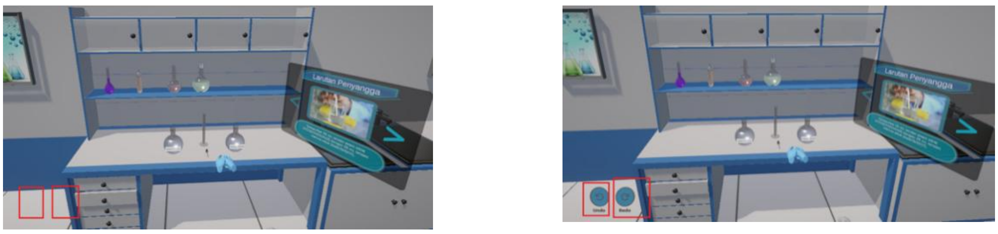


Fig 6. Recommendations for Improving the Undo and Redo Features

4.4. Evaluations Against Requirements

The final stage aims to evaluate the usability results after design improvements have been made. This test again involved the same 40 respondents as in the initial stage, using similar evaluation methods, namely the System Usability Scale (SUS) and User Experience Questionnaire (UEQ), so that the results before and after the improvements could be compared objectively [9] [10].

4.4.1. System Usability Scale

The final measurement of virtual reality usability was measured using the System Usability Scale with 40 respondents. The SUS statement weights ranged from 1 to 5, with the conversion of odd statement item weights eliminating the value 1 and even statement item weights eliminating the value 5. A summary of the scores in the final stage SUS can be seen in Table 7.

Table 7. Summary Scores Results Final Stage

Statistics	Values
Number of Respondents	40
Minimum SUS Score	52.5
Maximum SUS Score	85
Average SUS Score	73.25
Usability Category	Good, Acceptable

Based on the final testing with 40 respondents, the system achieved an average SUS score of 73.25. This score surpasses the standard threshold of 68, classifying the usability as "Good" and acceptable [32]. Scores ranged from 52.5 to 85, indicating that while minor obstacles persist, the majority perceive the system as easy to use. These results confirm that design improvements specifically the tutorial menu, tool information, and undo/redo features effectively enhanced usability. This aligns with previous research, demonstrating that interactive navigation and learning support features contribute significantly to improving learnability and user experience.

4.4.2. User Experience Questionnaire

The next usability stage is to test usability using the UEQ method. This test is conducted using the same procedure as in the initial stage, which is to distribute questionnaires to the same 40 respondents. The UEQ instrument consists of 26 statement items with a Likert scale of 1 to 7. The results of the mean, variance, and standard deviation calculations at this stage can be seen in Table 8.

Table 8. Mean, Variance, and Standard deviation Final Stage

UEQ Scale	Mean	Variance	Std. Dev.	Description
Attractiveness	1.80	1.77	1.33	Users perceived the system as attractive, pleasant, and visually improved.
Perspicuity	1.78	2.18	1.48	The system became easier to understand and learn, with clearer interaction flow.
Efficiency	1.90	1.63	1.28	Users experienced faster, smoother, and more efficient interaction.
Dependability	1.33	2.03	1.42	The system was perceived as more predictable, secure, and reliable.
Stimulation	1.45	2.08	1.43	Users felt more motivated and engaged while using the system.
Novelty	1.13	2.35	1.53	The system appeared more innovative and creative after improvements.

Based on the results of the final UEQ testing in Table 8, the mean values of the 26 items show a positive trend in all aspects of usability. The dimensions of Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty obtained scores ranging from good to very good, with mean values mostly close to 2 (positive). The relatively low variance and standard deviation indicate that the respondents' answers were quite consistent [34]. This indicates that after design improvements, the application is considered easier to use, enjoyable, practical, and supportive of the learning process.

Table 9. UEQ Scales Final Stage

UEQ Scales (Mean and Variance)		
Attractiveness	1,796	1,20
Perspicuity	1,763	1,47
Efficiency	1,888	1,03
Dependability	1,325	0,76
Stimulation	1,438	1,35
Novelty	1,106	1,30

Based on the final UEQ results in Table 9, it can be seen that all dimensions received positive scores. The dimension with the highest score was Efficiency (mean = 1.888), which indicates that the application is considered efficient and supports practical activities well. This was followed by Attractiveness (1.796) and Perspicuity (1.763), which indicate that the application is considered attractive and easy to understand. Meanwhile, the lowest scores were obtained in the Novelty (1.106) and Dependability (1.325) dimensions. This shows that although the system is considered quite innovative and reliable, there is still room for improvement, especially in providing a newer experience and increasing user trust [36]. Overall, these scores show that the application is in the positive category across all UEQ dimensions, so it can be concluded that design improvements have had a significant impact on the quality of the user experience.

Table 10. Benchmark Results Final Stage

Scale	Mean	Comparisson to benchmark	Interpretation
Attractiveness	1,80	Good	10% of results better, 75% of results worse
Perspicuity	1,76	Good	10% of results better, 75% of results worse
Efficiency	1,89	Excellent	In the range of the 10% best results

Dependability	1,33	Above Average	25% of results better, 50% of results worse
Stimulation	1,44	Good	10% of results better, 75% of results worse
Novelty	1,11	Above Average	25% of results better, 50% of results worse

Compared to similar VR-based educational studies, the substantial improvement in efficiency reflects the effectiveness of simplifying interaction flow and providing guided tutorials. Efficiency showed greater improvement than novelty and dependability because the redesign focused primarily on optimizing task execution rather than introducing fundamentally new system features. From an educational usability perspective, achieving high efficiency and clarity is considered more critical than novelty, particularly for learning-oriented VR environments.

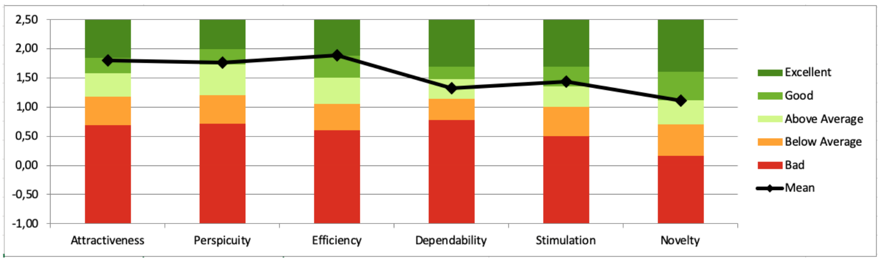


Fig 7. Benchmark Graph Results

Based on the final UEQ benchmark results in Table 10 and Figure 7, most dimensions fall into the "Good" to "Excellent" categories. The Efficiency dimension achieved the highest score of 1.89, earning an "Excellent" rating top 10%, demonstrating that users find the application highly efficient. Additionally, Attractiveness (1.80), Perspicuity (1.76), and Stimulation (1.44) were categorized as "Good," indicating the system is engaging and comprehensible. Meanwhile, Dependability (1.33) and Novelty (1.11) reached the "Above Average" level. While adequate, these specific scores suggest that reliability and innovation remain key areas for potential optimization. Overall, the application demonstrates strong usability with efficiency as its primary strength, though further development is recommended to further bolster system dependability and novelty [38].

4.4.3. Comparison of Initial and Final Usability

Table 11. Usability Average SUS Scores Comparison

Usability Average SUS Scores		
Preliminary Testing	Final Testing	Improvement
50,81	73,25	44,16%

Table 12. Usability UEQ Scores Comparison

Usability UEQ Scores			
UEQ Scale	Preliminary Testing	Final Testing	Improvement
Attractiveness	0,16	1,80	+1,64
Perspicuity	0,11	1,76	+1,65
Efficiency	0,05	1,89	+1,84
Dependability	-0,09	1,33	+1,42
Stimulation	-0,09	1,44	+1,53
Novelty	-0,04	1,11	+1,15

A comparison between the initial and final usability tests was conducted to measure the improvement after implementing UCD-based design changes [40]. The SUS score increased from 50.81 to 73.25, with an increase of 44.16%, indicating a significant improvement in the overall usability of the system, as shown in Table 11. Similarly, all UEQ dimensions experienced significant improvements, with the highest increase recorded in Efficiency (+1.84), followed by Clarity (+1.65) and Attractiveness (+1.64). Stimulation, Reliability, and Uniqueness also showed notable increases, indicating that the updated system is considered more attractive, reliable, and innovative, as shown in Table 12 [41].

5. Conclusion

Although learning outcomes were not quantitatively measured, improved usability is closely associated with enhanced learning experiences. Several participants reported increased confidence and engagement during virtual experiments after the redesign. This suggests that improved usability may indirectly support better conceptual understanding by allowing students to focus on learning tasks rather than interface navigation. This study is subject to several limitations. The evaluation was conducted at a single institution using specific VR hardware, which may influence generalizability. The experimental scenarios were limited to selected chemistry laboratory activities, and long-term learning effects were not assessed. Future research should involve multi-institutional samples, comparative studies with real laboratory environments, extended usage periods, and integration of learning performance metrics. This study demonstrates the effectiveness of the User-Centered Design (UCD) approach in enhancing the Metaverse Lab. Initial evaluations revealed significant usability issues, evidenced by a low System Usability Scale (SUS) score of 50.81 and "Bad" User Experience Questionnaire (UEQ) ratings. Following the implementation of targeted solutions including interactive tutorials, tool descriptions, and undo/redo functionalities performance improved substantially. The final evaluation recorded a 44.16% increase in the SUS score to 73.25, classifying usability as "Good". Furthermore, UEQ scores rose dramatically, with Efficiency reaching "Excellent". These results confirm that the iterative UCD process successfully transformed the Metaverse Lab into a highly effective and satisfying educational tool for chemistry learning.

Acknowledgements

The authors gratefully acknowledge the financial support from the Kementerian Pendidikan Tinggi, Sains, dan Teknologi of the Republic of Indonesia under the Fundamental Research Grant Scheme Contract Nos: 0419/C3/DT.05.00/2025, 129/C3/DT.05.00/PL/2025, 2166/LL8/AL.04/2025, and 010/INSTIKI.R1.D2/PM.03/06.2025. This funding enabled the successful completion of the study titled "Evaluation of Metaverse Lab Interactivity on Student Engagement and Learning Outcomes in Virtual Chemistry Laboratory Simulations". We appreciate the Ministry's commitment to advancing educational research. Additionally, we extend our sincere thanks to the students and faculty of **SMA Negeri 3 Mengwi** for their invaluable participation and cooperation.

References

[1] I. N. T. A. Putra, A. Yuniarti, H. Fabroyir, and I. P. B. G. P. Raharja, "Transformer Performance Evaluation in 3D Reconstruction of Balinese Mask Wood Carving," in *2024 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, 2024, pp. 285–289.

[2] I. N. T. A. Putra, K. S. Kartini, P. S. U. Putra, I. N. W. Adnyana, and N. K. N. N. Pande, "Design and Development of Interactive Media Application Based on Android Case Study of Hydrocarbon Chemical Lesson Materials," in *2021 6th International Conference on New Media Studies (CONMEDIA)*, IEEE, 2021, pp. 113–117.

[3] Y. Zhang and K. Choi, "Game Flow and Attitude in Computer-Based Business Simulations: Implications For Management Education," *Journal of Applied Learning & Teaching*, vol. 8, no. 2, Jul. 2025, doi: 10.37074/jalt.2025.8.2.8.

[4] L. Amaral, J. Santos, F. Soares, J. Felix, L. Cardoso, and T. H. Nascimento, "Interactive Educational Game for Chemistry in Virtual Reality with User Gesture Interactions via Smartwatches," in *Proceedings of the 26th Symposium on Virtual and Augmented Reality*, 2024, pp. 294–298.

[5] I. N. T. A. Putra, A. Yuniarti, and H. Fabroyir, "BaliMask3D Dataset for 3D Completion and Reconstruction of Traditional Balinese Masks," *Sci Data*, vol. 12, no. 1, p. 1233, 2025.

[6] A. Agfary, M. Muhallim, and S. Paembonan, "Augmented Reality Media Pembelajaran Reproduksi Tumbuhan Pada Siswa Kelas IX Sekolah Menengah Pertama (SMP) di SMP Negeri 8 Kota Palopo," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 1, 2025.

[7] E. Ay Kaya and C. Yangın Ersanlı, "Metaverse Effect: Exploring Students' Vocabulary Achievement, Retention, and Perceptions in Virtual Environments," *Journal of Metaverse*, vol. 5, no. 2, pp. 92–103, Apr. 2025, doi: 10.57019/jmv.1557771.

[8] K. S. Kartini, I. M. M. Yusa, I. N. W. Adnyana, and I. N. T. A. Putra, "Impact of Augmented Reality on High School Students' Motivation and Understanding of Acid-Base Titration Concepts," *Edelweiss Applied Science and Technology*, vol. 8, no. 6, pp. 7559–7569, 2024.

[9] A. S. D. Martha, S. Widowati, D. P. Rahayu, M. I. Nursyawal, and R. R. Hariz, "Examining Usability and Student Experience of ChatGPT as a Learning Tool on Moodle in Higher Education," in *2024 International Conference on Information Technology Systems and Innovation (ICITSI)*, IEEE, 2024, pp. 8–13.

[10] B. Sousa, M. Marques, and L. Pombo, "EDUCITY app user experience: School students' perspective using UEQ," in *EDULEARN24 Proceedings*, IATED, 2024, pp. 3392–3401.

[11] R. Siregar, A. Nasikun, and R. Ferdiana, "Immersive Virtual Reality-based Serious Game for Fire Drill Education," in *2024 16th International Conference on Information Technology and Electrical Engineering (ICITEE)*, IEEE, 2024, pp. 358–363.

[12] P. Acevedo, A. J. Magana, B. Benes, and C. Mousas, "A Systematic Review of Immersive Virtual Reality in STEM Education: Advantages and Disadvantages on Learning and User Experience," *IEEE Access*, vol. 12, pp. 189359–189386, 2024, doi: 10.1109/ACCESS.2024.3489233.

[13] R. Afandi and H. B. Santoso, "Improving the Usability and User Experience of a Ticketing and Knowledge Base Web Application: A Case Study of the ESDM 136 Contact Center," *The Indonesian Journal of Computer Science*, vol. 14, no. 1, 2025.

[14] F. A. F. Illahi, K. A. Laksitowening, and R. Nurtantyana, "User-Centered Design Approach in Mobile AR: Application for Hydrocarbon Visualization in Chemistry Course," *JURNAL INFOTEL*, vol. 17, no. 1, pp. 38–53, 2025.

[15] C. S. Chew, M. N. S. Zafarin, N. H. M. Zain, R. Aminuddin, T. G. Tan, and K. O. Chin, "Transforming Secondary School Chemistry Learning with Virtual Reality," in *2024 IEEE 22nd Student Conference on Research and Development (SCOReD)*, IEEE, 2024, pp. 482–487.

[16] R. van Dinther, L. de Putter, and B. Pepin, "Features of Immersive Virtual Reality to Support Meaningful Chemistry Education," *J Chem Educ*, vol. 100, no. 4, pp. 1537–1546, 2023.

[17] C.-H. Wang, "Education in the Metaverse: Developing Virtual Reality Teaching Materials for K–12 Natural Science," *Educ Inf Technol (Dordr)*, vol. 30, no. 7, pp. 8637–8658, 2025.

[18] E. Zilles Borba, A. G. Corrêa, R. de Deus Lopes, and M. Zuffo, "Usability in Virtual Reality: Evaluating User Experience With Interactive Archaeometry Tools in Digital Simulations," *Multimed Tools Appl*, vol. 79, no. 5, pp. 3425–3447, 2020.

[19] F. Tzortzoglou, S. Vosinakis, A. Gardeli, and A. Sofos, "Interaction Design in Mobile Augmented Reality for Education: Insights from a Systematic Review," in *Proceedings of the 3rd International Conference of the ACM Greek SIGCHI Chapter*, 2025, pp. 212–221.

[20] M. S. van de Warenburg, C. Kamphuis, S. Hummelink, D. J. O. Ulrich, and M. L. A. W. Vehmeijer-Heeman, "Augmented Reality for Medical Education in the Primary Survey of Burns: an Exploratory Study," *Injury*, p. 112747, 2025.

[21] A. Mirabdolah, M. Alaeifard, and A. Marandi, "User-Centered Design in HCl: Enhancing Ssability and Interaction in Complex Systems," *International Journal of Advanced Human Computer Interaction*, vol. 1, no. 1, pp. 16–33, 2023.

[22] B. Laugwitz, T. Held, and M. Schrepp, "Construction and evaluation of a user experience questionnaire," in *Symposium of the Austrian HCI and usability engineering group*, Springer, 2008, pp. 63–76.

[23] S. Gelbvaks, C. D. A. Bersot, and J. E. G. Pereira, "Usability and Educational Potential of a Virtual Reality Software for Training in a Target-Controlled Infusion Pump Handling: a Pilot Study," *Perioperative Anesthesia Reports*, vol. 3, p. 0, 2025.

[24] ISO 9241-210, *Ergonomics of Human-System interaction-Part 210: Human-Centred Design for Interactive Systems (ISO 9241-210: 2010)*. ISO, 2010.

[25] Q. Mao, P.-L. Teh, S. J. Wang, and H. Wang, "Freedom to Personalize Walking Aids: a User-Centric Design Framework for Age-Friendly Smart Canes," *Int J Hum Comput Interact*, pp. 1–16, 2025.
[26] M. L. Wong, C. W. Khong, and H. Thwaites, "Applied UX and UCD Design Process in Interface Design," *Procedia-Social and Behavioral Sciences*, vol. 51, pp. 703–708, 2012.
[27] H. Maulana, D. C. Kurniasari, L. S. P. Wardhani, T. S. Mustofa, and E. A. Sutrisni, "Implementation of Design Thinking in Designing UPN Wellness & Youth Support Applications for Student Mental Health," *Nusantara Science and Technology Proceedings*, pp. 831–842, 2025.
[28] A. Bianco, P. Farrugia, and N. Sammut Bartolo, "Bridging Human-Centred Design and Engineering Education: a Case Study of Pharmaceutical Packaging for the Elderly," 2025.
[29] O. Zechner, H. Schrom-Feiertag, R. Wespi, D. Pretolesi, Q. Nguyen, and M. Tscheligi, "Enhancing Mixed Reality Simulation Training Technology With Real-Time Performance Visualization: Mixed Methods Study With Medical First Responders," *JMIR XR and Spatial Computing (JMXR)*, vol. 1, no. 1, p. e57655, 2024.
[30] E. Salinas, R. Cueva, and F. Paz, "A Systematic Review of User-Centered Design Techniques," in *International Conference on Human-Computer Interaction*, Springer, 2020, pp. 253–267.
[31] S. Litwin, M. Mohabir, I. S. Kocak, and D. Singh, "Creating a Parent-Informed Pediatric Emergency Department Wait Time App: Human-Centered Design Approach to Creating an AI Health Care Tool," *J Particip Med*, vol. 17, p. e66644, 2025.
[32] Y. Choi, P. Thamjamrassri, Y. Yoon, and H.-J. Kong, "Usability of Virtual Reality Based Bathroom Safety Education System for Dementia Caregiver: A Nielsen Heuristics Approach," *Stud Health Technol Inform*, vol. 329, pp. 2006–2007, 2025.
[33] M. Schrepp, "User experience questionnaire handbook," *All you need to know to apply the UEQ successfully in your project*, vol. 10, 2015.
[34] S. Singh and A. Kaur, "Relative Assessment of Traditional and Cloud Enabled Augmented Reality Applications for Chemistry Teaching and Learning Practices," *SN Comput Sci*, vol. 6, no. 3, pp. 1–15, 2025.
[35] G. W. Young, S. Stehle, B. Y. Walsh, and E. Tiri, "Exploring Virtual Reality in the Higher Education Classroom: Using VR to Build Knowledge and Understanding," *Journal of Universal Computer Science*, no. 8, pp. 904–928, 2020.
[36] L. W. Ward, D. Spencer, D. Chauhan, and M. Oliver-Hoyo, "Lessons Learned: the Use of an Augmented Reality Application in Organic Chemistry Laboratories," in *Frontiers in Education*, Frontiers Media SA, 2024, p. 1384129.
[37] F. Bruni, V. Mancuso, and E. Pedroli, "How to Evaluate User Experience in Digital Health? A Scoping Review of Questionnaires in Virtual Reality Applications," *Front Digit Health*, vol. 7, p. 1561364, 2025.
[38] A. Somrak, M. Pogačnik, and J. Guna, "Suitability and Comparison of Questionnaires Assessing Virtual Reality-Induced Symptoms and Effects and User Experience in Virtual Environments," *Sensors*, vol. 21, no. 4, p. 1185, 2021.
[39] L. Emma, "User-Centered Design to Enhance Accessibility and Usability in Digital Systems," *no. December*, 2024.
[40] A. K. Darmawan, M. A. Hamzah, B. Bakir, M. Walid, A. Anwari, and I. Santosa, "Exploring Usability Dimension of Smart Regency Service With Indonesian Adaptation of The System Usability Scale (SUS) and User Experience Questionnaire (UEQ)," in *2021 International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE)*, IEEE, 2021, pp. 74–79.
[41] A. N. Darmawan, A. Faroqi, and A. Pratama, "Evaluation of Student UX on E-Learning Web Using User Experience Questionnaire (UEQ)," *bit-Tech*, vol. 8, no. 1, pp. 1120–1130, 2025.