

Evaluating the effectiveness of a collaborative robotic maintenance training platform in extended reality (XR): A comprehensive validation study

Batur Alp Akdoğan^{a*} , Yunus Emre Esen^a , Kübra Yayan^b , Uğur Yayan^b 

^a Lider Teknoloji Geliştirme, 26250, Eskisehir, Türkiye

^b Eskisehir Osmangazi University, Eskisehir, Türkiye

Suggested citation: Akdoğan, B. A., Esen, Y. E., Yayan, K., & Yayan, U. (2026). Evaluating the effectiveness of a collaborative robotic maintenance training platform in extended reality (XR): A comprehensive validation study *Journal of Educational Technology and Online Learning*, 9(1), 123-145

Highlights

- XR-based training reduced scenario completion time by up to 76.8% compared to traditional methods.
- The platform achieved a "Good" usability rating with a System Usability Scale (SUS) score of 78 and induced a "Moderate" cognitive load (NASA-TLX).
- Expert participants trained with XR completed real-world tasks 28.7% faster than those trained traditionally.
- The platform's no-code scenario creation tool was found to be accessible and effective for both novice and expert users.
- Validation findings confirm that XR technology significantly enhances skill transfer and operational efficiency in industrial robotic maintenance.

Abstract

Maintenance training for industrial robotic systems involves challenges such as high costs, safety risks, and hardware limitations. This study presents the validation outcomes of the Multiplayer No-Code Mixed Reality Editor solution developed to provide immersive, interactive, and scalable training. The platform features a no-code scenario creation tool that allows users to design detailed training scenarios without programming knowledge, offering a more hands-on experience than traditional formats. A two-phase research design was adopted, involving 25 undergraduate students (Phase 1) and 6 experienced robotics experts (Phase 2). Participant performance was evaluated using task completion times, pre/post-tests, the System Usability Scale (SUS), and the NASA Task Load Index (NASA-TLX). The findings showed significant advantages over traditional methods. Results indicated high usability (average SUS score of 78) and a balanced cognitive workload (average NASA-TLX score of 4.3). XR-trained experts completed virtual scenarios up to 76.8% faster and real-world tasks 28.7% faster than their traditionally trained counterparts, reporting higher clarity and applicability. These results prove that editor is an effective, efficient, and user-friendly training solution for both novice and expert users in industrial robotic maintenance.

Article Info: Research Article

Keywords: *Extended Reality (XR), Robotic Maintenance, Training Effectiveness, Validation, System Usability Scale (SUS), NASA-TLX*

1. Introduction

The increasing complexity of industrial robotic systems is driving the demand for qualified technicians capable of performing their maintenance and repair (Coupry et al., 2023). However, traditional training methods have significant limitations, such as restricted access to physical robots, high costs, the potential to disrupt production processes, and, most importantly, the safety risks they pose to operators (Söderlund

* Corresponding author. Lider Teknoloji Geliştirme, 26250, Eskisehir, Türkiye.

E-mail addresses: ba.akdogan@liderteknoloji.com, yeesen@liderteknoloji.com, kokur@ogu.edu.tr, ugur.yayan@ogu.edu.tr

Doi: <https://doi.org/10.31681/jetol.1772552>

Received 26 Aug 2025; Accepted 22 Jan 2026

ISSN: 2618-6586. This is an open Access article under the CC BY license.



et al., 2024). In this context, Extended Reality (XR) technologies offer a safe, cost-effective, and repeatable solution for industrial training by allowing users to interact with virtual objects in a real-world environment.

Developed to address this need, Multiplayer No-Code Mixed Reality Editor is a multi-user, collaborative XR training platform designed for industrial robot maintenance. One of the platform's most distinctive features is its intuitive, node-based visual editor, which allows instructors and domain experts to create customized training scenarios without any programming knowledge. This approach democratizes the training content development process, offering a flexible structure that can quickly respond to the specific needs of the industry.

The primary objective of this article is to present the findings of a comprehensive validation study conducted on two different user groups (university students and industry experts) to demonstrate the effectiveness, usability, and impact on cognitive load of the multiplayer no-code mixed reality editor. The study aims to quantitatively evaluate the advantages of XR-based training over traditional methods in terms of performance, learning, and user experience.

With these objectives, this study seeks to address key challenges repeatedly identified in recent XR training research. The literature indicates that XR authoring tools often remain inaccessible to non-programmer instructors and domain experts, not primarily due to hardware dependency or limitations of XR devices, but rather because of the complexity and technical burden associated with content creation and scenario authoring processes (Dengel et al., 2022). As a result, the scalability and adoption of XR-based training solutions are constrained. Furthermore, existing studies frequently emphasize usability-focused evaluations, while providing limited empirical evidence linking XR-based training to real-world task performance and skill transfer (Blattgerste et al., 2023). In alignment with these observations, the objectives of the present work focus on enabling no-code scenario authoring, supporting collaborative training workflows, and validating training effectiveness through a combination of subjective measures and performance-based evaluation on a physical robot task.



Figure 1 Multiplayer No-Code Mixed Reality Editor Product Picture

2. Literature

2.1. The Role of XR in Industrial Training

The application of XR technologies in industrial and vocational training has emerged as a transformative paradigm, addressing the limitations of conventional methods (Abidi et al., 2022). Traditional training,

often reliant on manuals, lectures, and limited hands-on practice, struggles to effectively convey complex, three-dimensional procedural tasks. XR environments, by contrast, provide an immersive and interactive learning space where trainees can engage in realistic simulations that closely resemble physical workflows, tools, and spatial constraints encountered on the shop floor. Research consistently demonstrates that VR and MR training can significantly reduce learning time, decrease error rates, and improve knowledge retention for complex assembly, maintenance, and operational tasks (Monetti et al., 2022; Papanastasiou et al., 2022). These benefits are particularly pronounced in contexts where direct access to physical equipment is limited due to safety risks, cost, or production constraints. The ability to practice procedures repeatedly in a safe, consequence-free environment is particularly beneficial for high-risk operations, strengthening both competence and confidence (Stary, 2023). Multiplayer no-code mixed reality editor builds on this foundation by utilizing Mixed Reality (MR), which overlays virtual information onto the real world, thus enhancing contextual awareness and minimizing the motion sickness sometimes associated with fully immersive VR systems. Recent systematic reviews further indicate that XR-based training environments are particularly effective when they support experiential and situated learning paradigms, enabling learners to actively engage with realistic tasks rather than passively consuming instructional content (Carvalho et al., 2025; Grassini et al., 2020). In industrial contexts, this alignment with experiential learning is critical for fostering procedural knowledge, spatial understanding, and task readiness.

2.2. Usability and Cognitive Load in XR Learning Environments

For any educational technology to be effective, it must not only deliver content but also provide positive user experience. The usability of an XR system, how easily and intuitively it can be used is a critical determinant of its adoption and success. The System Usability Scale (SUS) is a widely accepted, reliable tool for measuring perceived usability, providing a single score that allows for straightforward benchmarking (Brooke, 1996). A high usability score indicates that the system does not present a barrier to learning, allowing the user to focus on the training content rather than struggling with the interface. This is especially important in XR environments, where complex interaction techniques and unfamiliar devices can otherwise divert attention away from learning objectives.

Equally important is the concept of cognitive load, which refers to the total amount of mental effort being used in the working memory (Sweller, 1988). An effective training environment should impose an optimal level of cognitive load challenging enough to stimulate learning but not so overwhelming as to hinder it. In XR-based training, excessive cognitive load may arise from poor interaction design, information overload, or mismatches between virtual representations and real-world task logic. The NASA Task Load Index (NASA-TLX) is a multidimensional assessment tool used to measure the perceived workload of a task across six dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration (Hart & Staveland, 1988). Studies in XR training have shown that poorly designed interfaces or overly complex tasks can induce high cognitive load, negatively impacting learning outcomes (Choi et al., 2022). Therefore, evaluating both usability and cognitive load is essential to validate the pedagogical soundness of multiplayer no-code mixed reality editor, as these factors jointly influence user acceptance, learning efficiency, and the likelihood of successful skill transfer to real-world tasks. Prior studies emphasize that usability and workload metrics alone are insufficient unless they are interpreted in relation to training transfer and task execution in real-world settings, a limitation frequently reported in XR training literature (Grassini et al., 2020; Monetti et al., 2022).

2.3. The Gap: Collaborative Training and No-Code Authoring

Despite the advancements in XR training, two significant gaps persist in the literature and available platforms, limiting the practical deployment of XR solutions in real industrial training settings. First, most industrial maintenance procedures are not performed by individuals in isolation but by teams. However, the majority of existing XR training solutions are designed for single-user experiences, failing to address the critical need for collaborative skill development, communication, and coordinated action (Söderlund et al., 2024). This limitation is particularly problematic for maintenance and repair tasks, which often require shared situational awareness and role-based coordination. Second, the creation of high-quality XR training

content typically requires specialized programming and 3D modeling skills, making it a costly and time-consuming process inaccessible to most educators and domain experts (Wang et al., 2024). Prior reviews on XR and AR authoring toolkits highlight that this limitation stems less from hardware constraints and more from the complexity of authoring workflows, which often require scripting, engine-level configuration, and multidisciplinary technical expertise (Blattgerste et al., 2023; Dengel et al., 2022). Moreover, while several visual or low-code authoring tools have been proposed, the literature does not establish a clear, comparable baseline for scenario creation time using conventional coding-based pipelines, particularly for non-programmer instructors. As a result, efficiency claims in prior work are often qualitative or anecdotal, making it difficult to assess the practical accessibility of XR content creation.

This editor is specifically designed to address these two gaps. Its architecture supports real-time, multi-user sessions, allowing trainees to collaborate within a shared MR space. Furthermore, its integrated no-code, node-based authoring tool empowers non-programmers to create and modify complex, interactive training scenarios. This study, therefore, not only validates the effectiveness of an XR platform but also investigates the impact of these two innovative features collaboration and accessible authoring on the training process. By combining multi-user training with a no-code authoring approach, the proposed editor addresses limitations identified in prior XR studies that focus primarily on single-user execution or pre-defined scenarios, while offering a practical pathway toward scalable content creation without relying on conventional code-centric development workflows.

3. Methodology

This study employed a two-phase, pre-test and post-test research design to comprehensively evaluate the editor. The methodology was structured to assess the platform's performance across different user demographics and to compare its effectiveness against traditional training methods.

3.1. The Multiplayer No-Code Mixed Reality Editor

Multiplayer No-Code editor is a Mixed Reality training platform built on the Unity engine and deployed through the VIROO platform for multi-user capabilities. Its core feature is a visual, node-based editor that functions entirely within the MR environment. An instructor, wearing an MR headset, can create a training workflow by dragging and connecting nodes. Each node represents a specific action (e.g., Grab, Touch, Unscrew) or a logical step. These nodes are then linked to 3D models of robotic components placed in the virtual workspace. This process allows for the rapid creation of interactive, step-by-step procedures without writing a single line of code. During a training session, trainees see instructions and highlighted virtual objects overlaid on their real environment, guiding them through the maintenance task. The system tracks their progress and provides real-time feedback.

3.2. Participants and Research Design

The validation was conducted at the IFARLAB-EDIH (Eskisehir Osmangazi University Intelligent Factory and Robotics Laboratory- European Digital Innovation Hub). A total of 31 participants were recruited for the study, divided into two distinct phases:

- **Phase 1 (Single-User XR Group):** This phase included 25 undergraduate Computer Engineering students with limited to no prior experience in robotics maintenance. The objective was to evaluate the platform's usability, cognitive load, and learning effectiveness for novice users in a single-user setting.
- **Phase 2 (Control Group Comparison):** This phase involved six experienced professionals, all of whom had completed at least a master's degree and were actively pursuing Electrical and Electronic PhD studies in robotics-related fields. The participants were affiliated with IFARLAB and were actively involved in robotics research and development activities, including ongoing European Union-funded projects, ensuring a high level of domain expertise and familiarity with real-world industrial constraints. This cohort was deliberately selected to evaluate the platform's impact on users who already possess advanced technical knowledge and practical experience in robotics.

All participants had previously received formal training in robotic maintenance and setup through traditional methods, such as classroom-based instruction, technical documentation, and hands-on practice on physical robotic systems. This background ensured that the evaluation on how the proposed mixed reality platform supports experienced users in comparison to established training approaches. The group was randomly assigned to one of two experimental conditions:

- **Mixed Reality Group (n=3):** Received training exclusively through the Mixed Reality Editor.
- **Traditional Group (n=3):** Participants in this group received training exclusively through conventional instructional methods. For each maintenance scenario, they were provided with fixed-duration training consisting of a 10-minute high-quality instructional video, complemented by a separate, structured presentation explaining the same procedure step by step.

3.3. Validation Scenarios

The training content was standardized across all groups and focused on the Universal Robots UR10e platform. Three specific maintenance scenarios were developed:

1. **Scenario 1- Robot Arm Disassembly:** A procedural task involving the step-by-step removal of the wrist joint cover and end-effector. Figure 2 illustrates the flowchart of this scenario, detailing to provide a clear visualization of the task workflow.

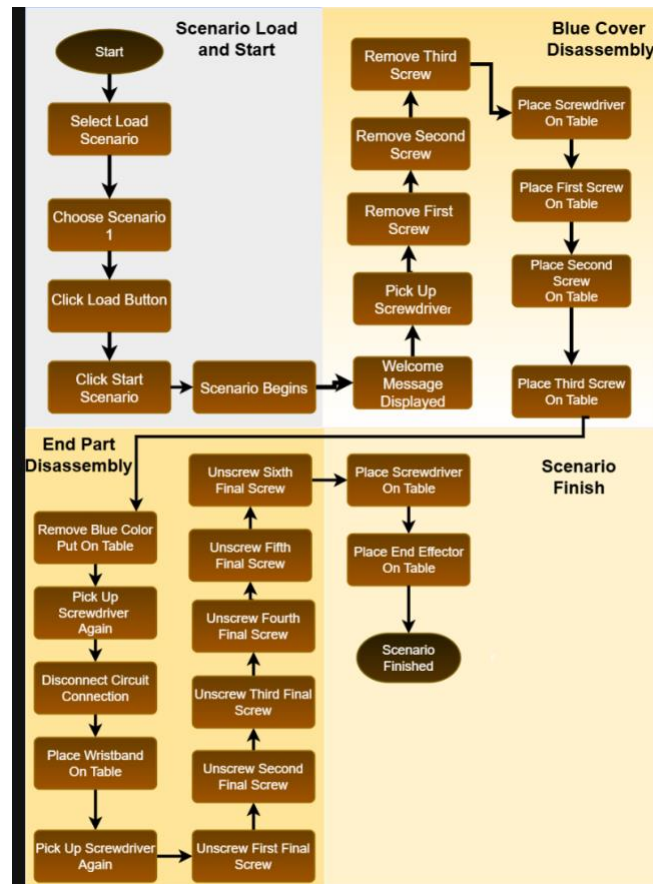


Figure 2 Scenario 1 Flowchart

2. **Scenario 2- Robot Assembly:** The reverse procedure, requiring users to reassemble the previously removed components in the correct sequence. Figure 3 illustrates the flowchart of this scenario, detailing to provide a clear visualization of the task workflow.

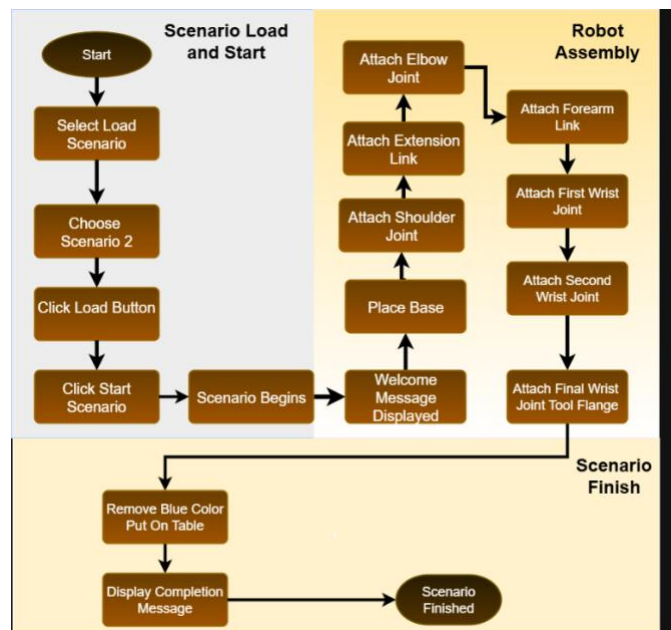


Figure 3 Scenario 2 Flowchart

3. **Scenario 3- Control Box Fan Replacement:** A component-level repair task requiring users to open the robot's control box, remove four screws, replace a fan unit, and secure the new component. Figure 4 illustrates the flowchart of this scenario, detailing to provide a clear visualization of the task workflow.

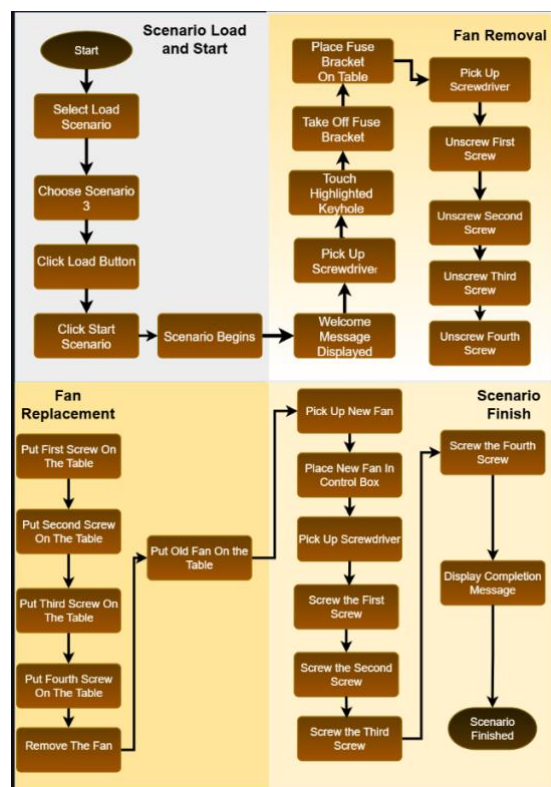


Figure 4 Scenario 3 Flowchart

In addition to these predefined scenarios, participants in Phase 1 were tasked with using the no-code editor to create a simple scenario of their own, in order to assess the authoring tool's intuitiveness.

3.4. Data Collection Tools

Pre-Tests

Pre-tests were designed to evaluate participants' baseline declarative and procedural knowledge of maintenance tasks before training. For **Phase 1**, a dedicated pre-test was created specifically for this group and is detailed in Table 1. For **Phase 2**, a standardized pre-test was administered to both traditionally trained participants and those trained on the editor, with all questions provided in **Table 2**.

Table 1 Phase 1 - Pre-Test Survey

No. Questions	Response Type	Scale anchors
1 Have you previously worked with any industrial robot, mobile robot, or a robotic maintenance scenario?	Likert 1–5	1 = No experience; 2 = Some; 3 = Moderate; 4 = Good; 5 = Advanced
2 Have you experienced an XR training environment before?	Likert 1–5	1 = No experience; 2 = Little; 3 = Moderate; 4 = Good; 5 = Very good
3 How beneficial would it be for you to perform work through XR?	Likert 1–5	1 = Definitely not beneficial; 2 = Slightly beneficial; 3 = Moderately beneficial; 4 = Beneficial; 5 = Definitely beneficial
4 Can you recognize potential risks when working with robotic systems?	Likert 1–5	1 = Not at all; 2 = Partly; 3 = Moderately; 4 = Yes; 5 = Definitely
5 Do you think XR-based training applications can contribute to the development of professional skills?	Likert 1–5	1 = Definitely no contribution; 2 = Slight contribution; 3 = Moderate contribution; 4 = Large contribution; 5 = Definite contribution
6 What challenges do you think you may face while training on such a platform?	Open-ended	—
7 In which aspect do you expect to benefit the most from this platform?	Open-ended	—
8 How motivated do you feel to use the Mixed Reality Editor?	Likert 1–5	1 = Not motivated at all; 5 = Highly motivated (<i>intermediate points unlabeled</i>)

Building upon this, Table 2 presents the standardized pre-test applied in **Phase 2**, ensuring consistent evaluation for both traditionally trained.

Table 2 Phase 2 - Pre-Test Survey

No. Questions	Response Type	Scale anchors
1 I expect this training (Traditional or Mixed Reality) will help me understand key concepts in robotic maintenance.	Likert 1–5	1 = Strongly Disagree, 2 = Disagree, 3 = Neither Agree nor Disagree, 4 = Agree, 5 = Strongly Agree
2 I expect the maintenance procedures will be presented clearly and be directly applicable to the real fan-removal task.	Likert 1–5	1 = Strongly Disagree, 2 = Disagree, 3 = Neither Agree nor Disagree, 4 = Agree, 5 = Strongly Agree
3 I expect to receive sufficient information to perform the fan removal task safely and effectively.	Likert 1–5	1 = Strongly Disagree, 2 = Disagree, 3 = Neither Agree nor Disagree, 4 = Agree, 5 = Strongly Agree
4 I expect this training will support my problem-solving and decision-making during maintenance.	Likert 1–5	1 = Strongly Disagree, 2 = Disagree, 3 = Neither Agree nor Disagree, 4 = Agree, 5 = Strongly Agree
5 I expect the training content to be relevant and useful for my future work in robotic maintenance.	Likert 1–5	1 = Strongly Disagree, 2 = Disagree, 3 = Neither Agree nor Disagree, 4 = Agree, 5 = Strongly Agree
6 I anticipate being confident about safety precautions while working on the control box.	Likert 1–5	1 = Strongly Disagree, 2 = Disagree, 3 = Neither Agree nor Disagree, 4 = Agree, 5 = Strongly Agree
7 What concerns or challenges do you anticipate during the upcoming hands-on fan removal task?	Open-ended	—
8 Which specific aspects of the training do you expect will benefit you the most during the real task?	Multi-Choice	—

Post-Test

Following completion of the training interventions, post-tests were administered to assess knowledge acquisition, procedural understanding, and skill transfer. For Phase 1, the post-test, detailed in Table 3, focused on clarity, interactivity, and perceived skill development after completing the XR-based scenarios and scenario creation task. For Phase 2, a common post-test shown in Table 4 was administered to both participant groups to facilitate direct comparisons between XR-based and traditional training outcomes.

Table 3 Phase 1 -Post Test Survey

No. Question	Response Type	Scale Anchors (if applicable)
1 The robotic maintenance scenarios on the Mixed Reality were clear and applicable.	Likert 1–5	1 = Strongly Disagree; 5 = Strongly Agree
2 My scenario experience in the XR environment was interactive and engaging.	Likert 1–5	1 = Strongly Disagree; 5 = Strongly Agree
3 Performing tasks in the XR environment was efficient.	Likert 1–5	1 = Strongly Disagree; 5 = Strongly Agree
4 During the application, I was able to identify safety risks related to robotic systems.	Likert 1–5	1 = Strongly Disagree; 5 = Strongly Agree
5 I believe the Mixed Reality Editor contributed to the development of my professional skills.	Likert 1–5	1 = Strongly Disagree; 5 = Strongly Agree
6 What were the main challenges you faced while using this platform?	Open-ended	—
7 In your opinion, in which area did you experience the most improvement after the Mixed Reality editor experience?	Multiple choice	Problem-solving skills; Robotic maintenance knowledge; Practical thinking and decision-making; Familiarity with XR technology
8 How motivating was it for you to use this application?	Likert 1–5	1 = Not motivating at all; 5 = Highly motivating

Transitioning to Phase 2, a common post-test as shown in Table 4 was developed to capture a comparable set of metrics, allowing meaningful evaluation of differences between the XR-based and traditional training experiences.

Table 4 Phase 2 - Post Test Survey

No. Question	Response Type	Scale Anchors (if applicable)
1 The training helped me understand key concepts in robotic maintenance.	Likert 1–5	1 = Strongly Disagree; 5 = Strongly Agree
2 The robotic maintenance scenarios were clear and directly applicable to real-world tasks.	Likert 1–5	1 = Strongly Disagree; 5 = Strongly Agree
3 The training provided sufficient knowledge to perform robotic maintenance tasks effectively.	Likert 1–5	1 = Strongly Disagree; 5 = Strongly Agree
4 The scenarios included in the training helped me improve my problem-solving and decision-making skills related to robotic maintenance.	Likert 1–5	1 = Strongly Disagree; 5 = Strongly Agree
5 I would recommend this training format to others learning about robotic maintenance.	Likert 1–5	1 = Strongly Disagree; 5 = Strongly Agree
6 The training content was important and useful for my future work in robotic maintenance.	Likert 1–5	1 = Strongly Disagree; 5 = Strongly Agree

Nasa-TLX

The NASA Task Load Index (NASA-TLX) is a standardized, multi-dimensional instrument for assessing subjective workload after a task. It captures six dimensions (Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration) to quantify how demanding an activity felt to a participant.

In Mixed Reality Editor, NASA-TLX is administered after participants try the application (post-exposure) and is applied to all participants to enable comparable analysis across scenarios and phases.

Scoring note. In this study, we use a 1–10 response scale and report the raw (unweighted) mean of the six subscales as the overall workload indicator. Higher scores indicate greater workload, except for Performance, where the direction is inverted, as shown in Table 5.

Table 5 Nasa-TLX Survey

No.	Subscale	Item wording	Response Type	Scale anchors
1	Mental Demand	How mentally demanding was the task?	1–10	1 = Very Low, 10 = Very High
2	Physical Demand	How physically demanding was the task?	1–10	1 = Very Low, 10 = Very High
3	Temporal Demand	How hurried or rushed was the pace of the task?	1–10	1 = Very Low, 10 = Very High
4	Performance	How successful were you in accomplishing what you were asked to do?	1–10	1 = Perfect/Excellent, 10 = Failure/Poor
5	Effort	How hard did you have to work to accomplish your level of performance?	1–10	1 = Very Low, 10 = Very High
6	Frustration	How insecure, discouraged, irritated, stressed, and annoyed were you?	1–10	1 = Very Low, 10 = Very High

In NASA-TLX, higher values indicate greater perceived workload. The Performance subscale is reverse-anchored (0 = Perfect/Excellent, 10 = Failure/Poor). The raw (unweighted) mean of the six subscales is reported on a 0–10 scale and can be interpreted as follows: 0.0–0.9 = Low, 1.0–2.9 = Medium, 3.0–4.9 = Moderate, 5.0–7.9 = High, and 8.0–10.0 = Very High. Universal pass/fail cut-offs are not defined; results should be interpreted in context and compared across tasks or conditions. These scale ranges are provided in Table 6.

Table 6 Nasa-TLX Calculation

Category	Nasa-TLX Score
Low	0.0–0.9
Medium	1.0–2.9
Moderate	3.0–4.9
High	5.0–7.9
Very High	8.0–10.0

System Usability Scale (SUS)

The System Usability Scale (SUS) is a standardized questionnaire developed by John Brooke in 1986 to assess the overall perceived usability of a system. It consists of 10 items rated on a 5-point Likert scale, yielding a single score between 0–100. Higher values indicate better usability. SUS is widely used in both industry and research because it is quick to administer, technology-independent, and applicable to software, hardware, and mixed reality systems.

In the editor validation, SUS was applied to quantitatively measure participants' perceptions of the usability of the XR-based robotic maintenance training platform. This facilitated direct comparisons across participant groups and training methods.

The SUS includes both positively and negatively worded statements. Positively worded questions are scored directly, while negatively worded questions are reverse-scored. The complete set of items and their scales are shown in Table 7.

Table 7 System Usability Scale

No	SUS Items	Scale Anchors
1	I think that I would like to use this system frequently.	1 = Strongly Disagree, 2 = Disagree, 3 = Neither, 4 = Agree, 5 = Strongly Agree
2	I found the system unnecessarily complex.	1 = Strongly Disagree, 2 = Disagree, 3 = Neither, 4 = Agree, 5 = Strongly Agree
3	I thought the system was easy to use.	1 = Strongly Disagree, 2 = Disagree, 3 = Neither, 4 = Agree, 5 = Strongly Agree
4	I think that I would need the support of a technical person to be able to use this system.	1 = Strongly Disagree, 2 = Disagree, 3 = Neither, 4 = Agree, 5 = Strongly Agree
5	I found the various functions in this system were well integrated.	1 = Strongly Disagree, 2 = Disagree, 3 = Neither, 4 = Agree, 5 = Strongly Agree
6	I thought there was too much inconsistency in this system.	1 = Strongly Disagree, 2 = Disagree, 3 = Neither, 4 = Agree, 5 = Strongly Agree
7	I would imagine that most people would learn to use this system very quickly.	1 = Strongly Disagree, 2 = Disagree, 3 = Neither, 4 = Agree, 5 = Strongly Agree
8	I found the system very cumbersome to use.	1 = Strongly Disagree, 2 = Disagree, 3 = Neither, 4 = Agree, 5 = Strongly Agree
9	I felt very confident using the system.	1 = Strongly Disagree, 2 = Disagree, 3 = Neither, 4 = Agree, 5 = Strongly Agree
10	I needed to learn a lot of things before I could get going with this system.	1 = Strongly Disagree, 2 = Disagree, 3 = Neither, 4 = Agree, 5 = Strongly Agree

The SUS score is computed through a standardized calculation procedure to ensure methodological rigor and comparability across participants. First, each questionnaire item is classified as positively or negatively worded. For positively worded items Table 7 (1, 3, 5, 7, 9), the contribution score is obtained by subtracting one from the response value. For negatively worded items Table 7 (2, 4, 6, 8, 10), the contribution score is calculated by subtracting the response from five. Summing these contributions yields a raw score between 0 and 40, which is then multiplied by 2.5 to convert it to a 0–100 scale.

According to established benchmarks, a score of **68** corresponds to average usability, with higher scores indicating superior usability and lower scores suggesting areas for improvement (Brooke, 1996). The full grading scheme and adjective rating interpretation are presented in Table 8.

Table 8 SUS Interpretation Scale

SUS Score Range	Grade	Adjective Rating
> 80.3	A	Excellent
68 – 80.3	B	Good
68	C	Okay
51 – 68	D	Poor
<51	F	Awful

As shown in Table 8, scores above 80.3 reflect excellent usability, while those below 51 indicate severe usability issues requiring significant design revisions

3.5. Procedure

The validation was conducted at the IFARLAB-EDIH. Upon arrival, all participants were given an introduction to the study and signed an informed consent form. They then completed the pre-test.

- **Phase 1 participants** received a brief tutorial on how to use the MR headset and controllers. They then proceeded to complete the three predefined scenarios individually, followed by the scenario creation task. After completing all tasks, they filled out the post-test, SUS, and NASA-TLX questionnaires.
- **Phase 2 participants** were assigned to their respective groups. The Mixed Reality group followed a similar procedure to Phase 1 participants (excluding the scenario creation task). The Traditional

group attended a 30-minute session where an expert trainer explained the three scenarios using slides and video clips. Immediately following their respective training sessions, both groups were tasked with performing the Scenario 3 (fan replacement) on a real, physical UR10e control box. After this practical test, they completed the post-test.

3.6. Data Analysis

The quantitative data collected were analyzed using descriptive and inferential statistics.

- **Task Completion Times and Test Scores:** Mean and standard deviation were calculated for each group. The percentage difference between the XR and Traditional groups in Phase 2 was calculated to determine the efficiency gain.
- **SUS Score Calculation:** The standard scoring methodology for SUS was applied. Responses to odd-numbered (positive) items had 1 subtracted from the score. For even-numbered (negative) items, the score was subtracted from 5. The sum of these new scores was then multiplied by 2.5 to yield a final score between 0 and 100.
- **NASA-TLX Score Calculation:** The raw, unweighted mean of the six subscale scores was calculated to provide an overall workload score out of 10.
- **Comparisons:** Due to the small sample size in Phase 2, formal statistical significance tests (e.g., t-tests) were not conducted. Instead, the analysis focuses on the descriptive differences and effect sizes between the groups, which are more appropriate for pilot and validation studies of this nature.

4. Conclusion and Suggestions

This section presents the main findings from the validation study, discusses their implications in the context of industrial training, and concludes with suggestions for future research.

4.1. Findings

The findings section summarizes the outcomes of the multiplayer no-code mixed reality editor validation study, highlighting measurable improvements in procedural knowledge, safety awareness, and task performance across both novice and expert participants. Data from pre- and post-tests, workload assessments, usability evaluations, and task completion metrics collectively demonstrate that this study delivers a clear, engaging, and effective training experience, offering significant advantages over traditional approaches.

4.1.1 Performance and Efficiency: Task Completion Times – Phase 1

This section presents the detailed analysis of task completion times for all scenarios in Phase 1, highlighting participant performance and complexity differences between training and content creation activities.

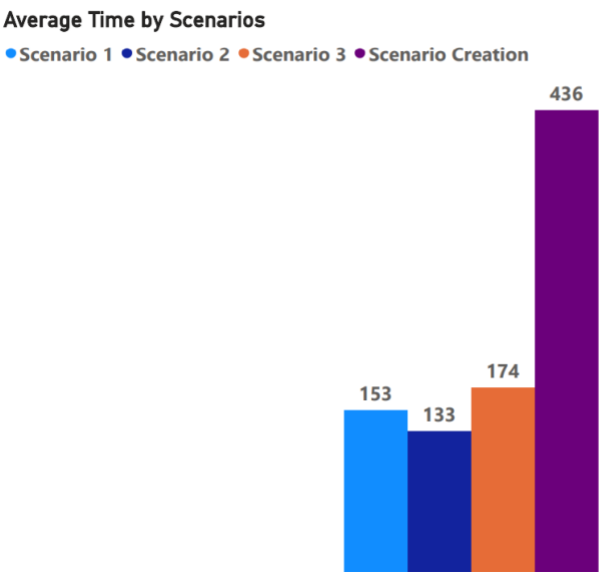


Figure 5 Average Time by Scenarios

Figure 5 Average Time by ScenariosFigure 5 shows the average task completion times for all Phase 1 activities. Scenario 1 (Robot Arm Disassembly) averaged **153 seconds**, Scenario 2 (Robot Assembly) **133 seconds**, and Scenario 3 (Control Box Fan Replacement) **174 seconds**. The Scenario Creation task required significantly more time, averaging **436 seconds**, reflecting its complexity and creative demands. This visualization highlights that execution-focused scenarios were relatively uniform in duration, while content authoring required more effort and decision-making.

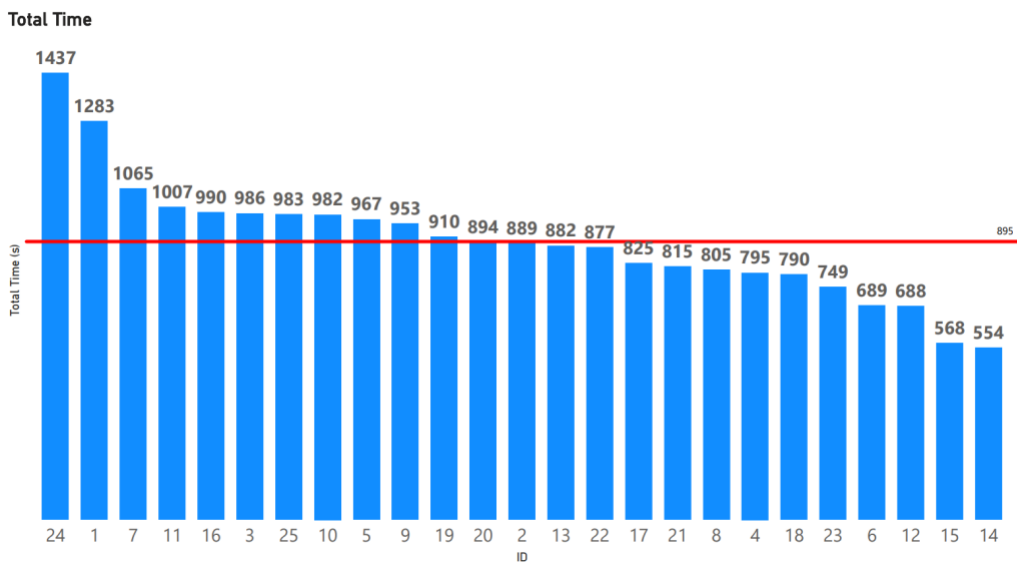


Figure 6 Total Time By Per ID

Figure 6 provides a participant-level overview of total time spent across all scenarios. The average total time was **895 seconds**, with participant times ranging from **554 seconds** to **1437 seconds**. This distribution illustrates variability in user interaction speed and adaptability, with a majority of participants clustering around the average time, indicating consistent usability and scenario flow.

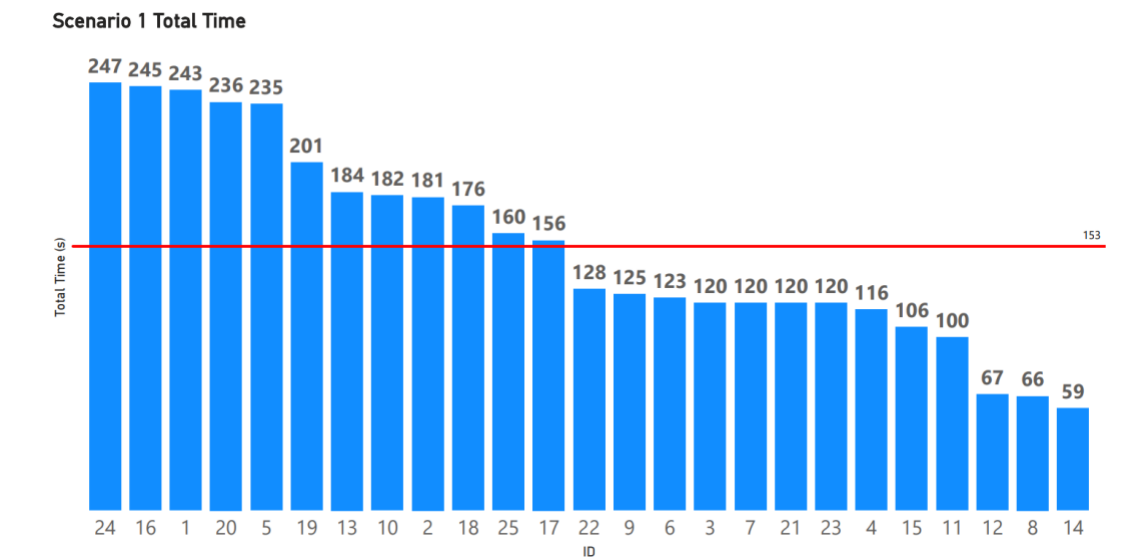


Figure 7 Scenario 1 Total Time by Per ID

Figure 7 breaks down Scenario 1 completion times per participant. The fastest completion time was **59 seconds**, and the slowest was **247 seconds**, with most participants completing within a narrow range close to the mean. This suggests Scenario 1 offered a well-balanced introductory task that users adapted to quickly.

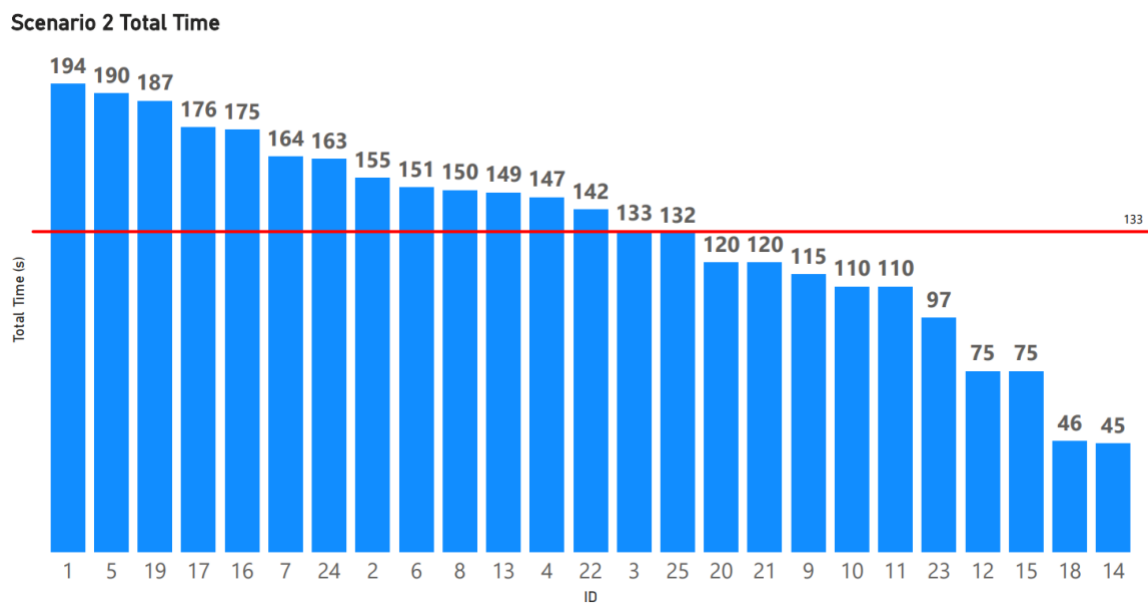


Figure 8 Scenario 2 Total Time by PER ID

Figure 8 shows participant performance in Scenario 2, where the average was **133 seconds**. Times ranged from **45 seconds** to **194 seconds**. The relatively small spread reflects consistent performance and minimal difficulty, demonstrating user familiarity with XR interactions by the second scenario.

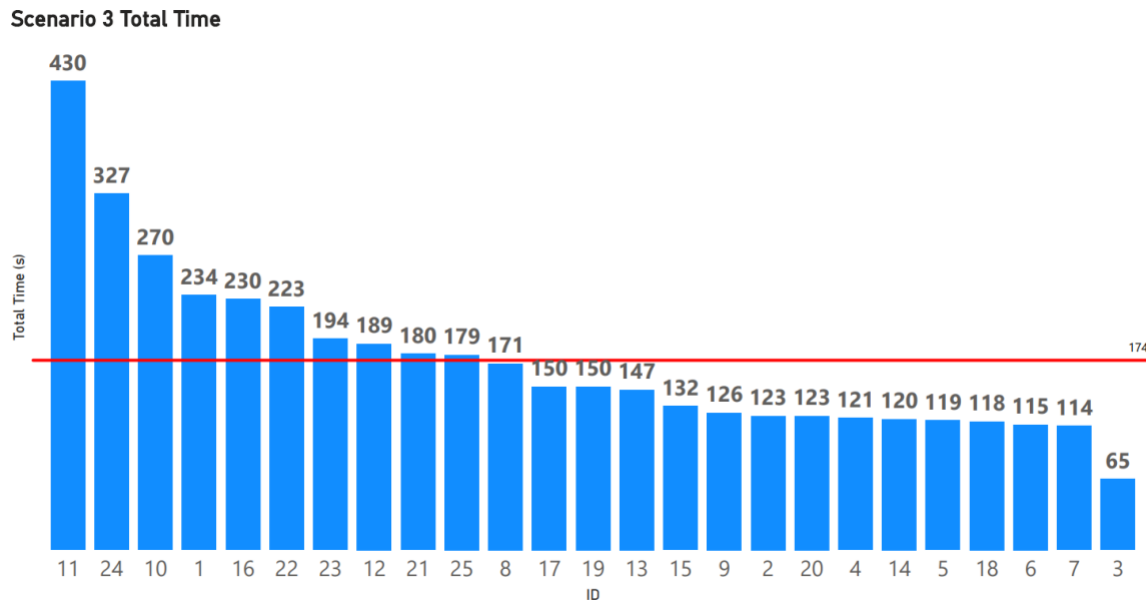


Figure 9 Scenario 3 Total Time by Per ID

Scenario 3 results, displayed in Figure 9, show a broader spread with times ranging from **65 seconds** to **430 seconds**. This scenario was more challenging, involving multiple steps and complex interactions. The data indicate some participants required additional time, underscoring the scenario's complexity and suitability for advanced testing.

Beyond the prepared scenarios, participants engaged in a Scenario Creation task that averaged **436 seconds**, where they built their own scenario consisting of **three nodes and two interactive objects**. This longer duration reflects the additional planning and decision-making involved in designing custom training flows using no-code editor. Despite the increased complexity, every participant successfully created a functional scenario, emphasizing the platform's usability for content authoring as well as training execution.

4.1.2 Performance and Efficiency: Task Completion Times – Phase 2

This section summarizes Phase 2 task completion times, comparing Mixed Reality training participants with those trained through traditional methods.

Participants trained using Mixed Reality completed Scenario 1 in **136 seconds**, Scenario 2 in **128 seconds**, and Scenario 3 in **180 seconds** on average, demonstrating consistent performance gains and strong familiarity with XR workflows. In contrast, traditionally trained participants received a **10-minute instructor-led session prepared by IFARLAB**, supported by detailed presentations and video demonstrations. While this approach offered structured theoretical knowledge, their task execution remained slower compared to MCR-trained peers.

Control Group Real-World Task Completion Times

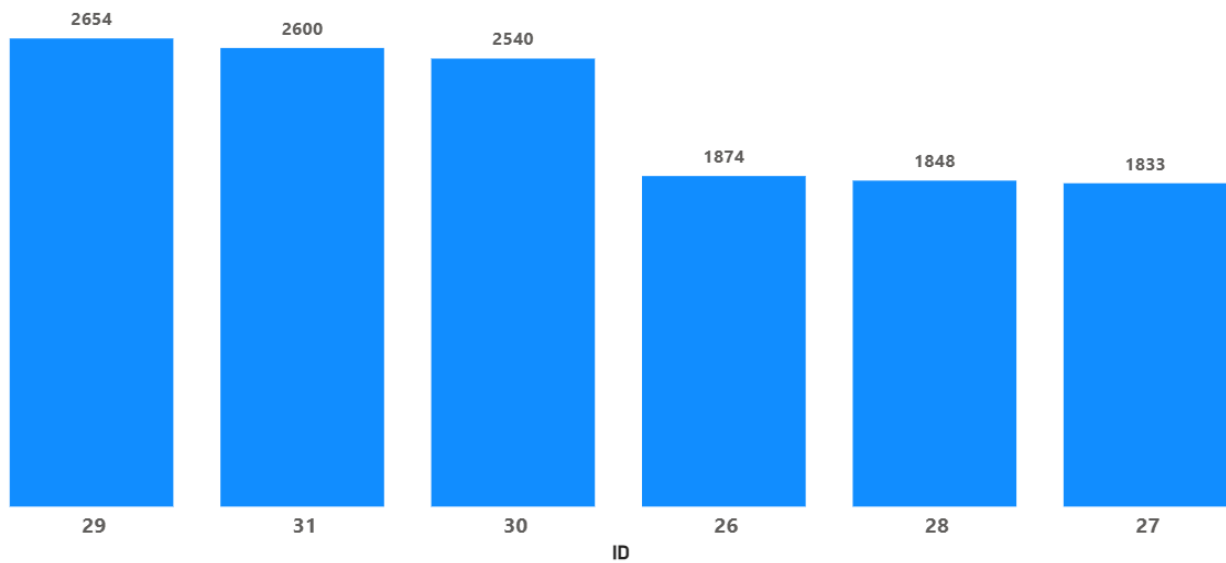


Figure 10 Control Group Real-World Task Completion Times

Figure 10 illustrates real-world task performance times, where MR-trained participants averaged **1852 seconds**, compared to **2598 seconds** for traditionally trained participants. This indicates a **~29% improvement in efficiency** with MR training. The combined overall average across all participants was **2225 seconds**. These findings provide strong evidence that immersive, hands-on training in mixed reality significantly enhances task execution efficiency.

4.1.2. User Experience: Usability and Cognitive Load

The platform's user experience was evaluated with the SUS and NASA-TLX questionnaires. The NASA Task Load Index (NASA-TLX) was employed to assess subjective workload across six subscales as defined in Table 5. Table 5 summarizes the scale anchors used for each subscale, where lower scores generally indicate lower demand, except for Performance, where a lower score signifies higher perceived success. Figure 16 presents the aggregated results, showing an overall average workload score of **4.3/10**, corresponding to a **Moderate** workload level. The following subsections highlight participant responses:

- **Mental Demand:** Most participants rated this subscale between moderate and slightly low, indicating that the tasks were cognitively engaging without being overwhelming, as shown in Figure 11.

NASA-TLX Mental Demand

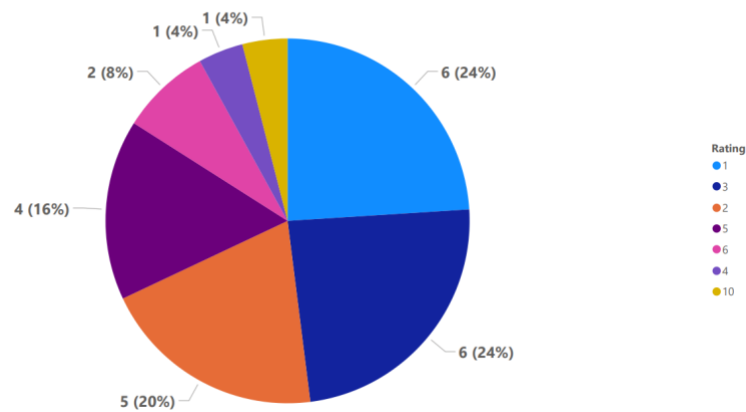


Figure 11 NASA-TLX Mental Demand

- **Effort:** Ratings clustered around moderate effort levels, confirming a balanced challenge, as shown in Figure 12.

NASA-TLX Effort

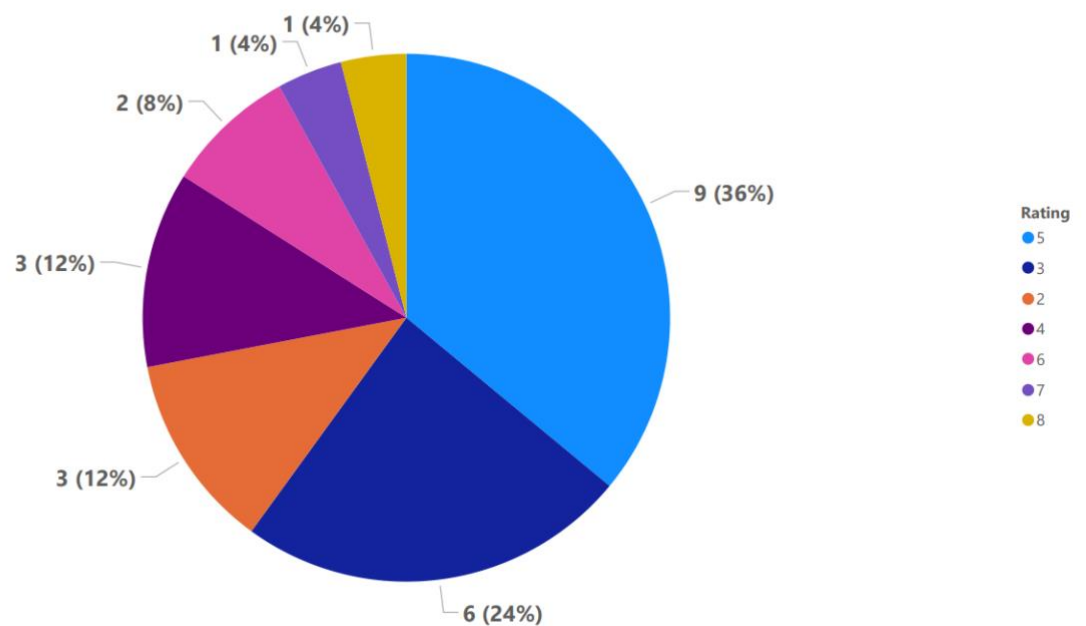


Figure 12 NASA-TLX Effort

- **Temporal Demand:** The majority reported low urgency, suggesting sufficient time allocation for each scenario, as shown in Figure 13.

NASA-TLX Temporal Demand

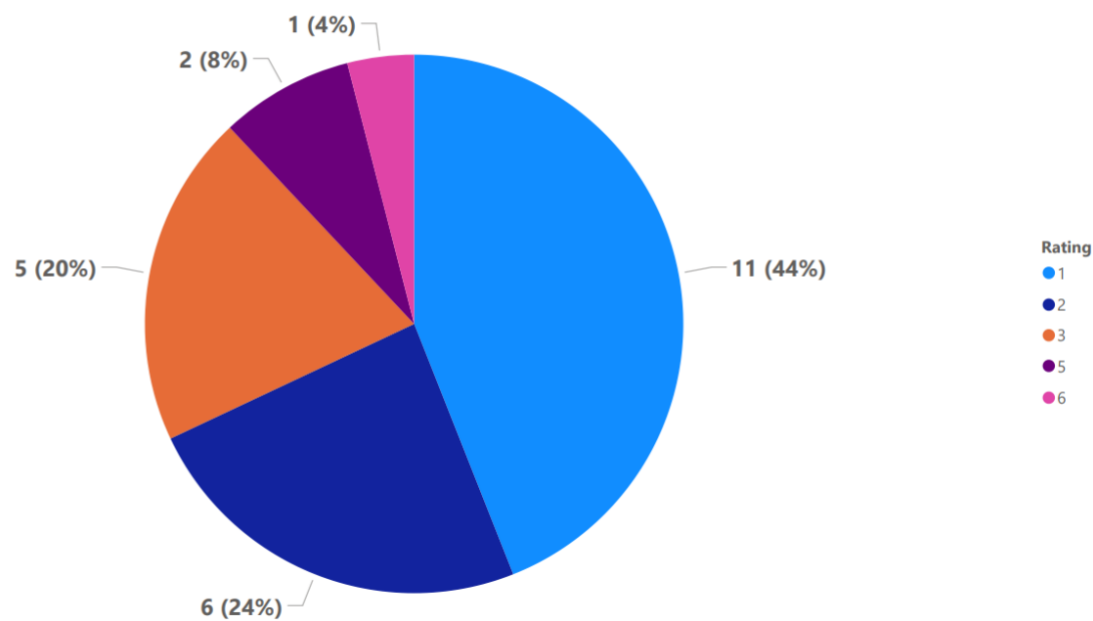


Figure 13 Nasa-TLX Temporal Demand

- **Performance:** The inverted scoring revealed a very high sense of success, with a mean score of 1.12, as shown in Figure 14.

NASA-TLX Performance

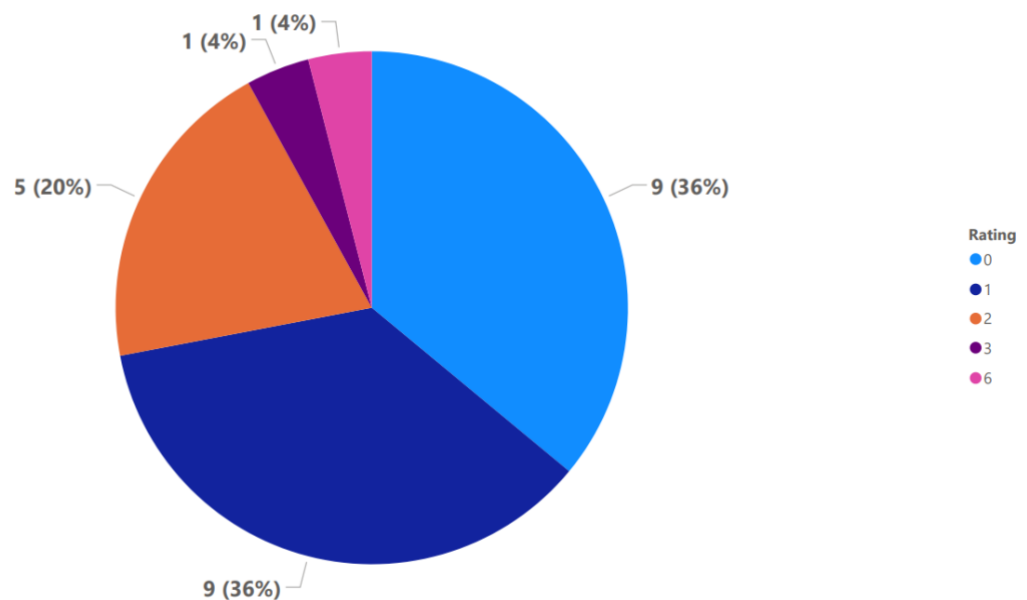


Figure 14 Nasa-TLX Performance

- **Physical Effort:** Feedback indicated manageable physical demand with no extreme strain, as shown in Figure 15.

NASA-TLX Physical Effort

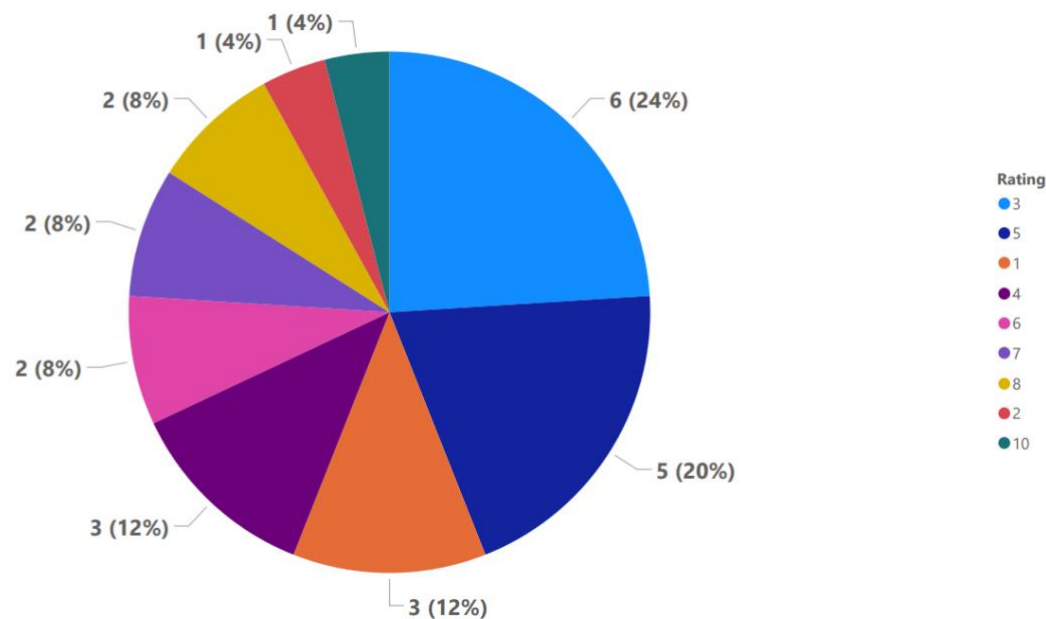


Figure 15 Nasa-TLX Physical Effort

- Frustration:** Participants consistently reported low frustration, with an average of **2.72/10**, as shown in Figure 16.

NASA-TLX Frustration

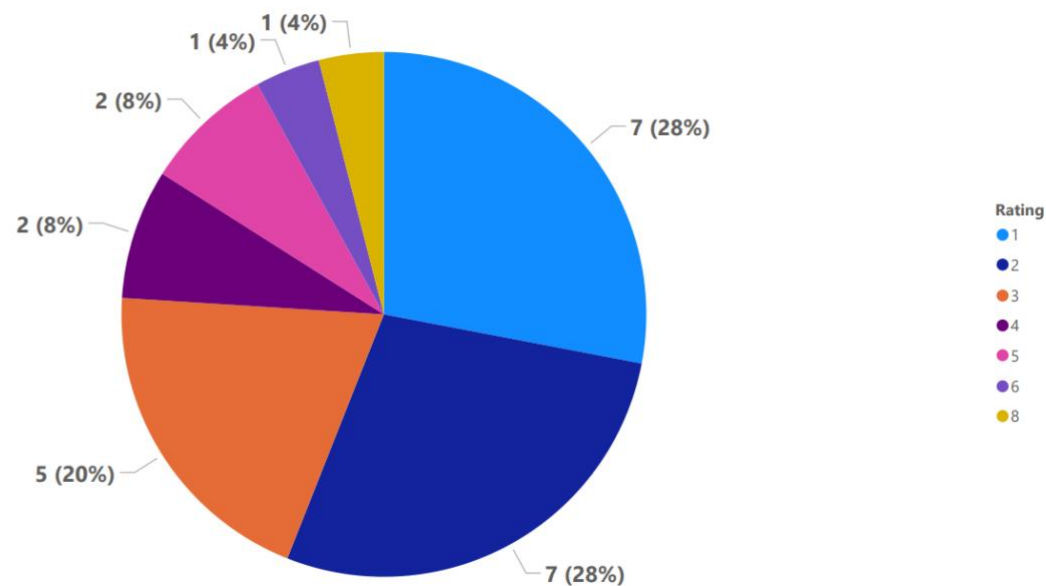


Figure 16 Nasa-TLX Frustration

Figure 17 presents the aggregated results, and after evaluating all categories individually, it summarizes the overall average workload score, offering a holistic picture of perceived task difficulty.

NASA-TLX Average Scores

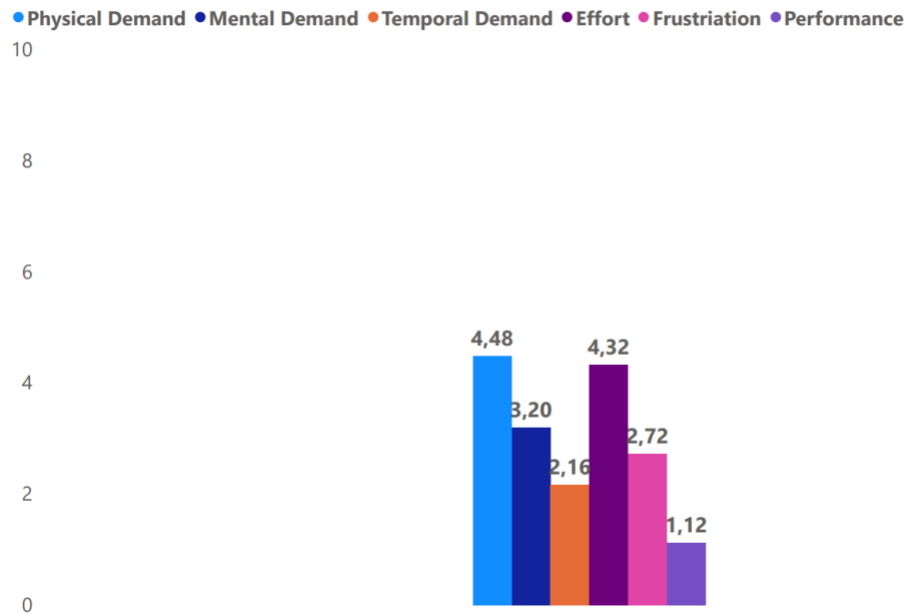


Figure 17 Nasa-TLX Average Scores

Together, these findings confirm that MR training scenarios maintain an optimal level of cognitive engagement, avoiding excessive difficulty while preventing disengagement from overly simple tasks.

Following the workload assessment, the usability of MR was evaluated using the SUS. Table 4 presents the questions used and participants' responses, demonstrating positive evaluations across usability dimensions such as learnability, confidence in use, and overall satisfaction. Based on the SUS categorization Table 8 and as shown in Figure 18, the platform achieved a **mean SUS score of 78**, placing it within the “**Good**” **usability** range, with a majority of participants rating the platform as Excellent or Good. This indicates a user-friendly, intuitive interface that supports efficient task execution and rapid learning.

SUS Categorization

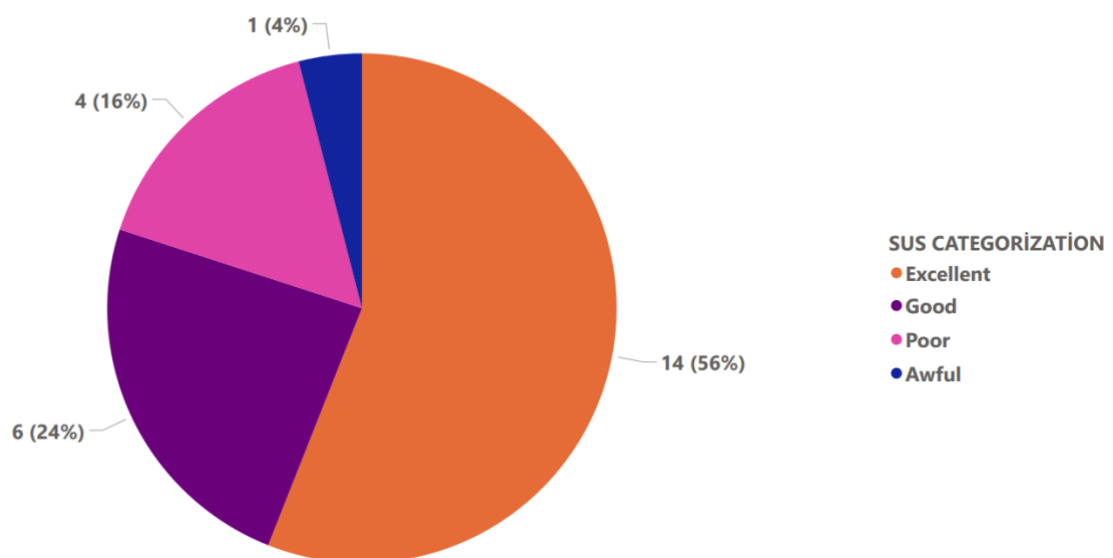


Figure 18 SUS Categorization

4.1.3. Learning Effectiveness: Pre-Test and Post-Test Comparison

The evaluation of MR's training effectiveness followed a structured approach, beginning with detailed pre-tests in both Phase 1 and Phase 2 to establish participant baselines, and concluding with post-tests that measured knowledge gains, procedural confidence, and perceived applicability of training. By connecting these stages, the study provides a clear trajectory of participants' learning experiences and highlights the value of immersive XR training.

Phase 1 – Pre-Test

Before engaging with the XR scenarios, all Phase 1 participants completed a comprehensive pre-test survey, as shown in Table 1, to capture baseline familiarity with robotics, previous XR experience, and expectations from the training. Results showed that **most participants had no previous robotics or XR training experience**, yet they demonstrated strong motivation and high expectations for the XR platform. Many anticipated significant contributions to their professional development, while responses on safety awareness revealed that only a minority felt confident in hazard recognition. These insights underscored the importance of integrating safety guidance and introductory content into the training. The Phase 1 pre-test results thus established a baseline of low prior exposure but high enthusiasm, setting the stage for evaluating learning gains.

Phase 2 – Pre-Test

In Phase 2, the pre-test, as shown in Table 2, targeted a smaller group of highly experienced participants, split equally between MR and traditional training. The survey assessed expectations around clarity, applicability, safety readiness, and problem-solving support. Even for this expert cohort, responses highlighted a strong appreciation for step-by-step guidance and safety preparedness, along with curiosity about the immersive advantages of XR compared to traditional methods. This pre-test established a high baseline of technical knowledge but varying anticipation of interactivity, enabling a meaningful comparison of training methods.

Phase 1 – Post-Test

Following completion of all three scenarios and the scenario creation task, Phase 1 participants provided detailed feedback, as shown in Table 3. Ratings confirmed the training's clarity and applicability (4.56/5), supported by engaging XR interactions (4.20/5), strong motivation (4.08/5), and positive contributions to professional skills (3.76/5). Perceived efficiency was also high (3.92/5). While safety awareness improved, its slightly lower score (3.44/5) indicated a need to reinforce safety messaging further. These findings demonstrate that MR effectively delivered a well-structured and engaging training experience, particularly impactful for participants with limited prior exposure.

Phase 2 – Post-Test

The post-test results in Phase 2, as shown in Table 4, provide a striking contrast between XR and traditional training. After completing the real-world fan replacement task, the XR group consistently rated clarity, applicability, and preparedness higher than their traditionally trained counterparts. Perfect scores in clarity and applicability (5.00 vs. 2.33) and substantial gains in knowledge sufficiency (4.67 vs. 2.67) and problem-solving support (4.33 vs. 2.33) illustrate XR's effectiveness in boosting procedural confidence and readiness. Ratings for future applicability (4.00 vs. 2.00) and training recommendation (4.33 vs. 3.00) further validated MR's advantages. These results confirm that even for experienced practitioners, immersive XR training fosters deeper engagement, more robust skill transfer, and higher confidence in task execution.

4.2. Discussion, Conclusion and Suggestions

The validation results demonstrate that the proposed Multiplayer No-Code Mixed Reality Editor constitutes an effective and robust solution for immersive, skill-based robotic maintenance training. Across both novice and expert user groups, participants demonstrated measurable improvements in procedural performance and task execution efficiency when compared to traditional training approaches (Monetti et al., 2022). In addition, users reported increased clarity and confidence when performing maintenance procedures, suggesting that immersive interaction supports a clearer understanding of task sequences and spatial relationships (Papanastasiou et al., 2022). These findings are consistent with prior studies indicating that XR-based training environments can significantly reduce execution time and improve operational readiness in industrial and robotic contexts, particularly for complex procedural tasks (Monetti et al., 2022).

A central contribution of this study lies in its combined evaluation of usability, cognitive workload, and real-world task performance. The System Usability Scale (SUS) score of 78 places the platform in the “Good” usability category, indicating that the interface design and interaction mechanisms do not introduce barriers to learning (Carvalho et al., 2025). This is a critical factor for XR-based systems, where complex interfaces can otherwise hinder adoption and negatively impact training effectiveness. Complementing this, the NASA Task Load Index (NASA-TLX) results indicate a moderate and well-balanced workload, suggesting that the training scenarios were cognitively engaging without overwhelming users (Grassini et al., 2020; Monetti et al., 2022). Prior research emphasizes that maintaining an optimal cognitive load is essential for effective learning in immersive environments, as excessive mental demand can impair knowledge acquisition and task performance (Brooke, 1996). The observed balance between immersion and workload therefore supports the pedagogical suitability of the proposed platform.

Beyond subjective evaluations, the most notable outcome of the validation is the demonstrated transfer of training benefits to a real-world robotic maintenance task. In Phase 2, expert participants trained using the mixed reality platform completed the physical fan replacement task in substantially less time than those trained through conventional methods. This finding directly addresses a recurring limitation in XR training literature, where many studies focus primarily on usability and user experience without providing empirical evidence of real-world performance improvements (Blattgerste et al., 2023; Stary, 2023). By incorporating a physical robot task conducted in an industrial laboratory environment, the present study enhances ecological validity and provides concrete evidence of skill transfer, aligning with prior experimental work on VR-based training for industrial robotics (Monetti et al., 2022).

The study also highlights the potential of no-code authoring tools to lower the barrier for XR content creation. Previous reviews consistently report that XR authoring workflows remain largely inaccessible to non-programmer instructors and domain experts due to technical complexity and reliance on specialized development skills (Dengel et al., 2022; Hart & Staveland, 1988). In contrast, the results of Phase 1 demonstrate that all participants were able to successfully create functional training scenarios using the proposed visual editor, despite having no prior experience in XR development. Although scenario creation required more time than execution-focused tasks, this difference primarily reflects the cognitive planning and structuring inherent to scenario design rather than usability limitations of the authoring interface itself (Blattgerste et al., 2023). However, the absence of a comparable baseline for traditional XR authoring workflows prevents a quantitative assessment of time savings. This limitation is not unique to the present study, as existing authoring-focused research also notes the lack of standardized benchmarks for measuring XR scenario creation effort, particularly for non-technical users (Blattgerste et al., 2023; Dengel et al., 2022).

Despite these positive outcomes, several limitations should be acknowledged. First, long-term skill retention was not assessed, which limits conclusions regarding the durability and persistence of learning effects over time. While immediate performance improvements were observed, the absence of longitudinal follow-up measurements constrains the ability to determine whether the observed gains reflect genuine learning and internalization of procedural knowledge or short-term recall and reproduction of demonstrated actions, and whether these skills are retained beyond the short-term evaluation period. This limitation reflects a broader methodological challenge in XR training research, where standardized and validated approaches for assessing long-term skill retention remain underdeveloped.

Second, the relatively small sample size in Phase 2 ($n = 6$) restricts the statistical power of comparisons between immersive and traditional training conditions. Small sample size prevents the application of more robust inferential statistical analyses. Future work should therefore involve larger participant groups to enable more detailed statistical testing and to improve the generalizability of findings.

Another limitation concerns the assessment of scenario authoring efficiency. Although participants were able to successfully create training scenarios using the no-code editor, the study lacks a quantitative baseline for comparison with traditional XR authoring workflows. A review of the literature indicates that no widely accepted benchmark currently exists for measuring the time and effort required to author XR training scenarios, particularly for non-technical users. As a result, claims regarding authoring efficiency remain qualitative and would benefit from controlled comparative studies.

Finally, the current platform does not incorporate haptic feedback, which may limit the realism of motor skill acquisition in certain maintenance tasks. While visual and spatial cues were sufficient to support procedural learning in the evaluated scenarios, the absence of tactile feedback may constrain fine motor skill development. Integrating haptic or multisensory feedback modalities therefore represents an important direction for future system development and validation.

In conclusion, this study provides comprehensive validation evidence that the Multiplayer No-Code Mixed Reality Editor is a user-friendly, effective, and scalable training solution for industrial robotic maintenance. By combining immersive multi-user training, accessible no-code scenario authoring, and performance-based validation on a physical robot, the platform addresses several key gaps identified in recent XR training research. Future work will focus on expanding participant numbers, improving statistical rigor through enhanced experimental design, integrating haptic feedback, and conducting longitudinal studies to assess long-term skill retention. These extensions have the potential to further establish the platform as a benchmark solution for XR-based industrial maintenance training.

Acknowledgments

The XR4MCR Project has received funding from the European Union Horizon Europe research and innovation programme under the Open Call #1 – MASTER (Grant Agreement No. 101093079). The validation activities described in this paper were conducted at the Intelligent Factory and Robotics Laboratory (IFARLAB), Eskişehir Osmangazi University, under the supervision of Assoc. Prof. Uğur Yayan, who also provided guidance and expertise throughout the project. Dr. Kübra Yayan contributed to both the validation studies at IFARLAB and the dissemination efforts, ensuring that the project's outcomes reached both academic and industrial audiences. Batur Alp Akdoğan actively participated in the validation process and served as the lead author, coordinating the manuscript's structure and overall content. Yunus Emre Esen led the software development efforts, contributing significantly to system integration and the preparation of this manuscript. The authors acknowledge and thank all contributors for their efforts and collaboration, which were instrumental to the successful execution of this research.

References

- Abidi, M. H., Al-Ahmari, A., Ahmad, A., Ameen, W., & Alkhalefah, H. (2022). A systematic review of the applications of augmented and virtual reality in maintenance. *Virtual Reality*, 26(4), 1435–1453.
- Barbalho, Y. G. S., Silva, C. M. D. S., Fernandes, C. S., & Funghetto, S. S. (2024). Designing and validating Health Unit in Focus: A serious game for enhancing undergraduate education on older adults' health. *Applied Sciences*.
- Blattgerste, J., Behrends, J., & Pfeiffer, T. (2023). TrainAR: An open-source visual scripting-based authoring tool for procedural handheld augmented reality trainings. *Information*, 14(4), 219.
- Brooke, J. (1996). SUS: A quick and dirty usability scale. In *Usability Evaluation in Industry* (pp. 189–194).

- Carvalho, S. A., Conceição, E. S., & Marques, I. C. P. (2025). The impact of virtual reality on employee training and learning in organisations: A systematic literature review. *Applied Sciences*, 15(19), 10459.
- Choi, S., Yoon, J., & Lee, J. (2022). The effects of cognitive load on user experience in virtual reality training environments. *Journal of Educational Technology & Society*, 25(3), 112–125.
- Coupry, C., Richard, P., Bigaud, D., Noblecourt, S., & Baudry, D. (2023). The value of extended reality techniques to improve remote collaborative maintenance operations: A user study. In *Proceedings of the International Conference on Construction Applications of Virtual Reality* (pp. 23–33).
- Dengel, A., Iqbal, M. Z., Grafe, S., & Mangina, E. (2022). A review on augmented reality authoring toolkits for education. *Frontiers in Virtual Reality*, 3, 798032.
- Grassini, S., Laumann, K., & Rasmussen Skogstad, M. (2020). The use of virtual reality alone does not promote training performance (but sense of presence does). *Frontiers in Psychology*, 11, 1743.
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Human Mental Workload* (pp. 139–183).
- Monetti, F. M., de Giorgio, A., Yu, H., Maffei, A., & Romero, M. (2022). An experimental study of the impact of virtual reality training on manufacturing operators on industrial robotic tasks. *Procedia CIRP*, 106, 33–38.
- Papanastasiou, G., Drigas, A., & Skianis, C. (2022). Virtual and augmented reality in industrial and manufacturing training: A systematic literature review. *Education and Information Technologies*, 27(8), 11489–11516.
- Söderlund, H., Zamola, S., Boström, J., Li, D., Mugur, P., Cao, H., & Johansson, B. (2024). The creation of a multi-user virtual training environment for operator training in VR. In *Sustainable Production Through Advanced Manufacturing, Intelligent Automation and Work Integrated Learning* (pp. 173–180).
- Stry, C. (2023). Skill transfer in virtual reality training: A meta-analysis. *Computers & Education*, 205, 104892.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285.
- Wang, P., Wu, P., Wang, J., Chi, H. L., & Wang, X. (2024). A critical review of the use of virtual and augmented reality in construction safety. *Automation in Construction*, 157, 105120.