

Article

Not peer-reviewed version

Usability of Virtual Reality Systems in Engineering Product Development: A Multi-Experiment Evaluation of Software, Hardware, and User Factors

[Ali Abughalia](#) and [Carsten Stechert](#) *

Posted Date: 6 May 2026

doi: 10.20944/preprints202605.0289.v1

Keywords: Virtual Reality (VR); engineering product development; usability evaluation; usability of VR systems; VR in product development; design review; ergonomic evaluation; PC-based vs. standalone VR



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC, OpenAlex.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Usability of Virtual Reality Systems in Engineering Product Development: A Multi-Experiment Evaluation of Software, Hardware, and User Factors

Ali Abughalia and Carsten Stechert *

Ostfalia University of Applied Science

* Correspondence: c.stechert@ostfalia.de

Abstract

This paper investigates how software configuration, hardware type, user background and context of use influence the usability of Virtual Reality (VR) systems in engineering product development. A VR usability assessment approach that combines task-based questionnaires, the System Usability Scale (SUS) and the NASA-TLX questionnaire was evaluated systematically across six experiments involving students, junior engineers and senior engineers in academic and industrial settings. The results demonstrate that user background and task context are at least as significant as the underlying hardware or software in influencing perceived usability and acceptance. Standalone headsets achieve higher usability scores with inexperienced users, whereas PC-based systems are still necessary for high-precision engineering tasks. Professional engineers primarily evaluate VR in terms of workflow integration, precision and return on investment, whereas students focus more on novelty and the interaction experience. Based on these findings, practical design recommendations have been derived for selecting a VR system, adapting interaction concepts, and implementing VR in product development processes. The study highlights that VR should not be deployed as a one-size-fits-all solution, but rather as a tool that is both context-specific and user-centered. It also demonstrates how systematic, iterative usability evaluation can directly support the successful industrial integration of VR technologies.

Keywords: Virtual Reality (VR); engineering product development; usability evaluation; usability of VR systems; VR in product development; design review; ergonomic evaluation; PC-based vs. standalone VR

1. Introduction

In recent years, Virtual Reality (VR) technology has emerged as a valuable tool in product development processes [1]. It has shown potential in supporting various phases of the process, including idea generation [2], design reviews [3], and digital prototyping [4]. The implementation of new technology such as VR in industry, particularly in product development, is typically driven by factors such as improving efficiency [5], enhancing innovation [6], and meeting sustainability goals [7]. Before adopting such technologies, companies must carefully evaluate critical factors to ensure that the decision aligns with their business objectives.

A key step in this adoption process is conducting a cost-benefit analysis to estimate the expected return on investment. It is also crucial to determine where VR should be implemented, which departments will benefit, and what form of the technology best suits the company's requirements [8], for instance, whether VR is implemented to support ideation, design reviews or ergonomic tests. One of the most influential factors in successful implementation is user acceptance. If users reject VR, for instance, due to health concerns, integration becomes significantly more difficult [9]. On the other hand, user resistance caused by factors such as lack of training [10], or skepticism about the technology usefulness [11] can often be addressed through targeted strategies such as training plans.

Another essential consideration is the usability of VR technology. Usability depends on several variables, such as the software and hardware employed [12]. Moreover, variables such as the characteristics of intended user groups and the specific circumstances of use could also influence the evaluation and will therefore be explored in this study. Furthermore, it is essential to determine whether a given VR system is appropriate for a particular application. For instance, standalone VR systems, which offer greater mobility, are more appropriate for field testing and on-site validation, whereas PC-based systems provide higher precision for complex tasks [13]. Similarly, software requirements vary depending on the purpose such as design review or ergonomic analysis and on the roles of the users, whether engineers, clients, or stakeholders.

This paper addresses this gap by conducting six structured usability experiments with three distinct user groups (students, junior engineers and senior engineers) using PC-based and standalone VR systems in academic and industrial product development scenarios. A combined usability evaluation approach (VR-specific questionnaires, SUS, NASA-TLX) is applied to:

1. quantify the impact of software version, hardware type and user background on the usability of VR systems in product development tasks;
2. identify recurring usability issues and missing functions that hinder VR adoption in engineering practice;
3. derive practical recommendations for context-specific VR implementation strategies in product development processes.

2. State of the Art

This chapter provides an overview of the development of VR technology, the current applications and use of this technology in product development processes, and the concept of usability in VR systems, including how it can be evaluated.

2.1. VR technology

VR technology has evolved significantly over time, with much of this progress driven by advances in hardware, particularly improvements in display resolution, graphics performance and design comfort. As a result, several types of VR systems have been developed, each with distinct characteristics and technical configurations.

Virtual Reality is defined as a computer-generated three-dimensional environment that enables users to experience immersive interaction with virtual objects and spaces, often through specialized interfaces such as head-mounted displays and motion-tracking systems, allowing real-time manipulation and exploration of the virtual environment [14].

Based on this definition, three key aspects characterize VR: immersion, interaction, and imagination. Immersion refers to the ability of users to feel present within the virtual environment and become completely separated from the physical world. Interaction describes the ability of users to manipulate and engage with the virtual environment through input devices and software interfaces. Imagination refers to the cognitive ability of users to perceive and interpret virtual elements as meaningful objects within the simulated environment.

These three aspects also distinguish VR from Augmented Reality (AR). While VR separates users from the physical world and provides a fully immersive experience, AR overlays virtual elements onto the real environment. In AR systems, users can perceive and interact with both physical and digital elements simultaneously. Consequently, the reliance on imagination is reduced, as virtual objects are directly integrated into the real-world context.

VR systems typically consist of several core components. The primary component is the head-mounted display (HMD), which renders three-dimensional visual content to the user. In addition, tracking systems such as sensors, cameras, or base stations are used to determine the position and orientation of the user in space. Interaction devices, commonly handheld controllers, allow users to manipulate objects and interact with the virtual environment.

The development of VR headsets has primarily focused on improvements in display resolution, field of view, refresh rate, and device weight. These factors directly influence the visual quality of the virtual environment and contribute to a higher level of immersion and improved user comfort.

Tracking technologies in VR systems are generally categorized into two main approaches: outside-in tracking and inside-out tracking. In outside-in tracking systems, external sensors or base stations (often referred to as “lighthouses”) send and receive signals to and from receivers integrated into the headset or controllers in order to determine their position and orientation. These systems require calibration between the tracking devices and the interaction hardware before use. Outside-in tracking is commonly employed in PC-based VR systems.

In contrast, inside-out tracking integrates the tracking components directly into the headset. These systems typically use cameras to detect visual features in the environment, such as edges, corners, or objects. Additionally, inertial measurement units (IMUs), which include gyroscopes and accelerometers, measure rotational and linear movements and provide rapid motion data. By combining visual and inertial data through Simultaneous Localization and Mapping (SLAM) algorithms, the system continuously maps the surrounding environment and determines the headset’s position within that map. Inside-out tracking is commonly used in standalone VR systems.

Interaction in VR environments is primarily achieved through handheld controllers. These devices enable hand tracking, haptic feedback, and user input. Controller designs vary between hardware providers, and more intuitive interaction devices generally contribute to a better user experience.

VR software provides a wide range of functionalities that allow users to interact with virtual environments. One fundamental interaction method is navigation, which enables users to move through the virtual space without physically walking long distances. This is commonly implemented through teleportation mechanisms controlled via the VR controllers.

Additional functionalities vary depending on the software provider and the application domain. In engineering-focused applications, functions such as measurement tools, section views, component assembly and disassembly, and object manipulation are particularly important. In contrast, entertainment-focused applications emphasize gameplay mechanics, interaction with virtual objects, and exploration of virtual environments. Consequently, each application domain requires a specific set of VR functionalities tailored to its objectives.

The two VR system types of particular relevance to this study are PC-based VR systems and standalone VR systems. PC-based VR systems include head-mounted displays such as the HTC Vive or Oculus Rift. These systems require external hardware, including a high-performance computer and tracking devices. After the hardware components are installed and calibrated with the software, users must remain within the designated tracking area. Relocating the system typically requires a new setup and calibration process.

The second type is standalone VR headsets, such as the Meta Quest series. In these systems, all tracking components and processing capabilities are integrated into the headset itself. This eliminates the need for external hardware and enables greater portability and ease of use, as the system can be deployed without complex installation procedures.

Both systems are commonly implemented in product development, depending on the development tasks and the objectives that need to be achieved.

2.2. VR in the Product Development Process

In the product development process, VR is implemented selectively and supports certain phases more effectively than others. For instance, during the idea generation phase, VR assists designers by enabling them to visualize and represent their concepts. At this stage, designers are able to sketch or model ideas in three-dimensional form. However, certain inefficiencies remain, such as limited measurement precision [2].

Another key area where VR is applied is the design review phase. Engineers are expected to use VR to collaboratively examine and evaluate a design in an immersive environment. This provides a

near-realistic representation of the product, which is particularly beneficial for distributed teams and for reviews involving both engineers and stakeholders. Such immersive reviews help verify design intent and ensure that requirements are addressed early in the development cycle.

VR is also valuable in the field of ergonomic evaluation. Instead of building costly and time-consuming physical prototypes [15], digital prototypes in VR allow teams to assess aspects such as reachability, accessibility, and visibility [16]. While VR currently supports only a limited range of ergonomic factors, it remains a useful tool for envisioning how a product will be experienced and interacted with by users [17].

While VR offers clear benefits across multiple phases of the product development process, its actual implementation depends on a range of organizational, technological, and environmental considerations that shape adoption decisions [18].

One important determinant in implementation decisions is financial considerations [19]. Beyond the potential added value of VR, organizations are required to evaluate the direct costs of acquisition, implementation, and personnel training against the expected return on investment (ROI) within a strategic or tactical timeframe [20]. The balance between costs and anticipated profitability often determines whether VR integration proceeds.

Another factor frequently discussed in the literature is company size. Some studies suggest that large enterprises are more inclined to adopt innovative technologies [21] because they possess greater resource availability and a stronger capacity to absorb risks associated with failure such as unsuccessful implementation or low ROI [22]. Conversely, other opinions highlight the advantages of medium-sized enterprises, which often demonstrate higher organizational agility and greater flexibility in restructuring processes and training personnel [22]. Small enterprises tend to benefit from relatively straightforward adaptation processes, but the financial risks associated with failed adoption are often disproportionately high [23].

Closely related to organizational size is management orientation. Leadership awareness such as openness to innovation plays a decisive role in adoption decisions. The environmental consciousness and digital expertise of leadership influence decisions regarding the adoption of new technologies [24].

In addition, competitive pressures from both industry rivals and customers act as important external drivers of adoption. Market trends, regulatory requirements (e.g., safety standards), and consumers' expectations for higher quality or advanced features often accelerate technology integration.

One of the critical factors influencing the integration of VR as a new technology in the product development process is user acceptance. Drawing on the Technology Acceptance Model (TAM) from [25], user acceptance is primarily shaped by two constructs: perceived usefulness and perceived ease of use. Users must believe that VR contributes meaningful value to their tasks and that its use does not impose an additional cognitive or operational burden.

Ease of use is enhanced through targeted training, while perceived usefulness is strongly influenced by the system's overall usability. Perceived usefulness refers to the extent to which users believe that a technology improves their performance, for example by reducing effort or increasing efficiency. Thus, achieving a high level of usability is essential to ensure strong perceptions of usefulness, which in turn supports the adoption process [26].

2.3. Usability of VR Systems

Usability refers to how effectively and efficiently users are able to accomplish their tasks using a system. According to the ISO 9241 standard, usability is defined as "the extent to which a system, product, or service can be used by specified users to achieve specified goals effectively, efficiently, and satisfactorily in a specified context of use."

While several usability evaluation methods exist, such as ISO 9241, System Usability Scale (SUS), and NASA Task Load Index (NASA-TLX), none of them are specifically designed for evaluating the usability of VR systems.

- ISO 9241 is a set of international standards that provides guidelines for the usability and ergonomics of human–system interaction, including software and hardware user interfaces.
- NASA-TLX (Task Load Index) is an assessment tool developed by NASA to measure perceived workload across different tasks.
- The System Usability Scale (SUS) is a questionnaire-based tool used to evaluate the usability of a wide range of products and systems, including software, websites, mobile applications, and hardware devices. It consists of ten statements that users rate on a Likert scale.

To address this gap, a long-term project was conducted to develop a suitable evaluation approach for assessing the usability of VR systems. This approach is presented in [26]. The methodology integrates elements from established usability evaluation standards and employs a tailored questionnaire divided into two categories. The first category is an inspection-based questionnaire designed for participants who observed the VR system without direct interaction. The second category is an empirical questionnaire intended for participants who actively engaged with and operated the system in a practical context.

The usability results are presented across seven dimensions that characterize the usability of a VR system (Figure 1): task appropriateness, expectation conformity, error tolerance, learnability, self-descriptiveness, controllability, and user commitment. Each dimension consists of several questions from two categories (inspection, empirical) that can be answered with either “yes” or “no.” This structure facilitates the calculation of the proportion of “yes” and “no” responses provided by all participants. Based on these responses, the usability proportion for each dimension can be determined. Finally, the overall usability degree is calculated by averaging the results across all seven dimensions.

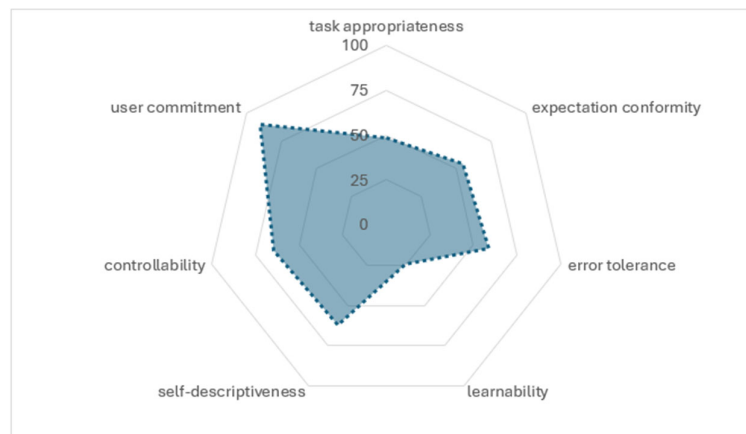


Figure 1. The seven dimensions of usability.

In addition, the evaluation approach incorporates two standardized assessment tools. First, the NASA Task Load Index (NASA-TLX) is used to measure the subjective workload and stress experienced by participants while performing specific tasks. Second, the System Usability Scale (SUS) is applied to assess the usability of the system by asking participants to rate a set of statements on a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The SUS evaluates several aspects of the user experience, including perceived usefulness, complexity, ease of use, the need for support, feature integration, learnability, and operational comfort.

The evaluation methods, processes, and resulting metrics used in this study are summarized in Table 1.

Table 1. The combined method used to derive the usability results.

Method	Process	Results
Survey	Survey calculation (empirical, inspection)	Seven dimensions of usability
Usability dimensions	Average calculation	Overall usability degree
NASA TLX	Responses review	Workload index
SUS	Score calculation	System usability score

Another essential factor in evaluating the usability of a system is the number of participants involved in testing. This number directly influences the problem discovery rate, i.e., the proportion of usability issues identified during the evaluation process.

Increasing the number of participants generally increases the number of problems discovered. However, the return on investment (ROI) of usability testing depends on balancing the number of participants and test iterations, since additional participants yield diminishing returns while consuming unnecessary resources and effort [27].

The proportion of discovered problems is estimated using the following model from [28]:

$$P = 1 - (1 - p)^n \quad (1)$$

Where:

P = proportion of problems discovered,

p = probability of discovering a single problem (problem discovery rate),

n = number of participants.

The problem discovery rate (p) is calculated as:

$$p = \frac{\text{Number of unique problems detected by one participant}}{\text{Total number of problems identified by all participants}} \quad (2)$$

When the required proportion of discovered problems (P) is predetermined, the equation (1) can be rearranged to estimate the necessary number of participants:

$$n = \frac{\ln(1-P)}{\ln(1-p)} \quad (3)$$

This formulation enables usability practitioners to determine the number of participants required to achieve a desired problem discovery rate. Conversely, for a specific number of participants, it allows estimation of the proportion of usability issues likely to be identified. The unknown parameter p can be determined through several approaches.

The first approach involves using an empirically derived value for p based on prior studies, such as [29], which found that the average probability of discovering a given usability problem during testing is approximately 0.31, the value denoted in their study by the symbol λ .

Another approach is to obtain the value p from a pilot testing. In this way, an initial usability test is conducted with a small group of participants (e.g., three to five), and all unique problems identified are recorded. The average number of unique problems discovered per participant is then divided by the total number of problems to estimate p . For example, if the participants collectively identify 20 distinct problems, and each participant discovers an average of four unique issues, then $p = 4/20 = 0.20$. This method is particularly useful for minimizing testing effort and for tailoring usability evaluations to specific participant groups. In such cases, practitioners typically conduct only two rounds of testing: a pilot test followed by a final test.

3. Methodology and Experiment Setup

In this section, six experiments were conducted to evaluate the usability of the VR system using the evaluation framework originally introduced in [30]. It specifies a five-step procedure for conducting VR usability tests in product development contexts:

1. First, the objective of the experiment and the tasks to be performed must be defined, such as conducting ergonomic evaluations or design reviews.
2. Second, the designated circumstances for the target group must be specified. In this step, the characteristics of the participants are defined according to factors such as level of experience, age, gender, familiarity with digital tools, professional background, number of participants, test location, and the hardware used.
3. Third, the specific circumstances of the actual test participants are recorded. These are later compared with the predefined target conditions in order to identify and quantify any deviations.
4. Fourth, the usability test is executed. During this phase, inspectors complete the inspection questionnaire, while the immersed participants complete the empirical questionnaire.
5. Finally, the responses of the participants are evaluated by analyzing the seven usability factors together with the results of the TLX and the SUS.

Across the six experiments, several common experimental conditions were maintained. A single VR software application was used throughout the project; however, it was continuously improved through a series of updated versions, with each updated version being implemented and evaluated in a subsequent experiment.

Two types of hardware were used in the experiments: a standalone headset (H1) and a PC-based headset (H2). The experiments also involved three different user groups:

- U1: bachelor's students in mechanical engineering without practical experience.
- U2: junior engineers with early industry experience.
- U3: senior engineers from an industrial development team.

The tasks performed in the experiments involved testing various VR application scenarios, such as ergonomic evaluations and design reviews. Each experiment aimed to optimize a specific phase or task within the product development process through the use of VR. Furthermore, the proportion of discovered usability problems in each experiment was calculated using Equation (1), which estimates the number of problems P identified by the participating users, assuming a detection probability of $p = 0.31$.

3.1. Experiments

In this section, each experiment and its setup are reviewed in order to clarify the respective objectives and the tested scenarios. Furthermore, the section explains how these objectives are achieved through planned tasks derived from product development activities and executed within the VR environment.

Experiment I

The objective of this experiment was to conduct a design review of a timekeeper, a cyber-physical device used for measuring disruption times, with dimensions of 10 cm × 10 cm (Figure 2). The purpose of the design review was to verify the design accuracy, support quality assurance, and improve traceability for future development iterations within the product development process.

To achieve this objective, participants performed several review tasks in the virtual environment. These tasks included dimensional measurements, geometric validation, assessment of surface characteristics and side counts, and systematic documentation of identified findings using a digital checklist workflow integrated into the VR system.

The experiment was conducted using the developed first version of the VR software and involved 23 participants divided into two groups based on the hardware configuration used. One group performed the experiment using a standalone VR headset (H1), while the second group used a PC-connected VR headset (H2). This setup allowed the evaluation of the system across different hardware platforms and interaction conditions.

Before starting the experiment, participants received a short introduction to the VR system and its interaction methods. They were then given a brief familiarization phase to practice navigation and

object interaction in the virtual environment. After this introduction, participants performed the assigned design review tasks individually.

During the experiment, participants used the VR system to inspect the virtual model of the timekeeper and identify potential design issues according to the defined checklist. The identified problems were recorded and later analyzed to determine the number of usability or design-related issues detected by the participants.

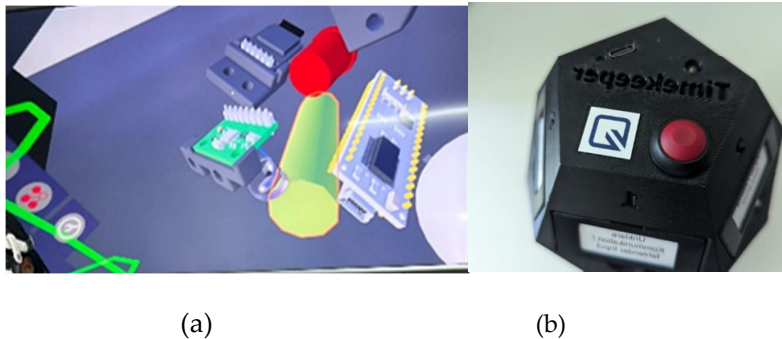


Figure 2. Timekeeper: (a) virtual representation (interior view); (b) physical device (exterior view).

Experiment II

The objective of this experiment was to evaluate the ergonomic and functional design of a train interior (Figure 3) using VR system. The primary goal of the design review was to identify potential improvements related to passenger comfort, accessibility, and spatial efficiency in future train cabin configurations.

To achieve this objective, participants interacted with a virtual model of the train interior and performed several evaluation tasks designed to simulate typical passenger activities and spatial interactions. The evaluation focused on key aspects of interior usability, including passenger comfort and available movement space, compatibility of storage areas with passengers' personal belongings, clarity of user orientation within the cabin, aesthetic perception of the environment, as well as safety and accessibility during passenger movement.

The experiment was conducted using an updated version of the VR software application, in which the usability issues identified during experiment I had been addressed and corrected. The improved software version aimed to enhance interaction efficiency, navigation stability, and measurement accuracy. The VR environment was operated using the PC-based headset (H2).

A total of 10 participants took part in the experiment. All participants were bachelor's students enrolled in the Ergonomics and Industrial Design course, representing users with theoretical knowledge of ergonomic principles but limited professional experience.

Before starting the experiment, participants received a short introduction to the VR system and were given time to familiarize themselves with the navigation and interaction methods. After this familiarization phase, each participant individually explored the virtual train interior and performed the assigned evaluation tasks.

During the session, participants inspected the spatial configuration of the interior, assessed comfort and movement possibilities, and identified potential design limitations. Observations and identified issues were documented using the integrated digital checklist system within the VR environment. The collected data were later analyzed to evaluate the usability of the VR-based ergonomic assessment approach.

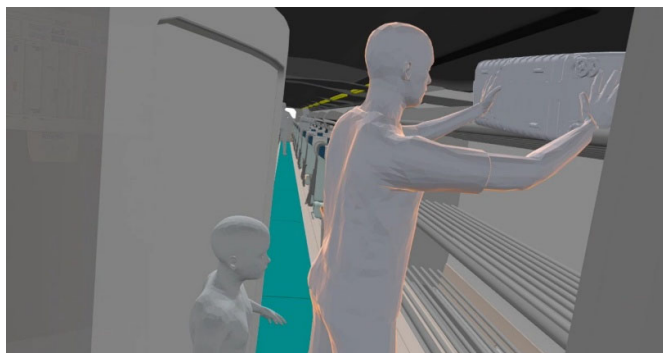


Figure 3. Evaluation of the ergonomic and functional design of a train interior.

Experiment III

The objective of this experiment was to conduct a collaborative design review of a cutting machine (Figure 4) using the multi-user functionality of the developed VR system. The purpose of this test was to evaluate how effectively development teams could perform design review within separated teams.

The review session was conducted as an online multi-user VR meeting, where several participants were connected simultaneously to the same virtual session. This setup allowed participants to collaboratively inspect the machine model, discuss design aspects in real time, and identify potential design improvements.

The evaluation focused on several important aspects related to the machine's operational and ergonomic performance. These included the analysis of material flow and operator ergonomics, the inspection of safety mechanisms and guarding elements, the accessibility of maintenance components, and the evaluation of loading and operational procedures during machine use.

The experiment utilized a further updated version of the VR software, building upon the improvements implemented in experiment II. This version additionally incorporated multi-user communication and synchronized interaction features, enabling real-time collaboration between participants. The system was operated using the PC-based headset (H2).

A total of five participants took part in the experiment. All participants were engineers from a product development team in a company specializing in freezer-cutting machines. Their professional background provided practical industry experience relevant to machine design, safety evaluation, and production processes.

Before the collaborative review began, participants received a short briefing on the VR system and the multi-user interaction features. Once connected to the virtual environment, participants explored the cutting machine model, discussed design elements, and identified potential issues related to ergonomics, safety, and operational efficiency.

Throughout the session, identified findings were documented and later analyzed to assess the effectiveness of the VR-based collaborative design review process within an industrial development context.



Figure 4. Example of a representative cutting machine as used in the design review [31].

Experiment IV

The objective of this experiment was to evaluate the routing of cables on the roof of a regional train using the developed VR system. The purpose of this assessment was to analyze the cable layout and identify potential spatial limitations while ensuring compliance with technical requirements such as minimum bending radii and safe separation distances between cables and surrounding components.

To achieve this objective, participants interacted with a virtual model of the train roof assembly containing the cable routing configuration. The evaluation focused on several key aspects of the installation and maintenance process. These included verifying that the cables could be installed smoothly without physical obstructions, confirming that assembly and maintenance procedures could be performed practically, preventing overcrowding within the cable routing paths, and ensuring compliance with relevant safety and technical standards.

The experiment was conducted using a developed version of the VR software, which incorporated additional improvements based on feedback and observations from the previous experiments. These improvements primarily focused on enhancing interaction stability, learnability, and the visualization of complex assemblies. The VR environment was operated using the PC-based headset (H2).

A total of five participants took part in the experiment. All participants were engineers from a product development team in a company operating in the railway industry, providing practical experience in train system design and technical installation processes.

Before starting the experiment, participants received a brief introduction to the VR system. However, they did not engage in any prior practice interaction because the available time for the employees was insufficient. After the introduction, participants individually inspected the cable routing configuration within the virtual environment. During the inspection, they assessed spatial feasibility, installation accessibility, and potential design issues related to safety and maintenance. Identified findings were documented and later analyzed to evaluate the effectiveness of the VR-based cable routing assessment approach.

Experiment V

The objective of this experiment was to investigate how VR tools can support and enhance design evaluation processes, particularly for assessing ergonomic and spatial characteristics of a train interior (Figure 5). The objective was to determine how effectively VR could be used to identify design issues related to accessibility, passenger interaction, and spatial configuration.

Participants interacted with a detailed virtual representation of the train interior and performed several evaluation tasks designed to simulate realistic user interactions within the cabin environment. The evaluation focused on multiple aspects of the design, including the visual inspection of internal

components and their accessibility, passenger movement and safety considerations, accurate dimensional analysis of the interior space, inspection of internal structural elements, and assessment of storage requirements and passenger behavior patterns.

The experiment utilized a modified version of the VR software developed in the previous experiment, incorporating additional improvements to visualization and interaction functions. Unlike the previous experiments, the system was operated using a standalone VR headset (H1), enabling a more flexible and portable VR setup.

A total of 40 participants took part in the study. The participants were junior engineers with limited industry experience, yet they possessed a solid background and familiarity with engineering design concepts.

Before starting the experiment, participants received a short introduction to the VR system and were allowed time to familiarize themselves with the navigation and interaction mechanisms. After this training phase, each participant individually explored the virtual train interior and completed the assigned evaluation tasks.

During the experiment, participants inspected the design from different viewpoints, assessed ergonomic aspects of the interior layout, and identified potential design improvements. Observations and identified issues were recorded using the integrated documentation tools of the VR system.



Figure 5: Evaluation of the ergonomic and spatial characteristics of a train interior.

Experiment VI

The objective of this experiment was similar to that of experiment V, focusing on the evaluation of ergonomic and spatial aspects of a train interior using the VR system (Figure 6). However, in this case the tasks were designed to be more complex and rigorous, providing a deeper assessment of the participants' ability to identify design issues within the virtual environment.

Participants were required to perform the same evaluation activities as in the previous experiment, including visual inspection of internal components, assessment of passenger movement and safety, dimensional analysis, structural inspection, and evaluation of storage requirements. The increased level of difficulty resulted from the fact that the experiment was conducted as part of a graded academic course requirement, requiring participants to perform a more detailed and systematic evaluation.

The experiment was conducted using a new version of the VR software, which incorporated additional refinements to improve usability and interaction performance. The system was again operated using the standalone VR headset (H1).

A total of ten participants took part in the experiment. The participants were bachelor's students enrolled in the Ergonomics and Industrial Design course, representing users with foundational knowledge of ergonomic analysis and product evaluation.

Before beginning the evaluation tasks, participants received an introduction to the VR system and completed a short familiarization session. They then individually performed the assigned tasks within the virtual train interior model.

During the experiment, participants analyzed the design in detail and documented any detected issues related to ergonomics, accessibility, spatial arrangement, or structural design.



Figure 6. Evaluation of ergonomic and spatial aspects of a train interior using the VR system.

An overview of the experiments, the tested scenarios, and the objectives of the evaluations is presented in Table 2.

Table 2. Overview of the application scenarios and evaluation tasks across all experiments.

Experiment	Application Scenario	Main Evaluation Focus
I	Design review of a Timekeeper device	Dimensional measurement, geometric validation, surface characteristics, documentation workflow
II	Ergonomic evaluation of a train interior	Passenger comfort, storage compatibility, orientation and aesthetics, safety and accessibility
III	Multi-user design review of a cutting machine	Material flow, ergonomics, safety mechanisms, maintenance accessibility, operational procedures
IV	Cable routing evaluation on the roof of a regional train	Cable layout feasibility, bending radii, safety distances, installation and maintenance procedures
V	Ergonomic and spatial evaluation of a train interior	Component accessibility, passenger movement, dimensional analysis, structural inspection, storage behavior
VI	Advanced ergonomic and spatial evaluation of a train interior	Same evaluation aspects as experiment V with increased task complexity and graded assessment

4. Analysis of the Participants' Responses

In this section, all experiments and their results are analyzed and discussed in detail. The objective of this section is to understand how each usability factor affects the overall acceptance of the system. In this analysis, the usability factors are the usability dimensions, SUS and TLX.

Experiment I

The first experiment investigated the usability of the VR system in a cross-cultural context, involving junior Engineers from two countries. The objective of this evaluation is to examine how cultural differences between teams influence the usability of the VR system for both inspectors and actively involved users. A total of 23 engineers participated in the usability test, conducted in collaboration with Ostfalia University of Applied Sciences (Germany) and Tshwane University of Technology (South Africa).

The overall usability average across both groups was 51.7%, with a notably low score in the learnability dimension, indicating challenges in system understanding. The SUS scores were comparable between universities, averaging 3.17 and 3.05, suggesting a consistent perception of usability across cultural groups. These findings indicate that, although the system was generally

usable, users experienced difficulties in quickly understanding and learning how to interact with the software effectively.

During the inspection, both participant groups provided largely consistent responses, indicating a high degree of objectivity in the evaluation process. The analysis confirmed that the VR system provided the core functionalities required to perform the assigned design review task, including object grouping, model scaling for inspection, and the selection of individual components within the virtual environment.

Despite this, the inspection phase also revealed several limitations. Specifically, the lack of supporting features such as error feedback and instructions to fix these errors. These missing elements were recognized as potential areas to be improved.

In contrast to the inspection results, the empirical survey revealed noticeable differences between the two participant groups. User responses showed variations in user interactions, particularly regarding the tasks completeness. While participants from one group generally considered the available functions sufficient for the assigned tasks, participants from the other group indicated the need for additional features. These differences may be partially attributed to variations in prior experience with VR technologies, which in turns influence user expectations and evaluation criteria.

The TLX results further highlighted differences in user opinions, particularly with regard to generally satisfaction. While some participants reported satisfaction with their performance, others expressed lower levels of satisfaction through challenges in system interaction.

In summary, the results of experiment I indicate that the used VR system provides the essential functionality required for design review tasks and is generally perceived as usable across different cultural contexts. However, limitations related to system learnability, task appropriateness, and self-description of the software were identified. These findings highlight the need for improvements in interface design and user support mechanisms to enhance usability and ensure a more consistent user experience across various user groups.

Experiment II

The second usability experiment focused on how users interacted with the VR system during their first encounter with a set of predefined, product-related tasks. The main aim was to examine the initial user experience, with particular emphasis on usability, interaction behavior, and perceived workload, rather than on task efficiency or complete functional performance. The study involved students enrolled in an ergonomics course, who completed structured tasks in a virtual train model. These tasks included object inspection, taking measurements, navigating the environment, and using basic interaction functions.

Overall, the usability evaluation indicated a higher level of usability than in the first experiment with an average score of 62.3%. The SUS resulted in a mean value of 3.18 on a five point scale, suggesting a generally acceptable usability perception among participants. While users were able to complete the assigned tasks successfully, several usability aspects revealed opportunities for improvement, particularly regarding system learnability and the clarity of certain interaction mechanisms. Analysis of the responses showed that several core interaction functions were clearly recognized by the participants. Object selection and manipulation were identified as available and simple, suggesting that the software supports interaction tasks required. In addition, the system response was perceived positively, as users experienced immediate feedback following their actions. However, other functional aspects revealed noticeable uncertainty among users. Features, such as error handling or object grouping, were not clearly recognized by many participants, and a number of users reported difficulties in evaluating these features. This suggests that such features were either not sufficiently visible within the interface or were not required during the assigned tasks. Similarly, system status information, such as battery status, was not noticed, indicating limitations in interface transparency.

The evaluation of usability dimensions based on established principles revealed mixed results across all categories. Task suitability was generally perceived positively, particularly with regard to

the availability of relevant functions for completing the as-signed tasks. However, supporting elements such as error messages or contextual help were considered less effective, indicating a need for improved user guidance during task execution. In terms of expectation conformity, the interface design was largely perceived as unintuitive. Menu structures and visual elements did not align with user expectations, and several graphical representations were difficult to understand. Learnability emerged as one of the weaker aspects of the system. Many participants experienced difficulties identifying features related to system guidance, or preview functions. Visual orientation within the interface was not consistently clear, indicating that first-time users may require additional instructional support or guided interaction mechanisms. Similarly, error tolerance and system controllability were not clearly perceived by users. Functions such as 'undo' or alternative input methods were not widely recognized, suggesting that these features were either insufficiently communicated or not encountered during the experimental tasks. Despite these limitations, several usability aspects received positive feedback. The system was generally perceived as self-descriptive, with users reporting a clear sense of control during interaction and an adequate understanding of icons. Furthermore, user engagement was notably high, as participants expressed a positive initial impression of the software. The results of the NASA TLX indicated a manageable level of workload during task execution. Participants described the tasks as moderately demanding, primarily due to the novelty of the VR environment and unfamiliar interaction techniques. Physical workload was perceived as low, and time pressure was considered appropriate for the experimental setup. Emotional responses varied, with some users reporting satisfaction and a sense of accomplishment, while others experienced temporary uncertainty, particularly when interacting with unfamiliar system features.

In summary, the results of experiment II indicate that the VR system provides a generally positive first user experience with moderate usability and manageable workload. Core interaction functions performed effectively and supported task completion. However, several usability aspects, including learnability, expectation conformity, and error tolerance, require further optimization to improve overall usability and reduce uncertainty for users with limited practical experience.

Experiment III

The third experiment investigated the usability of the VR system within a real industrial context, focusing on a multi-user design review of a cutting machine. The evaluation was conducted with experienced engineers and emphasized collaborative interaction, technical inspection, and ergonomic assessment like reachability aspect within a virtual environment. The targeted outcomes of this experiment focus on evaluating the usability of the VR system from the perspectives of only active users.

Overall, the findings suggest that the VR system is generally usable with 50.3% and accepted in a design review, although several usability limitations remain. The SUS results indicate a moderate usability score 2.77. The system was perceived as relatively simple to operate, with users indicating that most functions could be learned quickly. At the same time, the willingness to use the system regularly was rated higher, suggesting that further improvements are required to achieve long-term adoption.

The empirical evaluation identified several missing or insufficiently implemented features, particularly in the areas of learnability and error tolerance. Key shortcomings included the absence of flexible error message handling, lack of visible system status indicators (e.g., controller status), missing diagnostic tools, and the absence of 'undo' functionality. In addition, the system did not provide clear previews of actions or sufficient visual cues to indicate menu hierarchy levels. These limitations negatively affect the transparency of the system and increase the cognitive effort required for task execution.

The empirical questionnaire results were limited due to the small number of valid responses. As a result, no comprehensive quantitative conclusions could be drawn for most usability dimensions. However, a positive tendency in self-descriptiveness and user engagement was observed, indicating that participants recognized the potential value of the VR system for collaborative engineering tasks.

In addition, the CEO of the company stated clearly that they are planning to implement VR in their design-review process with clients, because it provides more clarity and allows clients to familiarize themselves with the machine, especially those who may not be able to understand CAD designs.

The NASA TLX results indicate that the perceived workload during task execution was generally low to moderate. Task complexity was rated between simple and moderately complex, primarily due to limited prior experience with VR systems and insufficient preparation. Physical workload was perceived as low, confirming that interaction with the VR system did not impose significant physical strain. Time pressure was not considered an issue by any participant, with the task pace described as appropriate or even slow.

User satisfaction and perceived performance varied among participants. While some users reported that tasks were easy and understandable, one participant found them more challenging and indicated that their performance could be improved with additional training. Overall effort was rated as low, although one participant reported difficulty in reaching their desired performance level. Emotional responses were mostly positive, with participants generally feeling relaxed; however, minor stress and frustration were reported in relation to technical issues such as audio communication problems and occasional system instability during the multi-user session.

Despite these limitations, participants demonstrated active engagement with the VR system and were able to complete the assigned collaborative tasks. The multi-user functionality enabled effective communication and joint inspection of the virtual model, highlighting the potential of VR for distributed design reviews. At the same time, the identified usability issues such as system feedback, learnability, and technical reliability, indicate that further refinement is necessary to ensure consistent performance in professional environments.

In summary, experiment III demonstrates that the VR system is functionally applicable and positively perceived in an industrial multi-user design review scenario. However, improvements in system robustness, user guidance, and feature transparency are required to enhance usability. These aspects will be taken into account in the next version of the software.

Experiment IV

The fourth experiment evaluated the application of the VR system within another real industrial design review scenario, focusing on the analysis of cable routing on the roof of a regional train. The objective of this evaluation was to assess system usability in a professional engineering context, with special focus on interaction quality, task support, and user acceptance in comparison to traditionally CAD software. The targeted outcomes of this evaluation involve a detailed analysis of the seven usability dimensions, as the experiment was conducted with an experienced industrial team.

Overall, the results indicate a moderate level of usability, with an assessed usability score of 54.5%. The SUS yielded a value of 2.9 on a five point scale, reflecting a rather critical perception of the system among professional users. Although participants were able to complete the assigned tasks, the results highlight several usability limitations that negatively affected efficiency, intuitiveness, and overall acceptance.

From a functional perspective, the system demonstrated strong capabilities in visualization and object interaction. However, ideal precision was not achievable with the available headset at the time. The participants were development engineers who conducted a design review in VR, following their usual review practices. Their feedback was strongly influenced by comparisons between VR and the CAD software they typically used. Many participants initially resisted the VR technology, citing the need for additional training before adoption. One notable comment from the team leader was:

“If I have to invest more money and time to prepare the workforce and adapt the process to implement a new software that only supports one phase of the process, while I can already perform all tasks with the current software, then I do not need it.”

Participants evaluated system responsiveness and the ability to manipulate and inspect complex geometries within the virtual environment negatively. Users reported insufficient flexibility in object selection and difficulties related to controller input, which reduced interaction efficiency.

The analysis of responses revealed deficiencies in system self-descriptiveness. While basic feedback mechanisms were present, important system states such as controller status or active interaction modes, were not continuously visible. This lack of transparency led to uncertainty during task execution, particularly when switching between different tools or interaction modes. In addition, inconsistencies in interaction logic were identified, as users were required to manually deactivate functions before activating new ones, which does not align with typical user expectations.

Evaluation of usability dimensions showed a mixed performance across categories. Task suitability was generally rated as adequate, as the system provided the core functions required for the design review tasks. However, the lack of supporting features, such as advanced measurement tools and precise representation of cable radii, limited the effectiveness of the system for detailed engineering analysis. This technical limitation had a direct negative impact on user trust and perceived reliability of the VR model.

Expectation conformity was only partially fulfilled. While some interface elements, such as menu structures and visual design, were considered understandable, the overall interaction concept was perceived as non-intuitive. Participants indicated that additional training would be required before the system could be effectively integrated into existing workflows.

The learnability of the system was identified as a critical weakness. Although some visual cues, such as color coding and icons, supported user orientation, these were not sufficiently clear or consistent. The absence of preview functions and limited guidance mechanisms made it difficult for users to anticipate the outcome of actions, increasing cognitive effort during task execution.

Similarly, error tolerance and controllability were limited. The system lacked essential features such as undo functionality and diagnostic feedback, restricting users' ability to recover from mistakes. While most participants were eventually able to perform the required interactions, the process was often inefficient and required additional effort.

User satisfaction results reflect these usability challenges. While participants acknowledged the high potential of VR for immersive visualization and collaborative design reviews, they also emphasized that the system is currently less efficient than conventional CAD tools. Resistance to adoption was observed, particularly from a managerial perspective, where the additional effort required for training and process integration was perceived as a barrier.

The NASA TLX results indicate a moderate workload. Physical demand was generally low, confirming that VR interaction does not impose significant physical strain. However, cognitive load and frustration levels were elevated in some cases, mainly due to interaction difficulties and system limitations. Time pressure was not considered a significant issue.

In summary, the results of experiment IV demonstrate that the VR system offers strong advantages in terms of visualization and spatial understanding, particularly for large and complex models. However, limitations in usability, interaction design, and technical accuracy significantly affect user efficiency and acceptance in a professional engineering context. To enable successful integration into industrial workflows, improvements are required in system intuitiveness, feature completeness, and reliability, as well as in reducing the gap between VR and established CAD-based processes.

Experiment V

The fifth experiment aimed to evaluate the VR system in a broader and more diverse user context, with a particular focus on identifying missing functionalities and collecting user-driven recommendations for improving the system. Due to the relatively large number of participants and their varied professional backgrounds across different engineering domains, this experiment emphasized qualitative insights into user needs alongside the assessment of usability across seven dimensions.

A total of 40 participants, junior engineers with practical experience in various industrial departments, took part in the evaluation. This heterogeneous background enabled a comprehensive assessment of the system from multiple professional perspectives, particularly regarding its applicability in real-world engineering tasks.

Overall, the usability evaluation yielded a score of 68.7%, representing the highest usability rating among all conducted experiments. The SUS resulted in a mean value of 3.0 on a five-point scale, indicating a generally positive perception of the system. Participants were able to complete the assigned tasks effectively, and the system demonstrated improved performance compared to earlier versions. Nevertheless, the primary outcome of this experiment lies in the identification of missing features and improvement potential.

A key result of this study is the identification of 18 missing functions required by users to effectively perform their tasks. These functions were derived from participants' direct interaction with the system and reflect practical requirements from different engineering domains. In addition, participants proposed 23 recommendations aimed at improving system usability, functionality, and integration into existing workflows. The feedback was notably detailed and critical, reflecting the participants' technical background and professional experience.

Analysis of the usability dimensions revealed a generally positive performance across most categories. In terms of task suitability, participants confirmed that the system provides the core functionalities required for design evaluation and spatial analysis. However, the absence of several advanced features limited the completeness and efficiency of task execution.

Regarding self-descriptiveness, the system was perceived as understandable, with users generally able to interpret system behavior and interaction outcomes. Nevertheless, some participants indicated that additional guidance and clearer system feedback would further improve usability, particularly for more complex tasks.

The expectation conformity dimension was evaluated positively overall. Interface elements, such as menus and visual structures, were largely consistent with user expectations. However, certain interaction mechanisms still deviated from conventional engineering software workflows, requiring adaptation by the users.

In terms of learnability, the system showed noticeable improvement compared to earlier experiments. Participants were generally able to familiarize themselves with the system within a short period. However, given the complexity of some tasks, additional onboarding support and training features were still considered beneficial.

The evaluation of controllability indicated that users were able to interact with the system and perform the required operations successfully. Interaction with objects and navigation within the virtual environment were generally perceived as manageable, although some users reported minor inefficiencies in control precision.

The error tolerance dimension remained an area with improvement potential. Participants noted the absence of certain features, such as undo functions and error handling mechanisms, which are essential for efficient and confident task execution in professional environments.

Finally, user engagement was rated highly. Participants expressed strong interest in the VR system and recognized its potential for supporting engineering tasks, particularly in visualization and interdisciplinary collaboration. The immersive nature of the system contributed positively to user motivation and acceptance.

The TLX results further support these findings, indicating low perceived workload across cognitive, physical, and temporal dimensions. Participants reported high levels of satisfaction and relatively low effort during task execution, suggesting that the system provides a comfortable and efficient interaction environment despite existing limitations.

In summary, the results of experiment V demonstrate that the VR system achieves a high level of usability and user acceptance in a diverse engineering context. The large and varied participant group enabled the identification of a substantial number of missing functions and practical improvement recommendations, which are critical for further system development. While the system performs well across most usability dimensions, targeted enhancements particularly in feature completeness and error tolerance are necessary to fully support professional engineering workflows.

Experiment VI

The sixth experiment investigates the usability of the VR system under conditions of increased task and time pressure. In this experiment, participants were required to complete predefined tasks within a limited time frame as part of a graded academic activity. The objective of this evaluation is to analyze how time pressure and performance requirements influence the usability across the seven defined dimensions, as well as their impact on perceived workload, user satisfaction, and suggested improvements.

A total of ten participants, all bachelor's students in mechanical engineering, took part in the experiment. Compared to previous experiments, the participants reported a higher level of experience with digital tools and VR systems. This provides a suitable basis to evaluate the system under more demanding conditions. Overall, the results indicate a moderate to good level of usability. Despite the imposed time constraints, participants were partially able to complete the assigned tasks, which shows that the system supports task execution even under pressure.

The analysis of the seven usability dimensions shows generally positive results, although some limitations become more visible under time pressure. Task appropriateness was generally not rated positively. Although participants confirmed that the system provides the necessary functions to complete the tasks, some users indicated that certain functions were missing, which affected the completeness of task execution. Expectation conformity is evaluated positively, since the interface structure, including menus and icons, was generally perceived as clear and understandable. Nevertheless, some participants reported that object manipulation was not fully intuitive, indicating differences between expected and actual interaction behavior.

Self-descriptiveness was predominantly evaluated negatively, as participants reported that they did not feel in control of the interaction and were unable to understand the system's behavior. In addition, not all users were able to clearly identify the next steps during task execution, which indicates that system guidance is still limited in more complex situations.

Learnability represents one of the weaker dimensions. Participants reported difficulties in understanding system functions, especially in relation to error messages and predictable system responses. This issue becomes more critical under time pressure, where the lack of guidance increases uncertainty.

Controllability is generally sufficient, as users were able to select and manipulate objects within the virtual environment. However, some inconsistencies in interaction precision were observed.

Error tolerance is identified as a weak aspect, since participants reported issues such as missing recovery functions and limited ability to correct mistakes. These limitations negatively influence user confidence, especially in time-constrained scenarios.

User commitment is generally positive, as most participants reported a good first impression and did not perceive the system as overly demanding. However, the perceived efficiency varies, indicating that time pressure influences the interaction performance.

In addition to the usability evaluation, participants provided several suggestions for system improvement. Frequently mentioned aspects include the integration of alternative interaction methods such as hand tracking, as well as the implementation of a tutorial or guided onboarding. Furthermore, improvements in object interaction, such as snapping functions and more interactive elements, were suggested. Participants also highlighted the need for better system adaptability, for example through adjustable user height or automatic detection. Additional features such as object scaling and coloring were also identified as relevant improvements. These suggestions indicate the need for a more intuitive, flexible, and user-adapted system.

The NASA TLX results show that the overall workload is high under time pressure. Cognitive demand is perceived as high due to the need to understand the system during task execution. Physical demand is low, and participants reported no significant physical strain. Time pressure is perceived as manageable, and the effort required to complete the tasks remains relatively low. Some participants reported dissatisfaction with their performance, and some experienced frustration due to unclear interaction elements.

The results of the SUS indicate a generally moderate score of 2.9 on a five-point scale. Participants reported that the system is relatively not easy to use and that its functions are not fully integrated. The system was also considered moderately learnable within a reasonable amount of time. However, some users reported a certain level of complexity and occasional inconsistencies in system behavior. The need for technical support was not dominant, but it was still present in more complex interaction scenarios.

In summary, the results of experiment VI show that the VR system maintains a stable level of usability under time pressure. Since the participants were students enrolled in the course, and the experiment was part of their assessment, it appears that some of them attempted to attribute unfavorable course evaluation outcomes primarily to the software. This interpretation is supported by the discrepancy between the responses provided in the open-text fields and those recorded in the scoring fields. In Table 3, a summary of all experiment setups and findings is presented.

Table 3. Experiment setups and usability findings.

Exp. No	Software S1	Year of test	Hardware	Test Group	Usability Degree	Number of participants	Proportion of discovered problems
I	Version 1.70	Oct. 2023	PC-Based	International Teams	51.7%	23	99.9%
II	Version 1.70.3	Apr. 2024	PC-Based	Students	62.3%	10	97.5%
III	Version 1.71	Jul. 2024	PC-Based	Senior Engineers	50.3%	5	84.4%
IV	Version 1.72	Mar. 2025	PC-Based	Senior Engineers	54.5%	5	84.4%
V	Version 1.72.1	May. 2025	Standalone	Junior Engineers	68.7%	40	99.9%
VI	Version 1.73	Jun. 2025	Standalone	Students	63.6 %	10	97.5%

5. Discussion

This study set out to investigate whether variables such as software, hardware, user background, and context of use affect the usability of VR systems within the product development process. Based on six experiments involving participants with different levels of experience, as well as varied hardware configurations and use cases, the findings indicate that VR provides demonstrable benefits in specific phases of product development, while its effectiveness remains highly context-dependent.

The comparative analysis across experiments II and VI, which involved inexperienced users, and experiments III and IV, which included senior engineers from a development team, clearly demonstrated the influence of user background on usability outcomes. Professional development teams were more concerned with technical precision and the integration of VR into existing workflows, resulting in lower acceptance when the system did not fully align with their operational requirements. The results indicate that participants' professional backgrounds and prior experience significantly shaped their expectations and perceived system needs. Consequently, professional users identified specific deficiencies and missing functionalities, as their feedback was closely tied to the practical requirements of their workplace tasks, for example when relying on a particular CAD software. This highlights the importance of tailoring usability assessments to the users' operational environments.

In addition, the usability ratings given by the development engineers in the third and fourth experiments were closely aligned. This again confirms that the user background influences both perceived usability and technology acceptance. In these cases, the engineers evaluated the technology

more critically in relation to their actual professional needs. When real decision-making and the potential profitability of an investment are at stake, the technology tends to be assessed more rigorously. In contrast, the student participants tended to evaluate the technology based on personal preferences, without considering profitability.

Regarding the software factor, the first experiment employed version 1.70, while the second used a slightly updated version (1.70.3). Usability scores improved in the latter case, reflecting the positive effect of addressing previously identified inefficiencies. A similar pattern was observed in experiments IV and V, where software updates again led to higher usability ratings.

The comparison between PC-based and standalone VR systems suggests that hardware configuration influences usability, particularly for inexperienced users. Usability improved slightly with the standalone devices, suggesting that such systems offer greater ease of use and flexibility, particularly for less experienced participants. However, PC-based systems remain necessary for high-precision engineering applications where graphical performance and model complexity are critical.

With respect to the use case factor, the TLX results were generally positive across all experiments, indicating low physical and cognitive workload. However, in the final two experiments which conducted under identical technical conditions and with the same tasks but differing user roles and objectives, the participants in experiment 6 exhibited higher levels of stress and cognitive effort. This was likely because the tasks in experiment 6 were performed as part of a formal course assessment, which introduced additional cognitive pressure and performance-related stress. These findings suggest that the specific use case and contextual purpose of the activity can significantly influence user acceptance and perceived task load.

Another finding from the comparison of the reactions of the two leaders in experiments III and IV concerned their openness to adopting new technology. The leader from the medium-sized enterprise (experiment3) was more receptive, whereas the leader from the large enterprise (experiment4) was more cautious. This observation aligns with the findings of [22], who reported in section 2.2 that medium-sized enterprises tend to be more open to new technologies.

It is evident that, after the first experiment, 99.9% of the system's problems had been identified. However, the objective of the subsequent usability tests was not only to detect problems, but rather to optimize the overall usability of the system as well as to evaluate the factors influence the usability.

It has been demonstrated that the application of a standardized usability evaluation contributes to the continuous improvement of the VR system. The progressive software enhancements are clearly observable and indicate a positive correlation between usability assessments and the iterative development process of the targeted system. This means that the advancements achieved in the software can be directly associated with improvements in usability. This finding underscores the effectiveness of a systematic evaluation approach.

It is important to distinguish the purpose of a usability test. When the primary objective is to identify system errors or to determine the required number of participants, it is recommended to conduct a minimal number of tests using the approach described in Section 2.3. However, if the objective is the continuous optimization of the system in order to enhance user satisfaction and technology acceptance, it is recommended to conduct iterative usability evaluations. In this case, each testing cycle should incorporate previously identified variables, such as user feedback, and involve new user groups, new scenarios, and updated versions of both the software and hardware.

Across several experiments, participants consistently identified missing functions required for performing domain-specific tasks. This indicates that usability of VR systems is not only determined by interaction quality but also by the completeness of task-relevant features. Particularly in professional environments, the absence of specialized features significantly reduces perceived usefulness and limits system acceptance, even when the underlying interaction mechanisms function correctly

The comparison between VR and conventional CAD tools emerged as a recurring theme, particularly among professional engineers. While VR was highly valued for its immersive visualization and spatial understanding of complex assemblies, participants emphasized that

traditional CAD systems still provide superior precision, feature depth, and workflow integration. This suggests that VR systems are currently better suited as complementary tools for design reviews and collaborative visualization rather than as direct replacements for established engineering software.

Several experiments revealed that first-time users required additional onboarding and guidance to interact effectively with the system. This was particularly evident in experiment IV, where participants received only a brief introduction without any practical VR familiarization. As a result, many users rejected the system; although this was not the only contributing factor, it clearly played a role. These findings indicate that training and onboarding procedures are essential for the successful adoption of VR tools in engineering contexts. Systems intended for industrial environments should therefore incorporate guided tutorials or training modules to reduce the initial learning curve and improve overall user acceptance.

In addition to the importance of the onboarding process prior to applying VR, the learnability dimension was evaluated predominantly negatively across most experiments. Although several optimizations were implemented, the system was still perceived as difficult to learn. A likely explanation is that the technology is relatively new and many participants were not yet familiar with it, which naturally increases the initial learning effort. Minor inconsistencies in the interaction design or limited exposure time may also have contributed to this perception.

The collaborative evaluation conducted in experiment III also highlights the potential of VR as a communication platform for distributed teams. Participants reported that the shared virtual environment facilitated discussion and joint model inspection, suggesting that VR can support collaborative decision-making processes in product development.

6. Conclusions

This study investigated how different factors (software configuration, hardware type, user background, and context of use) influence the usability of VR systems within the product development process. The analysis was based on a series of experiments involving participants with varying levels of expertise, different hardware configurations, and multiple product development scenarios.

The results demonstrate that VR can provide clear benefits in specific phases of product development, particularly in activities related to visualization, spatial analysis, and collaborative design review. However, the effectiveness and acceptance of VR systems are strongly dependent on the context in which they are applied. Differences in user expertise, professional expectations, and operational requirements significantly influenced usability perceptions and technology acceptance across the experiments.

From an organizational perspective, the findings indicate that VR adoption should follow a context-specific implementation strategy rather than being considered a universal solution. Companies aiming to integrate VR into their product development processes should therefore conduct targeted cost-benefit analyses, select hardware and software configurations appropriate to the specific development phase, and provide training programs adapted to the experience level of their users. In particular, the collaborative capabilities of VR environments offer considerable potential for improving communication and coordination within distributed development teams.

A key methodological contribution of this study is the demonstration of the value of systematic and iterative usability evaluation. The conducted experiments showed that applying standardized usability assessments enabled the identification of system limitations and guided successive software improvements. The progressive development of the VR system across the experiments indicates a clear relationship between usability feedback and system optimization, confirming the effectiveness of a user-centered, iterative evaluation process. These improvements were reflected not only in enhanced system performance but also in increased user acceptance and overall user experience. The following design and implementation principles for VR systems in product development are proposed based on the experiment analysis:

1. Match the level of hardware to the level of expertise and precision of the task: Standalone VR headsets are more suitable for early-phase reviews, whereas PC-based systems are better suited to high-precision engineering analyses.
2. Differentiate interaction concepts by user group: Students and junior engineers benefit from simplified menus and guided interaction, whereas senior engineers require direct access to precise measurement and inspection tools that are aligned with CAD workflows.
3. Integrate VR iteratively: Regular usability evaluations using standardized tools (SUS, NASA-TLX and task-based questionnaires) should accompany each software iteration to systematically improve usability and acceptance.
4. Consider the organizational context and ROI: For industrial adoption, improvements in usability must be communicated in terms of workflow integration, training effort and potential return on investment.
5. Use VR as a complement, not a replacement, for CAD: VR is most effective for immersive visualization, spatial understanding and collaborative design reviews, while CAD remains the primary tool for detailed design and documentation.

Furthermore, the study highlights the importance of clearly distinguishing the purpose of usability testing. When the objective is to identify the majority of critical usability problems, a limited number of participants may be sufficient, following established usability evaluation approaches. However, when the goal is the continuous improvement of system usability and technology acceptance, iterative testing cycles become essential. Such cycles should incorporate feedback from previous evaluations, involve new user groups, and consider updated versions of both software and hardware, as well as different application scenarios.

Despite the insights gained, the scope of this study is subject to certain limitations. The experiments were conducted with specific user groups and focused on a particular VR software within defined industrial contexts. Future research should therefore investigate long-term adoption patterns of VR technologies in product development environments, evaluate additional VR platforms and interaction methods, and include larger and more diverse industrial teams from different sectors. In addition, further studies should examine the economic impact and return on investment of VR integration in the engineering field.

Overall, the findings indicate that VR has considerable potential to support innovation and efficiency in product development. However, its successful implementation depends on aligning the technology with user needs, task requirements, and organizational capabilities. A structured, user-centered evaluation approach can play a critical role in achieving this alignment and enabling the effective integration of VR systems into industrial product development process.

Funding: “This research was funded by the German Federal Ministry for Economic Affairs and Climate Action, grant number KK5243803GR2.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data supporting the findings of this study, as well as the proposed method for evaluating the usability of VR software, are available from the corresponding author upon reasonable request.

Abbreviations

The following abbreviations are used in this manuscript:

VR	Virtual Reality
AR	Augmented Reality
SUS	System Usability Scale
TLX	Task Load Index
TAM	Technology Acceptance Model
SLAM	Simultaneous Localization And Mapping
ROI	Return On Investment

CAD Computer-Aided Design
HMD Head Mounted Display

References

1. Rademacher, M.H. *Virtual Reality in der Produktentwicklung*; Springer Fachmedien Wiesbaden: Wiesbaden, 2014, ISBN 978-3-658-07012-0.
2. Ekströmer, P.; Wever, R.; Wängdahl, J. *VIRTUAL REALITY SKETCHING FOR DESIGN IDEATION*, 2018.
3. Wolfartsberger, J. Analyzing the potential of Virtual Reality for engineering design review. *Automation in Construction* **2019**, *104*, 27–37, doi:10.1016/j.autcon.2019.03.018.
4. Aromaa, S.; Väänänen, K. Suitability of virtual prototypes to support human factors/ergonomics evaluation during the design. *Appl. Ergon.* **2016**, *56*, 11–18, doi:10.1016/j.apergo.2016.02.015.
5. Hung, L.C.; Chen, C.-M. The Impact of Digital Transformation in Manufacturing on Firm Performance: A Deleveraging Perspective. *JAEPS* **2025**, *15*, 23–29, doi:10.54254/2977-5701/2024.19902.
6. Värzaru, A.A.; Bocean, C.G. Digital Transformation and Innovation: The Influence of Digital Technologies on Turnover from Innovation Activities and Types of Innovation. *Systems* **2024**, *12*, 359, doi:10.3390/systems12090359.
7. Khan, M.I.; Yasmeen, T.; Khan, M.; Hadi, N.U.; Asif, M.; Farooq, M.; Al-Ghamdi, S.G. Integrating industry 4.0 for enhanced sustainability: Pathways and prospects. *Sustainable Production and Consumption* **2025**, *54*, 149–189, doi:10.1016/j.spc.2024.12.012.
8. Abughalia, A.; Stechert, C. A Decade of Virtual Reality in Product Development: A Literature Review of Effectiveness, Challenges, and Future Research. *Procedia CIRP* **2025**, *136*, 438–443, doi:10.1016/j.procir.2025.08.076.
9. Yang, M.; Miller, C.; Crompton, H.; Pan, Z.; Glaser, N. The Implementation of Virtual Reality in Organizational Learning: Attitudes, challenges, side effects, and affordances. *TechTrends* **2024**, *68*, 111–135, doi:10.1007/s11528-023-00917-y.
10. Abughalia, A.; Stechert, C. Immersive Onboarding: Designing a Training Framework for Effective Virtual Reality Integration in Product Development. *Procedia CIRP* **2025**, *136*, 432–437, doi:10.1016/j.procir.2025.08.075.
11. Merz, A.; Moser, I.; Bergamin, P.B. Performance expectancy and social influence drive the acceptance of immersive virtual reality for professional collaboration. *Virtual Real.* **2025**, *29*, doi:10.1007/s10055-025-01182-5.
12. Schon, C.; Huang, R.; Hessenmüller, H.; Przybyl, S.; Tümler, J. *Classification of the Topicality and Relevance of Evaluation Tools for VR Applications*, 2025.
13. Rendeovski, N.; Trajcevska, D.; Dimovski, M.; Veljanovski, K.; Popov, A.; Emini, N.; Veljanovski, D. PC VR vs Standalone VR Fully-Immersive Applications: History, Technical Aspects and Performance. In *2022 57th International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST)*. 2022 57th International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST), Ohrid, North Macedonia, 16–18 Jun. 2022; IEEE, 2022; pp 1–4, ISBN 978-1-6654-8500-5.
14. Shourangiz, E.; Ghafari, F.; Wang, C. Human-robot collaboration integrated with virtual reality in construction and manufacturing industries: A systematic review. *Virtual Reality & Intelligent Hardware* **2025**, *7*, 317–343, doi:10.1016/j.vrih.2024.08.004.
15. Winkler, I.; Murari, T.; Ferreira, C.; Freitas, F. *VR-based product development process: opportunities and challenges in the automotive industry*, 2022.
16. Lawson, G.; Herriotts, P.; Malcolm, L.; Gabrecht, K.; Hermawati, S. The use of virtual reality and physical tools in the development and validation of ease of entry and exit in passenger vehicles. *Appl. Ergon.* **2015**, *48*, 240–251, doi:10.1016/j.apergo.2014.12.007.
17. *Advances in Digital Human Modeling II*; Marshall, R.; Summerskill, S.; Harih, G.; Scataglini, S., Eds.; Springer Nature Switzerland: Cham, 2025, ISBN 978-3-032-00838-1.
18. Fares, O.H.; Aversa, J.; Lee, S.H.; Jacobson, J. Virtual reality: A review and a new framework for integrated adoption. *Int J Consumer Studies* **2024**, *48*, doi:10.1111/ijcs.13040.

19. Pöhler, L.; Teuteberg, F. Suitability- and utilization-based cost-benefit analysis: a techno-economic feasibility study of virtual reality for workplace and process design. *Inf Syst E-Bus Manage* **2024**, *22*, 97–137, doi:10.1007/s10257-023-00658-8.
20. Zolas, N.; Kroff, Z.; Brynjolfsson, E.; McElheran, K.; Beede, D.; Buffington, C.; Goldschlag, N.; Foster, L.; Dinlersoz, E. *Advanced Technologies Adoption and Use by U.S. Firms: Evidence from the Annual Business Survey*, Cambridge, MA, 2020.
21. Jalo, H.; Pirkkalainen, H.; Torro, O.; Pessot, E.; Zangiacomi, A.; Tepljakov, A. Extended reality technologies in small and medium-sized European industrial companies: level of awareness, diffusion and enablers of adoption. *Virtual Real.* **2022**, *26*, 1745–1761, doi:10.1007/s10055-022-00662-2.
22. Clemente-Almendros, J.A.; Nicoara-Popescu, D.; Pastor-Sanz, I. Digital transformation in SMEs: Understanding its determinants and size heterogeneity. *Technology in Society* **2024**, *77*, 102483, doi:10.1016/j.techsoc.2024.102483.
23. Tamvada, J.P.; Narula, S.; Audretsch, D.; Puppala, H.; Kumar, A. Adopting new technology is a distant dream? The risks of implementing Industry 4.0 in emerging economy SMEs. *Technological Forecasting and Social Change* **2022**, *185*, 122088, doi:10.1016/j.techfore.2022.122088.
24. Nakandala, D.; Yang, R.; Elias, A.; Fanousse, R. Effects of managers' environmental consciousness and digital expertise on their technology adoption intentions. *Journal of Cleaner Production* **2024**, *474*, 143558, doi:10.1016/j.jclepro.2024.143558.
25. Davis, F.D. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly* **1989**, *13*, 319, doi:10.2307/249008.
26. Lewis, J.R.; Sauro, J. Effect of Perceived Ease of Use and Usefulness on UX and Behavioral Outcomes. *International Journal of Human-Computer Interaction* **2024**, *40*, 6676–6683, doi:10.1080/10447318.2023.2260164.
27. Cazañas-Gordón, A.; Miguel, A.; Parra Mora, E. Estimating Sample Size for Usability Testing. *Enfoque UTE* **2016**, *8*, doi:10.29019/enfoqueute.v8n1.126.
28. Lewis, J.R. Evaluation of Procedures for Adjusting Problem-Discovery Rates Estimated From Small Samples. *International Journal of Human-Computer Interaction* **2001**, *13*, 445–479, doi:10.1207/S15327590IJHC1304_06.
29. Nielsen, J.; Landauer, T.K. A mathematical model of the finding of usability problems. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '93*. the SIGCHI conference, Amsterdam, The Netherlands, 24–29 Apr. 1993; Arnold, B., van der Veer, G., White, T., Eds.; ACM Press: New York, New York, USA, 1993; pp 206–213, ISBN 0897915755.
30. Balzerkiewitz, H.-P.; Dlamini, N.; Stechert, C.; Mpofo, K. Usability of VR-Systems in Cross-Cultural Product Development: A Case Study. *Procedia CIRP* **2024**, *128*, 399–404, doi:10.1016/j.procir.2024.03.019.
31. MAGURIT Freezing and Fresh-Cutting Machines Company. Cutting Machine Design. Available online: <https://www.magurit.de/drumcut/> (accessed on 20 April 2026).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.