

Article

Development and Evaluation of Training Scenarios for the Use of Immersive Assistance Systems

Maximilian Rosilius ^{1,*}, Lukas Hügel ², Benedikt Wirsing ¹, Manuel Geuen ³, Ingo von Eitzen ⁴,
Volker Bräutigam ¹ and Bernd Ludwig ⁵

¹ Institute of Digital Engineering, Technical University of Applied Sciences Würzburg-Schweinfurt, 97421 Schweinfurt, Germany

² T&O Unternehmensberatung GmbH, 80687 München, Germany

³ Department of Health Sciences, Technical University of Applied Sciences, 63743 Aschaffenburg, Germany

⁴ Regain Development, 22949 Ammersbek, Germany

⁵ Department of Information Science, University of Regensburg, 93053 Regensburg, Germany

* Correspondence: maximilian.rosilius@thws.de

Abstract: Emerging assistance systems are designed to enable operators to perform tasks better, faster, and with a lower workload. However, in line with the productivity paradox, the full potential of automation and digitalisation is not being realised. One reason for this is insufficient training. In this study, the statistically significant differences among three different training scenarios on performance, acceptance, workload, and technostress during the execution of immersive measurement tasks are demonstrated. A between-subjects design was applied and analysed using ANOVAs involving 52 participants (with a statistical overall power of 0.92). The ANOVAs were related to three levels of the independent variable: quality training, manipulated as minimal, personal, and optimised training. The results show that the quality of training significantly influences immersive assistance systems. Hence, this article deduces tangible design guidelines for training, with consideration of the system-level hardware, operational system, and immersive application. Surprisingly, an appropriate mix of training approaches, rather than detailed, personalised training, appears to be more effective than e-learning or ‘getting started’ tools for immersive systems. In contrast to most studies in the related work, our article is not about learning with AR applications but about training scenarios for the use of immersive systems.

Keywords: training; onboarding; immersive assistance systems; augmented reality; acceptance; performance; guideline; measurement accuracy; workload; stress



Citation: Rosilius, M.; Hügel, L.; Wirsing, B.; Geuen, M.; von Eitzen, I.; Bräutigam, V.; Ludwig, B.

Development and Evaluation of Training Scenarios for the Use of Immersive Assistance Systems. *Appl. Syst. Innov.* **2024**, *7*, 73. <https://doi.org/10.3390/asi7050073>

Academic Editors: Christos Douligeris and Friedhelm Schwenker

Received: 18 March 2024

Revised: 20 May 2024

Accepted: 20 August 2024

Published: 26 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the future world of working, employees will be empowered by assistance systems to expand their skills. The functions of immersive systems promise the smart operators the right information in the right place at the right time. This enables them to carry out the tasks assigned to them in less time with better quality and higher output [1]. One concrete example of this is a measurement application using augmented reality. This allows users to virtually interact with the real world around them and to annotate it. This system has the ability to geodetically map the real environment in real time [2], enabling real dimensions and references to be augmented virtually. Such a system can support employees, especially when measurements are difficult to access or out of reach for operators. These conditions could favour the selection of immersive systems compared to conventional measurement systems, as conventional systems are limited by their inherent measurement principles.

However, the productivity paradox approach of Schweikl and Obermaier [3] shows that the potential of modern information and communication technology (ICT) applied in assistance systems is not fully exploited. The principle states that thanks to modern automation and digitalisation, higher overall productive performance is expected than

what is currently measured. Due to fundamentally new interaction paradigms and the interplay of virtual assets within the real world, users face major obstacles in understanding the logic of immersive systems. Especially for beginners, the error rates are very high due to the participation gap. This leads to masking of the added value of the immersive features along the user journey. This dilemma poses major challenges for both developers and users. The question from our research association is whether the system design or the training design has a greater influence on user experience. According to Schweikl and Obermaier [3], the reasons for this paradox are multi-layered, including exaggerated expectations, adjustment delays, and misjudgements of the added value. However, a key aspect is the mismanagement of these technologies and the lack of complementary preconditions. The operators lack the necessary skill level to be able to use the systems effectively. Corresponding to [3], insufficient resources are invested in the necessary training. This training error, as outlined in the AR failure taxonomy [4], contributes to adverse effects on performance (as discussed by Bahaei et al. [4] and Simatupang and Saroyeni [5]) and acceptance (indirect correlation as discussed by Giovanni Mariani et al. [6] and Marshall et al. [7]). It also contributed to technostress [8–10], which negatively impacts overall productivity. In general, immersive technologies promise significant potential for education and training in the industrial environment [11–13], such as speeding up the reconfiguration of production lines, supporting shop-floor operations, or virtual training for the assembly of parts. Initial concepts have also been explored regarding how the use and handling of immersive systems can be trained in the context of training and teaching in a university environment [14]. The critical literature search revealed that there is no contribution that deals with training scenarios, evaluation methods, or guidelines specifically for learning the usage of immersive systems in an industrial context. However, there is a demand for research (see [7,15–19]) on how immersive assistance systems can be introduced to user groups effectively and efficiently in order to avoid negative effects. Indeed, there is a lack of applicable design guidelines and recommendations regarding training scenarios for specialised, professional use. In our research association, a research prototype of an industrial AR application was developed for layout planning, and its effects on long-term use were investigated. The first challenge was to tailor the application design towards the industrial requirements. The aim was to explore effective and efficient AR training scenarios to pilot tangible software prototypes under real conditions. In preliminary research for this study, an optimised training scenario was developed, the results of which were incorporated into the current research design as the factor level of an independent variable. Motivated by the prior experience of the research association, the study we present here aims to investigate the effects of the quality of training on performance, acceptance, and stress. This study contributes to the state of scientific knowledge in the following way: in contrast to the related work, this article addresses training scenarios in an industrial context. Additionally, unlike theoretical approaches, practical implications can be drawn from this work, e.g., how the right mix of training strategy (emphasising quality over quantity or intensity) delivers effective training success. Furthermore, a replicable experimental design (with penalty times) for measuring the efficacy of training scenarios is presented. In addition, the ‘performance’ index is introduced as a new metric (accounting for the time and correctness of a task). In summary, tangible design guidelines for training scenarios towards immersive assistance systems are presented. This topic may also be meaningful in the context of the industrial metaverse.

2. Related Work

Relevant findings have been researched and embedded in the context of ICT and immersive assistance systems.

Despite substantial investments in information and communication technologies, the anticipated improvements in productivity, as measured by output per worker or output per unit of capital, have not always been realised. According to the scientific discourse of a recent meta-review [3] about its effects, causes, and socio-technical contexts, the pro-

ductivity paradox also applies to immersive assistance systems. Research from the last three decades suggests that ‘adjustment delays, measurement issues, exaggerated expectations and mismanagement’ [3] are root causes for the recent deceleration in productivity growth. A critical examination of mismanagement highlights that insufficient introduction of ICT and missing organisational change can lead to substantial productivity loss. In addition, facilitating organisational change to fundamentally transform business processes is a necessary action for management to undertake [3].

The following needs for training to increase productivity were identified from preliminary work. Therefore, it is important for workers to continuously develop their skills and for companies to invest in training and development programs to ensure that their workforce has the necessary skills to effectively use new technologies [3]. There is a demand for providing training and education programs that equip workers with the necessary skills to use immersive technology correctly [3], thus allowing them to succeed in the changing workplace. Zoetbrood [20] observed that training does not have a significant impact on productivity, instead pointing to other interacting factors, including multiple deficiencies, e.g., integration into workflows and processes, linkage with other organisational changes, adaptation to employees’ needs, and continuity and sustainability.

According to Kurilovas [21], quality is defined as the extent to which an entity (e.g., a product or service) is able to meet specific requirements. In the context of training quality, quality refers to the effectiveness of training scenarios and their suitability to learners’ personal needs. It is emphasised that the quality of learning content should be assessed through both an expert-centred (top-down) and a learner-centred (bottom-up) approach. Using the top-down approach, external experts carry out the assessment by applying various methods of decision analysis. They draw up a quality model, determine the weighting of the quality criteria, and apply suitable evaluation methods. The bottom-up approach focuses on direct feedback and evaluation from the actual users of the training scenario. This can be achieved through surveys, interviews, case studies, or other methods to obtain a comprehensive understanding of how learners perceive the scenario and what improvements could be made from their perspective. In the context of ICT, the quality of training is regarded as the user’s acquisition of skills that enable them to perform their technical task effectively and efficiently together with the immersive assistance system [22].

Results regarding the impact of training on stress could be found in preliminary work. For instance, Korunka and Vitouch [23] and Day et al. [15] demonstrate that adequately trained workers involved in the implementation of new ICT experience less stress. The findings presented by Wang et al. [24] indicate that technostress can be reduced by fostering learning organisation and sufficient training.

Previous work has shown these outcomes regarding the influence of training on acceptance. To determine how to improve the implementation and adoption of new technologies, Marshall et al. [7] examined the impact of end-user training on technology acceptance in the context of oral surgeons. Both determinants, performance expectancy and effort expectancy, were positively correlated with training. The work by Alqahtani et al. [25] demonstrates the influence of educational quality, including the features and functionality of the e-learning system, on the acceptance factors perceived ease of use and perceived usefulness. As a summary of past studies, Giovanni Mariani et al. [6] (n = 497 participants) analysed the influence of training opportunities introducing new IT tools by selecting some factors of the TAM model. The results show that providing training opportunities indirectly impacts the intention to use (acceptance). Training opportunities were shown to have a direct positive impact on IT self-competence, job satisfaction, perceived ease of use, and perceived usefulness. The last two determinants, in turn, have a direct impact on the intention to use. There were also other effects with moderate factors such as the organisational area, gender and age, educational level, IT experience, and learning strategies.

In the following, learning factors, methods, and approaches for ICT, as well as for augmented reality, are presented. In line with Korpelainen and Kira [22], two basic learning strategies are differentiated: formal and informal. Formal learning is structured training

guided by a trainer. It includes presenting help through user training sessions organised by the organisation's support staff. These sessions are designed to offer employees a structured and guided approach on how to use the ICT system. Formal training may also include self-study courses (e.g., e-learning) or training exercises with peers. In contrast, informal learning is not structured or guided by a trainer. It consists of practical application, seeking help from written material, or interpersonal help from peers. Practical application involves trying things out by trial and error, exploring different functions, and learning by doing. Written material, representing manuals or email information from the organisation's IT support department, must be studied by the participants. The interviews conducted by Korpelainen and Kira [22] show that the test subjects favoured informal learning strategies, such as practical experiments, using helpdesk services, or help from peers. These informal strategies offer employees flexibility and autonomy in the learning process and are, therefore, recommended according to Korpelainen and Kira [22]. Giovanni Mariani et al. [6] recommend that companies consider various learning strategies to maximise the effectiveness of training initiatives and meet the individual learning needs of employees. This includes creating opportunities for informal learning, promoting learning with others, and taking into account the different learning strategies utilised by employees. The following implications can be drawn from the field of agile learning formats and practice-based learning environments: learners should use competency-based agile learning formats, learning environments with incentives and opportunities to apply their knowledge through a mixture of learning elements, case studies, and practical sessions [26].

Additional learning approaches for industrial use include blended [27] and flipped learning [28,29]. Both approaches aim to enhance the effectiveness of learning by integrating various teaching methods and technologies. Blended learning emphasises the integration of digital learning tools, e.g., e-learning, video tutorials, manuals, and different learning formats, such as face-to-face training, practical sessions, and so on. Flipped learning focuses on reversing the traditional teaching structure by providing learning materials and practical sessions before personal training. Both approaches can be leveraged in practice to develop effective training programs and meet the requirements of ICT.

Dörner and Horst [14] present relevant methods, discussions, and inferred best practices for overcoming challenges when teaching hands-on courses about mixed reality in higher education. The key method is the Circuit Parcours Technique [14], which provides a structured mixture of hands-on experiences with theoretical backgrounds and evaluations for higher education. It involves role-based learning and group collaboration and aims to achieve pedagogical goals, such as active learning and peer teaching. This method can be integrated into the curriculum to offer a comprehensive learning experience and includes reflection and evaluation phases for continuous improvement. Overall, it offers an interactive approach for educators to promote practical learning and skill development in VR and AR.

The preceding study by Bradley et al. [30] discusses how the learning curve of task difficulty with new technologies looks schematically over time; see Figure 1a. Accordingly, the novice user starts their conventional work task at a certain difficulty level. If the novice starts their learning journey from this point, they must invest a certain amount of effort until they reach maximum learning pain. As soon as they overcome this turning point, they will have acquired so much learning experience that the task difficulty decreases over time. With increasing training and experience, the user masters the assistance system, thus benefitting from the potential of the easiness opportunity when carrying out their activity in combination with the assistance system.

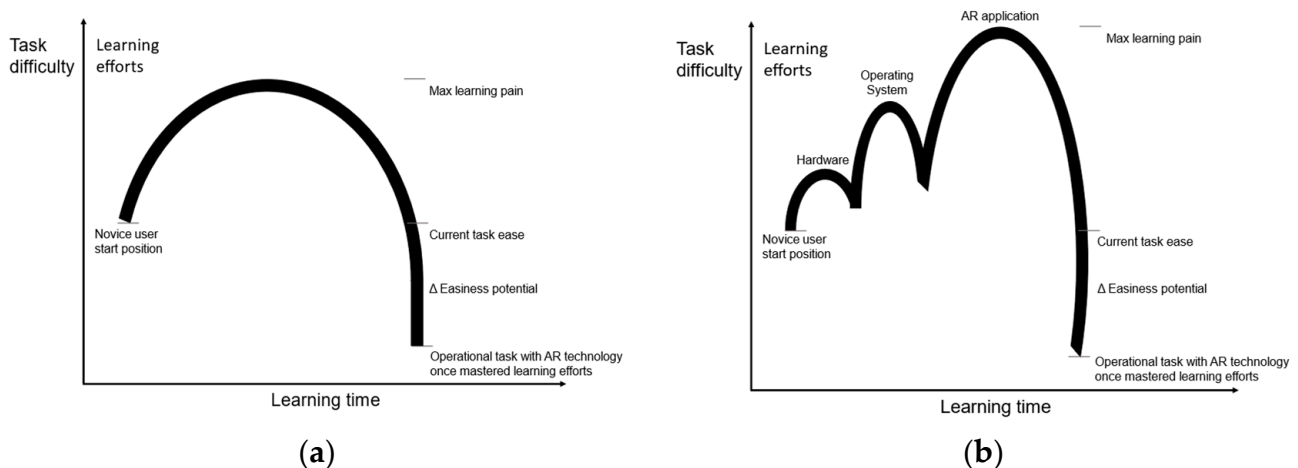


Figure 1. (a) Schematic learning process for a novice adopting new technologies based on [30,31]; (b) schematic dedicated learning process for a novice using an immersive assistance system according to our implications. Our approach requires distinctive learning scenarios for each system level.

In accordance with Dunleavy [32], three design principles were formulated for the development of AR learning experiences. The first principle is to enable and then challenge, which involves providing support for positive interdependence among learners to achieve a common goal in a physical space [32]. The second principle is driven by gamification, which involves direct player interaction and learning through gamified narratives to enhance engagement and learning outcomes [32]. The third principle is to spark curiosity by enabling the exploration of spatial awareness features through AR technology for immersive learning experiences [32]. Due to the high density of previous work on training with AR, the findings and pedagogical concepts can be generalised to the learning of immersive systems. Mystakidis et al. [33] recommended engaging novices in playful activities to promote critical thinking and reflection, utilising AR and technical skill development. The most commonly used instructional technique involves learners constructing knowledge by interacting with AR-enhanced content [33]. This empowers novices to take control of their learning through hands-on activities [33]. Combining various AR approaches provides diverse and engaging learning experiences [33]. Implementing playful designs and strategies enhances novices' engagement and motivation [33].

The following results were obtained concerning the learning and adoption of technologies. According to Barnard et al. [31], the specific relationship between acceptance and the adoption of new technologies was discussed. According to Barnard et al. [31], two basic conditions were identified that must be met for users to eagerly learn a new technology: intention to use and usability. It is, therefore, important to strengthen the determinants of these two basic conditions when training new technologies in order to ensure that they are well accepted by users. For the intention to use, we refer to the UTAUT acceptance model [7] and its four determinants for the intention to use: performance expectation, effort expectation, Cohen's influence, and facilitating conditions. According to the definition of ISO9241 [34], the usability of a technology is the perceived effectiveness, efficiency, and satisfaction with which users achieve their goal, where effectiveness refers to accuracy and completeness. Contrastively, Nielsen [35] shows that usability is characterised by the following five attributes: easy to learn, efficient to use, easy to remember, minimal errors, and subjectively appealing. In turn, Rogers et al. [36] identify five factors that influence the adoption of innovations, including relative advantages, compatibility, complexity, testability, and observability. These factors combine user and system aspects and emphasise the importance of user experience in the technology acceptance process. Another study by Barnard et al. [31] also justifies the relevance of experimenting with technology [37] using the incorporation phase of the STAM model, in which the user should be given an understanding of the usefulness of the system. Based on these findings from the litera-

ture, Barnard et al. [31] conducted two qualitative case studies to gain insights into the experiences, opinions, and challenges of older adults in using technology. One study dealt with the use of mobile technologies when walking, the other with errors that older users experience when using a tablet computer for the first time. The researchers used interviews, open discussions, and experimental methods to collect and analyse the data. Derived from these data, together with the results of related work, two theoretical models were postulated, which include the influencing factors of the ease of learning perspective and the system and user perspective [31]:

- 1 Model of technology acceptance from an ease of learning perspective [31]: Due to the perception derived from perceived self-efficacy and perceived difficulty, the users can ascertain an individual's attitude to learning. These attitudes are also affected by the social environment and the availability of support. They then lead to the crucial decision-making point: intention to learn, which is essentially the determination to engage with the technology. In line with this intention, the user may proceed through experimentation and exploration, encountering either a barrier of learning, which can lead to rejection, or if the difficulty is manageable and the experience is positive, to acceptance and satisfaction.
- 2 Model of technology acceptance from system and user perspective [31]: In short, some selective factors influence the perceived difficulty of learning when a user explores a new technology. These factors include the actual difficulties of the system, the user's experience with technology, the transfer of learning experiences, feedback, error recovery, quality of training and training materials, and support from the social environment. The influence of these factors on the perceived difficulty of learning were given, which can ultimately lead to acceptance or rejection of the system.

Our study continues the preliminary work in the context of the human-machine interaction loop model [19]. This model investigates the errors and their effects on immersive assistance systems in industrial work environments. The design was derived from the Human-Cyber-Physical System (HCPS; see [38–40]) and included a control loop approach for human-computer interaction. This research model was created to analyse the effects of data errors, visibility errors, interaction errors, and training errors on user performance, acceptance, and stress. For the industrial environment, the initial results show that erroneous speech interaction has a negative impact on performance, acceptance, and stress in human-machine interaction.

Our work investigates the training error within the augmented reality error taxonomy by Bahaei et al. [4]. The model by Rosilius et al. [19] was consequently motivated by the theoretical base of a taxonomy of faults. This taxonomy encompasses training inside the categories' personnel and organisational faults, which refer to the lack of skills and knowledge required to perform a task. This highlights the importance of technology adoption through appropriate training and skill development to mitigate human error in AR-enabled systems.

The following research demand was identified with respect to the related work. Summarising the above, it is essential for research to explore the interplay of organisational support, such as resources for training new technologies that take the specific needs of individual workers into account [22]. 'End-user training appears to be an important and understudied factor in technology acceptance' [7] (p. 5). According to Palanque et al. [16], errors within the human-computer interaction loop can be avoided by choosing suitable training content. Relevant research needs were identified regarding industrial applications and acceptance factors for augmented reality [17]. Relevant research demand was highlighted concerning industrial use and acceptance factors for augmented reality [17]. As noted by Graser and Böhm [18], 'There is a particular of models that reflect application conditions in the field of corporate training outside schools and academic institutions' [18] (p. 7). Based on the literature reviews conducted, it is apparent that in the field of immersive assistance training, studies on performance, acceptance factors, and technostress, as well as

practitioner notes, are appreciably underrepresented. In compliance with Rosilius et al. [19], further studies could focus on the research design to investigate inferior training.

3. Methodology

As stated above, our objectives, research questions, and hypotheses were derived from previous work. This study was conducted as part of an overarching research approach investigating the effects of faults on immersive assistance systems in an industrial context [19]. As a contribution to this proposed approach, the presented study analyses the effects of training errors. We developed a dedicated research design to meet the following research questions:

Research Question 1. *What effect does the quality of training have on the performance of an immersive assistance system?*

Research Question 2. *What effect does the quality of training have on the human factors of acceptance and stress in an immersive assistance system?*

Based on the research demand mentioned earlier and the findings of the related work, and in accordance with the overarching research approach [19], the following hypotheses were deduced (the corresponding hypotheses $H_{X,0}$ are negated accordingly):

Hypothesis $H_{A,1}$. *Insufficient training has a negative impact on the performance of an immersive assistance system.*

Hypothesis $H_{B,1}$. *Insufficient training has a negative impact on the acceptance of an immersive assistance system.*

Hypothesis $H_{C,1}$. *Insufficient training leads to higher workload and stress levels for an immersive assistance system.*

3.1. Explorative Pre-Study for an Optimised Training Scenario

As preparative work, an optimised AR training scenario was carried out in the research association using an explorative study in a mixed-methods design [41]: 10 subjects were trained personally (detailed) or minimally (insufficient) and subsequently asked to perform a subtask from the corresponding system level (hardware, operation system, and AR prototype app), as described below in Section 3.2. Following the methodology of Barnard et al. [31] created using case studies on the use and error evaluation, the participants were asked in a questionnaire to evaluate the quality of the training as well as the respective potential for improvement per system level. Additionally, a qualitative failure mode and effect analysis was carried out with the input of the participants, and corresponding actions were identified. These two types of training were evaluated regarding the system levels both quantitatively (performance) and qualitatively (results of the questionnaires and sources of errors). Combining this evaluation with the practical experience of the research project and the implications of the related work led to the development of an optimised training scenario. In addition, specific design guidelines for AR training were derived; see Appendix A.

Critical discourse from related work in terms of approaches and methods for developing the optimised training scenario as a result of our preliminary work includes the following.

According to Giovanni Mariani et al. [6], training opportunities should generally be offered by the organisation. Giovanni Mariani et al. [6] and Korpelainen and Kira [22] state that informal strategies and the adaptation of learning content to individual needs are also recommended. Based on previous experience from the research project, it became evident that formally, i.e., didactically, structured learning strategies with a guided practical component are effective for immersive assistance systems. This recommendation was also

confirmed via the survey. The specific preference (presentation via e-learning, video, trainer, or guided learning by doing) depended on the system levels.

The preference for blended learning and the selection of the method depending on the system level were confirmed [27].

The approach of flipped learning, represented by self-study via annotated video and subsequent guided practice, also proved to be advantageous for immersive assistance systems, depending on the system level [28,29].

Practical experience from the collaborative industrial research project has shown that the methods of the Circuit Parcours Technique [14] cannot be used in industrial AR applications. Trails with group training or topic islands of any kind failed. The industrial users (aged 40–55) could only be successfully taught in individual sessions. This fact clearly shows that tried and tested learning models from higher education cannot be transferred to industry. It is important that any inhibitors that arise during practical training are recognised and resolved immediately; otherwise, immersive systems are likely to be rejected. For this reason, among others, such as social shame (technostress), the Circuit Parcours Technique is unsuitable for this target group.

The general model of the learning curve (see Figure 1a; [30]) has only limited applications for immersive assistance systems. The experience of the industrial research project has shown that the learning curves of immersive assistance systems are not characterised by sequential single-stage challenges. As a consequence, a dedicated approach for immersive assistance systems is presented on the basis of this contribution; see Figure 1b. The novice users face challenges at each system level. At first, they must be familiar with the hardware, i.e., position the HMD correctly on their head, which has a major influence on the success of the application and the quality of service [42] (e.g., visibility, hand interaction, consistency of the virtual assets, etc.). Once the user has successfully mastered this learning effort, they are faced with the challenge of the operation system and the AR interaction paradigm. As soon as they have mastered the basic principles of interaction, such as logging in and starting and exiting applications, they can approach the user journey of immersive applications. Finally, the user must overcome the application-specific challenges of the respective immersive applications and their complexities. It is important that the competence levels are mastered in ascending order: first hardware, then the operational system with the interaction paradigm, representing basic AR competencies, and then subsequently increasingly more specific application-specific competences. Only once the user is sufficiently familiar with all three competence levels can they utilise the easiness potential of operational tasks. Learning content should be designed accordingly. It is recommended that learning content is taught in sections and that these competences are first internalised with practice before further learning content from the next system level.

Determinants and models regarding the acceptance and adoption of technology were presented in accordance with Barnard et al. [31]. In our experience, for an optimised training scenario, it is always important to emphasise the benefits of technology at an early stage, such as the determinants of usability (effectiveness and completeness [7,34], perceived performance, effort expectancy, and facilitating conditions [30]). According to our findings, social influence cannot be influenced by the training scenario. Furthermore, the training scenario should be designed in such a way [35] that it is easy to learn (quickly overcoming obstacles), and the use case should point out recognisable efficiency potentials. One should also make sure that the extent and nature of the training is entertaining and memorable and, if possible, that the trainer avoids mistakes during the demonstration. It is also important for the training to convey a sense of enjoyment at an early stage and for the trainer to positively reinforce the initial training successes.

Training should emphasise both the specific benefits and the complementarity of the immersion system with the operational task. Furthermore, the trainer should adapt individually to the novice users' progress and quickly reduce the perceived complexity. During the training, the learner should always have a good situational awareness of the immersive scene (e.g., via a shared AR stream) and the opportunity to internalise what

they have learned by watching the supervisor, promptly and iteratively through practical exercises [36].

Factors of the model of technology acceptance for ease of learning [31] have been discussed in the bullet points above.

The model of technology acceptance from a system and user perspective [31] was also taken into account for the development of the optimised training scenario. The scenario should be transparent, clearly structured, and comprehensible. Attention must always be paid to appreciative feedback and, most importantly, queries must be dealt with in a committed manner. If errors occur during the training scenario, they must be resolved in a confident and explainable way. Through active interaction of the trainer with the user, related technologies, as well as features (e.g., analogies of the operating system Windows Mixed Reality to Windows 10/11), should be explained. It should also always be ensured that the previously acquired competences are recapitulated.

The effects of the different training qualities are analysed in the following experimental design.

3.2. Design of Experiment

In the current study, participants were initially trained on how to use an immersive assistance system for measuring various reference dimensions. Afterwards, the subjects were asked to perform various tasks on a HoloLens2 (HL2; e.g., open a website, enter a promocode, etc.), including the immersive measurement of several markers and reference objects in real space. Table 1 shows the subtasks of the respective system levels that the participants had to perform.

Table 1. Overview of didactic concepts and scenarios of training related to subtasks, OP, and system levels.

OP	System Level	Subtask	MT	DT	OT
01/15	hardware	placing the HMD correctly on the head, start and shutdown the operating system	TP	P	AV
02/15	operation system	system login via user credentials	TP	P	AV + GLBD
03/15		screensharing via Miracast with a monitor	TP	P	AV + GLBD
04/15		connected with Wi-Fi	TP	P	AV + GLBD
05/15		starting the MS Edge browser application	NTBI	NTBI	NTBI
06/15		entering URL to the website and ‘promo-code’	NTBI	NTBI	NTBI
07/15		closing the MS Edge browser application	NTBI	NTBI	NTBI
08/15		starting the research prototype ‘PlanAR’ application	AV	P	AV + GLBD
09/15		user login via credentials	AV	P	AV + GLBD
10/15	AR prototype app	opening hand menu dialog and activating the hand-ray measurement application	AV	P	P + GLBD
11/15		measuring the first reference via the hand-ray measurement application	AV	P	P + GLBD
12/15		measuring the second reference via the ‘hand-ray’ measurement application	AV	P	P + GLBD
13/15		switching to the ‘Bounding-Box’ measurement feature and measuring the physical reference asset	AV	P	P + GLBD
14/15		switching to the ‘point-to-point’ measurement feature and measuring dedicated references	AV	P	P + GLBD
15/15		close every application, shutdown the operating system, and take off the HMD	NTBI	NTBI	NTBI

Footer: TP = text and picture; AV = annotated video; GLBD = guided learning by doing; NTBI = not trained but instructed.

The research methodology implemented in this study is a between-subjects design. The performance of the subtasks was assessed using two dependent variables: operational progress and operational time, which were also used to calculate the new index. Acceptance was measured using the metrics derived from the UTAUT2 questionnaire [43]. Additional data were collected for further analysis, including the perceived workload as determined by the NASA TLX questionnaire [44] and physiological parameters to ascertain the level of stress (refer to technostress).

Quality of Training (QoT): The independent variable in this study was the quality of training, which was partitioned into three levels: minimal, detailed, and optimised training. These levels can be characterised as follows:

- Minimal Training (MT): the minimum training is a training scenario in an e-learning format, where a step-by-step tutorial consisting of a mix of text and picture descriptions is enriched with example videos. In addition, HL2's native 'Getting started' application 'Tips' was used for the interaction training to enable the user to study interactively. Each participant had the opportunity to internalise the tutorial as often and for as long as necessary. Table 1 shows the training content for the subtasks, the respective system level, and the corresponding QoT. This MT scenario as an insufficient quality of training should be considered as the training error.
- Detailed Training (DT): with this type of training, a supervisor personally presented and explained the content (see Table 1) in detail to the participant. The participant was able to see the live stream of the HL2 view on a monitor. The participant also had the opportunity to ask questions directly. This type of training also attempted to anticipate previous technological experience, age, and level of education.
- Optimised Training (OT): in this scenario, various methods were combined in a favourable way. The interaction paradigm was again conveyed to the test person interactively via the 'Tips' application. In accordance with Table 1, the basic hardware features were explained via video tutorials. The subtasks of the operating system were trained via annotated video tutorials. In addition, the participant was given the opportunity to practise the content once again via guided learning by doing. The subtasks of the AR application were first trained in a face-to-face session. HL2 was then handed over to the participant, and the content was expanded via guided learning by doing. The supervisor, meanwhile, accompanied the participant on the AR stream.

Operational progress (OP) is calculated from the number of subtasks the respondent was able to complete successfully up to the termination criterion; refer to Equation (1). Quitting was defined as a termination criterion (refer to the accuracy of the system [45]).

$$OP = \text{count of accomplished sub-tasks} / \text{total of sub-tasks}, \quad (1)$$

Total Operational Time (TOT) describes the full amount of time a participant needs to complete all subtasks successfully. Refer to Equation (2) and the metric for evaluations of AR systems [45].

$$TOT = \text{total time to accomplish sub-tasks}, \quad (2)$$

Total Operational Time with Penalties (TOTP) is the total time taken by the subject to complete the last successful subtask plus the penalty times for each unsuccessful subtask; refer to Equation (3). The penalty time per subtask is defined as the 75th percentile, which is calculated across all subjects for each subtask; see Equation (4).

$$TOTP = \text{total time of accomplished sub-tasks}_a + \text{sum of penalties for not completed subtasks}_{(1-a)}, \quad (3)$$

$$\text{Penalty}_k = \text{Percentile (0.75) of operational time for sub-task}_k \text{ about all participants}, \quad (4)$$

Performance: The two dependent variables, OP and TOT, and their respective extension, TOTP, are common metrics for the evaluation of immersive systems; refer to the meta-review by Dey et al. [45] and Cao et al. [46]. To evaluate the quality of service, as a superordinate metric as discussed by Moller et al. [47], the performance can be derived based on the speed–accuracy trade-off [48] according to Equation (5). This index is derived analogously to the visibility index by Rosilius et al. [42]. Accordingly, the effective performance is rated as high if a high OP is achieved by taking a low TOTP.

$$\text{Performance} = \text{OP} / \text{TOTP}, \quad (5)$$

UTAUT2 questionnaire (UTAUT2): The UTAUT2 questionnaire is used to determine the effects of QoT on the acceptance dimensions. Behavioural intention (BI) is used as the dependent variable and as an indicator of acceptance of the immersive system [49]. In accordance with the UTAUT2 questionnaire [43], further acceptance factors are also taken into account, including the following: the effects of the training error on performance expectancy (PE), effort expectancy (EE), hedonic motivation (HM), social influence (SI), habit (HT), and facilitating conditions (FC) are to be analysed. The UTAUT2 item ‘price value’ was consciously not asked. In the industrial context, monetary considerations do not play a role for the user because they rarely make financial decisions.

NASA TLX questionnaire (NASA TLX [44]): The NASA TLX questionnaire is surveyed to determine the impact of QoT on the workload. The perceived workload comprises the mean value of the corresponding subscales: mental demand (MD), physical demand (PD), temporal demand (TD), performance (P), effort (E), and frustration (F). This self-assessment of the workload is also used to evaluate perceived stress. Accordingly, MD is the psychological response as a component of the measurement of stress [50].

Physiological parameter: The Empatica E4 wearable device measured the level of stress as a physiological response [50] via the heart rate level (HR) and the skin conductance level (SCL). These metrics had to be compared to baselines measured before the experiment. To isolate the effects of stress, HR_Diff and SCL_Diff were calculated by subtracting the baselines from the action phases [51].

3.3. Experimental Setup

The experiment was conducted in a laboratory (8.32 m × 5.5 m). Figure 2 shows the setup schematically. At the beginning of this study, the participants were invited into the laboratory and sat down at the table (bottom right), where they answered the questionnaires on a laptop. In a subsequent relaxation phase, the physiological baselines were recorded. The live stream of HL2 was transmitted to the monitor via a Microsoft Miracast HDMI stick. The theoretical training session was spent on the left-hand side of the desk. There, the participant was either trained through e-learning (text and picture (TP), annotated video (AV)) or through personal training (PT). Within PT, the participant could follow both the supervisor acting and the AR stream on the screen. Participant questions were answered directly by the supervisor.

For subtask 11/15, two reference dimensions, 169 mm and 1077 mm, which were marked on the whiteboard using a marker pen, had to be measured via a ‘hand-ray’ measurement application (HRMA). The HRMA allowed the subject to interact directly with both hands. As an extension of their two forearms, a raycast was immersively projected from each of the participant’s hands against the wall (blue). The collision of the hand array with the wall was augmented as a cursor using a 2D ring. The next largest references (1342 mm and 2462 mm) were measured for subtask 12/15. For subtask 14/15, all two-point dimensions, including 8319 mm (from wall to wall), had to be measured using the ‘point-to-point’ measurement application (PTPMA). With this feature, measurement points can be arbitrarily translated and positioned as virtual assets (small cubes) according to the MRTK interaction paradigm [2]. A virtual measuring tape is annotated between the small cubes (see measuring tape 2462 mm in Figure 2).

Subtask 13/15 involves measuring the flight case by precisely scaling and positioning the virtual 'bounding box' (blue) measurement application (BBMA) over the flight case as an enveloping object using the interaction paradigm of the MRTK [2]. The Cartesian dimensions are shown by the blue 'bounding box'.

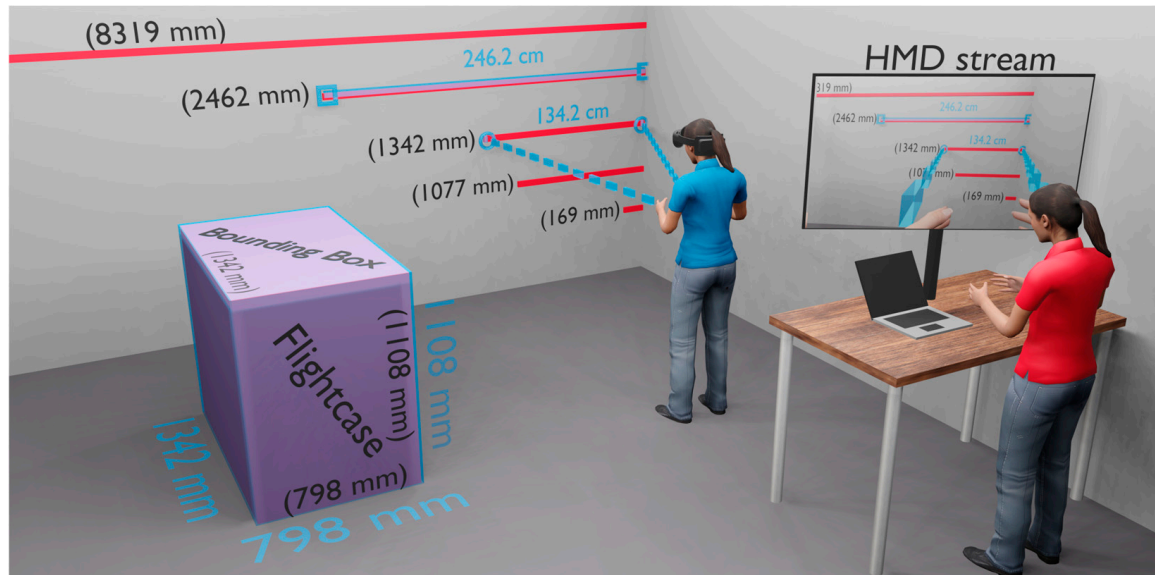


Figure 2. Schematic representation with red indicating physical references, purple indicating a physical flight case, and black information providing a description to the readers; all blue annotations were virtually augmented via HL2.

3.4. Experimental Plan

This study was conducted according to the experimental flowchart; see Figure 3. At the beginning of the experiment, the participants were asked to answer a socio-demographic questionnaire. QoT was randomly assigned to each participant. The subject was trained by the supervisor according to the resulting type of training. The participant was briefed on each subtask and asked to perform the subtask accordingly. Both the time taken to complete (see TOT) and the success (see OP) of the subtask were manually documented by the supervisor. As soon as the participant successfully completed the subtask, the next subtask was instructed in the same way as described above. This procedure was continued until either all 15 subtasks were successfully completed or until 1 subtask was not finished successfully. Subsequently, the so-called relaxation phase (baseline) of the physiological parameters was measured for 5 min using the wearable Empathica E4. At the end of the experiment, the participant was asked to complete the UTAUT2 and the NASA TLX questionnaires.

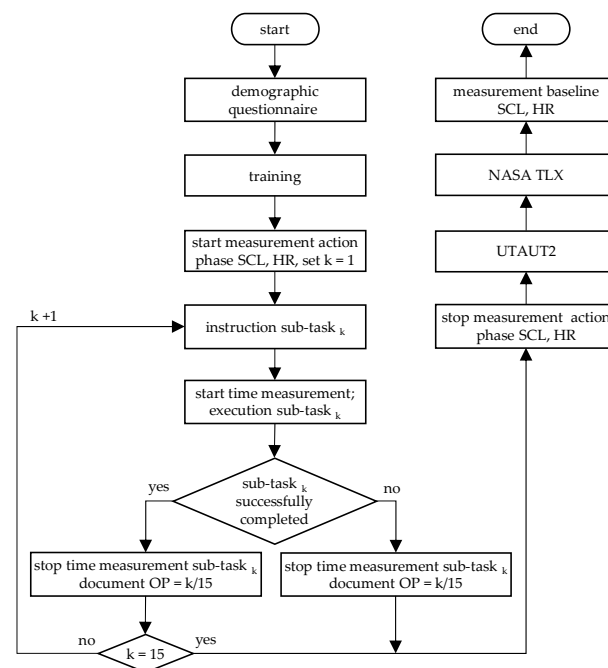


Figure 3. Experimental flowchart for each quality of training.

3.5. Experimental Procedure

Since no estimate of the sample size could be derived from the related work, an a priori calculation via G*Power V3.1.9.7 was performed from the data of the exploratory preliminary study (in week 43 of 2021 with $n = 20$ probands); see Section 3.1 and Table 2.

Table 2. Calculated sample size via G*Power.

Statistical Parameter	Performance	BI (UTAUT2)
η^2_p	0.3	0.124
Power	0.8	0.8
α	0.05	0.05
Number of groups	2	2
Calculated sample size	22	58

This study was conducted during weeks 25 and 26 of 2022. All data sets were recorded anonymously. The participants joined this study voluntarily and received a monetary compensation of EUR 12. Each participant was informed about the guidelines for good ethical research and had to give their explicit consent to participate. The experiment's duration was approximately 45 min. The equipment was completely disinfected after each participant.

3.6. Statistical Procedure

The statistical analysis was run in Jamovi 2.4.80. To analyse the influence of QoT on the dependent variables (OP, TOTP, performance, acceptance, and physiological parameters), individual ANOVAs were performed with subsequent post hoc tests [52]. Due to the high sensitivity of the SCL and HR, as well as to avoid an over-interpretation of anomalies in the statistical procedure, outliers were cleansed by means of Winsorising (Equations (6) and (7)) [53,54] as follows:

$$\text{upper boundary} = Q3 + 1.5 \times IQR, \quad (6)$$

$$\text{lower boundary} = Q1 - 1.5 \times IQR. \quad (7)$$

As an assumption check, homogeneity and normality were tested. In case of violation of homogeneity, Welch's ANOVA was carried out. If the normality was not given, the ANOVA could be considered to be robust because a sample size higher than 30 ($n = 52$ probands) was met [55,56]. To guarantee statistical validity, dedicated alpha corrections were applied to the post hoc tests. The Tukey test is a conservative method applicable when group variances and the sample size are equal [57]. The Scheffe test, although very conservative, is suitable for scenarios with varying group variances and sample sizes [58].

4. Results

4.1. Analysis of Data

Demographic data: A total of $n = 52$ participants attended this study, 90% male and 10% female. All participants were enrolled as students at a technical university of applied sciences. The participants were aged between 19 and 39 years, and the average age was 25.8 years. Figure 4 shows the results of the experience level in dealing with the mixed reality technologies regarding the self-assessment of the participants.

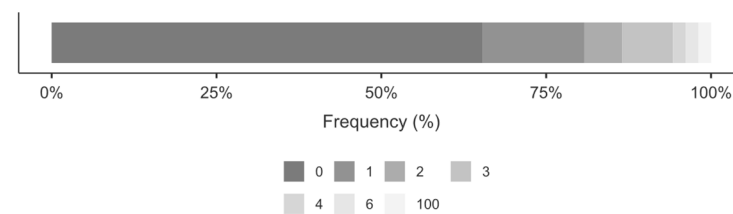


Figure 4. Distribution (%) of experience level in hours.

Duration of Training: Descriptively, the duration of training for MT had a mean of 849 s ($SD = 83.3$), for DT, a mean of 813 s ($SD = 246$), and for OT, a mean of 1329 s ($SD = 189$).

QoT: Every condition of the independent variable QoT led to significant changes in all dependent variables, which are presented in detail below and highlighted with a star (*).

Performance: According to the performance, a statistically significant difference was shown for QoT $F(2, 29.6) = 12.83$; $p < 0.001$; $\eta^2 p = 0.40$; and power = 0.99. The conditions of normal distribution and the variance homogeneity were violated. Welch's ANOVA was calculated as compensation, which is robust for $n > 30$ with the absence of a normal distribution. The post hoc comparison with the Scheffe alpha correction shows statistically significant differences between MT and OT $t(49) = -5.41$; $MD = -2.47\%/min$; $p < 0.001$; and $d = -1.828$, as well as between DT and OT $t(49) = -4.12$; $MD = -1.88\%/min$; $p < 0.001$; and $d = -1.392$. Figure 5 shows the corresponding boxplots.

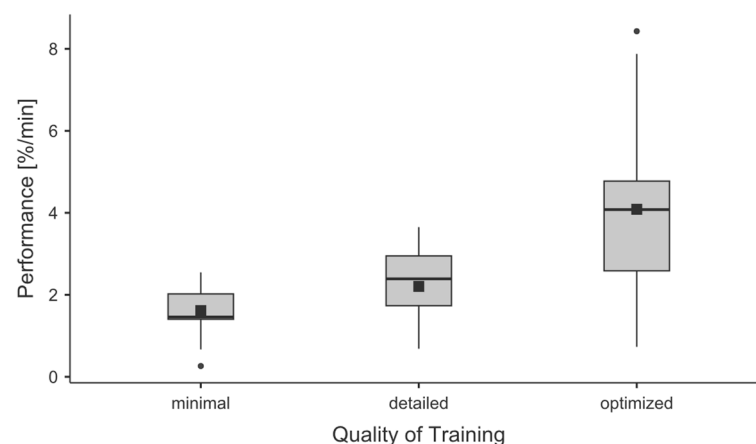


Figure 5. Boxplot of the performance (%/min)–quality of training (QoT).

OP: For the OP, a statistically significant difference within the independent variable QoT $F(2, 49) = 5.28$; $p = 0.008$; $\eta^2 p = 0.177$; and power = 0.84 could be proven via ANOVA.

Here, the variance homogeneity is met, while the normal distribution is violated. Due to the large sample size, the violation can be compensated for. The post hoc comparison with the Tukey alpha correction demonstrates a statistically significant difference between MT and OT $t(49) = -3.103$; $MD = -21.46\%$; $p = 0.009$; and $d = -1.049$. Figure 6 represents the corresponding boxplots with the mean, median, and IQR.

TOTP: The statistical significance of the difference between QoT could also be proven for TOTP $F(2, 28.8) = 21.67$; $p < 0.001$; $\eta^2p = 0.51$; and power = 0.99. Since the assumption check of homogeneity was violated, Welch's ANOVA was applied. Moreover, the distribution of normality was given. The post hoc comparison with the Scheffe alpha correction indicates statistically significant differences between MT and OT $t(49) = 6.592$; $MD = 593.3$ s; $p < 0.001$; and $d = -2.230$, as well as between DT and OT $t(49) = -5.688$; $MD = 511.9$ s; $p < 0.001$; and $d = -1.924$. Figure 7 shows the corresponding boxplots with the mean, median, and IQR.

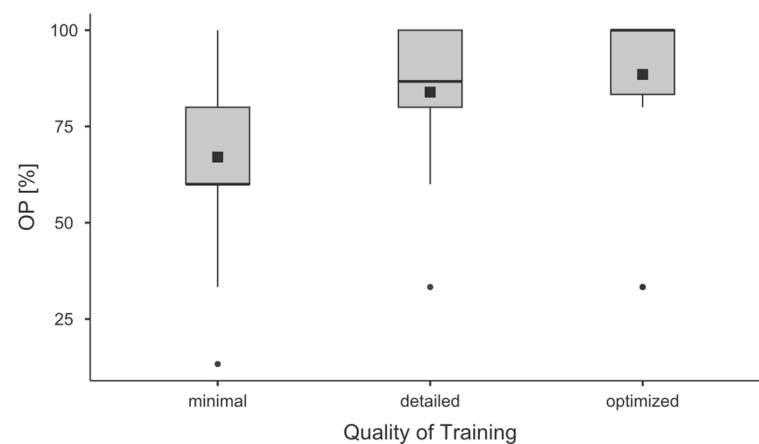


Figure 6. Boxplots of the OP–QoT.

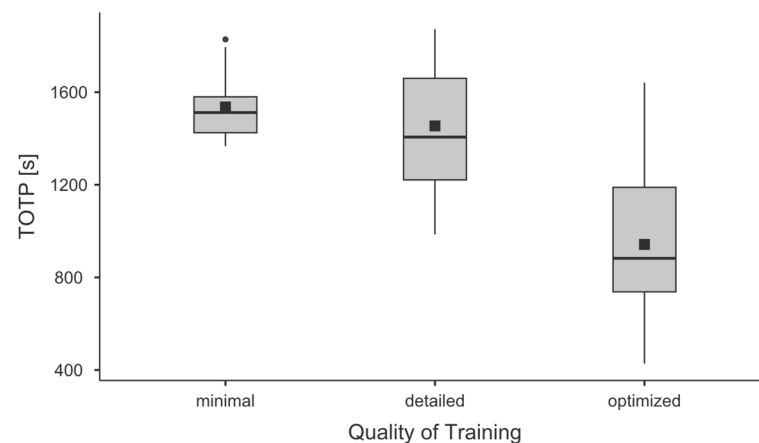


Figure 7. Boxplots of the TOTP (s)–QoT.

UTAUT2: The subscales of UTAUT2 were tested for differences between the three levels of the independent variable QoT using ANOVA; see Table 3. The condition of variance homogeneity is met for all subscales. The normality distribution is only violated for the scales SI, FC, and HM. However, an ANOVA is robust against this violation if the sample size is large enough (see above), as in this case with $n = 52$. The primary focus is on behavioural intention. The results show a statistically significant difference between the training scenarios. Training also influenced the acceptance factors PE, EE, FC, and HM.

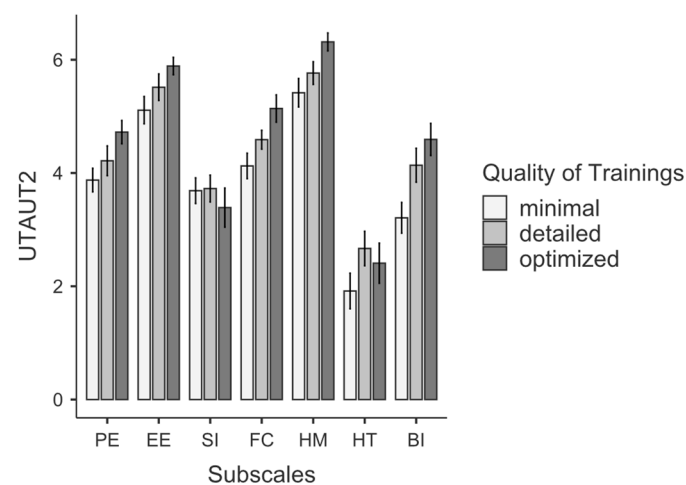
Table 3. ANOVAs of the subscales of UTAUT2–QoT. (significant changes are highlighted with a ‘*’).

Subscales	df1, df2	F	p	η^2p	Power
PE *	2, 48	3.53	0.037	0.128	0.66
EE *		3.41	0.041	0.124	0.65
SI		0.445	0.643	0.018	0.12
FC *		4.384	0.007	0.188	0.86
HM *		4.96	0.011	0.171	0.81
HT		1.31	0.280	0.052	0.29
BI *		5.99	0.005	0.200	0.89

A post hoc comparison with Tukey alpha corrections was also carried out. For the subscales PE, EE, FC, HM, and BI, statistically significant differences were found between MT and OT. The corresponding statistical parameters can be found in Table 4. Figure 8 shows the corresponding bar chart.

Table 4. Subscales of the UTAUT2 post hoc comparison with the Tukey alpha correction. (significant changes are highlighted with a ‘*’).

Subscales	QoT		Mean Diff	t	p _{Tukey}	Co’d
PE	minimal	optimised	−0.847	−2.63	0.030 *	−0.903
EE	minimal	optimised	−0.780	−2.61	0.032 *	−0.898
FC	minimal	optimised	−1.014	−3.33	0.005 *	−1.143
HM	minimal	optimised	−0.898	−3.11	0.009 *	−1.068
BI	minimal	optimised	−1.384	−3.42	0.004 *	−1.174

**Figure 8.** Bar chart (with standard errors) of the subscales of UTAUT2–QoT.

NASA_TLX: To assess the perceived overall stress, the Total_Mean was calculated across all subscales of NASA TLX. The ANOVA $F(2, 48) = 8.15$; $p < 0.001$; $\eta^2p = 0.254$; and power = 0.96 shows statistically significant differences in workload due to QoT. The prerequisites of normal distribution and variance homogeneity are fulfilled here. Furthermore, corresponding ANOVAs across all subscales were calculated; see Table 5. The requirements for normal distribution of the data are met for all subscales, except for physical effort and performance. When testing for variance homogeneity, the requirements for all subscales, except for performance, were met. As before, Welch’s ANOVA was used for the analysis due to violations of the requirements. Table 6 shows the details of the corresponding post hoc comparison. For the subscale MD, statistically significant differences were found between MT–OT and DT–OT. According to the subscale P, a significant difference was shown between the factor levels MT and OT. For subscale E, a significant difference was

identified between DT and OT. For subscales F and Total_Mean, statistically significant differences were demonstrated between MT–OT and DT–OT. Figure 9 demonstrates the corresponding bar chart.

Table 5. ANOVA of the subscales of NASA TLX–QoT. (significant changes are highlighted with a ‘*’).

Subscales	df1; df2	F	<i>p</i>	η^2p	Power
MD	2; 48	4.84.	0.012 *	0.167	0.81
PD	2; 48	0.369	0.693	0.015	0.11
TD	2; 48	0.879	0.422	0.035	0.20
P	2; 30	5.130	0.001 *	0.282	0.98
E	2; 48	4.96	0.011 *	0.171	0.82
F	2; 48	5.79	0.006 *	0.194	0.87

Table 6. Subscales of the NASA TLX post hoc comparison with the dedicated alpha correction. (significant changes are highlighted with a ‘*’).

Subscales	Alpha Correction	QoT		Mean Diff	t	<i>p</i>	Co’d
MD	Tukey	minimal	optimised	17.19	2.464	0.045 *	0.846
		detailed	optimised	19.56	2.848	0.017 *	0.963
P	Scheffe	minimal	optimised	35.3	4.33	<0.001 *	1.488
E	Tukey	detailed	optimised	20.46	3.081	0.009 *	1.042
		minimal	optimised	29.42	3.162	0.008 *	1.086
F	Tukey	detailed	optimised	23.96	2.616	0.031 *	0.885
		minimal	optimised	20.32	4.106	<0.001 *	1.411
Total_Mean	Tukey	minimal	optimised	16.97	3.484	0.003 *	1.178
		detailed	optimised				

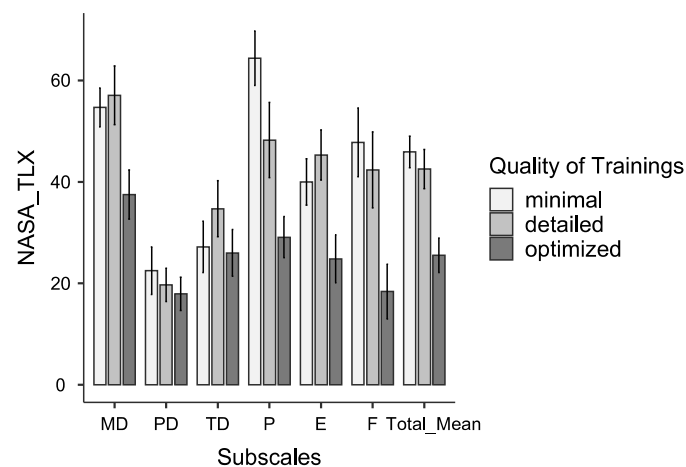


Figure 9. Bar chart (with standard errors) of the subscales of NASA TLX–QoT.

Physiological Parameters: The statistically significant influence of training quality on physiological responses was demonstrated via ANOVAs using the relative mean HR_Diff $F(5, 45) = 12.2$; $p < 0.001$; $\eta^2p = 0.352$; and power = 1.00 and the relative SCL_Diff $F(2, 45) = 5.20$; $p = 0.009$; $\eta^2p = 0.188$; and power = 0.83. The conditions for variance homogeneity were met by both the HR and SCL, while the condition for normal distribution was met only by the SCL. However, the violation of the normal distribution can be compensated for by the sample size of $n = 48$. The post hoc comparison with the Tukey correction shows statistically significant differences for HR_Diff between MT–OT $t(45) = 2.88$; MD = 4.91 BPM; $p = 0.045$; and $d = -1.036$ and DT–OT $t(45) = 4.920$; MD = 8.11 BPM; $p = 0.017$; and $d = 1.712$. For SCL_Diff, a statistically significant difference between DT and OT $t(45) = 3.130$; MD = 1.309 μS ; $p = 0.008$; and $d = -1.090$ could be demonstrated. Figure 10 shows the corresponding boxplots with the means, median, and IQR.

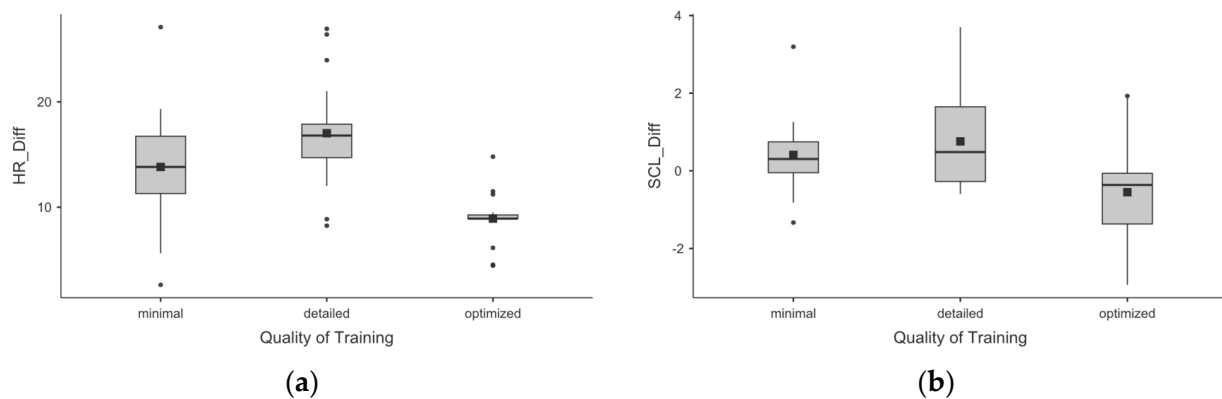


Figure 10. Boxplots of the (a) HR_Diff (BPM) and (b) SCL_Diff (μ S)–QoT.

Figure 10 shows the corresponding boxplots.

The measurement accuracies of the HRMA, PTPMA, and BBMA measurement functions from the operational task are documented in Appendix B.

4.2. Interpretation of the Results

The results of this study allow the following conclusions to the hypotheses:

Hypothesis H_{A.1}. *Insufficient training has a negative impact on the performance of an immersive assistance system.*

Proof of Hypothesis H_{A.1}. is confirmed and, thus, Hypothesis H_{A.0} is rejected. $\square$

In detail, it can be shown with significance that the mean OP between OT and MT decreases by 21.46% ($p = 0.009$; Co'd = -1.049) and between DT and MT by 16.86% ($p = 0.050$; Co'd = -0.825). From the perspective of TOTP, however, the mean from OT to MT can be significantly increased by 593.3 s ($p < 0.001$; Co'd = 2.230) and from OT to DT by 511.9 s ($p < 0.001$; Co'd = 1.924). The mean of the index performance is significantly reduced by the switch from OT to MT by 2.47%/min ($p < 0.001$; Co'd = -1.828) and from OT to DT by 1.88%/min ($p < 0.001$; Co'd = -1.392). Assuming that the MT has an inferior training quality and thus represents faulty training, the post hoc comparisons of the OP, TOTP, and performance show a statistically significant deterioration between the optimised and faulty training scenarios. Due to the high statistical power and high effect sizes, Hypothesis H_{A.1} can be confirmed.

Hypothesis H_{B.1}. *Insufficient training has a negative impact on the acceptance of an immersive assistance system.*

Proof of Hypothesis H_{B.1}. is confirmed and, thus, Hypothesis H_{B.0} is rejected. $\square$

The key significance for acceptance lies in the BI item, whose ANOVA $F(2, 48) = 5.99$; $p = 0.005$; $\eta^2 p = 0.200$; and power = 0.89 shows a significant difference within the independent variable QoT. In detail, a significant reduction in the mean by 1384 units (based on a seven-point Likert scale) from the factor level OT to MT was demonstrated via post hoc tests. This statement is also supported by significant differences in other subscales of UTAUT2 (see Tables 3–6, A1 and A2. and Figure 8). The faulty training scenario also reduces the subscales (compared to OT) of the PE (Mean_Diff = -0.847 ; $p = 0.030$; Co'd = -0.903), EE (Mean_Diff = -0.780 ; $p = 0.032$; Co'd = -0.898), FC (Mean_Diff = -1.014 ; $p = 0.005$; Co'd = -1.143), and HM (Mean_Diff = -0.898 ; $p = 0.009$; Co'd = -1.068). This leads to the conclusion that the user expects lower performance (output and quality) and less enjoyment combined with greater effort with a worse training scenario. According to the subscale FC,

a poorer training scenario also means that conditions such as resources, expertise, previous experience, and troubleshooting are assessed as poorer. As expected, the subscales SI and HT show no statistically significant differences, which is also logically understandable. At this time, the quality of training should have no influence on the social environment or the habits of the probands. HT and SI can only arise when the participants are working with AR and in conversation with each other.

Hypothesis H_{C.1}. *Insufficient training leads to higher workload and stress levels for an immersive assistance system.*

Proof of Hypothesis H_{C.1}. is confirmed and, thus, Hypothesis H_{C.0} is rejected. □

The ANOVA of the Total_Mean (perceived total stress) $F(2, 48) = 8.15$; $p < 0.001$; $\eta^2p = 0.254$; and power = 0.96 across all subscales of the NASA TLX shows a significant difference within the independent variable QoT. The mean was increased by 20.32 units (scale 0–100) from OT to MT (insufficient training) and by 16.97 units from OT and DT. This statement is also supported by significant differences in other subscales of the NASA TLX (see Table 5 and Figure 9). The mean MD rose significantly between OT and MT by 17.19 units ($p = 0.045$; Co'd = 0.846) and between OT and DT by 19.56 units ($p = 0.017$; Co'd = 0.963). The inverted mean performance increased from OT to MT by 35.3 units ($p < 0.001$; Co'd = 1.488). The mean F increased statistically significantly from OT to MT by 29.42 units ($p = 0.008$; Co'd = 1.086). The mean E increased significantly from OT to DT by 20.46 units ($p = 0.009$; Co'd = 1.042) and from OT to MT by 15.17, but this could not be statistically verified ($p = 0.073$; Co'd = 0.772). As expected, no statistically significant effects of insufficient training on PD and TD could be identified. According to Alsuraykh [50], MD is a psychological response is representative of stress, which has been shown to be statistically significant. The suggestion that the training scenario has an effect on stress is also confirmed by the physiological response of the means HR_Diff $F(5, 45) = 12.2$; $p < 0.001$; $\eta^2p = 0.352$; and power = 1.00 and SCL_Diff $F(2, 45) = 5.20$; $p = 0.009$; $\eta^2p = 0.188$; and power = 0.83 via ANOVAs. In detail, the HR_Diff is significantly increased from OT to MT by 4.91 Hz ($p = 0.045$; Co'd 1.036) and from OT to DT by 8.11 ($p = 0.017$; Co'd = 1.712). Moreover, SCL_Diff also shows a significant increase from DT to MT by 1.309 μS ($p = 0.008$; Co'd 1.09).

Research Question 1. *What effect does the quality of training have on the performance of an immersive assistance system?*

Answer 1. In short, QoT has a significant impact on performance. For the dependent variables, statistically significant differences between QoT could be analysed using ANOVAs OP $F(2, 49) = 5.28$; $p = 0.008$; $\eta^2p = 0.177$; and power = 0.84 and TOTP $F(2, 28.8) = 21.67$; $p < 0.001$; $\eta^2p = 0.51$; and power = 0.99, as well as the index performance $F(2, 29.6) = 12.83$; $p < 0.001$; $\eta^2p = 0.40$; and power = 0.99. Within the sample, a performance of 1.61%/min was achieved for the operational tasks with QoT corresponding to the MT scenario, i.e., *insufficient* training. A performance of 2.20%/min was achieved in the DT scenario, which corresponds to detailed, personalised training. In contrast to the OT scenario, the subject was able to perform their task at 4.09%/min, which corresponds to an improvement by a factor of approximately 2.5 in relation to the erroneous training.

Research Question 2. *What effect does the quality of training have on the human factors of acceptance and stress in an immersive assistance system?*

Answer 2. According to the results, it was shown that QoT has a significant influence on acceptance (BI) and the subscales PE, EE, FC, and HM. This demonstrates that insufficient training not only reduces the acceptance of the immersive assistance system but also increases the fun and enjoyment of using it. QoT also has a significant influence on perceived stress and frustration when performing operational tasks. It was also shown

that QoT has a significant influence on stress during operational tasks with an immersive assistance system.

5. Discussion

In contrast to most existing AR studies [11–13], our contribution does not focus on learning with AR applications but on training scenarios for immersive systems. The critical literature review revealed that there is no contribution that deals with training scenarios, evaluation methods, or guidelines specifically for learning the usage of immersive systems in the industrial context. Dörner et al. [14] developed an AR learning course for the university environment, but its implications are only transferable to industrial use to a very limited extent.

The uncertainty of preliminary work, whether and what effect end-user training has on performance and acceptance, is addressed by the research questions and associated results of our study. Our results contribute to this research gap. In general, it is widely understood that end-user training is an important factor that requires further research [7,16–19,22]. But, contrary to the results of Zoetbrood [20], our study shows that training indeed has a significant influence on productivity ($p = 0.009$; $\text{Co'd} = -1.049$). First, Zoetbrood's correlation matrix [20] showed only a very small, non-statistically significant effect (Pearson = 0.69; $p > 0.05$) between training practices and productivity. Second, the hypothesis 'The more training and development practices used in an organisation the greater the productivity benefits because of ICT investments' could not be confirmed on the basis of the multivariate analysis ($\beta = 0.27$, $t(98) = 1.003$, $p = 0.318$). In particular, the development of hypothesis H2 by [20] is based only on indirect and weak assumptions, so further investigations are necessary. In addition, the research design by [20] and the related work do not allow any conclusions to be drawn about a differential analysis of the quality and quantity of the inherent training, which, in turn, shows a relevant difference to our DOE. But, we agree with the described interaction effects that performance can only be exploited effectively if other (facilitating) conditions, such as the integration of training measures into process structures, linked with organisational changes, and the necessary continuity, are also implemented.

The sensitivity of performance with regard to QoT supports the postulated mismanagement as a cause of the productivity paradox [3], indicating that insufficient training in ICT and missing organisational changes can lead to substantial productivity loss. The statement by Schweikl and Obermaier [3] that companies should invest in training and development programmes to ensure the correct use of new technologies and strengthen their workforce can also be confirmed. Due to the differences in the technologies, DOE, and hypotheses used, the results of the preliminary work could not be directly comparable to our study. The preliminary work could, therefore, only be considered as a basis for the hypotheses. In line with the results from the oral surgeons' systems [7], the influence of end-user training on the acceptance factors, PE (Pearson = 0.769 with significance) and EE (0.840 with significance), was confirmed when transferred to immersive assistance systems according to our results with PE ($p = 0.030$; $\text{Co'd} = -0.903$) and EE ($p = 0.032$; $\text{Co'd} = -0.898$). Both studies exhibit relatively high effect sizes, but [7] does not allow any conclusions to be drawn about the scope and quality of the training scenarios. Similarly, Alqahtani et al. [25] showed that educational quality has an influence on the acceptance factors' perceived ease of use and perceived usefulness for e-learning systems. We also confirm the influence of training opportunities on intention to use (=BI) in the field of immersive assistance systems. In contrast to Giovanni Mariani et al. [6], we did not conduct a confirmatory study based on the TAM model. However, the results can be transferred to the extent that the influence of training opportunities on the determinants, EE (which corresponds to perceived ease of use) and PE (which corresponds to perceived usefulness), can be confirmed by the UTAUT2 model.

In line with the implications of Korunka and Vitouch [23] and Wang et al. [24], our results confirm that the appropriate QoT can reduce the stress levels caused by ICT. This preliminary work indicates the need for appropriate training scenarios, but no proof has

been provided yet. Bala and Venkatesh [10] investigated the effects of training effectiveness on perceived threat, which showed highly significant effects. In our research, the perceived total stress respective workload (Total_Mean) across various subscales of the NASA TLX increased by 20.32 units from OT to MT and by 16.97 units from OT to DT. Other subscales of the NASA TLX also exhibited significant differences (see Table 5). The Mean MD as the psychological response of stress increased significantly between OT and MT by 17.19 units ($p = 0.045$; $Co'd = 0.846$) and between OT and DT by 19.56 units ($p = 0.017$; $Co'd = 0.963$). The physiological responses also supported the impact of insufficient training. HR_Diff increased significantly from OT to MT by 4.91 Hz ($p = 0.045$; $Co'd = 1.036$) and from OT to DT by 8.11 Hz ($p = 0.017$; $Co'd = 1.712$). SCL_Diff exhibited a significant increase from DT to MT by 1.309 μS ($p = 0.008$; $Co'd = 1.09$). But, the results of the SCL_Diff show a conspicuous difference in detail, as DT compared to MT ($p = 0.008$) shows a statistically significant difference, but MT compared to OT does not ($p = 0.077$). The analyses show negative values for Mean SCL_Diff at OT, i.e., that the SCL was higher in the baseline (resting) phase than in the action phase. This could be explained by the high susceptibility of the measurement system to interference, including electrode wear, skin care products, disinfection of the hands or incorrect positioning, sequence effects during measurement, etc.

It should be emphasised that the teaching methods and models from the educational sector are not directly transferable to the industrial context. According to Barnard et al. [31] and Giovanni Mariani et al. [6], age, level of education, technology affinity, intention to learn, and organisational conditions, such as the expectation of productive success, etc., play a pivotal role as differentiating factors. This can also be deduced from the large standard deviation in the mean performance in the OT scenario of 2.09%/min. These factors were also confirmed by our own experience from the industrial research project. In summary, we recommend that training for immersive assistance systems should always be based on the user's specific tasks. There should also be as few barriers as possible and a good trainer for troubleshooting. A mix of formal methods with blended learning approaches, personal guidance, and guided learning-by-doing sessions is recommended. It is important to differentiate between system levels, e.g., for hardware and the operational system. It seems advantageous to present the learning content via structured videos with analogies to familiar knowledge (Windows operations). In general, it has been found that when there is a lot of input, learners tend to forget important learning content and become overwhelmed or confused. These findings contradict the implications of Korpelainen and Kira [22], Dörner and Horst [14], and Giovanni Mariani et al. [6]. Figure 1b presents postulated special learning curves for immersive systems with different task difficulties and learning efforts per system level. Handling the hardware and the operational system is relatively simple. The main challenge with the operational system (HL2) is the interaction paradigm. The similarity to Windows means that the user journey is easy to manage, based on previous experience. Here, the user encounters a customised user interface, incorporating a demanding interaction paradigm and spatial dependencies, among others. It can, therefore, be indirectly concluded that errors within the human–computer interaction loop can be avoided through suitable training content [16]. Our study addresses the demand for acceptance research with augmented reality in an industrial context [17] and with the aim of finding training scenarios for corporate training outside schools and academic institutions [18].

A critical examination of our study allows the following claim to be made: with more effort, the levels of the independent variable QoT could have been designed more appropriately. In order to reproduce realistic corporate conditions, the MT scenario used correct content but with very simplified learning and insufficiently designed training methods. Naturally, there is no limit on how incorrectly this scenario can be designed. The intermediate DT level was designed so that a trainer endeavours to convey the training content as effectively as possible in person during a session. Other scenarios are conceivable here, such as a trainer accompanying a GLBD session throughout, which could theoretically

deliver promising results. As described in Section 3.1, a mixed-methods optimisation approach was chosen. Of course, there is still further research potential in the OT scenario.

In contrast to conventional studies [45,46] in which performance was always assessed separately via duration and accuracy, which can lead to unreliable results, the performance index was refined for this DOE. Performance is derived using the speed–accuracy trade-off [48], following Equation (5). This index is analogous to the visibility index proposed by Rosilius et al. [42]. In relation to the overarching research approach by Rosilius et al. [19], our contribution has shown that a deficient training scenario has a negative impact on performance, acceptance, and stress. In the case of erroneous speech interaction, a significant difference in performance, acceptance, and frustration was found, but not in terms of strain and stress. Hereby, a contribution was made within this research context, and a dedicated research design for the independent variable QoT was developed, which provides an opportunity for further research activities. In quantitative terms, the BI could be reduced by an average of 1.4 units due to an incorrect training scenario. Compared to Rosilius et al. [19], the BI was degraded by 2.3 units on average due to an erroneous speech interaction. These differences in mean values or weightings of different errors in human–machine interaction need to be analysed further using inferential statistics, as these purely qualitative observations do not allow any general statements to be made.

6. Conclusions and Further Research

This paper addresses the general research need for adapted training regarding emerging technologies. It also addresses the primary research gap on the impact of the quality of immersive assistance system training scenarios on performance, acceptance factors, and technostress. Based on the confirmation of all three hypotheses, the results of this study show that QoT has a significant influence on performance, acceptance, workload, and the physiological parameters, the HR and SCL. The selected research design has high statistical validity with high effect sizes and resulting statistical power (mean = 0.92). As a secondary result, the performance index offers a new metric for evaluating the effectiveness and comparability of training scenarios. In addition, the DOE has a high degree of reproducibility; it thus may be used for further research in this area. This article also presents an approach on how an optimised training scenario for immersive systems can be evaluated. Relevant findings can be derived from this, including the need to accommodate the different system levels with different training methods. Indirectly, a causal hierarchy has been discovered: the operator must first master the hardware level (including the correct positioning of the HMD on their head) before focussing on the downstream basic functionalities, such as the interaction paradigm and the operating system. Consecutively, the learner should only learn dedicated functions of immersive applications after the operating system level. Generalised design guidelines based on the findings of the OT are provided in Appendix A. Another important finding is that it is not the personalised training (refer to DT) that yields the best results but the mixture of methods for the respective system level.

Further research demands can be derived from the findings of this DOE and the results.

- Verification of the results obtained and the general DOE regarding performance, acceptance, workload, and technostress by third parties;
- The mixture of methods per system level proves to be advantageous and could be used for further research. The constraints and areas of validity should be further specified;
- The OT should be investigated in more detail and the scenario further optimised;
- The dependencies of the appropriate training scenario regarding the specific application, the assistance system, and the specialised task must be investigated further;
- It is necessary to investigate in detail what weightings a faulty assistance system and a faulty training scenario have on the dependent variables. One research question could be whether the performance or loss of acceptance of the immersive system can be compensated for by adapted training;
- In addition, dependencies of the OT on hardware should be investigated.

The results of the operational measurement task are presented as an added value in this article in Appendix B. These are intended to serve both as proof of the current technological maturity of the assistance system and as a motivation for the industrial sector to deploy comparable immersive assistance systems. The research association has shown that the trade-off between measurement deviation and measurement effort (e.g., time required) is indirectly correlated with the accessibility (reachability) of the measurement object. In other words, the more difficult it is for the operator to reach real reference points, the more advantageous the immersive measuring application is compared to conventional measuring methods. For this reason, the research question and the study design were examined in an industry-related context because, in the short to medium term, the industry is facing the challenge of applying practicable training scenarios.

Summarising the findings, it can be concluded that QoT has a significant impact on performance, acceptance, workload, and technostress. A faulty training scenario can have a significant negative impact on these metrics. According to the current state of research, this type of error has been significantly underrepresented to date. Nevertheless, the effect of training should be consciously taken into account as a disruptive influence when assessing the potential for the productivity of the human–machine interface. This leads to the conclusion that it is not the design of the assistance system that may be inadequate but the training scenario that has not yet been effectively implemented. This research topic is particularly relevant to the operationalisation of the so-called industrial metaverse and could be one of the key factors in the productivity paradox. Moreover, it is important to investigate whether these implications are also valid for new generations of immersive headsets, such as Meta Quest 3, Apple Vision Pro, and other future HMDs.

Author Contributions: Conceptualisation, M.R., L.H., V.B. and B.L.; methodology, M.R., M.G., L.H., B.W., I.v.E., B.L. and V.B.; software, M.R. and L.H.; validation, M.G., V.B., B.W. and B.L.; formal analysis, L.H., M.G. and M.R.; investigation, M.R., L.H. and B.W.; resources, M.R., B.W. and V.B.; data curation, M.R. writing—original draft, M.R., L.H. and B.W.; writing—review and editing, V.B. and B.L.; visualisation, M.R., B.W. and L.H.; supervision, V.B. and B.L.; project administration, M.R., L.H. and V.B.; funding acquisition, V.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the publication fund of the Technical University of Applied Sciences Würzburg-Schweinfurt.

Institutional Review Board Statement: Ethical review and approval were waived for this study due to the design of the study in accordance with the rules of the Ethical Committee of the University of Regensburg.

Informed Consent Statement: Written informed consent has been obtained from the participants to publish this paper.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

Conflicts of Interest: The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results. Author Lukas Hügel was employed by the company T&O Unternehmensberatung GmbH and Author Ingo von Eitzen was employed by the company Regain Development. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Appendix A

Table A1. Design guidelines of industrial training for immersive assistance systems.

No.	System Level	Task	Didactic Methodology
1	All	General	
During training, it is important for the learner to have a clear understanding and awareness of the immersive scene and to be able to apply what they have learned through practical exercises. The trainer should individualise their methods and make the training less complex for the novice user. Feedback should be appreciated and addressed in a committed manner. Any errors that occur during the training scenario should be corrected confidently and explained. The training should be engaging, memorable, and as flawless as possible. It is also important to convey the joy of use and appreciate initial training successes. Additionally, the training should highlight the advantages of the immersive system and its usefulness with the operational task.			
No.	System Level	Task	Didactic Methodology
2	Hardware	General, getting started	Video tutorial
Before using AR HMD for the first time, users should be made aware of the opportunities and risks through a video. In addition to the possibilities offered by the technology, there is also a risk of triggering health problems, such as dizziness. Users also need to be sensitised to the correct way to use this technology. Due to the limitation of hardware resources, for example, the system takes time in certain situations to load and process the retrieved content correctly.			
No.	System Level	Task	Didactic Methodology
3	Hardware	Switching on and off	Video tutorial
Switching the HMD on and off should be explained in a video tutorial in which an instructor holds the HMD in their hands in front of the camera. The instructor must clearly show where the ON/OFF button is located and explain how long it must be pressed to start and shut down the system (switch on: press briefly for approximately one second; switch off: press and hold for approximately five seconds until all lights go out and a signal tone is played).			
No.	System Level	Task	Didactic Methodology
4	Hardware	Correct wearing position	Video tutorial
A video tutorial should be used to show how to put on and take off the AR HMD and position it correctly on the user's head. The instructor can be seen from the side performing and commenting on the process (to put on the AR HMD, tilt the visor upwards, put on the HMD and adjust to the head circumference using the adjustment wheel, tilt the visor all the way down, and make sure the glasses are positioned correctly over your visual aid. To remove the AR HMD, tilt the visor up, turn the adjustment wheel, and take off the HMD).			
No.	System Level	Task	Didactic Methodology
5	Operating system	Generally (getting started)	Guided learning by doing
To familiarise yourself with the HoloLens 2 and its interaction paradigms for the first time, we recommend using the default application 'Tips'. This provides information on basic functions, and the necessary gestures are practised in self-learning phases. This includes interacting with objects (select, move, zoom in, zoom out, rotate, etc.) both at close range and from a distance. Subsequently, the most important functions should be intensified through further training outside the learning application.			
No.	System Level	Task	Didactic Methodology
6	Operating system	Starting menu	Video tutorial, guided learning
The functionality for opening and closing the start menu should be shown in a video tutorial and then practiced by the user. In addition, the purpose and capabilities of the start menu must be explained by an instructor in a video tutorial. This includes searching for and opening applications or settings, such as connecting to the Internet and other devices to mirror the view through the AR glasses. These functions should not only be explained in one video but should also be available as individual videos.			
No.	System Level	Task	Didactic Methodology
7	Operating system	Interaction paradigm	Video tutorial, guided learning by doing
The interaction paradigms, also known as 'near tap' and 'far tap', should be demonstrated in a video using the example of a 'bounding box'. The operator's fingers must be visible, and the distance required to use the far interaction must be specified. The change between near and far interaction should be carried out several times. The user should also repeat these procedures multiple times individually. It is important to note that the finger tap is best recognised when the thumb and pointer finger are rotated to the side and not along a horizontal axis. The interactions required to operate the functions must be made clearly visible to users but must also be tested in collaboration with an instructor or following a video tutorial. A distinction must always be made between near and far interaction. In the sense of 'best practice', negative examples can also be used here to illustrate how the fingers should be positioned in the best case. The detailed description of interactions by a trainer is particularly important when using an industrial AR application for the first time.			

Table A1. Cont.

No.	System Level	Task	Didactic Methodology
8	Operating system	Close application	Video tutorial, guided learning by doing
The closing of an application should be explained in a video tutorial by an instructor. A distinction should be made between closing and opening 2D windows using the cross in the top right-hand corner and closing a 3D application using the 'Home' button in the start menu. A dedicated 'guided learning by doing' phase is also recommended for this.			
No.	System Level	Task	Didactic Methodology
9	AR application	Generally (at first contact)	Individual training, guided learning by doing
On first contact with the application, users should be introduced to the basic functions by means of one-to-one training. The instructor first wears the glasses and mirrors their view to a screen that the learner observes. In this way, the application can be introduced with questions answered at the same time. The learner then wears the glasses and independently explores the outlined functions. The instructor is at their side and gives them tips on how to perform this better. If further functions are added to the AR application after the introduction, these can be learned via video tutorial.			
No.	System Level	Task	Didactic Methodology
10	AR application	Functions of an AR application	Video tutorial, guided learning by doing
The explanations of the various specific functions of an AR application must be available in separate, individually accessible video tutorials. This will clearly show learners which function is described in the tutorial. Care must be taken to ensure that the contents are formulated. This is because a tutorial must provide users with training content that is as precise as possible ([59]). In addition, excessively long tutorials can lead to cognitive overload. Furthermore, individually accessible videos offer the possibility of completing the training step by step and being able to try out key functions again at different entry points.			
No.	System Level	Task	Didactic Methodology
11	AR application	Functions of an AR application	Individual training, video tutorial
For better understanding, the training content for a specific application function should be reviewed again in fast-forward mode at the end. This form of repetition in conjunction with feedback enables users to review and self-reflect and increases their learning success ([60]).			
No.	System Level	Task	Didactic Methodology
12	AR application	Functions of an AR application	Individual training, video tutorial
Users must be given the exact name of a function and be informed about its possible uses. For example, use cases for the various measurement functions must be named. The symbols for the functions in the AR application must also be displayed appropriately. This makes it easier for users to memorise the function and associate the symbol with the name.			
No.	System Level	Task	Didactic Methodology
13	AR application	Functions of an AR application	Individual training, video tutorial
In case of trouble while using certain functions, learners must be shown in advance how they can solve or work around them. Users should also be aware of some knowhow for troubleshooting common issues, including freezing, connection loss, software updates in the background, light reflections, malfunction of hand recognition, and unplanned downtimes. For example, it should be pointed out that virtual content can be removed and reloaded.			
No.	System Level	Task	Didactic Methodology
14	AR application	Language	Individual training, video tutorial
It is important to use appropriate but simple language at a moderate pace. This should be adapted to the linguistic abilities of the majority of users in the industrial environment. If technical terms must be used, they should be clearly explained. Using the example of the 'bounding box measurement function', it can be demonstrated that the terms 'zoom in', 'zoom out', 'rotate', and 'move' are used instead of 'scale, rotate and manipulate'. In addition, the technical term 'handle bar' can be renamed 'edge' and 'corner', for example.			

Appendix B

Table A2. Accuracy of the applied measurement applications.

Reference Values	HRMA				PTPMA				BBMA			
	16.9	107.7	134.2	246.2	16.9	107.7	134.2	246.2	831.9	79.8	110.8	134.2
Systematic Errors (WOI)	0.29	0.49	0.09	0.16	0.04	0.51	0.61	1.04	5.65	1.20	0.68	0.98
Standard Errors (WOI)	0.16	0.51	0.50	1.37	0.10	0.26	0.26	0.41	0.57	1.26	0.70	0.75
Systematic Errors (OI)	0.59	2.16	2.99	1.64	0.24	0.51	0.50	0.75	6.81	3.9	−3.1	2.4
Standard Errors (OI)	0.92	1.44	2.69	3.14	0.64	1.99	1.11	1.92	4.16	4.96	5.25	7.46
Total Error (WOI)	0.45	0.99	0.57	1.523	0.14	0.77	0.87	1.45	6.22	2.47	1.38	1.74
Relative Total Error (WOI) [%]	2.67	0.93	0.44	0.62	0.82	0.72	0.65	0.59	0.75	3.09	1.25	1.29
Relative Mean Error (WOI) [%]			1.17				0.70				1.88	
Total Error (OI)	1.51	3.60	5.69	4.78	0.89	2.50	1.60	2.68	10.97	8.83	2.17	9.82
Relative Total Error (OI) [%]	8.92	3.34	4.24	1.94	5.24	2.32	1.20	1.08	1.32	11.06	1.96	7.31
Relative Mean Error (OI) [%]			4.61				2.23				6.78	

Footer: OI = operator involved; WOI = without operator involvement.

References

- Fink, K. Cognitive Assistance Systems for Manual Assembly throughout the German Manufacturing Industry. *J. Appl. Leadersh. Manag.* **2020**, *8*, 38–53.
- Semple, K.; Olminkhof, A.; Patel, M.; Tieto, V.; Wen, Q. MRTK2-Unity Developer Documentation—MRTK 2. Available online: <https://learn.microsoft.com/en-us/windows/mixed-reality/mrtk-unity/mrtk2/?view=mrtkunity-2022-05> (accessed on 12 March 2024).
- Schweikl, S.; Obermaier, R. Lessons from Three Decades of IT Productivity Research: Towards a Better Understanding of IT-Induced Productivity Effects. *Manag. Rev. Q.* **2020**, *70*, 461–507. [\[CrossRef\]](#)
- Bahaei, S.S.; Gallina, B.; Laumann, K.; Skogstad, M.R. Effect of Augmented Reality on Faults Leading to Human Failures in Socio-Technical Systems. In Proceedings of the 2019 4th International Conference on System Reliability and Safety (ICSRS), Rome, Italy, 20–22 November 2019; pp. 236–245.
- Simatupang, A.C. The Effect of Discipline, Motivation and Commitment to Employee Performance. *IOSR J. Bus. Manag.* **2018**, *20*, 31–37.
- Giovanni Mariani, M.; Curcuruto, M.; Gaetani, I. Training Opportunities, Technology Acceptance and Job Satisfaction: A Study of Italian Organizations. *J. Workplace Learn.* **2013**, *25*, 455–475. [\[CrossRef\]](#)
- Marshall, B.; Mills, R.; Olsen, D. The Role of End-User Training in Technology Acceptance. *RBIS* **2008**, *12*, 1–8. [\[CrossRef\]](#)
- Bloom, A.J. An Anxiety Management Approach to Computerphobia. *Train. Dev. J.* **1985**, *39*, 90–92.
- Berger, M.; Schäfer, R.; Schmidt, M.; Regal, C.; Gimpel, H. How to Prevent Technostress at the Digital Workplace: A Delphi Study. *J. Bus. Econ.* **2023**, *1*–63. [\[CrossRef\]](#)
- Bala, H.; Venkatesh, V. Adaptation to Information Technology: A Holistic Nomological Network from Implementation to Job Outcomes. *Manag. Sci.* **2016**, *62*, 156–179. [\[CrossRef\]](#)
- De Giorgio, A.; Monetti, F.M.; Maffei, A.; Romero, M.; Wang, L. Adopting Extended Reality? A Systematic Review of Manufacturing Training and Teaching Applications. *J. Manuf. Syst.* **2023**, *71*, 645–663. [\[CrossRef\]](#)
- Martins, B.R.; Jorge, J.A.; Zorzal, E.R. Towards Augmented Reality for Corporate Training. *Interact. Learn. Environ.* **2023**, *31*, 2305–2323. [\[CrossRef\]](#)
- Garcia Fracaro, S.; Glassey, J.; Bernaerts, K.; Wilk, M. Immersive Technologies for the Training of Operators in the Process Industry: A Systematic Literature Review. *Comput. Chem. Eng.* **2022**, *160*, 107691. [\[CrossRef\]](#)
- Dörner, R.; Horst, R. Conveying Firsthand Experience: The Circuit Parcours Technique for Efficient and Engaging Teaching in Courses about Virtual Reality and Augmented Reality. *Eurographics* **2021**, *7*. [\[CrossRef\]](#)
- Day, A.; Barber, L.K.; Tonet, J. Information Communication Technology and Employee Well-Being: Understanding the “iParadox Triad” at Work. In *The Cambridge Handbook of Technology and Employee Behavior*; Landers, R.N., Ed.; Cambridge University Press: Cambridge, UK, 2019; pp. 580–607, ISBN 978-1-108-64963-6.
- Palanque, P.; Cockburn, A.; Gutwin, C. A Classification of Faults Covering the Human-Computer Interaction Loop. In *Computer Safety, Reliability, and Security*; Casimiro, A., Ortmeier, F., Bitsch, F., Ferreira, P., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2020; Volume 12234, pp. 434–448, ISBN 978-3-030-54548-2.
- Quandt, M.; Freitag, M. A Systematic Review of User Acceptance in Industrial Augmented Reality. *Front. Educ.* **2021**, *6*, 700760. [\[CrossRef\]](#)
- Graser, S.; Böhm, S. A Systematic Literature Review on Technology Acceptance Research on Augmented Reality in the Field of Training and Education. In Proceedings of the CENTRIC 2022 The Fifteenth International Conference on Advances in Human oriented and Personalized Mechanisms, Technologies, and Services, Lisbon, Portugal, 16–20 October 2022; Volume 12.

19. Rosilius, M.; Spiertz, M.; Wirsing, B.; Geuen, M.; Bräutigam, V.; Ludwig, B. Impact of Industrial Noise on Speech Interaction Performance and User Acceptance When Using the MS HoloLens 2. *Multimodal Technol. Interact.* **2024**, *8*, 8. [CrossRef]
20. Zoetbrood, M. Do Training and Development Practices Help in Overcoming the “IT Productivity Paradox”? Master’s Thesis, Radboud University, Nijmegen, The Netherlands, 2021.
21. Kurilovas, E. Evaluation of Quality and Personalisation of VR/AR/MR Learning Systems. *Behav. Inf. Technol.* **2016**, *35*, 998–1007. [CrossRef]
22. Korpelainen, E.; Kira, M. Employees’ Choices in Learning How to Use Information and Communication Technology Systems at Work: Strategies and Approaches. *Int. J. Train. Dev.* **2010**, *14*, 32–53. [CrossRef]
23. Korunka, C.; Vitouch, O. Effects of the Implementation of Information Technology on Employees’ Strain and Job Satisfaction: A Context-Dependent Approach. *Work. Stress.* **1999**, *13*, 341–363. [CrossRef]
24. Wang, K.; Shu, Q.; Tu, Q. Technostress under Different Organizational Environments: An Empirical Investigation. *Comput. Hum. Behav.* **2008**, *24*, 3002–3013. [CrossRef]
25. Alqahtani, M.A.; Alamri, M.M.; Sayaf, A.M.; Al-Rahmi, W.M. Exploring Student Satisfaction and Acceptance of E-Learning Technologies in Saudi Higher Education. *Front. Psychol.* **2022**, *13*, 939336. [CrossRef] [PubMed]
26. Fischer, S.; Rosilius, M.; Schmitt, J.; Bräutigam, V. A Brief Review of Our Agile Teaching Formats in Entrepreneurship Education. *Sustainability* **2022**, *14*, 251. [CrossRef]
27. Makarova, I.; Pashkevich, A.; Shubenkova, K. Blended Learning Technologies in the Automotive Industry Specialists’ Training. In Proceedings of the 2018 32nd International Conference on Advanced Information Networking and Applications Workshops (WAINA), Krakow, Poland, 16–18 May 2018; pp. 319–324.
28. Rahmadani; Herman, T.; Dareng, S.Y.; Bakri, Z. Education for Industry Revolution 4.0: Using Flipped Classroom in Mathematics Learning as Alternative. *J. Phys. Conf. Ser.* **2020**, *1521*, 032038. [CrossRef]
29. Chang, S.-C.; Hwang, G.-J. Impacts of an Augmented Reality-Based Flipped Learning Guiding Approach on Students’ Scientific Project Performance and Perceptions. *Comput. Educ.* **2018**, *125*, 226–239. [CrossRef]
30. Bradley, M.D.; Barnard, Y.; Lloyd, A.D. Digital Inclusion: Is It Time to Start Taking an Exclusion Approach to Interface Design? In Proceedings of the Annual Conference of the Institute of Ergonomics and Human Factors on Contemporary Ergonomics and Human Factors 2010, Keele, UK, 13–15 April 2010; pp. 549–553.
31. Barnard, Y.; Bradley, M.D.; Hodgson, F.; Lloyd, A.D. Learning to Use New Technologies by Older Adults: Perceived Difficulties, Experimentation Behaviour and Usability. *Comput. Hum. Behav.* **2013**, *29*, 1715–1724. [CrossRef]
32. Dunleavy, M. Design Principles for Augmented Reality Learning. *Techtrends Tech. Trends* **2014**, *58*, 28–34. [CrossRef]
33. Mystakidis, S.; Christopoulos, A.; Pellas, N. A Systematic Mapping Review of Augmented Reality Applications to Support STEM Learning in Higher Education. *Educ. Inf. Technol.* **2022**, *27*, 1883–1927. [CrossRef]
34. ISO 9241-20:2021(En), Ergonomics of Human-System Interaction—Part 20: An Ergonomic Approach to Accessibility within the ISO 9241 Series. Available online: <https://www.iso.org/obp/ui/#iso:std:iso:9241:-20:ed-2:v1:en> (accessed on 18 March 2024).
35. Nielsen, J. *Usability Engineering*; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 1994; ISBN 978-0-08-052029-2.
36. Rogers, E.M.; Singhal, A.; Quinlan, M.M. Diffusion of Innovations. In *An Integrated Approach to Communication Theory and Research*; Routledge: Oxfordshire, UK, 2014; pp. 432–448.
37. Renaud, K.; Van Biljon, J. Predicting Technology Acceptance and Adoption by the Elderly: A Qualitative Study. In Proceedings of the 2008 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on IT Research in Developing Countries: Riding the Wave of Technology, Wilderness, South Africa, 6 October 2008; pp. 210–219.
38. Zhou, J.; Zhou, Y.; Wang, B.; Zang, J. Human–Cyber–Physical Systems (HCPs) in the Context of New-Generation Intelligent Manufacturing. *Engineering* **2019**, *5*, 624–636. [CrossRef]
39. Wang, B.; Li, X.; Freiheit, T.; Epureanu, B.I. Learning and Intelligence in Human-Cyber-Physical Systems: Framework and Perspective. In Proceedings of the 2020 Second International Conference on Transdisciplinary AI (TransAI), Irvine, CA, USA, 21–23 September 2020; pp. 142–145.
40. Hadorn, B.; Courant, M.; Hirsbrunner, B. *Towards Human-Centered Cyber-Physical Systems: A Modeling Approach*; University of Fribourg: Fribourg, Switzerland, 2016.
41. Dawadi, S.; Shrestha, S.; Giri, R.A. Mixed-Methods Research: A Discussion on Its Types, Challenges, and Criticisms. *JPSE* **2021**, *2*, 25–36. [CrossRef]
42. Rosilius, M.; Wilhelm, M.; Seitz, P.; Von Eitzen, I.; Wirsing, B.; Rabenstein, M.; Decker, S.; Brautigam, V. Equalization of the Visibility Loss between AR and Real Stimuli Sizes. In Proceedings of the 2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), Singapore, 17–21 October 2022; pp. 821–826.
43. Venkatesh, V.; Thong, J.Y.; Xu, X. Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. *MIS Q.* **2012**, *36*, 157–178. [CrossRef]
44. Hart, S.G. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2006**, *50*, 904–908. [CrossRef]
45. Dey, A.; Billinghurst, M.; Lindeman, R.W.; Swan, J.E. A Systematic Review of 10 Years of Augmented Reality Usability Studies: 2005 to 2014. *Front. Robot. AI* **2018**, *5*, 37. [CrossRef]

46. Cao, Y.; Qian, X.; Wang, T.; Lee, R.; Huo, K.; Ramani, K. An Exploratory Study of Augmented Reality Presence for Tutoring Machine Tasks. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 21 April 2020; pp. 1–13.
47. Moller, S.; Engelbrecht, K.-P.; Kuhnel, C.; Wechsung, I.; Weiss, B. A Taxonomy of Quality of Service and Quality of Experience of Multimodal Human-Machine Interaction. In Proceedings of the 2009 International Workshop on Quality of Multimedia Experience, San Diego, CA, USA, 29–31 July 2009; pp. 7–12.
48. Standage, D.; Wang, D.-H.; Heitz, R.P.; Simen, P. Toward a Unified View of the Speed-Accuracy Trade-Off. *Front. Neurosci.* **2015**, *9*, 144369. [\[CrossRef\]](#)
49. Venkatesh, V.; Morris, M.G.; Davis, G.B.; Davis, F.D. User Acceptance of Information Technology: Toward a Unified View. *MIS Q.* **2003**, *27*, 425–478. [\[CrossRef\]](#)
50. Alsuraykh, N.H.; Wilson, M.L.; Tennent, P.; Sharples, S. How Stress and Mental Workload Are Connected. In Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare, Trento, Italy, 20 May 2019; pp. 371–376.
51. Giorgi, A.; Ronca, V.; Vozzi, A.; Sciaraffa, N.; Di Florio, A.; Tamborra, L.; Simonetti, I.; Aricò, P.; Di Flumeri, G.; Rossi, D. Wearable Technologies for Mental Workload, Stress, and Emotional State Assessment during Working-like Tasks: A Comparison with Laboratory Technologies. *Sensors* **2021**, *21*, 2332. [\[CrossRef\]](#) [\[PubMed\]](#)
52. Park, E.; Cho, M.; Ki, C.-S. Correct Use of Repeated Measures Analysis of Variance. *Ann. Lab. Med.* **2009**, *29*, 1–9. [\[CrossRef\]](#)
53. Yang, J.; Rahardja, S.; Fränti, P. Outlier Detection: How to Threshold Outlier Scores? In Proceedings of the International Conference on Artificial Intelligence, Information Processing and Cloud Computing, Sanya, China, 19 December 2019; pp. 1–6.
54. Frey, B.B. *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation*; SAGE Publications, Inc.: Thousand Oaks, CA, USA, 2018; ISBN 978-1-5063-2615-3.
55. Blanca, M.J.; Alarcón, R.; Arnau, J. Non-Normal Data: Is ANOVA Still a Valid Option? *Psicothema* **2017**, *29*, 552–557. [\[CrossRef\]](#) [\[PubMed\]](#)
56. Schmider, E.; Ziegler, M.; Danay, E.; Beyer, L.; Bühner, M. Is It Really Robust?: Reinvestigating the Robustness of ANOVA against Violations of the Normal Distribution Assumption. *Methodology* **2010**, *6*, 147–151. [\[CrossRef\]](#)
57. Lee, S.; Lee, D.K. What Is the Proper Way to Apply the Multiple Comparison Test? *Korean J. Anesth.* **2018**, *71*, 353–360. [\[CrossRef\]](#)
58. Staffa, S.J.; Zurakowski, D. Strategies in Adjusting for Multiple Comparisons: A Primer for Pediatric Surgeons. *J. Pediatr. Surg.* **2020**, *55*, 1699–1705. [\[CrossRef\]](#)
59. Li, G.; Lu, T.; Yang, J.; Zhou, X.; Ding, X.; Gu, N. Intelligently Creating Contextual Tutorials for GUI Applications. In Proceedings of the 2015 IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom), Beijing, China, 10–14 August 2015; pp. 187–196.
60. Gerth, S.; Kruse, R. VR/AR-Technologien im Schulungseinsatz für Industrieanwendungen. In *Virtual Reality und Augmented Reality in der Digitalen Produktion*; Orsolits, H., Lackner, M., Eds.; Springer Fachmedien Wiesbaden: Wiesbaden, Germany, 2020; pp. 143–179, ISBN 978-3-658-29008-5.

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.