

Gear it up! Human Potential in Manual Assembly
-
**Subjective Evaluations and Performance Effects
of Augmented and Virtual Reality in Industrial Training**

Von der Philosophischen Fakultät der Rheinisch-Westfälischen Technischen Hochschule Aachen zur Erlangung des akademischen Grades einer Doktorin der Philosophie genehmigte Dissertation

vorgelegt von

Lea Marleen Daling, M.Sc.

Berichterinnen: Univ.-Prof. Dr. Sabine Schlittmeier

Apl.-Prof. Dr. Ingrid Isenhardt

Tag der mündlichen Prüfung: 14.05.2024

Diese Dissertation ist auf den Internetseiten der Universitätsbibliothek online verfügbar.

List of Scientific Publications

- Daling, L. M., Abdelrazeq, A., & Isenhardt, I. (2020). A comparison of augmented and virtual reality features in industrial trainings. In J.Y.C. Chen, G. Fragomeni (Eds.), *Lecture Notes in Computer Science: Vol. 12191. Virtual, Augmented and Mixed Reality. Industrial and Everyday Life Applications* (pp. 47–65). Springer. https://doi.org/10.1007/978-3-030-49698-2_4
- Daling, L., Abdelrazeq, A., Sauerborn, C., & Hees, F. (2020). A comparative study of augmented reality assistant tools in assembly. In T. Ahram, C. Falcão (Eds.), *Advances in intelligent systems and computing: Vol. 972. Advances in usability and user experience* (pp. 755–767). Springer. https://doi.org/10.1007/978-3-030-19135-1_74
- Daling, L. M., & Schlittmeier, S. J. (2024). Effects of augmented reality-, virtual reality-, and mixed reality-based training on objective performance measures and subjective evaluations in manual assembly tasks: a scoping review. *Human Factors*, 66(2), 589–626. <https://doi.org/10.1177/00187208221105135>
- Daling, L. M., Tenbrock, M., Isenhardt, I., & Schlittmeier, S. J. (2023). Assemble it like this! – Is AR-or VR-based training an effective alternative to video-based training in manual assembly?. *Applied Ergonomics*, 110, 104021. <https://doi.org/10.1016/j.apergo.2023.104021>

Danksagung

Die vorliegende Arbeit entstand im Rahmen meiner Tätigkeit als wissenschaftliche Mitarbeiterin und Forschungsgruppenleiterin am Lehrstuhl für Informations-, Qualitäts- und Sensorsysteme in der Produktion (WZL-IQS) sowie in enger Kooperation mit dem Lehr- und Forschungsgebiet Arbeits- und Ingenieurpsychologie des Institutes für Psychologie der RWTH Aachen University.

Ein besonderer Dank gilt Prof. Dr. Sabine Schlittmeier und Prof. Dr. Ingrid Isenhardt für die Betreuung dieser Arbeit. Liebe Sabine, vielen Dank, dass du dich ohne Zögern bereit erklärt hast, mich und diese Arbeit zu begleiten und mir im Team der Arbeits- und Ingenieurpsychologie eine „zweite Heimat“ zu geben. Danke für deinen kritischen Blick, deine Ausdauer und Geduld, deinen hohen Anspruch und dein Verständnis für die Herausforderungen in den verschiedensten Lebenslagen. Liebe Ingrid, auch bei dir möchte ich mich für die intensive Begleitung, die vielen erkenntnisreichen Diskussionen, dein Vertrauen, deine Unterstützung und dein unermüdliches Engagement in deiner Funktion als akademische Direktorin unseres Institutes bedanken.

Ein großer Dank gilt natürlich auch meinen Kolleginnen und Kollegen, die mich über die letzten Jahre in verschiedensten Konstellationen und Forschungsgruppen hinweg begleitet haben. Anas, bei dir möchte ich mich für dein Engagement und deine Begeisterung bedanken, mit der du unsere Gruppe geleitet und mich für das Thema AR/VR gewonnen hast. Danke an Kathrin und Paulina, die mich in allen Höhen und Tiefen begleitet, mir beharrlich Feedback gegeben und Mut gemacht haben. Danke an Samira, Johannes und Antonia: es ist eine wahre Freude, mit euch zu arbeiten und das Thema der Human Technology Interaction zu gestalten. Danke an Anders, Felix, Tseegi, Marisa, Chris und Kate – ohne euch wäre Vieles gar nicht erst möglich gewesen! Ein besonderer Dank gilt auch Johanna, nicht nur für die spannende und lehrreiche Erfahrung der gemeinsamen Forschungsgruppenleitung, sondern auch fürs Rücken freihalten, dein Verständnis und deine Weitsicht. Nina und Sarah, bei euch möchte ich mich für die vielen schönen Stunden in- und außerhalb des Institutes bedanken sowie bei Esther, Dennis und vielen weiteren KollegInnen, auch aus der IT und dem Office Bereich, für die immer konstruktive, bereichernde und wertvolle Zusammenarbeit. Ein großer Dank gilt auch meiner Mentorin Christina, die mich immer wieder bestärkt und mir geholfen hat, den Überblick über meine Handlungsmöglichkeiten zu gewinnen.

Zuletzt möchte ich mich bei meiner Familie und meinen Freunden bedanken. Danke an Mama, Harry, Papa, Ulrike, Harry, Julian, Doreen, Dominik und Astrid mit Franka, Joshi und Wanda für die bedingungslose Unterstützung und euren festen Glauben an mich. Danke an Willi, Ursel und Oma Maria für den Rückhalt und euren unermüdlichen Einsatz. Danke an Cathi, Max, Lisa und Michael sowie an Alissa, Line und Lena – ich bin so froh, dass ihr ein fester Bestandteil meines Lebens seid. Von Herzen danke ich außerdem meinen Mädels Anni, Annika, Caro, Lena, Anke, Jana und Charly – ihr wart und seid für mich der Fels in der Brandung. Mein größter Dank gilt meinem stärksten Kritiker, besten Freund und wichtigstem Menschen an meiner Seite. Danke, Stefan, dass du immer für mich da bist, mich begleitest, unterstützt und hinterfragst. Und Danke dafür, dass du ein großartiger Vater für unsere wundervolle Tochter Mila bist und ihr mich in dieser intensiven Zeit immer wieder an das Wesentliche im Leben erinnert.

Table of Contents

I	Abstract	2
II	Zusammenfassung	3
1.	Introduction	4
1.1	Defining Manual Assembly Tasks: Humans at the Heart of Production	6
1.1.1	Preparing Humans for Manual Assembly Tasks	6
1.1.2	A Human-centered Perspective on Training Success in Manual Assembly Tasks	8
1.2	AR and VR Technologies in Manual Assembly Training	9
1.3	Research Questions	12
2.	Synopsis of the Scientific Publications	13
2.1	Subjective Evaluation of AR- and VR-Based Training by Workers From the Industry (RQ1)	15
2.1.1	Initial Insights into Workers' Subjective Evaluations Using AR-based training in Manual Assembly (<i>Publication 1, Daling, Abdelrazeq, Sauerborn, & Hees, 2020</i>)	16
2.1.2	A Comparison of Workers' Subjective Evaluations Using AR- and VR-based Training in Manual Assembly (<i>Publication 2, Daling, Abdelrazeq, & Isenhardt, 2020</i>)	19
2.1.3	Answering RQ 1: Optimized Hardware Can Make AR- and VR-Based Training a Valuable Supplement to Training at the Workplace	21
2.2	Effects of AR- and VR-based Training Compared to Traditional Training in Terms of Subjective Evaluations and Performance Measures (RQ 2)	23
2.2.1	Findings From a Scoping Review on the Effects of AR- and VR-based Training on Users' Subjective Evaluations and Human Performance in Manual Assembly (<i>Publication 3, Daling & Schlittmeier, 2024</i>)	24
2.2.2	Answering RQ 2: AR- and VR-Based Training are at Least as Good as Traditional Training in Terms of Their Effects on Subjective Evaluations and Performance Measures	27
2.3	Comparing AR-Based, VR-Based and Traditional Training in Terms of Short- and Long-Term Training Success (RQ3)	28
2.3.1	Findings From an Experimental Study Comparing the Effects of AR-, VR- and Video-Based Training on Short- and Long-Term Training Success (<i>Publication 4, Daling et al., 2023</i>)	29
2.3.2	Answering RQ 3: Users Subjectively Perceive AR- and Video-Based Training as the Best, but All Three Training Formats Are Comparable in Terms of Their Impact on Short- and Long-Term Human Performance	31
3.	Discussion	32
3.1	The Potential of AR- and VR-Based Training for Improving Workers' Subjective Evaluations	33
3.1.1	Key Findings on Workers' Subjective Evaluations of AR- and VR-Based Training	33
3.1.2	Implications for Further Research on Workers' Subjective Evaluations of AR- and VR-Based Training	34
3.2	Effects of AR- and VR-Based Training on Human Performance in Assembly	36
3.2.1	Key Findings on the Effects of AR- and VR-Based Training on Human Performance in Assembly	37
3.2.2	Implications for Further Research on the Effects of AR- and VR-Based Training on Human Performance in Assembly	38
4.	Conclusion and Resulting Implications for Practice	39
5.	References	43
6.	Manuscripts	51

Abstract

Digital technologies such as augmented (AR) and virtual reality (VR) are increasingly being used in the context of manual assembly to prepare humans for their tasks in a safe, flexible and effective manner. However, there is a lack of scientific research into the effects of these new training methods on users and their performance. In this dissertation, I investigated the psychological effects of AR- and VR-based training on users' subjective evaluations and human performance in manual assembly in four scientific publications. The first two publications show that workers from the industry positively evaluated the use of AR and VR in training. Nevertheless, they rated the usability of the technologies and their impact on the perceived task load less favorably compared to traditional training.

Within the third publication, I conducted a literature review comparing the effects of AR- and VR-based and traditional training on subjective user evaluations and on human performance in manual assembly. The results indicate comparable or superior usability of AR- and VR-based training to traditional training methods, with a reduction in the perceived task load through AR-based training. As for human assembly performance, AR- and VR-based training showed comparable effects to traditional training. However, there was no consistent advantage of AR- and VR-based training over traditional training.

Within the fourth publication, I investigated and compared the short- and long-term effects of AR-, VR- and traditional video-based training by means of an experimental study. Results reveal comparable outcomes in short- and long-term assembly performance across training methods. Subjectively, users preferred AR-based and traditional video-based training over VR-based training.

The results of this dissertation indicate that a combination of AR-based training with real assembly components promises potential for successfully training manual assembly tasks, both in terms of users' subjective evaluations and in terms of the short and long-term effects on human performance. The thesis discusses implications for further research in this field and derives recommendations for the practical application of AR- and VR-based training in manual assembly.

Zusammenfassung

In der manuellen Montage werden Technologien wie Augmented (AR) und Virtual Reality (VR) zunehmend eingesetzt, um Menschen sicher, flexibel und effektiv auf ihre Aufgaben vorzubereiten. Jedoch fehlt es an wissenschaftlichen Auseinandersetzungen mit der Frage, welche Effekte diese neuen Trainingsmethoden auf die Nutzenden und ihre Leistungsfähigkeit haben. In dieser Dissertation untersuchte ich in vier wissenschaftlichen Publikationen die psychologischen Effekte AR- und VR-basierter Trainings auf die subjektive Wahrnehmung von NutzerInnen sowie auf deren Leistung in der manuellen Montage. Die ersten beiden Publikationen zeigen die positive Einstellung der NutzerInnen zu AR und VR im Trainingskontext. Dennoch wurden die Benutzerfreundlichkeit der Technologien und deren Auswirkungen auf die wahrgenommene Aufgabenbelastung im Vergleich zu traditionellem Training negativer bewertet.

In der dritten Publikation zeigt ein Literaturreview, dass die Benutzerfreundlichkeit von AR- und VR-basiertem Training vergleichbar oder besser bewertet wurde als traditionelles Training. AR-basiertes Training führte darüber hinaus zu einer Reduktion der wahrgenommenen Aufgabenbelastung. In Bezug auf die menschliche Montageleistung erzielten AR- und VR-basierte Trainings im Vergleich zu traditionellen Trainings überwiegend vergleichbare Effekte und keinen durchgängigen Vorteil.

In der vierten Publikation zeigt eine experimentelle Studie vergleichbare Effekte kurz- und langfristiger Montageleistung nach AR-, VR- und traditionellem Video-basierten Training. Die subjektiven Bewertungen des AR- und des Video-basierten Trainings waren jedoch signifikant besser als die des VR-basierten Trainings. Die Ergebnisse der Dissertation zeigen, dass insbesondere eine Kombination aus AR-basiertem Training mit realen Montagebauteilen sowohl in Bezug auf die subjektiven Bewertungen der NutzerInnen als auch in Bezug auf die kurz- und langfristigen Auswirkungen auf die menschliche Leistung Potenzial für das erfolgreiche Anlernen manueller Montageaufgaben verspricht. Die Arbeit diskutiert Implikationen für die weitere Forschung in diesem Bereich und leitet Empfehlungen für die praktische Anwendung von AR- und VR-basiertem Training in der manuellen Montage ab.

1. Introduction

Despite the increasing digitalization and automation of processes in production brought about by the fourth industrial revolution (Industry 4.0), humans remain irreplaceable in many segments of the industry. Because of their experience, their explicit and implicit knowledge and cognitive and motor skills, humans make the best contribution where machines reach their limits (Comberti et al., 2020). This holds particularly true in the field of manual assembly where the workers and their performance are crucial for enabling production that aligns with the growing trends towards individualization, customization, and smaller lot sizes (Merkel et al., 2018). At the same time, workers in manual assembly are facing new challenges in trying to develop new skills and prepare themselves for an increasing variety of product types in terms of the rising number of parts, their connections to each other and their variability over time (Bornewasser, 2020; Brolin et al., 2017, Funk et al., 2016).

To prepare workers for increasingly complicated and frequently changing assembly tasks, it is essential to develop suitable and effective training methods (Brolin, 2016; Doolani, Wessels, et al., 2020). In this respect, digital training formats based on augmented (AR) and virtual reality (VR) technologies are increasingly being used to prepare workers for their tasks. The expected advantages of AR- and VR-based training over traditional training methods, such as paper-based manuals, relate to their flexibility in terms of time and place of training, adaptability, and safety (Oestreich et al., 2020). However, from a human-centered perspective, AR- and VR-based training should, above all, have a positive impact on the users and improve their learning experiences (Riva et al., 2012). This not only means that workers in the industry should perform well, i.e., accurately and fast, in manual assembly after training (Blattgerste et al., 2017; Doolani, Wessels, et al., 2020; Gavish et al., 2013), but also that they subjectively perceive the technologies as helpful during the learning process and show a positive attitude towards them (Doolani, Owens, et al., 2020; Murcia-Lopez & Steed, 2018). Although considering the requirements and needs of the users is an important prerequisite for the successful implementation of the technologies (Langer, 2020), this has not been addressed

sufficiently in previous research on AR- and VR-based training in manual assembly (Neumann et al., 2021).

This dissertation addresses this issue and focuses on the psychological effects of AR- and VR-based assembly training. In this respect, there is a particular focus on the subjective evaluations of the users, e.g., how they experience the use of AR- and VR-based training and how they evaluate the perceived support in learning new cognitive and motor skills within manual assembly tasks. Additionally, to answer whether AR- or VR-based training is an effective alternative to traditional training, it is necessary to examine the effects of AR- and VR-based training on human performance in manual assembly. This dissertation thus provides a comprehensive, human-centered view on the potential of the method and points out the challenges of AR- and VR-based training in manual assembly tasks. Just as AR and VR technologies represent a key interface between humans and technology, research into the effectiveness of AR- or VR-based training is a bridge between research and practice. This dissertation considers this multi-faceted challenge by combining application-oriented findings on the needs of end users in industry with findings from scientific literature and experimental research.

The following chapter defines manual assembly tasks and discusses how the success of respective training methods can be measured. It is followed by an overview of AR- and VR-based training methods and possible applications of these technologies for training manual assembly tasks. In order to identify the extent to which workers can be supported in learning assembly tasks through AR- or VR-based training, three main research questions were developed and addressed through four scientific publications as part of the dissertation. The synopsis of the scientific publications includes a summary of the answers to each respective research question. Furthermore, implications for further research are discussed, leading to an overall conclusion with practical implications for the use of AR- and VR-based training in manual assembly at the end of the thesis.

1.1 Defining Manual Assembly Tasks: Humans at the Heart of Production

Until now, humans and their cognitive and physical abilities have been irreplaceable in manual assembly tasks (Neumann et al., 2021). Manual assembly tasks refer to manufacturing processes in which human workers physically put together parts and components to produce a finished product (Bornewasser, 2020; Lotter, 2006). They typically include activities such as joining, handling, inspecting and adjusting parts with a geometrically defined shape using tools and equipment (Lotter, 2006). Human workers make assembly more adaptable for the ongoing trends towards customization, high precision and smaller lot size production where automated systems are reaching their limits (Hinrichsen et al., 2016; Fletcher et al., 2020). Consequently, improvement and optimization within manual assembly needs to focus on humans, for example by improving their working conditions and increasing human performance (Brolin, 2016). One important factor in this respect is how workers are prepared and trained to develop the necessary performance skills for new and frequently changing assembly tasks (Holding, 1989; Kaplan et al., 2021).

1.1.1 Preparing Humans for Manual Assembly Tasks

The purpose of industrial training for manual assembly tasks is to enable workers to acquire the necessary knowledge, skills and techniques so they can successfully perform assembly tasks (Doolani, Wessels, et al., 2020; Gavish et al., 2011; Kagermann et al., 2013). Thus, training typically involves a combination of procedural skills, i.e., motor skills and habits through hands-on practical experience, as well as cognitive skills, i.e., knowledge about what, how and in which order the components must be assembled (Bannat et al., 2008; Henderson et al., 2011; Koziol & Budding, 2012).

Training is provided either to experienced assembly workers learning to assemble new elements or product types, or to novices who work as career changers, temporary or vacation assembly workers. The main objective of manual assembly training is that the workers trained can transfer what they had learned into practice after the training. In contrast to assistance systems that permanently support the workers in the assembly process (Hinrichsen et al.,

2016; Wolfartsberger et al., 2019), the term training in the context of this dissertation is used when the actual work process is separated from the learning process, i.e., there is a clear separation between training and the actual execution of the assembly task. In the best-case scenario, workers learn new tasks on-site exactly where the task is to be carried out. On-site training offers learning by performing the actual assembly tasks under the guidance of experienced assembly workers or trainers (Funk et al., 2016; Werrlich, Lorber, et al., 2018). This supports the workers in getting a feel for the materials, tools, and processes involved.

However, training at the workplace is not always possible because, for example, training within the work process requires an extreme number of human resources, i.e., providing training from experienced assembly workers and supervisors. This may slow down the production flow, for instance when the assembly tasks are scheduled into a timed production process, meaning that cycle times of the machines must be reduced during the training process (Sautter & Daling, 2021). This, in addition to the increasing complexity and repetitiveness in manufacturing, can put intense pressure on trainees and increase the risk of errors (Brolin et al., 2017, García-Acosta et al., 2021). In addition, there are certain assembly tasks in which training at the workplace can cause hazards for the workers as long as the individual steps are not perfectly mastered (Lilis et al., 1982).

For these reasons, different training methods and media are used to support trainers and enable workers to prepare themselves more flexibly, safely, and independently for new tasks (Doolani, Wessels et al., 2020). In the past and to some extent still today, traditional paper-based manuals and technical documentation or video tutorials were widely used to support the training process (Funk et al., 2016; Kaplan et al., 2021; Murcia-Lopez & Steed, 2018). In the context of the digitalization of production, innovative training formats such as AR and VR technologies have also increasingly found their way into manual assembly training (Guo, 2015; Koumaditis et al., 2019; Werrlich, Lorber, et al., 2018; Werrlich, Nguyen, & Notni, 2018).

The selection of a suitable training method is based on certain success criteria. The overall training success is mostly evaluated by measuring the speed, efficiency, and

transferability of training (Kaplan et al., 2021; Webel et al., 2011). However, since the European Commission formally called for more human-centricity in production in the 'Industry 5.0' (European Commission, Directorate-General for Research and Innovation, 2021), the question of which training method can best support workers in learning new assembly tasks and ensure the transfer of acquired knowledge and skills has become increasingly important. Thus, this dissertation focuses on the human-centered factors that can be used to evaluate training success as they can be seen as an important prerequisite for the evaluation of overall training success (Kulyk et al., 2017). These are outlined below.

1.1.2 A Human-centered Perspective on Training Success in Manual Assembly Tasks

From a human-centered perspective, the success of a training method covers two factors. On the one hand, the user's subjective evaluations of the training method can be decisive for whether the medium is accepted and will be used in the future (Langer, 2020; Venkatesh et al., 2016). These subjective evaluations provide important insights into workers' attitudes, perceptions, satisfaction and encountered difficulties while using a certain training method (deLone & McLean, 2002). In the context of this dissertation, the factors of subjectively perceived training success include the perceived *usability* of a training method (Brooke, 1996) and the perceived support in the training process, for example through the perceived *task load* (e.g., mental, physical and temporal demand, performance, effort and frustration) during training (Hart & Staveland, 1988). In the best-case scenario, usability should be as high as possible, while the perceived task load should be reduced by the training method.

On the other hand, the transfer of what has been learned to the actual task is one of the most important factors for training success (Kaplan et al., 2021). In the context of manual assembly, this training transfer can be measured by the performance that workers achieve after training in the actual assembly task, i.e., how accurately or error-free and how quickly they can complete the task (deMoura & Sadagic, 2019; Doolani, Owens, et al., 2020; Webel et al., 2013). These objective measures of human performance, namely *task accuracy* and

task completion times, make it possible to determine how well the training has prepared the workers (Carlson et al., 2015; DeMoura & Sadagic, 2019; Doolani, Owens et al., 2020; Webel et al., 2013; Werrlich, Lorber et al., 2018).

By defining training success of a training method in manual assembly tasks through positive subjective user evaluations and enhancing human performance, the following questions arise: to what extent can AR- and VR-based training fulfil these factors and are they an effective alternative to traditional methods? To this end, it is necessary to gain a deeper understanding of the definition of AR- and VR-based training. I address this in the following segment where I also explain some of the findings from the relevant literature to date about the effects of AR- and VR-based training on users' positive subjective evaluations and on enhancing human performance in assembly.

1.2 AR and VR Technologies in Manual Assembly Training

AR and VR describe technologies that create a link between real and virtual worlds (Skarbez et al., 2021). Different approaches are taken in defining and classifying the technologies, which have changed and evolved over the years (Cipresso et al., 2018; Lindemann & Noma, 2007; Mackay, 2000; Normand et al., 2012). Put simply, AR involves augmenting the real-physical world with virtual elements, while VR describes a computer-generated, three-dimensional (3D) environment in which users can interact with interactive and multisensory elements in real time (Langer, 2020). While the two technologies were considered separately from each other in the early stages of development (Langer, 2020; Milgram & Kishino, 1994), today there is much more focus on the overlap between the two (Skarbez et al., 2021). Thus, experts from the field started to use an overarching term to describe the functionalities of AR and VR, such as different forms and levels of user interaction, the extent of details modelling the real world (extent of world knowledge), or the plausibility and coherence of virtual behavior (immersion and fidelity), along a common continuum (Skarbez et al., 2021). Here, there are many partly contradictory discussions about a generic term, which is either called term XR (partly as an abbreviation for extended reality, partly as X for the

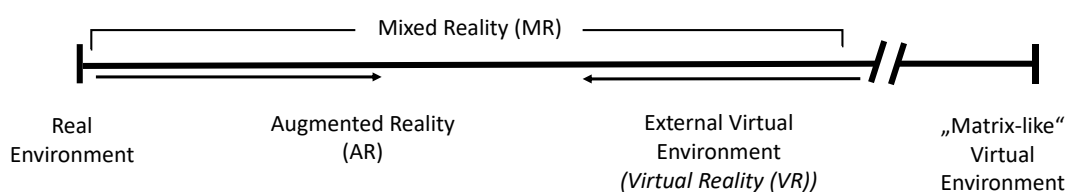
unknown variable of XReality; Rauschnabel et al., 2022) or as term MR (mixed reality) technologies (Skarbez et al., 2021). This dissertation follows the latter term of *MR technologies* and thus understands AR and VR as instances of a continuum between real and virtual environments (see Figure 1).

In this sense, AR-based training is referred to when a real-physical environment is used (e.g., a real assembly cell) in which the user's field of view is enriched in real time with virtual elements and objects, e.g., through see-through glasses or head mounted displays (HMD) or tablets (e.g., instructions, animations, tools or objects). Users can navigate and interact using voice commands, hand gestures or by clicking a remote control. VR-based training is referred to when a 3D modelled training environment is used while the real world is hidden. Users interact with virtual elements and objects (e.g., through closed or stereoscopic glasses) using controllers or mouse control. According to the abovementioned definition of the MR continuum, AR- and VR-based training are considered separate entities in this thesis, but both fall under the umbrella category of MR-based training. A specific definition of AR- and VR-based training is discussed in more depth in the third Publication of the thesis (Daling & Schlittmeier, 2024).

AR and VR technologies have been researched for decades (Boud et al., 1999; Reinhart & Patron, 2003; Wang et al., 2016) and they are expected to be able to meet the demands for a flexible, safe, and efficient training method (Doolani, Wessels, et al., 2020; Kaplan et al., 2021). However, when looking at the training success of AR- and VR-based training methods in the sense of positive subjective evaluations by users and enhancing human performance in assembly, it can be stated that promising, yet partly inconsistent findings on

Figure 1

Mixed Reality as an Umbrella Term for AR and VR technologies (based on Skarbez et al., 2021)



the effects of MR-based training have been published so far (Ganier et al., 2014; González-Franco et al., 2016; Webel et al., 2013). Although positive effects for users in terms of both positive subjective evaluations and increased human performance have been demonstrated in other MR-based training formats (e.g., physical training of balance skills (Kaplan et al., 2021; Prasertsakul et al., 2018; Rose et al., 2000)), no clear advantage of MR-based training over traditional training has yet been proven in the field of manual assembly tasks (Kaplan et al., 2021). Moreover, the fact that the interdisciplinary research community in the field of manual assembly uses different operationalizations to record subjective evaluations (e.g., using self-created items and scales instead of standardized questionnaires to assess the usability or experienced task load, for example) or performance measures (e.g., defining or assessing the severity of assembly errors or evaluating rework of assembly steps) presents an additional challenge (Daling & Schlittmeier, 2024). This weakens the validity of the results and makes the results difficult to compare. As a result, a consistent statement has hardly been possible to date. In addition, research on MR-based training is rarely conducted in the field and the controlled settings of laboratory studies are hardly transferable to realistic use cases in industry. It is currently not possible to state whether AR- and/or VR-based training in the area of manual assembly tasks can ensure training success with regard to workers' subjective evaluations and human performance or even offer an advantage over traditional training.

Consequently, the aim of this dissertation is to investigate whether using AR- or VR-based training can be successful in terms of providing users with the best possible support during training (concerning usability and reduced task load during the training process) and facilitating the transfer of what they had learned to the actual assembly task (by ensuring high task accuracy and short task completion times). By incorporating the perspective of users from the industry, analyzing and comparing results from the interdisciplinary research community and investigating the psychological effects by means of experimental studies, the presented research gaps will be tackled to bridge the currently prevailing gap between research and practice.

1.3 Research Questions

In this thesis, three main research questions are considered to find out whether AR- or VR-based training can ensure the training success in terms of users' subjective evaluations and human performance in assembly. First, as a crucial factor for the success of training, one main research question focuses on how workers in the industry perceive training using AR and VR and how they subjectively evaluate it (RQ 1). For this purpose, the workers' perception is evaluated within two realistic use cases from the industry, where AR- or VR-based training has already been implemented. From this practical consideration, the requirements for the implementation and transferability of controlled laboratory studies are derived.

The second important factor for the success of training was identified as measures of human performance in assembly to determine whether AR- and VR-based training can contribute to fast task completion and error-free assembly. In this respect, it is particularly important to consider the comparison of AR- and VR-based training methods with traditional training methods. Thus, the second research question extends the results on subjective measures to include findings on the effects of AR- and VR-based training on human performance in assembly compared to traditional training (RQ 2). For this purpose, current research work from the interdisciplinary research community is examined within a literature review and important findings on the effectiveness of AR- and VR-based training in manual assembly tasks are derived.

Based on the findings of the first two research questions, which primarily focus on the comparison between MR-based training and traditional training, the third research question further examines whether AR- or VR-based training is more successful in terms of short-, but also long-term training success in a direct comparison (RQ 3). In the context of an experimental study in a controlled setting, but with an industry-related assembly task, it becomes clear which insights for the use of AR- and VR-based training can be derived from science for practical use.

Taken as a whole, this thesis seeks to make a substantial contribution to the thorough investigation of AR- and VR-based training in manual assembly tasks by bringing

psychological research and industrial practice closer together. The results arising from the empirical investigations conducted to answer the research questions are briefly summarized here and translated into statements on each research question. Ultimately, directions for future research efforts and implications and recommendations for industry practitioners are derived from the results.

2. Synopsis of the Scientific Publications

Within the scope of the dissertation, four independent research papers were published to answer the presented research questions. Two of the publications were presented and published in the application-oriented community of engineering psychology and the other two were published in more research-oriented psychological journals. Table 1 shows the connection between scientific publications and the research questions. The first two publications (Daling, Abdelrazeq, & Isenhardt, 2020; Daling, Abdelrazeq, Sauerborn, & Hees, 2020) provide insight into the subjective evaluations of workers using AR and VR technologies in industrial assembly training (RQ 1). The third publication (Daling & Schlittmeier, 2024) extends the findings on the subjective evaluations of users with findings on the effects of AR- and VR-based training on human performance in assembly compared to traditional training by means of a literature review (RQ 2). Finally, the fourth publication (Daling et al., 2023) brings new insights into a direct comparison of AR- and VR-based training regarding short- and long-term training success in manual assembly tasks (RQ 3). Within the fourth publication, transferable findings for practice are derived based on the results of a controlled experimental study.

Table 1*Connection Between Scientific Publications and the Research Questions*

Number	Publication	Research Question	Method
1	Daling, Abdelrazeq, Sauerborn, & Hees, 2020	RQ 1: How is AR- and VR-based training for manual assembly tasks subjectively evaluated by workers from the industry?	Comparison between the subjective evaluations of (a) nine participants from the industry (field study, evaluation of the technology with posttest-only design) and (b) 15 participants from an academic sample (laboratory study, within-subject design) using AR-based training in manual assembly tasks.
2	Daling, Abdelrazeq, & Isenhardt, 2020		Comparison of the subjective evaluations of 16 participants from the industry using (a) AR-based and (b) VR-based training in a within-subject design.
3	Daling & Schlittmeier, 2024	RQ 2: To what extent can AR- and VR-based training be successful compared to traditional training in terms of their effects on users' subjective evaluations and on human performance in assembly?	Scoping review of 24 selected articles and analysis of their results on MR-based training in assembly tasks regarding users' subjective evaluations and objectively measured performance.
4	Daling et al. 2023	RQ 3: Comparing AR-based, VR-based and traditional training: Which training method is most successful in terms of its effects on users' subjective evaluations and on short- and long-term human performance in assembly?	Comparison of three training groups (n = 96) using either AR-, VR- or traditional training regarding users' subjective evaluations and objectively measured performance directly after training and after two weeks in a 3x2 repeated measures experimental design.

Note. The detailed list of publications is presented at the beginning of the dissertation.

2.1 Subjective Evaluation of AR- and VR-Based Training by Workers From the Industry (RQ1)

The subjective evaluation of a training process by the trainees is an important factor in assessing the overall success of training. Especially when new technologies are implemented, the users' assessment can have an impact on whether training is accepted and used in the future (Venkatesh et al., 2016). The overall aim of the first two publications (Daling, Abdelrazeq, & Isenhardt, 2020; Daling, Abdelrazeq, Sauerborn, & Hees, 2020) is to evaluate the subjective experience of assembly workers, i.e., their evaluation of usability (Brooke, 1996) and task load (Hart & Staveland, 1988) using AR and VR in assembly training.

The first publication focused on how workers from the industry evaluate AR-based training (Daling, Abdelrazeq, Sauerborn, & Hees, 2020). The findings from this study are important for determining whether workers would accept the technologies and perceive them as useful in their day-to-day work. Furthermore, important insights are gained about the execution of studies in the field which were taken into account in the second publication.

The second publication examines a further use case of using both AR- and VR-based training in the industry. The aim of this study was to increase the reliability of the findings on the subjective evaluations of workers in the industry from the first study. Moreover, the focus was on which specific MR-based functionalities within AR- and VR-based training were perceived as helpful by assembly workers (Daling, Abdelrazeq, & Isenhardt, 2020). The identified research gaps and improvements from the first field study were addressed when conducting the study.

The studies conducted within these publications were examined in different realistic use cases in German engineering companies in which AR and VR technologies were used for training manual assembly tasks. The investigation in the field provided important insights into the requirements and needs of the workers and their assessment of the potentials and challenges of AR and VR in manual assembly training, which were considered in the subsequent research of the topic as part of the thesis. Both publications and their key findings

are summarized below. Moreover, the contributions of the two publications on answering RQ 1 are discussed and open questions and research gaps identified are reflected upon.

2.1.1 Initial Insights into Workers' Subjective Evaluations Using AR-Based Training in Manual Assembly (Publication 1; Daling, Abdelrazeq, Sauerborn, & Hees, 2020)

The first publication (Daling, Abdelrazeq, Sauerborn, & Hees, 2020) provides insight into the subjective evaluation of workers using AR technology for training realistic assembly tasks in the industry. In a use case of a German industrial company from the field of drive and control technology, a mixed-methods approach of quantitative and qualitative methods was used to investigate the users' perspective on the potential of AR-based assembly training. In order to ensure both practical relevance and reliability of the results, the usability of an AR head-mounted display (AR HMD-based training) and its effects on user experience were evaluated within two different use cases of assembly tasks, one of which was conducted as a field and one as a laboratory study.

Within the field study, an AR-based training session was evaluated using a posttest-only design with nine production workers who were trained to perform manual assembly of a supporting bearing. Within the laboratory study, AR-based training was evaluated by 15 laypersons with an academic background who had no previous experience in the field of assembly. They were trained in gearbox assembly that was conducted in collaboration with a robot. To be able to make more differentiated statements about the use of AR HMD-based training in the laboratory study, we compared it to the training with a tablet-based AR and a 2D touchscreen in a within-subject design.

In both studies, the analysis of the subjective evaluations included the perceived usability (ease of use, enjoyment, media self-efficacy and usefulness) and the perceived task load (mental, physical and temporal demand, performance, effort and frustration) in relation to the training medium, i.e., the hardware of HoloLens 1 (Microsoft, 2021). Furthermore, it was also considered how the AR application, i.e., the software, was perceived by the users –

independently from which hardware was used. Due to the adaptation of the AR application to the respective assembly task, only the evaluation of AR HMD (hardware) was compared between the field and the laboratory study. The main results on the hardware evaluation are summarized in Table 2.

As for the usability of AR-based training, it can be highlighted that the participants from the field study showed significantly higher scores in the perceived ease of use of the HMD, but significantly lower scores in the perceived enjoyment than the participants from the laboratory

Table 2

Main Results From the Laboratory and the Field Study: Users' Subjective Evaluations of AR-Based Training

Evaluation Criteria	Results from the laboratory study	Results from the field study
Usability	<i>Ease of use:</i> AR HMD was perceived as the least easy to use.	<i>Ease of use:</i> Participants from the field study found it easier to use AR HMD than the participants from the laboratory study.
	<i>Perceived Enjoyment:</i> No differences between AR HMD, tablet-based AR and touchscreen.	<i>Perceived Enjoyment:</i> Participants from the field study had less fun using AR HMD.
	<i>Media Self-Efficacy:</i> No differences between AR HMD, tablet-based AR and touchscreen.	<i>Media Self-Efficacy:</i> No difference between the laboratory and the field study.
	<i>Perceived usefulness:</i> AR HMD and tablet-based AR were perceived as less useful than the touchscreen.	<i>Perceived Usefulness:</i> No difference between the laboratory and the field study.
Task Load	AR HMD reached the highest frustration, effort, temporal and mental demand. 2D touchscreen received the highest subjective performance scores.	No statistically significant differences between the field and the laboratory study.
Qualitative Comments	Positive evaluation of the innovative and intuitive character of the AR technology. Negative comments were mostly related to the lack of wearing comfort and limited vision using AR HMD.	Positive comments were related to the opportunity to learn new tasks effectively. Negative comments were related to the weight of hardware and incompatibility with visual aids.

Note. The results from the laboratory study include within-subject evaluations of AR HMD-based training, tablet-based AR and training using a 2D touchscreen. Results from the field study are reported in comparison to the results from the laboratory study on evaluating AR HMD. For further information on the results, see Daling, Abdelrazeq, Sauerborn, & Hees, (2020).

study. Within the laboratory study, the comparison of AR HMD with a tablet-based AR and a 2D touchscreen showed that AR HMD reached significantly lower usability values than the other two technologies. It was noticeable that participants from both the field and the laboratory study rated the technology as innovative and suitable for learning new assembly tasks. However, considerable shortcomings in the hardware, i.e., concerning wearing comfort and the incompatibility with visual aids, became apparent. As for the perceived task load, descriptive data indicated that the participants from the field study reported having to expend less effort to perform the task using HMD than the participants from the laboratory study. This difference, however, was not statistically significant.

Despite the heavy criticism of the AR HMD hardware, the evaluation of the AR application (software) revealed that most of the participants from the field study (77.78%) and from the laboratory study (86.67%) stated that they would prefer AR-based training over traditional manuals. This highlights its practical relevance and potential for the field of manual assembly tasks – especially if the hardware improves in the future.

The results provide initial insights into the subjective evaluation of users of AR-based training. However, it must be noted that the results of the first study were affected by the limitations of a small sample, which is a well-known problem when conducting studies in the field, as the time and cost involved in providing personnel is very high (Yin, 2009). The statements generated were therefore based on the subjective perception of a small number of individuals. In addition, individual usability scales and not a fully validated questionnaire were used to evaluate the usability in this study. In order to ensure transferability and comparability with other scientific study results, validated questionnaires using multi-item scales should be used (Diamantopoulos et al., 2012). Furthermore, no further control group could be used in the field. To counteract these issues, the research design and method of the second publication were adapted accordingly so as to minimize the effects of individual attitudes despite small samples in the field. Moreover, only the evaluation of AR-based training was considered in the first study. Thus, features of both AR- and VR-based training should be considered to better answer RQ 1.

The second study presented below incorporates the methodological and technical improvements by applying a within-subject design using both quantitative and qualitative measures in the evaluation of both AR- and VR-based training in another practical use case of MR-based training for manual assembly tasks.

2.1.2 A Comparison of Workers' Subjective Evaluations Using AR- and VR-Based Training in Manual Assembly (Publication 2; Daling, Abdelrazeq, & Isenhardt, 2020)

The second publication (Daling, Abdelrazeq, & Isenhardt, 2020) aimed to evaluate the use of both AR and VR technologies in manual assembly training in the industry. Considering the previous results, a within-subject design was used to investigate the user's perspective on AR- and VR-based training in an assembly task within a German industrial company in the field of pneumatic and electric drive technology. In contrast to the first study, a training scenario was used in which training is carried out away from the workplace to avoid slowing down production workflows and processes.

For the assembly of a pneumatic cylinder, training was developed with eleven steps that could be executed using either AR HMD (Microsoft HoloLens 1; Microsoft, 2021) or VR HMD (Acer Windows Mixed Reality headset; Acer, 2019). The technologies differed in the interaction possibilities (control by pointing and clicking using a remote control in the AR-based training; control by means of controllers in the VR-based training) for the users as well as in the technical specifications (i.e., the extent of details modelling the real world). It is important to note that the AR-based training in this case deviated from augmenting a real-world training environment with virtual instructions. Instead, a virtual replication of the complete training environment was projected into the user's field of view while standing in an empty room, thus enabling complete relocation of the training session without the need for additional equipment. In AR, the entire machine was projected into the real world, while in VR, a 360° view of the real environment was applied in which the virtually modelled machine was installed. The information presented and the training procedure were identical.

By using a within-subject design, the study aimed to counteract the problem of small samples and find out which of the two training formats was preferred by the users for learning the assembly task and why. Both quantitative (i.e., validated questionnaires on system usability (SUS, Brooke, 1996) and task load (NASA-TLX Hart & Staveland, 1988)) and qualitative methods (i.e., open questions) were used to assess the subjective evaluation of the workers ($N = 16$) from the industrial company.

The results of this study are summarized in Table 3. It became clear that the VR-based training was rated with a significantly higher system usability score than the AR-based training. The evaluation of the perceived task load during the training process revealed no differences between the two systems. Answers to the open questions allowed deeper insights into the participants' perception of the two training methods: most of the participants felt well prepared

Table 3

Main Results from a Field Study: Users' Subjective Evaluations Comparing AR- and VR-based Training

Evaluation Criteria	AR-based training	VR-based training
System Usability	SUS Score of 63.63 (usability below average)	Significantly higher SUS Score of 72.75 (good to acceptable usability)
Task Load	No differences between AR- or VR-based training regarding the perceived task load.	
Qualitative Comments	<p>Lack of orientation and haptics.</p> <p>The participants felt well prepared for the upcoming task.</p> <p>The method is perceived as suitable for initial familiarization with the process.</p>	<p>Very realistic due to bimanual interaction with the virtual objects.</p> <p>Caused dizziness.</p> <p>The majority of the participants felt well to very well prepared for the upcoming task.</p> <p>Advantage of VR over videos is questionable.</p>

Note. SUS Scores are ranging from 0 to 100 (Brooke, 1996), where 68 is considered as average usability score (Sauro, 2011). For further information on the results see Daling, Abdelrazeq, and Isenhardt (2020).

for the realistic assembly task, given that they would be able to repeat the training process several times in advance. The answers to open questions further indicated that the AR-based training in particular was criticized for the lack of integration of the users' hands and a poor orientation within the scenario. The VR-based training, in contrast, was criticized for causing dizziness and a lack of intuitiveness when grasping objects using the controllers. However, the use of controllers to grasp and move components using the left or right hand and the high level of realism of the environment were mentioned as helpful.

However, the participants emphasized that overall, both AR- and VR-based training would be more suitable for initial familiarization with the process and that an additional personal introduction into the real assembly would be favorable. As the object to be mounted could not be physically handled, both systems were perceived merely as a supplement to, but not a replacement for instructions on the real-physical machine. Interestingly, the question about preference for one of the training formats showed a very clear result: 14 out of 16 participants stated that they would select the VR-based training. The preference for VR-based training indicates that in this case, a direct interaction with the object to be mounted seems to be an extremely important feature in MR-based training. The AR environment implemented here, in which no real objects are used and the virtual objects cannot be moved directly but only via clicking and selecting, does not seem to be optimally suited for training for this assembly task. In the following segment, the contribution of the two publications presented to answering the first research question is discussed.

2.1.3 Answering RQ 1: Optimized Hardware Can Make AR- and VR-Based Training a Valuable Supplement to Training at the Workplace

The aim of the first two publications (Daling, Abdelrazeq, & Isenhardt, 2020; Daling, Abdelrazeq, Sauerborn, & Hees, 2020) was to assess how workers from the industry subjectively evaluate AR- and VR-based training of a manual assembly task. Overall, the subjective evaluations showed that the potential of MR-based training, especially in comparison to traditional paper-based manuals, was assessed positively. Nevertheless, the

hardware currently in use was rated as less user-friendly and difficult to use than other traditional methods, such as instructions via touchscreen. The results showed that considerable relevance of MR-based training was seen by the workers in the initial familiarization with a task. Like with any other training method, it was clear that MR should not replace actual training with a trainer in the real workplace. Thus, MR-based training should be seen as a supplement to, but not a replacement for training in the real workplace. Nevertheless, the advantages of MR-based training compared to other (digital) traditional methods, such as video-based training, remain unclear. The specific results of both publications on the workers' evaluations of AR- or VR-based training, methodological aspects of investigating in the field as well as points for further research are discussed below.

With regard to AR-based training, the results of the first publication show that AR potentially offers high practical relevance for users. However, besides the perceived potential of the AR software, the currently used hardware (i.e., Microsoft HoloLens 1 as AR HMD) was still heavily criticized. Moreover, while the quantitative analysis of subjective evaluations in the laboratory study revealed no advantage of AR-based training compared to traditional training, the findings of the qualitative analysis uncovered that the application of AR-based training for learning new assembly tasks is nevertheless considered useful by both novices and experienced workers from the industry.

The results of the second publication (Daling, Abdelrazeq, & Isenhardt, 2020) support the findings already obtained in the first publication (Daling, Abdelrazeq, Sauerborn, & Hees, 2020), indicating that both AR- and VR-based training are generally perceived as positive and helpful by workers from the industry. However, an initial comparison of the subjective evaluations of AR- and VR-based training showed that the workers' subjective evaluations of VR-based training were significantly more positive than those of AR-based training. This can be explained by the fact that the bimanual interaction possibilities using controllers in VR was perceived as particularly realistic and transferable to the actual task. At the same time, poorer evaluations of the AR-based training can be associated with the absence of a real object or components that can be interacted with (Daling, Abdelrazeq, & Isenhardt, 2020). The

availability of real objects should be taken into account when designing AR-based training in the future.

From a methodological perspective, it became apparent that the use of validated questionnaires, such as the SUS or the NASA TLX (Brooke, 1996; Hart & Staveland, 1988), would allow higher levels of comparability across several studies when assessing subjective evaluations and should therefore be preferred to the collection of individual scales. In addition, the first study provides important insights into the implementation and transferability of studies with laypersons in comparison to assembly experts from the industry. It appeared that both field and laboratory samples had similar perceptions of the technology, with the difference that experienced workers found it easier to deal with the new technology, but not as exciting as laypersons. These results of the first publication point out that researching new training technologies on non-expert samples is feasible, but there is a risk that the novelty effect of a technology may exaggerate the results (Wells et al., 2010). In order to test the applicability of a technology in day-to-day work, end users from the field should be involved.

Overall, the results from the first two publications support the findings of other pieces of research, stating that the evaluation of MR-based training has repeatedly produced promising, but also diverse and sometimes inconsistent results (Kaplan et al., 2021). The latter are largely attributed to variations in hardware, assembly tasks and methodological approaches employed to investigate the effects of MR-based training. In order to be able to compare the evaluation of AR- and VR-based training with traditional training and also to assess the extent to which they can ensure assembly performance, a larger database is required. This is addressed in the third publication.

2.2 Effects of AR- and VR-Based Training Compared to Traditional Training in Terms of Subjective Evaluations and Performance Measures (RQ 2)

The generally positive, yet critical evaluation of MR-based training by workers in the industry provided an important basis for in-depth research into its effects on training success.

The third publication aimed to expand the knowledge on workers' subjective evaluations by investigating the extent to which AR- and VR-based training are successful in terms of their effects on training transfer, i.e., objective assembly performance criteria. At the same time, the aim was to look at how AR- and VR-based training perform in comparison to traditional training. In order to obtain meaningful results, a large database should be used. Optimally, this is achieved by analyzing previous research in this area. By conducting a literature-based scoping review, the third publication makes an important contribution to identifying scientific robust findings on users' subjective evaluations of AR- and VR-based training compared to traditional training and their effects on human performance in manual assembly.

2.2.1 Findings From a Scoping Review on the Effects of AR- and VR-Based Training on Users' Subjective Evaluations and Human Performance in Manual Assembly (Publication 3; Daling & Schlittmeier, 2024)

The scientific community in the field of innovative training methods for manual assembly tasks is characterized by a high degree of interdisciplinarity and different research approaches. Within the third publication (Daling & Schlittmeier, 2024), a scoping review was conducted that aimed to transform the different results published within the diverse research landscape into coherent findings regarding the effects of AR- and VR-based training on users' subjective evaluations and human performance and to compare the two to traditional training methods. For this purpose, 24 studies were identified and analyzed, investigating five review questions (RevQ) about (1) the needs and expected benefits of using MR in assembly training, (2) the MR technologies and features being used, (3) the relevant outcome measures, (4) the effects and outcomes of MR-based training in comparison with traditional training as well as (5) the relevant research gaps. The key findings for each RevQ (1-5) are summarized below.

Regarding the first review question (RevQ 1) it became clear that the initial need for well and frequently trained employees results from the developments around Industry 4.0, namely the increasing influence of multi-variant and individualized products in manual assembly. In this context, MR technologies are expected to be able to provide safe and low-

cost training in a flexible, customizable environment. Within the framework of the content analysis, these needs and expectations could be combined into three dimensions, all of which are related to the goal of overall increased productivity. The first dimension referred to the direct influence MR potentially has on the employees' perception of performance during training, i.e., reducing mental workload, enhancing enjoyment as well as preventing misunderstanding and errors. A more objective perspective is provided in the context of the second dimension, which referred to the success of the training outcome itself by capturing objective performance measures. The third dimension referred to the saving of resources through the use of MR (i.e., reduction of production downtime, saving resources and personnel costs, reduction of work, shortening of ramp-up times, or creating interfaces to existing data). It can be concluded that training must meet all three needs in order to be considered successful. However, there are currently barely any studies that consider all these aspects together and holistically.

In terms of the hardware and features used in AR and VR or MR-based training (RevQ 2), it was noticeable that the designation and classification of technologies, even when using the same hardware, was done differently by some authors. A new classification was therefore proposed for the articles under consideration. Based on the consideration of the software and hardware used, as well as the functionalities related to their extent of world knowledge, immersion and fidelity, three categories could be developed for AR-based training (AR projectors, AR handhelds, and AR HMD) and two categories for VR-based training (screen-based VR and VR HMD).

The effects of AR- and VR-based training (RevQ 4) were categorized along with the identified relevant outcome measures (RevQ 3). These measures were divided into users' subjective evaluations and objective performance measures in assembly. The main results are summarized in Table 4. Subjective evaluations were assessed with questionnaires that capture, for example, the perceived system usability and task load during training. Results indicated that AR-based training was consistently rated with high usability scores and resulted in a perceived reduction in task load. VR was found to be rated with higher usability scores

Table 4

Main Results From the Literature: Users' Subjective Evaluations of AR- and VR-Based Training and the Effects on Human Performance in Assembly Compared to Traditional Training

Evaluation Criteria	AR-based training	VR-based training
System Usability	Usability was as good as or better than the usability of different traditional training methods.	Usability was higher than paper-based training, but not as good as training with physical objects.
Task Load	Same or lower task load compared to traditional training.	Was barely assessed, individual results showed no difference compared to paper-based manuals.
Objective Performance Measures	As good as or better effects on task completion times and task accuracy than traditional training.	Inconsistent effects on task completion times and task accuracy compared to traditional training.

Note. AR-based training includes AR projectors, AR handhelds and AR HMD. VR-based training includes screen-based VR and VR HMD. For further information on the results, see Daling & Schlittmeier, 2024.

than paper-based training, for instance, but was perceived as more difficult to use than training with real-physical objects. Objective performance measures were mostly assessed by measuring the time of completing a task (TCT) as well as counting the number of errors (task accuracy, TA) during the final assembly. The results of the review showed that all AR-based training formats overall showed systematic advantages over traditional paper-based training concerning both objective performance measures. However, screen-based VR or VR-HMD-based training formats were mostly, but not consistently, as good as paper-based training. Compared to other more digital training formats such as video-based training, AR- and VR-based training did not show consistent advantages. Interestingly, VR-based training proved to be better with easier assembly tasks, i.e., with less steps and less components.

At this point it should be emphasized that the statements about the training success only referred to the assembly performance immediately after training. The few findings on long-term effects of the trainings refer only to VR-based trainings, but show promising initial results,

especially when additional cues (tactile, visual or audio cues) or storytelling approaches were used. Examining whether and when AR- or VR-based training is more beneficial for manual assembly training as well as investigating their long-term effects in realistic use cases were mentioned as important aspects for further research (RQ 5). Below is discussed the contribution of the presented results to answering the second research question.

2.2.2 Answering RQ 2: AR- and VR-Based Training Are at Least as Good as Traditional Training in Terms of Their Effects on Subjective Evaluations and Performance Measures

Considering the results obtained in the third publication (Daling & Schlittmeier, 2024), it transpired that MR-based training is for the most part able to ensure training success in terms of both subjective evaluations and human performance in manual assembly. AR-based training in particular proved to have a positive effect on users' subjective evaluations and improved objective performance measures compared to traditional training. VR-based training was evaluated with equal or higher usability scores and mostly, but not consistently, led to comparable assembly performance as traditional training. These results suggest that AR-based training has potential to become a reliable method for ensuring training success.

However, the statements developed here were all subject to the influence of different measuring instruments and procedures used in the articles. Subjective evaluations played a subordinate role in the studies examined and in some cases were not surveyed at all. Furthermore, the transferability of the study results through practical experiments involving actual users or their characteristics and adapting the task difficulty to realistic assembly tasks has not been adequately addressed so far. In future studies, the assembly activities to be performed should ideally include all activities of manual assembly (joining, handling, inspecting and adjusting). The results of the review also show that a differentiation of traditional training is necessary. While comparable or better effects of AR- and VR-based training were shown in comparison to paper-based training, there was no clear advantage in comparison to digital, e.g. video-based formats. In addition, it became clear that the long-term effects of AR- and VR-

based training were important for training success, although little research has been conducted into this to date. It also became evident that there is still too little knowledge on whether AR- or VR-based training is more helpful for the workers training to perform new assembly tasks. In order to select the best training method for the workers that optimally prepares them for manual assembly tasks, a direct comparison between AR- and VR-based training is necessary. These aspects are addressed within the fourth publication.

2.3 Comparing AR-Based, VR-Based and Traditional Training in Terms of Short- and Long-Term Training Success (RQ3)

Although research has already provided promising results about the use of AR- and VR-based training for manual assembly tasks, important aspects were identified that need to be considered to answer the overarching question of this dissertation. Four of these aspects were discussed in the course of the fourth publication: (1) When evaluating AR- and VR-based training, both subjective user evaluations and objective measures of human performance were collected. (2) The assembly task to be taught was designed as realistically as possible to correspond to a real assembly task in terms of its level of difficulty. When recruiting the sample, attention was paid to reaching a broader spectrum than that of young academics in order to reflect the diversity of workers in assembly. (3) While both AR- and VR-based systems have each been only compared to traditional training methods before, a direct comparison of AR- with VR-based training was carried out. Moreover, a digital format, i.e. video-based training, was chosen for the comparison group of the traditional training (4). The long-term effects of AR- and VR-based training on assembly performance were investigated. In the following segment, the research design and main findings from the fourth publication are described.

2.3.1 Findings From an Experimental Study Comparing the Effects of AR-, VR- and Video-Based Training on Short- and Long-Term Training Success (Publication 4; Daling et al., 2023)

The fourth publication (Daling et al., 2023) compared short- and long-term effects of AR-, VR- and traditional training in terms of subjective evaluations and objectively measured human performance. For all three training methods, a comparable training environment was developed in which the assembly of a LEGO MINDSTORMS EV3 robot with several different components was explained. The assembly of the LEGO MINDSTORMS EV3 robot comprised ten higher-level steps and included all aspects of realistic assembly (joining, handling, inspecting, adjusting). The study was conducted in a controlled laboratory setting with high industrial reference by using a real assembly cell.

The training scenarios were implemented for training three groups using either AR HMD-, VR HMD- or video-based training representing a more advanced method of traditional training. In the AR-based training, real components were used and holographic instructions as well as 3D animations were applied, which were projected into the participants' field of view using the HoloLens 1 (Microsoft, 2021). Navigation was done via voice commands and/or the use of a remote control. The video-based training took place in the same real training environment and the participants were guided using videos on a tablet. Navigation was done by touching the display. In the VR-based training, the entire training environment was simulated (high level of world knowledge) using the Oculus Rift (Meta, 2024). Virtual components were interacted with controllers and 3D animations were shown to navigate the steps.

In a 3x2 repeated measures experimental design, a total of 96 participants between 18 and 73 years of age successfully completed the training with one of the technologies and were asked to perform the assembly task from memory both immediately afterwards and two weeks later. The following indicators for training success were used for comparison: the perceived system usability and perceived task load (subjective evaluations) as well as assembly errors

and task completion time (TCT) of the assembly task (objective performance). The main results are summarized in Table 5.

The results of the subjective evaluations showed that the VR-based training resulted in a significantly higher perceived task load than the AR- and video-based training. As for usability, it transpired that the participants rated the VR-based training with significantly lower usability scores than the participants using AR- and video-based training.

In terms of objective measures of human performance in assembly, the results showed no difference between the three groups neither in the short-term nor in the long-term comparison. This leads to the conclusion that both AR- and VR-based training are as effective as traditional video-based training. However, an exploratory analysis showed that user variables, such as prior experience using the technology and age, did have a significant impact on these results. When participants had prior experience, for example, they finished the

Table 5

Main Results From an Experimental Study: Users' Subjective Evaluations of AR-, VR- and Video-Based Training and the Short- and Long-Term Effects on Human Performance in Assembly

Evaluation Criteria	Video-based Training	AR-based training	VR-based training
System Usability	Not different from each other, but significantly higher than VR.		Significantly lower than AR and video.
Task Load	Not different from each other, but significantly lower than VR.		Significantly higher than AR and video.
Short-term Effects on Objective Performance Measures	Not different from each other with regard to both TCT and TA.		
Long-term Effects on Objective Performance Measures	Not different from each other with regard to both TCT and TA.		

Note. For further information on the results, see Daling et al., 2023.

assembly faster than the participants with less prior experience. This highlights the importance of considering prior experience as a relevant human factor affecting individual performance (Abubakar and Wang, 2019; Schwarz et al., 2020). Higher age contributed to longer assembly times in the short-term and more errors in the long-term, which was discussed as a known effect of age-related declines in working memory (Abubakar & Wang, 2019; Morrell & Park, 1993). Interestingly, after adjusting the data for the age of the participants, the AR group's performance was significantly faster in the short-term and the group showed significantly fewer errors in the long-term than the VR group. This particularly highlights the suitability of AR technologies in terms of training for manual assembly tasks and ensuring training success in the long-term. However, compared to conventional training, VR-based training also led to comparable results.

In summary, this means that for training manual assembly tasks that include joining, handling, inspecting and adjusting activities, AR-based training is preferable to VR-based training, while both training formats are equivalent, yet not more effective alternatives to traditional video-based training.

2.3.2 Answering RQ 3: Users Subjectively Perceive AR- and Video-Based Training as the Best, but All Three Training Formats Are Comparable in Terms of Their Impact on Short- and Long-Term Human Performance

Within the scope of the fourth publication (Daling et al., 2023), the previously identified research gaps were addressed. Thus, the aim of the experimental study was to investigate short- and long-term effects of AR-, VR- and traditional video-based training on users' subjective evaluations and objective performance measures in a realistic assembly task using a diverse sample. The findings revealed that AR-, like video-based training, was particularly suited to support users in learning the procedural process of assembling real-physical components. When controlling for the age, the VR-based training performed less well than the AR-based training, which was also reflected in the subjective evaluations. Nevertheless, it can be stated that compared to traditional training, both training methods were able to contribute

to comparable training success in both the short and the long term. This leads to the conclusion that depending on the application and the context of the assembly task (e.g., location, time or user-related requirements and characteristics), both technologies can be effective alternatives to traditional training methods.

3. Discussion

The aim of this dissertation was to investigate whether AR- or VR-based training can ensure the training success in terms of positive subjective evaluations by users and human performance in assembly. To answer this overarching question, three research questions were examined within four peer-reviewed publications. As part of the first research question, two studies conducted in the field provided new insights into the subjective evaluations, namely the perceived usability and the perceived task load, of MR-based training by workers from the industry. This showed that MR-based training was perceived as a promising supplement to workplace training, but the benefits concerning usability and task load compared to traditional formats would have to be proven before workers in the industry would consider it useful.

In the context of the second research question, a scoping review extended the findings on users' subjective evaluations by examining the effects of MR-based training on objective measures of human performance in manual assembly, i.e., task completion times and task accuracy, and comparing them with those of traditional training. Concerning users' subjective evaluations, the review showed that the usability of AR- and VR-based training in particular was comparable to or better than traditional paper-based training and that the task load was reduced by AR-based training. In the context of objective measures, it was shown that AR- and VR-based training mostly, but not consistently, achieved the same effects as traditional paper-based training. Compared to more digital formats of traditional training, however, the results were less consistent.

In the context of the third research question, it was investigated based on an experimental study examining which training method (AR-based, VR-based or traditional training) is most suitable in direct comparison for achieving positive user effects and improving

assembly performance not only in the short-, but also in the long-term. Here, especially with regard to subjective evaluations, AR- and video-based training showed better results than VR. In terms of the effects on short- and long-term assembly performance, however, all three training methods showed comparable results.

In the following segment, the specific research results of this dissertation and possible implications for further research are discussed. Subsequently, a conclusion of the findings is presented and practical implications and recommendations for the use of AR- and VR-based training are derived.

3.1 The Potential of AR- and VR-Based Training for Improving Workers' Subjective Evaluations

The positive subjective evaluation should be a main objective in the design of a technology-based training. In the course of this dissertation, subjective evaluations were captured by measuring the usability of the technology and the perceived task load while using a training technology. The main findings on the effects of both AR- and VR-based training on users' subjective evaluations are presented below. Consequently, the significance and limitations of the results are discussed and important implications for follow-up research are derived.

3.1.1 Key Findings on Workers' Subjective Evaluations of AR- and VR-Based Training

Based on the findings of the four publications within this dissertation, the following results on users' subjective evaluations of AR-based training can be summarized. Compared to traditional training methods, several formats of AR-based training (i.e., using HMD, tablet or projectors) showed higher usability and lower task load when compared to paper-based training (Daling & Schlittmeier, 2024), while no differences were observed when compared to 2D screen-based (Daling, Abdelrazeq, Sauerborn, & Hees, 2020) or video-based training (Daling et al., 2023).

VR-based training (i.e., screen-based VR and VR HMD) was perceived with comparable (Daling & Schlittmeier, 2024; Daling et al., 2023) or higher (Daling & Schlittmeier, 2024) usability than paper- and video-based training, but lower usability compared to training with real-physical objects (Daling & Schlittmeier, 2024) and AR-based training using real-physical assembly objects (Daling et al., 2023).

In a direct comparison between AR-, VR- and video-based training, the use of VR was associated with a significantly higher task load of the users. However, when AR was used without real-physical objects (Daling, Abdelrazeq, & Isenhardt, 2020), the interaction possibilities of VR (e.g., two-handed handling of the virtual objects by controllers) were preferred by the users. Experienced workers in particular noted that the bimanual interaction with real-physical or virtual components in combination with virtual instructions or animations is helpful for learning assembly tasks (Daling, Abdelrazeq, & Isenhardt, 2020).

From the results of this dissertation, it can be concluded that both AR- and VR-based methods are an adequate alternative to traditional paper- or video-based formats to ensure training success in terms of positive subjective evaluations. However, it transpired that the use of AR in combination with real-physical objects was preferable to VR-based training.

3.1.2 Implications for Further Research on Workers' Subjective Evaluations of AR- and VR-Based Training

Based on the findings obtained within this dissertation on workers' subjective evaluations of AR- and VR-based training, important starting points for further research can be derived. These are discussed in the following. First of all, it should be noted that the results obtained on the subjective evaluations of users were highly dependent on the technical and didactical design of the training, which in turn might have had an impact on the results (Blattgerste et al., 2021; Zender et al., 2018). For example, the training methods in the fourth publication (Daling et al., 2023) were all designed very similarly to enable comparability. This in turn could also have influenced the fact that the training sessions achieved comparable effects. Exploiting the full potential of the technologies (e.g., through the integration of direct

haptic feedback in VR) could lead to a difference in their effects on subjective evaluations and possibly also on human assembly performance. Further research should therefore verify these results by investigating the effects of different functionalities and specifics of MR-based training. For better comparability of different study results, it is essential to ensure transparency about the materials used and the technical specifications. This also includes a more precise differentiation of the comparison groups of traditional training, i.e. information on the design and content of paper-based or more digital formats such as video-based training. In the case of this dissertation, all materials used in the fourth publication were made available to enable the replication of the study (Daling et al., 2023)

In addition, further research should investigate the influence of the novelty effect of the technologies, which may also have had an influence on the results (Wells et al., 2010). It is known from other areas of research on VR environments that the unfamiliarity of the virtual environment can contribute to a high cognitive load and thus to lower user satisfaction (Miguel-Alonso et al., 2023). The extent to which the poorer subjective evaluation of VR decreases after more frequent use of the technology should therefore be investigated. Moreover, the long-term effects of regular use of the technologies in day-to-day work on users should be examined in this context.

As a further aspect, the combination of quantitative and qualitative data will be essential for a successful interaction between research and practice in future research work. Research with small samples and less controlled settings in the field, as it was the case in the first two publications, is always subject to a high degree of subjectivity on the part of the participants. Even if this limits the transferability of the results, it became clear how important it is to go beyond the recognized measurement methods to obtain the opinions and impressions of users as well as to survey real end users who were less impressed by the novelty of the technology (Daling, Abdelrazeq, Sauerborn, & Hees, 2020). Thus, future research should include, for example, open-ended questions or thinking-aloud methods in addition to the measurement instruments selected here for collecting quantitative data on subjective evaluations (system

usability and task load). Moreover, depending on the application, other factors such as flow or the feeling of presence might be a relevant extension of the criteria (Carlson et al., 2015).

Finally, the extent to which the users' subjective evaluations of MR-based training methods have correlation with or influence on assembly performance is an important follow-up question that should be focused on in further research. In the context of this dissertation, the subjective evaluations were considered only separately from the effects of MR-based training on objective measures of human performance in assembly. Moreover, the consideration of different user-related variables, such as prior knowledge, expert levels and other user factors (age and also other inclusive aspects, such as limitations in vision or hearing that should also be taken into account), should play a key role in the research of AR- and VR-based training in the future.

3.2 Effects of AR- and VR-Based Training on Human Performance in Assembly

In addition to their impact on users' subjective evaluations, assembly training methods should of course also ensure the success of the training by effectively transferring what has been learned to the specific task. Within the context of this dissertation, the transfer of a training session was captured by measuring objective performance, namely the time of completing a task (TCT) as well as counting the number of errors (task accuracy, TA) during the actual assembly task (Daling & Schlittmeier, 2024). The findings on the effects of MR-based training on human performance in manual assembly were obtained based on a scoping review of 24 articles (Daling & Schlittmeier, 2024) and an experimental controlled study (Daling et al., 2023). The main findings on the effects of AR- and VR-based training on short- and long-term performance in manual assembly are presented below. Subsequently, their implications are discussed with regard to possible limitations and connection points for further research.

3.2.1 Key Findings on the Effects of AR- and VR-Based Training on Human Performance in Assembly

The results of the conducted studies revealed that AR-based training formats led to consistently better assembly performance regarding TCT and TA than paper-based training (Daling & Schlittmeier, 2024) and to comparable results as video-based training in the short-term (Daling & Schlittmeier, 2024; Daling et al., 2023) and long-term (Daling et al., 2023). Screen-based VR or VR-HMD-based training formats were mostly, but not consistently, as good as traditional training formats (Daling & Schlittmeier, 2024; Daling et al., 2023). Interestingly, VR-based training proved to be better than paper-based training especially in cases of easier assembly tasks (Daling & Schlittmeier, 2024).

Concerning the long-term training success, there were various indications that AR-based training contributed to at least equivalent or better results than traditional training methods (Daling & Schlittmeier, 2024; Daling et al., 2023). VR-based training methods combined with storytelling approaches or additional cues (haptics, audio or visual cues) were even able to compensate for initially worse effects in the long-term (Daling & Schlittmeier, 2024). In a direct comparison between AR- and VR-based training, it was found that AR-based training led to faster assembly times in the short-term and better accuracy in the long-term when the data was adjusted for the age of participants.

The results of this dissertation suggest that both AR- and VR-based training were shown to be adequate alternatives to traditional paper-based training for ensuring positive subjective evaluations and assembly performance in manual assembly tasks. More specifically, it can currently be stated that AR-based methods in particular achieve at least comparable effects as video-based training, and even better effects than paper-based methods. Compared to AR-based training that is combined with real-physical assembly objects, VR-based training was shown to be less effective, while still providing as good results as traditional training formats.

3.2.2 Implications for Further Research on the Effects of AR- and VR-Based Training on Human Performance in Assembly

The significance of the findings is discussed below and important follow-up questions for further research are identified. In the course of the dissertation, it became clear how important it is to generate objective, reliable and valid results in the field of new and innovative technologies. The results on the effects of MR-based training on objective performance measures were obtained on the basis of a broad database from a literature review of 24 articles and an experimental study. Nevertheless, when interpreting the results, it should be mentioned that the publication of research on new technologies is often subject to the ever-present publication bias (Dickersin, 1990). This means that original research, e.g., on interventions or new technologies, is more likely to be published if the results are statistically significant, which often leads to the exclusion of important results gained in the field (Burdett et al., 2003). This has a strong influence on the evaluation of reviews or meta-analytical procedures (Bernard et al., 2014), which also affects the results found in the scoping review (Daling & Schlittmeier, 2024). Ensuring peer review procedures, also in the context of application-oriented research, therefore plays an important role in generating meaningful results in this interdisciplinary field of research.

Second, future research should attach importance to conducting studies in a methodologically sound manner in order to improve the quality of the data. In the context of the interdisciplinarity of the research field, which offers enormous opportunities for a holistic view of MR-based training, some challenges also arose in the context of the dissertation. For example, the differences in the quality of the data and evaluation of the research work examined did not allow for meta-analytical procedures to be carried out. For future research, this points towards a clear need for the introduction of standards in the selection of measurement methods for objective performance metrics (such as reporting effect sizes on task accuracy and task completion times) as well as for subjective evaluations (such as using validated questionnaires to capture system usability or task load, for example), without disregarding qualitative surveys. At the same time, research outside the field should ensure

that the results from a controlled setting are transferable to practical application. This can be achieved through the selection of the assembly task (consideration of different assembly activities as well as degrees of complexity) as well as through the composition of the sample (e.g., no purely academic sample) in future studies.

This dissertation investigated the effects of MR-based training on human potential in manual assembly. As mentioned above, this is an important prerequisite for assessing the overall success of AR- and VR-based training, i.e., its impact on efficiency factors such as time, personnel and material resources in production (Kaplan, 2021; Webel et al. 2011). To find out to what extent the use of MR-based training increases the overall efficiency, i.e. whether it helps save resources and optimize time schedules, the entire process of using MR-based training should be considered and evaluated in the future. This includes the technical and didactical development process of the training media (Khodaei et al., 2020), the implementation and long-term effects of MR-based training on users in day-to-day work as well as its impact on the production process, i.e. by avoiding the production of rejects.

4. Conclusion and Resulting Implications for Practice

This dissertation offers new insights into the effects of MR-based training in manual assembly tasks on workers' subjective evaluations and short- and long-term human assembly performance. From the conducted research on MR-based training it can be concluded that the use of AR and VR as training technologies produces positive effects for their users. In comparison to various traditional training formats, both AR- and VR-based training achieve, on average, equally good outcomes in terms of assembly performance. It was shown that short- and long-term training success in manual assembly can be equally ensured by AR-based and video-based training, while AR in combination with real-physical assembly objects showed advantages over purely virtual training in VR. These results and the experience gained during the research work enable me, the author of this dissertation, to derive concrete implications for the implementation of AR- and VR-based training in practice. In the next section, I conclude

how the insights gained in this work can be taken into account when introducing AR- or VR-based training methods.

During the field studies (Daling, Abdelrazeq, & Isenhardt, 2020; Daling, Abdelrazeq, Sauerborn, & Hees, 2020), it became clear that the context and the specification of the assembly task have a decisive impact on the choice of training method. Thus, as a first step of implementing MR-based training, it is recommended to further analyze and consider task- and context-related specifications: When using AR- or VR-based training, it should be clarified where and when the training is to take place (i.e., directly at the workplace or in replicated (virtual) workplaces) and which aspects are particularly important for the respective assembly task. Depending on the specifics of the task, it might be more or less important to interact with the real-physical object, to focus more on process flows or to consider certain safety aspects. For example, despite its poorer evaluation in usability and task load, VR could be particularly beneficial when conventional training is not feasible as it can provide enhanced safety in hazardous training environments (Daling & Schlittmeier, 2024). Furthermore, it is important to point out that MR systems should be considered as a technical complement to the instructions given by trainers in the physical workplace, which was also emphasized by workers in the industry (Daling, Abdelrazeq, & Isenhardt, 2020).

Moreover, as another important process step, the characteristics of the user group should be considered during the development of a training. Within the course of this dissertation, it became clear that both the age of the users and their previous experience with the technology played a decisive role in training success (Daling et al., 2023). According to the results, prior experience with the respective technology in training influenced immediate assembly performance, while the increasing age of participants even had a negative impact on assembly performance in the long term. Thus, it is imperative to give older and inexperienced users enough time to try out the technologies before applying them in assembly training. The end-users should also be accompanied in the use of the technology in order to be able to adequately address individual difficulties (e.g., dizziness or visual impairments) and needs (e.g., pace of training) (Daling, Abdelrazeq, & Isenhardt, 2020; Daling, Abdelrazeq,

Sauerborn, & Hees, 2020). In addition, it could be relevant whether only novices or different expert levels are trained in the same task. In any case, involving the users in the development and implementation process at an early stage seems to make sense and could positively influence the acceptance of the training.

Finally, the selection of the training technology should follow the above-mentioned considerations. Provided that the context-, task- and user-related factors are adequately addressed, it can be summarized that AR-based training in the real workplace can be recommended to ensure training success in terms of subjective satisfaction of users and objective assembly performance. However, it can also be deduced from the results of the studies that AR and VR should not simply be seen as two distinct technologies, but their functionalities and features along the MR spectrum should be explored. For example, if the focus is on the requirements for interaction with the assembly object, this can be solved either by using controllers in VR or by using a real-physical object in combination with AR. This will become increasingly important in the future, as technologies are also increasingly merging and, in some cases, both AR and VR components can be combined in one device. In the course of this dissertation, it was shown that a combination of interaction with the real-physical object and guidance via AR-based functionalities or video is often more effective than using VR-based training and a pure virtual interaction with the assembly object (Daling, Abdelrazeq, Sauerborn, & Hees, 2020; Daling & Schlittmeier, 2024, Daling et al., 2023). Additional interaction aids (e.g., haptic, visual or audio cues) were shown to improve the effects of VR-based training (Daling & Schlittmeier, 2024).

The insights gained in this dissertation indicate that AR- and VR-based training can increase human potential in manual assembly, while AR in combination with real-physical assembly components seems to be particularly suitable for training manual assembly tasks. However, it is important to emphasize that AR- or VR-based training is not generally better than traditional methods. This might also affect areas outside the human factors of training success examined here. For instance, traditional training might still retain its advantage particularly in terms of time and resource investments in development and implementation

compared to the extensive workflow of developing MR-based training. This should be considered when deciding for or against the use of MR-based training in manual assembly. In the future, hardware improvements of MR systems may contribute to increase the benefits of MR-based training. In any case, MR-based training can be seen as a promising supplement to, but not a substitute for, training by other human experts.

5. References

- Acer. (2019, April). *Acer Windows Mixed Reality Headset Setup Instructions*. Acer. <https://community.acer.com/en/kb/articles/10119-acer-windows-mixed-reality-headset-setup-instructions>
- Bannat, A., Wallhoff, F., Rigoll, G., Friesdorf, F., Bubb, H., Stork, S., Müller, H.J., Schubö, A., Wiesbeck, M., & Zäh, M. F. (2008). Towards optimal worker assistance: A framework for adaptive selection and presentation of assembly instructions. *Proceedings of the 1st international workshop on cognition for technical systems, Cotesys*. Springer.
- Bernard, R. M., Borokhovski, E., Schmid, R. F., & Tamim, R. M. (2014). An exploration of bias in meta-analysis: The case of technology integration research in higher education. *Journal of Computing in Higher Education*, 26, 183–209. <https://doi.org/10.1007/s12528-014-9084-z>
- Bornewasser, M. (2020). Montage und Komplexität. *Informatrische Assistenzsysteme in der variantenreichen Montage: Theorie und Praxis*, (pp. 43–64). https://doi.org/10.1007/978-3-662-61374-0_3
- Blattgerste, J., Strenge, B., Renner, P., Pfeiffer, T., & Essig, K. (2017). Comparing conventional and augmented reality instructions for manual assembly tasks. *Proceedings of the 10th international conference on pervasive technologies related to assistive environments*, (pp.75–82). <https://doi.org/10.1145/3056540.3056547>
- Boud, A. C., Haniff, D. J., Baber, C., & Steiner, S. J. (1999). Virtual reality and augmented reality as a training tool for assembly tasks. *1999 IEEE international conference on information visualization (Cat. No. PR00210)*, (pp.32–36). IEEE. <https://doi.org/10.1109/IV.1999.781532>
- Brolin, A. (2016). *An investigation of cognitive aspects affecting human performance in manual assembly* [Doctoral dissertation, Loughborough University]. Skövde University Library. <http://his.diva-portal.org/smash/record.jsf?pid=diva2%3A1044531>
- Brolin, A., Thorvald, P. & Case, K. (2017). Experimental study of cognitive aspects affecting human performance in manual assembly. *Production and Manufacturing Research: An Open Access Journal*, 5(1), 141–163. <https://doi.org/10.1080/21693277.2017.1374893>
- Brooke, J. (1996). SUS: A quick and dirty usability scale. *Usability evaluation in industry* (pp. 189–194). Taylor & Francis. <https://doi.org/10.1201/9781498710411-35>

-
- Burdett, S., Stewart, L. A., & Tierney, J. F. (2003). Publication bias and meta-analyses: A practical example. *International Journal of Technology Assessment in Health Care*, 19(1), 129–134. <https://doi.org/10.1017/S0266462303000126>
- Carlson, P., Peters, A., Gilbert, S. B., Vance, J. M., & Luse, A. (2015). Virtual Training: Learning Transfer of Assembly Tasks. *IEEE Transactions on Visualization and Computer Graphics*, 21(6), 770–782. <https://doi.org/10.1109/tvcg.2015.2393871>
- Cipresso, P., Giglioli, I. A. C., Raya, M. A., & Riva, G. (2018). The past, present, and future of virtual and augmented reality research: A network and cluster analysis of the literature. *Frontiers in Psychology*, 9, 2086. <https://doi.org/10.3389/fpsyg.2018.02086>
- Comberti, L., Demichela, M. & Leva, M. C. (2020). Human skills assessment as a support to human factor management. *Proceedings of the 30th european safety and reliability conference and 15th probabilistic safety assessment and management conference*. https://doi.org/10.3850/978-981-14-8593-0_3914-cd
- Daling, L. M., Abdelrazeq, A., & Isenhardt, I. (2020). A comparison of augmented and virtual reality features in industrial trainings. In J.Y.C. Chen, G. Fragomeni (Eds.), *Lecture Notes in Computer Science: Vol. 12191. Virtual, Augmented and Mixed Reality. Industrial and Everyday Life Applications* (pp. 47–65). Springer. https://doi.org/10.1007/978-3-030-49698-2_4
- Daling, L., Abdelrazeq, A., Sauerborn, C., & Hees, F. (2020). A comparative study of augmented reality assistant tools in assembly. In T. Ahram, C. Falcão (Eds.), *Advances in intelligent systems and computing: Vol. 972. Advances in usability and user experience* (pp. 755–767). Springer. https://doi.org/10.1007/978-3-030-19135-1_74
- Daling, L. M., & Schlittmeier, S. J. (2024). Effects of augmented reality-, virtual reality-, and mixed reality–based training on objective performance measures and subjective evaluations in manual assembly tasks: a scoping review. *Human Factors*, 66(2), 589–626. <https://doi.org/10.1177/00187208221105135>
- Daling, L. M., Tenbrock, M., Isenhardt, I., & Schlittmeier, S. J. (2023). Assemble it like this! – Is AR-or VR-based training an effective alternative to video-based training in manual assembly?. *Applied Ergonomics*, 110, 104021. <https://doi.org/10.1016/j.apergo.2023.104021>

- deMoura, D. Y., & Sadagic, A. (2019). The effects of stereopsis and immersion on bimanual assembly tasks in a virtual reality system. *2019 IEEE conference on Virtual Reality and 3D user interfaces (VR)*, (pp. 286–294). IEEE. <https://doi.org/10.1109/VR.2019.8798112>
- deLone, W. H., & McLean, E. R. (2002). Information systems success revisited. *Proceedings of the 35th annual Hawaii international conference on system sciences* (pp. 2966–2976). IEEE. <https://doi.org/10.1109/HICSS.2002.994345>
- Diamantopoulos, A., Sarstedt, M., Fuchs, C., Wilczynski, P. & Kaiser, S. (2012). Guidelines for choosing between multi-item and single-item scales for construct measurement: A predictive validity perspective. *Journal of the Academy of Marketing Science*, 40(3), 434–449. <https://doi.org/10.1007/s11747-011-0300-3>
- Dickersin, K. (1990). The existence of publication bias and risk factors for its occurrence. *JAMA: The Journal of the American Medical Association*, 263(10), 1385–1389. <https://doi.org/10.1001/jama.1990.03440100097014>
- Doolani, S., Owens, L., Wessels, C., & Makedon, F. (2020). VIS: An immersive virtual storytelling system for vocational training. *Applied Sciences*, 10(22), 8143. <http://doi.org/10.3390/app10228143>
- Doolani, S., Wessels, C., Kanal, V., Sevastopoulos, C., Jaiswal, A., Nambiappan, H., & Makedon, F. (2020). A review of extended reality (XR) technologies for manufacturing training. *Technologies*, 8(4), 77. <https://doi.org/10.3390/technologies8040077>
- European Commission, Directorate-General for Research and Innovation. (2021). Industry 5.0: Towards a sustainable, human-centric and resilient European industry. <https://data.europa.eu/doi/10.2777/308407>
- Fletcher, S., Johnson, T. L., Adlon, T., Larreina, J., Casla, P., Parigot, L., Alfaro, P. J. & Del Mar Otero, M. (2020). Adaptive automation assembly: Identifying system requirements for technical efficiency and worker satisfaction. *Computers & Industrial Engineering*, 139, 105772. <https://doi.org/10.1016/j.cie.2019.03.036>
- Funk, M., Kosch, T., & Schmidt, A. (2016). Interactive worker assistance: Comparing the effects of in-situ projection, head-mounted displays, tablet, and paper instructions. *Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing*, (pp. 934–939). <https://doi.org/10.1145/2971648.2971706>

-
- Ganier, F., Hoareau, C., & Tisseau, J. (2014). Evaluation of procedural learning transfer from a virtual environment to a real situation: A case study on tank maintenance training. *Ergonomics*, 57(6), 828–843. <https://doi.org/10.1080/00140139.2014.899628>
- García-Acosta, A., De La Riva-Rodríguez, J., Sánchez-Leal, J. & Reyes-Martínez, R. M. (2021). Neuroergonomic stress assessment with two different methodologies, in a manual repetitive task-product assembly. *Computational Intelligence and Neuroscience*, 2021, 1–13. <https://doi.org/10.1155/2021/5561153>
- Gavish, N., Gutierrez, T., Webel, S., Rodriguez, J., Peveri, M., Bockholt, U., & Tecchia, F. (2013). Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks. *Interactive Learning Environments*, 23(6), 778–798. <https://doi.org/10.1080/10494820.2013.815221>
- Gavish, N., Seco, T. G., Webel, S., Rodríguez, J., Peveri, M. & Bockholt, U. (2011). Transfer of skills evaluation for assembly and maintenance training. *BIO Web of Conferences*, 1, 00028. <https://doi.org/10.1051/bioconf/20110100028>
- González-Franco, M., Cermeron, J., Li, K., Pizarro, R., Thorn, J., Hutabarat, W., Tiwari, A. & Bermell-Garcia, P. (2016). Immersive augmented reality training for complex manufacturing scenarios. *arXiv preprint arXiv:1602.01944*. <https://doi.org/10.48550/arXiv.1602.01944>
- Guo, Q. (2015). Learning in a mixed reality system in the context of 'Industrie 4.0'. *Journal of Technical Education*, 3(2), 92–115. <https://doi.org/10.48513/joted.v3i2.60>
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology* (Vol. 52, pp. 139–183). [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
- Henderson, S. J., & Feiner, S. (2011). Augmented reality in the psychomotor phase of a procedural task. *10th IEEE international symposium on mixed and augmented reality*, (pp. 191–200). <https://doi.org/10.1109/ismar.2011.6092386>
- Holding, D. H. (1989). *Human Skills* (2nd ed.). John Wiley & Sons Incorporated.
- Hinrichsen, S., Riediger, D., & Unrau, A. (2016). Assistance systems in manual assembly. *Production Engineering and Management*, 01, 3–14. <https://www.th-owl.de/elsa/record/584>

-
- Kagermann, H., Helbig, J., Hellinger, A. & Wahlster, W. (2013). Recommendations for implementing the strategic initiative INDUSTRIE 4.0: Securing the future of German manufacturing industry; final report of the Industrie 4.0 Working Group. *Forschungsunion*. <https://gateway-bayern.de/BV041628398>
- Kaplan, A. D., Cruit, J., Endsley, M., Beers, S. M., Sawyer, B. D., & Hancock, P. A. (2021). The effects of virtual reality, augmented reality, and mixed reality as training enhancement methods: A meta-analysis. *Human Factors*, 63(4), 706–726. <https://doi.org/10.1177/0018720820904229>
- Khodaei, S., Sieger, J. L., Abdelrazeq, A. & Isenhardt, I. (2020). Learning goals in 360 virtual excursion - media creation guideline for mixed-reality-based classes. In *ICERI2020 proceedings* (pp. 2552–2561). IATED. <https://doi.org/10.21125/iceri.2020.0600>
- Koziol, L.F., & Budding, D.E. (2012). Procedural Learning. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (pp. 2694–2696). Springer. https://doi.org/10.1007/978-1-4419-1428-6_670
- Koumaditis, K., Venckute, S., Jensen, F. S., & Chinello, F. (2019). Immersive training: Outcomes from small scale AR/VR pilot-studies. *2019 IEEE conference on virtual reality and 3D user interfaces (VR)* (pp. 1–5). IEEE. <https://doi.org/10.1109/VR44988.2019.9044162>
- Kulyk, O., Kosara, R., Urquiza, J., & Wassink, I. (2007). Human-Centered aspects. In: Kerren, A., Ebert, A., Meyer, J. (Eds.) *Human-Centered visualization environments. Lecture Notes in Computer Science*, (Vol. 4417, pp. 13–75). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-71949-6_2
- Langer, E. (2020). *Medieninnovationen AR und VR*. Springer Vieweg. <https://doi.org/10.1007/978-3-662-60826-5>
- Lilis, R., Valciukas, J. A., Kon, S., Sarkosi, L., Campbell, C. E. & Selikoff, I. J. (1982). Assessment of lead health hazards in a body shop of an automobile assembly plant. *American Journal of Industrial Medicine*, 3(1), 33–51. <https://doi.org/10.1002/ajim.4700030107>
- Lindeman, R. W., & Noma, H. (2007). A classification scheme for multi-sensory augmented reality. *Proceedings of the 2007 ACM symposium on virtual reality software and technology* (pp. 175–178). Association for Computing Machinery. <https://doi.org/10.1145/1315184.1315216>

-
- Lotter, B. (2006). Einführung. In B. Lotter, & H. P. Wiendahl (Eds.), *Montage in der industriellen Produktion: Ein Handbuch für die Praxis* (pp. 1–8). Springer. https://doi.org/10.1007/3-540-36669-5_1
- Mackay, W. E. (2000). Augmented reality: Dangerous liaisons or the best of both worlds? *Proceedings of DARE 2000 on designing augmented reality environments* (pp. 170–171). Association for Computing Machinery. <https://doi.org/10.1145/354666.354697>
- Merkel, L., Atug, J., Berger, C., Braunreuther, S., & Reinhart, G. (2018). Mass customization and paperless assembly in the learning factory for cyber-physical-production systems: Learning module ‘from paperbased to paperless assembly’. *2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT)* (pp. 270–271). IEEE. <https://doi.org/10.1109/ICALT.2018.00130>
- Microsoft. (2021, November). *HoloLens 1 (1st gen) hardware*. Microsoft. <https://learn.microsoft.com/en-us/hololens/hololens1-hardware>
- Miguel-Alonso, I., Rodriguez-Garcia, B., Checa, D., & Bustillo, A. (2023). Countering the novelty effect: A tutorial for immersive virtual reality learning environments. *Applied Sciences*, 13(1), 593. <https://doi.org/10.3390/app13010593>
- Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, 77(12), 1321–1329.
- Murcia-Lopez, M., & Steed, A. (2018). A comparison of virtual and physical training transfer of bimanual assembly tasks. *IEEE Transactions on Visualization and Computer Graphics*, 24(4), 1574–1583. <https://doi.org/10.1109/tvcg.2018.2793638>
- Neumann, P., Winkelhaus, S., Grosse, E. H. & Glock, C. H. (2021). Industry 4.0 and the Human Factor – A systems framework and analysis methodology for successful development. *International Journal of Production Economics*, 233, 107992. <https://doi.org/10.1016/j.ijpe.2020.107992>
- Normand, J. M., Servièeres, M., & Moreau, G. (2012). A new typology of augmented reality applications. *Proceedings of the 3rd augmented human international conference* (pp. 1–8). Association for Computing Machinery. <https://doi.org/10.1145/2160125.2160143>
- Meta (2024). *[Oculus Rift becomes Meta Quest]*. Meta Quest, Meta. https://www.oculus.com/rift-s/?utm_source=www.google.com&utm_medium=oculusredirect

-
- Oestreich, H., Wrede, S. & Wrede, B. (2020). Learning and performing assembly processes. *Proceedings of the 13th ACM international conference on pervasive technologies related to assistive environments* (pp. 1–8). <https://doi.org/10.1145/3389189.3397977>
- Prasertsakul, T., Kaimuk, P., Chinjenpradit, W., Limroongreungrat, W., & Charoensuk, W. (2018). The effect of virtual reality-based balance training on motor learning and postural control in healthy adults: A randomized preliminary study. *Biomedical Engineering Online*, 17(1), 1–17. <https://doi.org/10.1186/s12938-018-0550-0>
- Rauschnabel, P. A., Felix, R., Hinsch, C., Shahab, H. & Alt, F. (2022). What is XR? Towards a framework for augmented and virtual reality. *Computers in Human Behavior*, 133, 107289. <https://doi.org/10.1016/j.chb.2022.107289>
- Reinhart, G., & Patron, C. (2003). Integrating augmented reality in the assembly domain-fundamentals, benefits and applications. *CIRP Annals*, 52(1), 5–8. [https://doi.org/10.1016/S0007-8506\(07\)60517-4](https://doi.org/10.1016/S0007-8506(07)60517-4)
- Riva, G., Baños, R. M., Botella, C., Wiederhold, B. K., & Gaggioli, A. (2012). Positive technology: using interactive technologies to promote positive functioning. *Cyberpsychology, Behavior, and Social Networking*, 15(2), 69–77. <https://doi.org/10.1089/cyber.2011.0139>
- Rose, F. D., Attree, E. A., Brooks, B. M., Parslow, D. M., & Penn, P. R. (2000). Training in virtual environments: Transfer to real world tasks and equivalence to real task training. *Ergonomics*, 43(4), 494–511. <https://doi.org/10.1080/001401300184378>
- Sauro, J. (2011). *A practical guide to the System Usability Scale: Background, benchmarks, & best practices*. Measuring Usability LLC.
- Sautter, B., & Daling, L. (2021). Mixed reality supported learning for industrial on-the-job training. *Proceedings of the Conference on Learning Factories*. <http://doi.org/10.2139/ssrn.3864189>
- Skarbez, R., Smith, M., & Whitton, M. C. (2021). Revisiting Milgram and Kishino's reality-virtuality continuum. *Frontiers in Virtual Reality*, 2, 647997. <https://doi.org/10.3389/frvir.2021.647997>
- Wang, X., Ong, S. K., & Nee, A. Y. (2016). A comprehensive survey of augmented reality assembly research. *Advances in Manufacturing*, 4, 1–22. <https://doi.org/10.1007/s40436-015-0131-4>

- Webel, S., Bockholt, U., Engelke, T., Gavish, N., Olbrich, M., & Preusche, C. (2013). An augmented reality training platform for assembly and maintenance skills. *Robotics and Autonomous Systems*, 61(4), 398–403. <https://doi.org/10.1016/j.robot.2012.09.013>
- Webel, S., Bockholt, U. & Keil, J. (2011). Design criteria for AR-Based training of maintenance and assembly tasks. *Lecture Notes in Computer Science* (pp. 123–132). Springer. https://doi.org/10.1007/978-3-642-22021-0_15
- Wells, J. D., Campbell, D. E., Valacich, J. S. & Featherman, M. (2010). The effect of perceived novelty on the adoption of information technology innovations: A risk/reward perspective. *Decision Sciences*, 41(4), 813–843. <https://doi.org/10.1111/j.1540-5915.2010.00292.x>
- Werrlich, S., Lorber, C., Nguyen, P. A., Yanez, C. E. F., & Notni, G. (2018). Assembly training: Comparing the effects of head-mounted displays and face-to-face training. In J. Y. C. Chen, & G. Fragomeni (Eds.), *Virtual, augmented and mixed reality: Interaction, navigation, visualization, embodiment, and simulation* (Vol. 10909, pp. 462–476). Springer. https://doi.org/10.1007/978-3-319-91581-4_35
- Werrlich, S., Nguyen, P.-A., & Notni, G. (2018). Evaluating the training transfer of head-mounted display based training for assembly tasks. *PETRA '18: Proceedings of the 11th Pervasive Technologies Related to Assistive Environments Conference* (pp. 297–302). Association for Computing Machinery. <https://doi.org/10.1145/3197768.3201564>
- Wolfartsberger, J., Hallewell Haslwanter, J. D., & Lindorfer, R. (2019). Perspectives on assistive systems for manual assembly tasks in industry. *Technologies*, 7(1), 12. <https://doi.org/10.3390/technologies7010012>
- Venkatesh, V., Thong, J. Y. & Xu, X. (2016). Unified Theory of Acceptance and use of technology: A synthesis and the road ahead. *Journal of the Association for Information Systems*, 17(5), 328–376. <https://doi.org/10.17705/1jais.00428>
- Yin, R. (2009). *Case Study Research (4th ed)*. SAGE Publications, Thousand Oaks, California. ISBN 978-1-4129-6099-1
- Zender, R., Weise, M., von der Heyde, M., & Söbke, H. (2018). Lehren und Lernen mit VR und AR–Was wird erwartet? Was funktioniert? (Teaching and Learning with VR and AR - What Is Expected? What Works?). *Proceedings of DeLFI workshops at the 16th e-learning conference of the German computer society*. https://ceur-ws.org/Vol-2250/WS_VRAR_paper5.pdf

6. Manuscripts

Publication 1:

Daling, L., Abdelrazeq, A., Sauerborn, C., & Hees, F. (2020). A comparative study of augmented reality assistant tools in assembly. In T. Ahram, C. Falcão (Eds.), *Advances in Intelligent Systems and Computing: Vol. 972. Advances in Usability and User Experience* (pp. 755-767). Springer. https://doi.org/10.1007/978-3-030-19135-1_74

Reproduced with permission from Springer Nature.



A Comparative Study of Augmented Reality Assistant Tools in Assembly

Lea Daling¹(✉), Anas Abdelrazeq¹, Christoph Sauerborn²,
and Frank Hees¹

¹ IMA & IfU, RWTH Aachen University, Dennewartstraße 27,
25068 Aachen, Germany

{Lea.Daling, Anas.Abelrazeq,
Frank.Hees}@ima-ifu.rwth-aachen.de

² RWTH Aachen University, Templergraben 55, 52062 Aachen, Germany
Christoph.sauerborn@rwth-aachen.de

Abstract. Digital learning formats using Augmented Reality (AR) technologies have already been successfully used in industry. AR opens new possibilities to guide employees systematically through assembly processes, ensuring the practical relevance of the learning content and a saving of time in the learning process. In order to enhance this technology's acceptance, user-centred designed AR applications are crucial. This paper aims at deriving conclusions on the usability of AR Head Mounted Displays (HMD) in the field of production. As the transferability of lab-based usability studies is still lacking, we conducted a mixed method approach in a laboratory as well as a field setting. This paper presents the current state of the art, the development of the AR applications and the main results of the study.

Keywords: Augmented Reality · Head Mounted Displays · On the job training · Assembly processes · Industry 4.0

1 Introduction

As a consequence of an increasingly networked and digitized production, the need for fast and time-effective trainings is growing [1]. One of the possible solutions is trainings on the job, ensuring the practical relevance of the learning content and time efficiency in the learning process [2]. In order to achieve this, the integration of new technology, such as Augmented Reality (AR) as digital learning format is needed to guide employees systematically through e.g. assembling processes.

Using AR as instructional assistant tool is expected to be a success factor of digital training programs [3] and first studies in this field showed promising results [4–6]. Head Mounted Displays (HMD) are often used to enable a hands-free working process. However, as the transferability of user studies conducted in laboratory settings is still lacking, there is still need for expansion of evidence-based knowledge about the use of HMD in realistic scenarios considering the target group of actual users at operator level [7]. The evaluation of developed solutions and used hardware remains an important field of research, especially with regard to acceptance and usability factors. This paper aims at deriving conclusions on the usability of HMD in the field of assembly processes.

In Sect. 2, we present the current state of the art of usability studies in this area to identify research gaps that are currently still prevalent in the steadily increasing number of publications in this field. Based on this, we present two different use cases, where AR was used to instruct the execution of an assembly process in a laboratory (Sect. 3) and a field setting (Sect. 4). The procedure of the AR development, the mixed-method approach, necessary adaptations for the respective use cases, as well as main results of the studies are presented. In Sect. 5, important findings on usability and transferability are summarized for both settings and the challenges of conducting field studies in this area are highlighted.

2 The Use of Head Mounted Displays in Manufacturing

The development of AR tools in the industry is a very fast growing and now integral part of different subspecialties of today's industry. To date, the use of AR in the industrial field is mainly related to maintenance and manufacturing or assembly related tasks [7] providing additional information or instructions regarding the learning subject. By projecting virtual objects into the user's real physical environment, AR enables assembly instructions to be directly connected to the products or machines [8]. Since both hands are used simultaneously, HMDs are the most commonly used device and several studies show that the use of HMD in on the job trainings can increase quality and efficiency [7].

Nevertheless, there are still numerous problems that prevent the potential of AR from being optimally exploited in the field of assembly [9]. A central point is that many studies already focus on objective key indicators (e.g., time, error rate, accuracy) as dependent variables, whereby basic usability factors such as acceptance, perceived usefulness or enjoyment as well as technical issues (e.g., display and tracking technology, calibration techniques, interfaces to the operating machines) have not yet been sufficiently considered [9]. For example, excessive reliance on AR-generated signals and displays can negatively impact user performance by disrupting the attention [10]. In addition, the usability evaluation barely distinguishes between the hardware and software to be tested, which is why reported positive effects of HMDs are not clearly attributable to the system or the medium itself [11].

Furthermore, there are some positive findings about the use of HMDs in lab-based environments, but the actual use in the field is rarely tested. Dey, Billingham, Lindemann and Swan reported on 10 years of user studies published in AR papers. They reviewed 291 papers and 369 individual studies and reported that the majority of the papers (76%) conducted the experiments in controlled laboratory environments, while only 44 papers (15%) conducted the experiments in a natural environment or as a field study [7]. Within the industrial application area, only six out of 29 studies were executed in a field setup. In addition to that, most of the studies were conducted with young participants, mostly university students. The authors state that performing more industrial AR studies in the natural environment would lead to more-usable results, as "controlled environments may not expose the users to the issues that they face in real-world setups" [7].

Moreover, several studies are mostly based on simple assembly tasks while the actual requirements for trainees are much more complex [5, 12]. This issue is the key reason for HMD being in laboratory use instead of actual job training. To extend the field of use of HMD, Werrlich and colleagues carried out a study with more complex tasks to compare the effects of HMD-based and paper-based instructions [5]. The results showed that the assembly training with the HMD reduced the error rate, but actually took more time than the paper-based version, which might be caused by the unfamiliarity with AR. The extra time needed by using HMD will probably be reduced by integrating this technology into the everyday life of assembly training [5]. In addition, the general research area of usability of AR applications shows a clear trend from HMD to handheld devices [7]. The potential of other devices in the assembly field in comparison to HMDs should therefore be further researched.

Taking this into consideration, we defined two use cases for the usability evaluation of HMDs in assembly, which should contribute to closing the existing gap between laboratory and field studies in this area. In order to control interfering and other influencing factors, one of the use cases was performed in a laboratory study. The second use case represents a realistic assembly scenario in production. In addition to the presentation of the use case requirements and the AR development, the study design tailored to the respective use case and the results are presented. In the later course, the challenges and opportunities with regard to the transferability of laboratory studies to the practical field will be discussed.

3 Use Case A: Laboratory Study

In the first use case, AR is used to give instructions for mounting a gear drive. In order to meet the requirement of a more complex task, the gearbox is not assembled manually but in cooperation with a robot (UR-5). The worker performs five steps, while the robot performs a total of four steps. The steps are executed in an assembly cell with a width of about two meters. First, the worker is instructed to position a base plate and a back plate in a holder. The robot inserts four screws and positions the back plate on the base plate while the human operator assembles two gear sets. In the last step, these are mounted on the base plate which was prefabricated by the robot. The collaborative assembly cell is part of an industry 4.0 demonstrator at the university and thus corresponds to a setting in a controlled laboratory environment. This application case makes it possible to evaluate the developed AR solution in an experimental within-subject study design. Here, it was possible to test not only HMDs but also other devices such as handhelds (tablet) and, as a control run, a non AR-supported solution with a touchscreen. In the following section, requirements for the development of the AR solution are derived. Then the study design is described and the results of the evaluation are presented.

3.1 Use Case Requirements and AR Application Design

Based on usability heuristics [13] and the reported challenges of AR applications in industry, the following requirements have been defined as a base for the development process:

1. The application should function on head-mounted (i.e. Microsoft HoloLens) and handheld (i.e. Android-based tablet) AR devices.
2. The application should be able to recognize the working space (i.e. assembly table in its entire width) and localize itself continuously during the operation.
3. The application should illustrate different step cues for the worker. Cues could be specified using text, 2D images and 3D animations.
4. The worker should be able to interact with the application in order to either navigate with touch or voice control through a number of steps.
5. The AR application should be able to communicate with the robot and be aware of its status.

Based on the listed requirements, an AR application has been developed targeting both, Microsoft HoloLens and Android-based Tablet with a desk mounted arm (i.e. Samsung Tab S2 T710 with 8.0" display) as testing devices. To ease the porting process for different platforms, Unity 3D game engine was adapted for the software development process [14]. The proposed AR application is composed from several modules (i.e. Game Objects) that are organized within a Unity Scene (see Fig. 1). Each module realizes some functionalities as the following:

Localization and Tracking. This component is based on Vuforia due to its ability for marker tracking and real-time processing [15]. In order to localize the device, one to two different marker have been located in a pre-defined location on the cell (see Fig. 1). Both HoloLens and the tablet are able to recognize whenever the markers are in the sight of their cameras. In case tracking was completely lost, the application asks the user to reposition the device where the marker is shown. In addition, arrows are used to indicate the localization of the next step.

Scenario Management. For our scenario, we defined a configuration “json” structure which enables dynamic and spontaneous changes without a need to rebuild the application. The configuration file can define the following cues for each step: Step description text, extra helping text, 2D leading photos for different tools, tools tags and names, 3D animations for the intended step, and information to act based on the robot status.

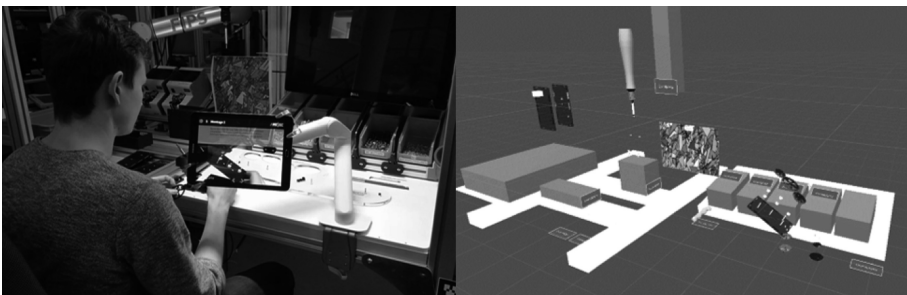


Fig. 1. Development of an AR-unity scene for the assembly cell.

Interactive User Interface. The interface allows navigation through the steps either by touching (i.e. clicking) for the tablet or by voice control for the HoloLens. The interface is divided into different regions indicating different information and notifications.

Robot Communication. This component is based on ROS in order to communicate with the robot either to push notification to the robot asking for initiating a step or to wait for the robot until it finalizes a step. In latter case, the interface will show a window to hold until the robot is ready.

3.2 Method

Study Design and Procedure. Both studies follow a mixed method approach combining qualitative and quantitative methods. This ensures a maximum gain of knowledge even with relatively small samples. In the first study, participants were randomly assigned to different test conditions in a within-subject design. Other media in addition to HMD were used to carry out the assembly process. Every participant conducted three rounds of the assembly task – each being instructed by different media: HoloLens and tablet are used for the AR solution; the fixed touchscreen using non AR 3D animations is used as a control trial. The order of the instruction media was randomized to control repetition effects [16]. In total, the study took about 60 min to complete. First, the participants filled out a pre-test questionnaire. Afterwards, they were asked to familiarize themselves with the assembly cell and the technological equipment. After each round of assembly, they were asked to fill out a post-test questionnaire referring to the respective medium they used. At the end of the study, the AR application itself was evaluated ensuring that the medium and embedded software are evaluated separately. During the whole process, the participants were asked to express their thoughts and concerns aloud (Thinking Aloud Method [17]), which was recorded with a voice recorder.

Participants. The sample consisted of 8 men and 7 women within the first study ($N_{\text{lab}} = 15$). The average age of the participants was 25 years ($MW = 25.07$, range = 20–32). A total of eleven students and four employed persons took part. Seven participants stated that they had a general university entrance qualification as their highest educational attainment, the remaining eight already had an academic degree ($n_{\text{Bachelor}} = 4$, $n_{\text{Master}} = 3$, $n_{\text{Doctorate}} = 1$). Only one person had already taken part in a study on the collaborative assembly cell.

Questionnaires. The pre-test questionnaire was composed of different questions concerning the participants' affinity for technology (e.g., "My enthusiasm for technology is...") and the "locus of control for technology (KUT)" questionnaire to assess general control beliefs while dealing with technology on a six-level scale [18]. In addition, we asked for demographic data and the technical usage behaviour of the participants (e.g., which media is used, frequency of use, type of use).

The composition of the post-test questionnaire will only be briefly summarized here - for further information on the scales and items used we refer to earlier works of the

authors [11]. In general, the assessment of instructional media was carried out separately from the assessment of the AR application. To evaluate the instructional media, we used the “NASA Task Load Index (NASA TLX)” [19] to measure the experienced demand during the task. Further constructs such as *perceived usefulness* (Cronbachs $\alpha = .91-.93$), *media self-efficacy* (Cronbachs $\alpha = .85-.95$), *perceived enjoyment* (Cronbachs $\alpha = .66-.92$), and *perceived ease of use* (Cronbachs $\alpha = .560-.89$) were selected and operationalized on the basis of existing questionnaires (e.g., TAM 3 [20], meCue2.0 [21], IsoMetrics [22]) and their face validity for the evaluation of the developed application. In addition, the post-test on the instructional media also contains open questions: “*What did you particularly like about the teaching medium you used?*”, “*What would have to be changed in the medium to make the assembly process even easier?*”, and “*Please rank the teaching media, where 1 is your strongest preference, 2 is your second choice, etc. Please give reasons for your decision.*”

To evaluate the AR application at the end of the study, we used the following constructs: *Perceived usefulness* (Cronbachs $\alpha = .96$), *aesthetic and layout* (Cronbachs $\alpha = .60$), *appropriateness of functions* (Cronbachs $\alpha = .72$), as well as *terminology and terms* (Cronbachs $\alpha = .65$). The open questions “*What did you particularly like about the AR application?*”, “*What would need to be changed in AR application to make the assembly process even easier?*” were supplemented here by the question whether and why they would prefer AR to classical manuals.

Analysis. Analysis in both studies was conducted using SPSS. Qualitative data from voice recordings and open questions was analysed with MAXQDA Software. Due to the small sample size, Friedman’s ANOVA for dependent samples and a post-hoc analysis according to Dunn-Bonferroni was conducted to explore existing group differences between the comparison groups “HoloLens”, “tablet”, and “touchscreen” [16]. Above all, the descriptive and qualitative findings especially with regard to the usability of the HoloLens are highlighted.

3.3 Results

Pre-test. Three participants used the HoloLens before conducting the experiment. The mean technical affinity of the sample ($N = 15$) is 4.61 (min = 3.4; max = 5.60; SD = 0.71) and the general control beliefs using technology has a mean of 4.73 (min = 3.00; max = 5.75; SD = 0.72; 1 = *very low*; 6 = *very high*).

Post-test - Instructional Media. Results of the NASA TLX (scale from 0 to 10) indicate that the HoloLens reached the highest mean level of *frustration* (HoloLens: 4.67; tablet: 4.53; touchscreen: 3.13), *effort* (HoloLens: 4.67; tablet: 4.13; touchscreen: 3.27), *temporal demand* (HoloLens: 4.20; tablet: 3.80; touchscreen: 2.93), and *mental demand* (HoloLens: 6.13; tablet: 5.07; touchscreen: 0.40). The highest *physical demand* was reported using the tablet = 5.27 (HoloLens: 4.33; touchscreen: 4.13), the highest *performance* was reached using the touchscreen = 7.33 (HoloLens: 6.40; tablet: 6.33).

Concerning the usability scales, all descriptive statistics can be seen in Table 1. No significant group differences were shown in *media self-efficacy* ($\chi^2_r(2) = 4.545$,

$p = .103$, $n = 15$) and *perceived enjoyment* ($\chi^2_r(2) = 2.980$, $p = .225$, $n = 14$). Results of Friedman’s ANOVA and the subsequently performed Dunn-Bonferroni test show a statistically significant group difference for *perceived ease of use* ($\chi^2_r(2) = 21.088$, $p < .000$, $n = 15$). Differences between all media used with a corrected alpha = 0.17 indicate that the HoloLens is considered to be the least easy to use, while the touchscreen reaches the highest value ($z = -3.425$, $p = .001$). The analysis for *perceived usefulness* shows a statistically significant difference between the groups ($\chi^2_r(2) = 10.67$, $p = .005$, $n = 15$). The subsequently performed Dunn-Bonferroni tests with a corrected alpha = .017 show that both the perceived usefulness between tablet and touchscreen differ statistically significantly ($z = -2.641$, $p = .008$), as well as the perceived usefulness between HoloLens and touchscreen ($z = -2.548$, $p = .011$), indicating that HoloLens and tablet are perceived as less useful as the touchscreen.

The ranking shows that the HoloLens was chosen three times a first choice (tablet 2 times, touchscreen 10 times), three times as second and nine times as third choice. Written comments and verbal expression were categorized in positive and negative comments, while individual comments were coded multiple times. Of 69 overall positive comments on the media, the HoloLens was 18 times positively evaluated referring to the intuitive operation and the innovative and supportive character of the technology: *“The animations show exactly what needs to be done. The process is very intuitive after a short familiarization phase.”* 38 of 68 negative remarks were related to the HoloLens, criticizing the lack of wearing comfort and limited vision: *“If you have these glasses on and don’t look at the animation, but at what you’re doing, you don’t see it very well. I’d rather try to look under my glasses”*.

Table 1. Descriptive statistics – instructional media

Scales	Descriptive statistics					
		<i>N</i>	<i>Mean</i>	<i>SD</i>	<i>Min.</i>	<i>Max.</i>
Media self-efficacy	HoloLens	15	4.33	1.51	2.75	6.00
	tablet	15	4.80	.95	3.00	6.00
	touchscreen	15	5.10	.96	3.00	6.00
Perceived enjoyment	HoloLens	14	4.38	.66	3.33	5.83
	tablet	14	4.29	.88	2.67	6.00
	touchscreen	14	4.98	.66	3.83	6.00
Perceived ease of use	HoloLens	15	3.78	.96	1.80	5.20
	tablet	15	4.60	.60	3.60	5.60
	touchscreen	15	5.17	.59	4.20	6.00
Perceived usefulness	HoloLens	15	3.85	1.11	2.00	5.50
	tablet	15	3.87	.96	2.50	5.75
	touchscreen	15	4.87	.93	2.25	6.00

Post test - AR Application. The descriptive statistics of the assessment of the AR application show that *perceived usefulness* was rated with a mean of 4.40 (min = 2.25; max = 5.75; SD = 1.12), *aesthetic and layout* reached a mean of 3.97 (min = 2.83;

max = 5.00; SD = .67), *appropriateness of functions* had a mean of 4.00 (min = 2.40; max = 5.40; SD = .93), and *terminology and terms* was rated with a mean value of 4.29 (min = 3.17; max = 5.67; SD = .71).

When asked about their preference, 13 out of 15 participants indicated that they would prefer AR-based trainings over traditional manuals. Particularly mentioned was the high level of interaction with the technology, the active learning process, the small steps and the perceived enjoyment. Those who preferred classic manuals mentioned above all the possibility of looking through manuals at one's own pace.

4 Use Case B: Field Study

The second use case examined in the field study took place in a plant of a mechanical engineering company in Germany. It deals with the manual assembly of a supporting bearing. The assembly process of a bearing, which is made out of 20 components, can be divided into 37 steps. These steps have to be executed in an assembly cell with a width of about three meters. The center of the cell is an assembly table, which is surrounded by shelves on the left, right and front. These shelves contain the required components. The process was to be stabilized with the help of AR for two reasons: first of all, inexperienced employees are often assigned to the assembly line and thus have to be trained to perform the task. Secondly, various process steps are often forgotten. Components affected by those mistakes are faulty and are subject to complaints by customers. Therefore, the AR application should provide a higher process stability. Due to the fact that all participants work professionally in the assembly process, a more complex but manual assembly process was specified. The application was changed in several ways.

4.1 Use Case Requirements and AR Application Design

In contrast to the laboratory study, the application in the field study was developed solely for a head-mounted AR device (i.e. Microsoft HoloLens). The following characteristics of the use case had to be considered during development:

- (1) The assembly steps have different characteristics, such as the location of execution and the required degree of accuracy. Therefore, the visualization should adapt to the changing conditions.
- (2) The AR-device used has a limited field of view. Therefore, an orientation aid should be implemented.
- (3) There is background noise in the production hall. This must be taken into account when using auditory elements.

The assembly steps were visualized with the help of 3D models of components and tools, with videos, 2D information graphics and geometrical 3D objects (e.g. arrows). All content is placed at predefined locations in relation to the context (i.e. context aware). The visualization depends on the degree of awareness of the location of a work step. If the exact location is known, digital twins are displayed wherever the real components have to be placed. If the location is only roughly known, a video of the

work step is displayed (i.e. comfortable height and angle) in the employee's ergonomically optimal radius of action. Two systems serve as an orientation aid for the employee. On the one hand, a virtual compass is displayed at a fixed position during the entire assembly process. It always points to the point of interest (POI). On the other hand, headphones attached to the HoloLens simulate a sound source with the POI as origin. Here we take advantage of the ability of humans to locate a sound source within short time. For navigation between assembly steps, the employee can either jump back and forth using the voice commands "Next" and "Back" or use a hardware device (e.g. HoloLens Clicker) to jump to the next work step.

4.2 Method

Study Design and Procedure. In the second study, the research question was transferred to a different setting in the field. However, both studies have some substantial differences in the methodology used. As the employees were released to participate in the study, there was only a limited timeframe available for conducting the experiment. Workers at operator level were invited to carry out the quasi-experimental study and were assigned to various test times by the factory manager. Since in this case time resources were limited, only the execution with the HMD was tested. As described later, the questionnaire was modified according to the circumstances of the field study.

Participants. After excluding one participant because of immense visual problems, a total of seven men and two women took part in the second study ($N_{\text{field}} = 9$). The average age of the field study participants was 37 years ($MW = 37.22$, range = 24–55). Five participants indicated to have middle school diploma, three had general or specific high school graduation as highest education degree, one remaining had completed lower secondary school. All participants work in assembly cells. None had used a HoloLens before.

Questionnaires. In the field study, the same questionnaires were used as presented above. Nevertheless, the established questionnaire was linguistically adapted to the lower level of education compared to the academics in the laboratory study (e.g., to explain "effectivity" we added "I could do more" in german language). Moreover, three self-generated items for both the virtual compass and spatial hearing scales were added. All items had the aim to question if the user perceived the placed elements as useful or disruptive (e.g. "It was easy for me to see where I had to do something." or "the virtual compass has irritated/distracted me", both in german language). Since no robot was included in this use case, the item "Application keeps me informed about what the robot is doing" (in german language) was omitted. Since only the HoloLens was used, participants were not asked for a ranking of the media.

4.3 Results

Pre-test. The participants had an average technical affinity of 4.53 ($N = 9$; min = 3.6; max = 6.00; $SD = 0.77$). General control beliefs while dealing with technology ranged

between min = 3.75 to max = 5.88 (N = 8; mean = 4.78; SD = 0.66) within the sample. Only one participant already used an HMD before this study.

Post-test - Instructional Media. Results of the NASA TLX show a mean *performance* level using the HoloLens of 7.56 (min = 3; max = 10; SD = 2.07) on a scale from 0 to 10. *Frustration* was rated in average with a value of 4.89 (min = 2; max = 8; SD = 1.9), while *effort* has a mean of 2.78 (min = 1; max = 7; SD = 1.72). *Temporal demand* shows a mean of 2.56 (min = 1; max = 7; SD = 1.68), *mental demand* was rated on average with 6.11 (min = 1; max = 10; SD = 2.93). *Physical demand* has a mean value of 4.56 (min = 2; max = 7; SD = 1.67). As the HoloLens was evaluated here without referring to the AR Application used, it was possible to compare the groups from the lab-based and the field study. Mann-Whitney-U-Test shows a statistically marginal significant difference between the groups referring to *effort* (“How hard did you have to work (cognitively and physically) to reach your performance level?”; $U = 35.5, p = .055$), indicating that the participants in the lab-based study had to spend more effort conducting the task.

In this study, *media self-efficacy* has a mean of 4.72 (min = 2.00; max = 6.00; SD = 1.18), *perceived enjoyment* shows a mean value of 3.66 (min = 2.33; max = 5.50; SD = .94), *perceived ease of use* has a mean of 4.60 (min = 3.60; max = 5.60; SD = .62), *perceived usefulness* shows a mean of 3.00 (min = 1.50; max = 5.00; SD = 1.06), and the added scale of *spatial hearing* shows a mean value of 2.93 (min = 1.67; max = 4.67; SD = 1.00). Mann-Whitney-U-Test revealed statistically significant differences between the lab-based and the field study within *perceived enjoyment* ($U = 28.5, p = .028$), indicating that the participants had more fun using the HoloLens, as well as within the *perceived ease of use* ($U = 32.5, p = .035$) in a way that participants from the field study found it easier to use the HoloLens.

Written comments and verbal expressions showed nine positive comments (e.g., “When assembling a product unknown to you, it makes sense to use the HoloLens”), 18 negative comments (e.g., “As a wearer of glasses, as always, unpleasant. As soon as you hit the glasses once and they slip slightly, they press on your nose.”) and seven suggestions for improvement (e.g., “Better adaptation to eyeglass wearers”) including comments that the music was perceived as unpleasant and disturbing for the function of the spatial hearing.

AR Application. The descriptive statistics are shown in Table 2. No group comparisons were conducted, as there were substantial differences in the design of the applications.

Table 2. Descriptive statistics – AR application.

Scales	Descriptive statistics				
	<i>N</i>	<i>Mean</i>	<i>SD</i>	<i>Min.</i>	<i>Max.</i>
Perceived usefulness	9	3.42	1.14	1.50	5.25
Aesthetic and layout	9	4.54	.59	3.83	5.67
Appropriateness of functions	9	4.20	.56	3.60	5.00
Virtual compass	9	4.19	1.00	2.33	5.00
Terminology and terms	9	4.42	.65	3.80	5.60

Seven participants stated that they would prefer AR over traditional manuals, because “*the principle of learning by doing leads more effectively and faster to the learning of new skills and abilities, especially in strongly practice-oriented fields of activity*”. One participant preferred classical manuals, as they do not require any training on the new technology and another one was indecisive.

5 Limitations and Discussion

Limitations. The variation of the use cases was accompanied by a substantial change of the AR application. Thus, a major limitation of the study is the lack of comparability between the two settings, which represents a general problem in transferring laboratory-based results into the practical field. Accordingly, the influence of the additional component of the robot in the first use case cannot be clearly determined. A further point of criticism, besides the very small sample, is the difference of the participants. While the participants in the laboratory study were still rather young and from an academic environment, the average age of the workers in the field study was considerably higher and the previous education was generally lower and widely diversified. These findings are consistent with the observations of Dey and colleagues who found that the samples from the reviewed studies correspond more to a young population, which is not really representative for the implementation in practice. In addition, they claim that 12 to 18 participants per study is a typical range in the AR community [7].

Moreover, a tailor-made usability catalogue was used for the survey, which supported the evaluation of the selected use cases, but did not make use of empirically validated scales [11]. The verification and validation of the instrument is an important factor for the transferability of the results in the course of further studies.

Discussion. This paper contributes to closing the research gap between theoretical and laboratory-based work and field studies in a practical environment. Furthermore, comparisons between different instructional media are carried out as far as the context of the use case allows.

The results of the NASA TLX show that the HoloLens requires a high mental demand value compared to other media. Overall, there were no significant differences in the perception of the HoloLens between the two groups tested. The marginal difference in the *effort* factor can be explained by the fact that the test persons in the laboratory study had no experience in the (assembly) task to be performed. For almost all factors a high standard deviation occurred in the field study, which indicates an interpersonally strongly varying assessment of the HoloLens. It would be interesting to connect individual prerequisites (e.g. control beliefs while dealing with technology) with the assessment of the task load when using the HoloLens.

In the area of usability factors, it can be shown that HoloLens performs rather poorly compared to other media. In the laboratory study, there was no difference in the perceived enjoyment when using the media. A difference can be seen to the participants of the field study, who had altogether less joy in the execution with the HoloLens. This can be attributed in particular to ergonomic aspects - almost all older participants said

that they had difficulties with the HMD in combination with visual disorders and visual aids. Surprisingly, the field sample's assessment of the ease of use was more positive, which could be due to their familiarity with assembly processes in general or to the fact that they could not test any other media for comparison. The qualitative statements support this impression.

Overall, the AR application is perceived as positive and supportive in both studies, independently of the medium used. The virtual compass added after the first study is generally perceived as helpful by the participants of the field study. The use of the spatial hearing, on the other hand, was perceived as strengthening and irritating in noisy environments. It can be concluded, that despite the different types of samples (laypersons vs. experts) and different starting variables (e.g. age or experience with the media), the results of the qualitative analysis show similar findings. Particularly with regard to ergonomic aspects, the hardware (the HMD) was still heavily criticized (weight, field of vision, combination with other visual aids), but the software was on average perceived as helpful and supportive.

References

1. Ahrens, D., Spöttl, G.: Die Vision Industrie 4.0 und ihre sozialen Herausforderungen: Digitalisierung industrieller Arbeit. In: Hirsch-Kreinsen, H., Ittermann, P., Niehaus, J. (eds.) p. 184. Nomos, Baden-Baden (2015)
2. Schuh, G., Gartzten, T., Rodenhauser, T., Marks, A.: Promoting work-based learning through industry 4.0. *Procedia CIRP* **32**, 82–87 (2015)
3. Guo, Q.: Learning in a mixed reality system in the context of industry 4.0. *J. Tech. Educ.* **3**, 92–115 (2015)
4. Funk, M., Kosch, T., Schmidt, A.: Interactive worker assistance. In: Lukowicz, P., Krüger, A., Bulling, A., Lim, Y.-K., Patel, S.N. (eds.) *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp 2016*, pp. 934–939. ACM Press, New York (2016)
5. Werrlich, S., Austino, D., Ginger, A., Nguyen, P.-A., Notni, G.: Comparing HMD-based and paper-based training. In: 2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 134–142. IEEE, Munich (2018)
6. Tang, A., Owen, C., Biocca, F., Mou, W.: Comparative effectiveness of augmented reality in object assembly. In: Cockton, G., Korhonen, P. (eds.) *Proceedings of the conference on Human factors in computing systems - CHI 2003*, pp. 73–80 ACM Press, New York (2003)
7. Dey, A., Billinghamurst, M., Lindeman, R.W., Swan, J.E.: A systematic review of 10 years of augmented reality usability studies: 2005 to 2014. *Front. Robot. AI* **5**, 161 (2018)
8. Webel, S., Bockholt, U., Engelke, T., Gavish, N., Tecchia, F.: Recent trends of mobile collaborative augmented reality systems. In: Alem, L., Huang, W. (eds.) pp. 69–82. Springer, New York (2011)
9. Wang, X., Ong, S.K., Nee, A.Y.C.: A comprehensive survey of augmented reality assembly research. *Adv. Manuf.* **1**(4), 1–22 (2016)
10. Akçayır, M., Akçayır, G.: Advantages and challenges associated with augmented reality for education: a systematic review of literature. *Educ. Res. Rev.* **20**, 1–11 (2017)
11. Daling, L., Abdelrazeq, A., Haberstroh, M., Hees, F.: Usability evaluation of augmented reality as instructional tool in collaborative assembly cells. In: *The Twelfth International Conference on Advances in Computer-Human Interactions (ACHI 2019)*. IARIA (2019, in press)

12. Tümler, J., Doil, F., Mecke, R., Paul, G., Schenk, M., Pfister, E.A., et al.: Mobile augmented reality in industrial applications: approaches for solution of user-related issues. In: 7th IEEE/ACM International Symposium on Mixed and Augmented Reality, pp. 87–90. IEEE, Cambridge (2008)
13. Nielsen, J.: Enhancing the explanatory power of usability heuristics. In: Adelson, B. (eds.) Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 152–158. ACM, New York (1994)
14. Abdelrazeq, A., Shehadeh, M., Schroeder, S., Richert, A., Jeschke, S.: A development kit for augmented reality applications in healthcare. In: Proceedings of IASTEM International Conference (2017)
15. Chiu, P.-H., Tseng, P.-H., Feng, K.-T.: Cloud computing based mobile augmented reality interactive system. In: IEEE Wireless Communications and Networking Conference (WCNC), pp. 3320–3325. IEEE, Piscataway (2014)
16. Field, A., Hole, G.: How to Design and Report Experiments. Sage, Los Angeles (2013)
17. Sarodnick, F., Brau, H.: Methoden der Usability Evaluation: Wissenschaftliche Grundlagen und praktische Anwendung. Huber, Bern (2016)
18. Beier, G.: Kontrollüberzeugungen im Umgang mit Technik. Ein Persönlichkeitsmerkmal mit Relevanz für die Gestaltung technischer Systeme. Zugl.: Berlin, Humboldt-Univ., Diss., dissertation.de, Berlin (2003)
19. Hart, S.G.: Nasa-task load index (NASA-ZLX); 20 years later. Proc. Hum. Factors Ergon. Soc. Annu. Meet. **50**, 904–908 (2006)
20. Davis, F.D.: A technology acceptance model for empirically testing new end-user information systems: theory and results. Ph.D. dissertation, Massachusetts Institute of Technology (1985)
21. Minge, M., Thüring, M., Wagner, I., Kuhr, C.V.: The meCUE questionnaire: a modular tool for measuring user experience. In: Advances in Ergonomics Modeling, Usability and Special Populations, pp. 115–228. Springer, Cham (2017)
22. Figl, K.: ISONORM 9241/10 und Isometrics: Usability-Fragebögen im Vergleich. In: Wandke, H., Kain, S., Struve, D. (eds.) Mensch & Computer 2009: Grenzenlos frei!?, pp. 143–152. Oldenbourg, Munich (2009)

Publication 2:

Daling, L. M., Abdelrazeq, A., & Isenhardt, I. (2020). A comparison of augmented and virtual reality features in industrial trainings. In J.Y.C. Chen, G. Fragomeni (Eds.), *Lecture Notes in Computer Science: Vol. 12191. Virtual, Augmented and Mixed Reality. Industrial and Everyday Life Applications* (pp. 47-65). Springer. https://doi.org/10.1007/978-3-030-49698-2_4

Reproduced with permission from Springer Nature.



A Comparison of Augmented and Virtual Reality Features in Industrial Trainings

Lea M. Daling^(✉), Anas Abdelrazeq, and Ingrid Isenhardt

Chair of Information Management in Mechanical Engineering (IMA),
RWTH Aachen University, Aachen, Germany
lea.daling@ima.rwth-aachen.de

Abstract. Short-term qualification for temporary workers is a constant challenge for manufacturing companies. Cycle times of machines often have to be reduced for training processes, which demands time and financial resources. This increases the need for near-the-job trainings without manipulating cycle times of the machine. Digital visualization tools using Mixed Reality (MR) promise opportunities for application-oriented practical training. However, especially for industrial applications, where procedural knowledge has to be transferred, it is not clear which MR technology should be used for which purpose. In order to answer this question, this paper examines the underlying MR-features of the technology. In an experimental setting, the same virtual training for the assembly of a pneumatic cylinder is examined with an augmented reality/augmented virtuality (AR/AV) based application in comparison to a virtual reality (VR) based application. Based on the carried out study, there are significant differences in the evaluation of the system usability, but no differences in the evaluation of the ergonomics and the perceived task load during the task. Out of 16 test persons, 14 would choose the VR system in the final analysis. The results are discussed in the paper and recommendations for the design of MR based systems in an industrial context are given.

Keywords: Industrial trainings · Augmented reality · Virtual reality · System usability · Mixed reality features

1 Introduction

1.1 Training Procedures in Production and Manufacturing

In the course of increasingly digitalized production, the evolvement of complex, digitally networked systems also poses challenges to their operators [1]. In this context, machine operation and associated assembly processes require safe and effective interaction between man and machine [2]. This also changes the way of corresponding training procedures. Face-to-face or text-based training procedures are increasingly being replaced by more flexible digital methods such as blended-learning or virtual learning environments [3]. This becomes particularly relevant if inexperienced or temporary workers have to be trained for new (assembly) processes [4].

By using digital (i.e. virtual) near-the-job trainings, inexperienced employees can be trained in a standardized way without manipulating the cycle time of a machine [5].

This not only improves the efficiency of the training by reducing costs and time, but also increases safety and enables the consideration of different learning paces and previous knowledge [6, 7]. Such trainings play a crucial role in improving the human capital of the workforce [4]. Virtual learning environments are particularly suitable for industrial application, since they offer practicing real-life activities of manufacturing processes in a virtual environment [8]. In addition to instructional videos and learning platforms, the use of Mixed Reality (MR) technologies has become a promising procedure in order to visualize and illustrate workflows and processes near-the-job [2, 9].

Although the benefits of MR technologies are empirically proven and widely acknowledged from workers and operators [3, 10, 11], there are currently only few cautious attempts to apply these technologies in a long-term and targeted manner [7]. There is also a lack of knowledge about which technology should be used for certain tasks or how the features should be specified: “each has unique strengths and limits for aiding learning, so understanding how to choose the right medium for a particular educational situation is an important next step in realizing the potential of immersive media in learning” [12]. Neglecting these different potentials can provoke unexpected effects, e.g. that paper-based training in experimental studies leads to better results than MR technologies [13]. Possible reasons for this are missing or not optimally used features (e.g., gesture control or hand recognition) of the used technology in the respective task [14]. To determine which technology or which features are particularly suitable for a certain task, the following section presents a definition of MR as well as a brief overview of the features of the technologies and devices used in this context.

1.2 Mixed Reality

The term MR describes a continuum between reality and virtuality, involving the merging of real and virtual worlds [15]. Within the reality-virtuality continuum, a distinction is made between the real environment, Augmented Reality (AR), Augmented Virtuality (AV) and the virtual environment (Fig. 1).

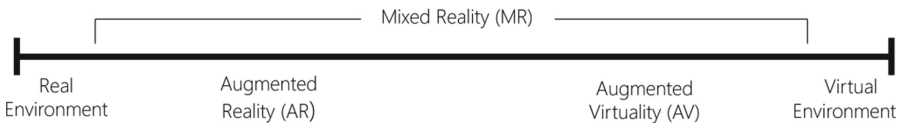


Fig. 1. Reality-virtuality continuum [15]

AR describes the immersion in reality and the handling or interaction with virtual objects. Thus, an AR system “combines real and virtual objects in a real environment; runs interactively and in real time and registers (aligns) real and virtual objects with each other” [16]. The use of AR in training of managing industrial equipment gives orientation of the correct utilization of the instruments, which in turn omits errors and production rejects [17]. AR enables the user to manage the equipment from screen or device through the recognition of current state of the equipment [17].

AV describes the immersion into a virtual world, which is extended by reality, while the user manipulates mainly virtual objects. Especially in collaborative use cases and settings AV promises to be successful. In industrial context, for example, virtual meetings can be held in which engineers can jointly manipulate or change 3D models in real time [18]. Compared to AR, immersion plays a more prominent role in AV [19]. Nevertheless, the boundary between AR and AV remains weak and depends strongly on the application and use [20].

Although not directly mentioned in the definition of Milgram and Colquhoun in 1999, Virtual Reality (VR) is a widely used technology that most closely corresponds to the virtual environment and thus is associated with the right side of the continuum [10]. In this paper, VR is understood as part of Mixed Reality. VR describes complete immersion in a virtual world, while the real environment is hidden [15]. Thus, VR offers a very comprehensive view of the structures and installations of a respective environment [21]. Providing the user with perceptive information about the structure and functioning of a real system is a cost-effective advantage of VR [8]. The advantage of using VR as industrial training is to prevent user from exposing risks of the real process equipment and familiarize with the tool and work steps at the same time through virtual immersion [22].

Many authors criticize the apparent separation of these terms in the continuum, as the boundaries of the terms become indistinct in most use cases [20]. The capabilities of technologies and devices are becoming more and more extensive and are increasingly overlapping. In the course of the discussion about the classification of Mixed Reality, suggestions came up to make the distinction rather on the basis of the technological specification (technology-centered taxonomies) [23], the preconditions of the user (user-centered taxonomies) [24], the specification of the presented information (information-centered taxonomies) [25] or the interaction between user and object (interaction-centered taxonomies) [26].

However, the applicability and usefulness of these classifications depends strongly on the use case. In the field of industrial assembly or machine operation, for example, the consideration of task characteristics is of enormous importance, which can hardly be found in any of the existing taxonomies [27]. The intransparency and overlapping of content within existing taxonomies contribute significantly to the fact that MR technologies are not yet part of the everyday tools used in the industry [14]. There is little research on when which technology (e.g. AR vs VR) is more suitable in an industrial context [6]. This impedes their implementation. Thus, companies are faced with the challenge of choosing the right methodology and the right technological tool for the respective task.

This paper aims to bridge the gap between theory-based taxonomies and practical use cases. It will answer the question which MR feature is most suitable for a specific assembly related task. Section 2 presents an industrial use case for a training process in production and defines the characteristics of the task. Based on this, the required features and specifications for the design of the based teach-in procedure are derived. Section 3 describes the study design in which two different technologies for the training process are compared. The results are presented in Sect. 4 and discussed afterwards.

2 Use Case Description

2.1 Assembly of a Pneumatic Cylinder

The Use Case describes the manual assembly of a pneumatic cylinder before its further processing in a partially automated machine. The whole assembly process consists of eleven steps. For each step it is specified whether the right or left hand shall be used. The steps are described in the following: First, the piston rod should be grasped from the colored left box and inserted into the workpiece fixture. Secondly, the piston buffer must be removed and mounted on the piston rod. Then, a piston has to be grasped from the upper left box and mounted on the piston rod. The fourth step includes taking the rod guide ring from the box. Subsequently, the ring magnet has to be grasped and a new ring magnet should be ordered via button. The ring magnet must be placed on the piston rod with the marked red side up and the rod guide ring must be mounted on the ring magnet. The mounting bush and sealing ring should then be placed on the piston rod. Then, the piston from the right box has to be mounted on the piston rod. The buffer piston must then be screwed onto the piston rod. The assembled piston rod should be inserted into the workpiece carrier and turned until it clicks. At the last step, the cylinder tube should be placed on the workpiece carrier, which will be sent to the respective station by pressing the start button.

The learning process for these steps in a face-to-face training that takes 40–70 min depending on the previous knowledge of the worker. The cycle time of the machine must then be reduced for several hours at a time.

An MR prototype has already been tested in a preliminary study. A 3D visualization of the plant was constructed and the sequence of individual work steps was visualized using Microsoft HoloLens. An overview on the virtual plant and the workspace can be seen in Fig. 2.

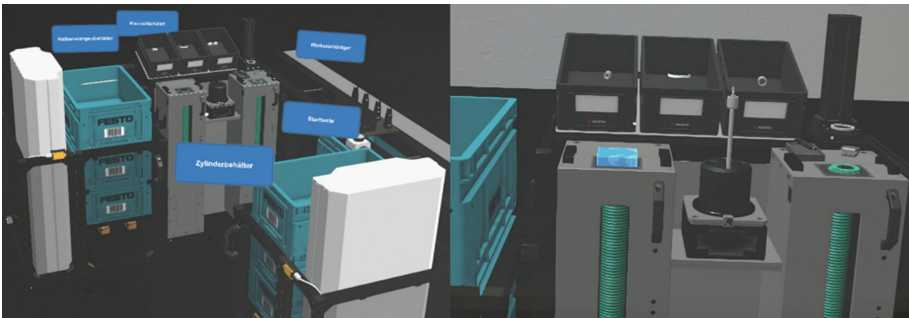


Fig. 2. 3D visualization of the plant

The machine was projected into real space in order to train the workers before they work on the real system. The first heuristic usability evaluation of the experts was positive. However, some important remarks concerned missing possibilities for interaction with the individual objects, as well as a possible irritation caused by an overlap

of the virtual objects with the real environment in which the training was conducted. Based on the preliminary study, the required functionalities for the present use case are derived in the following section, which will define the test scenarios.

2.2 Specifications of MR Features

Based on theory-based taxonomies [20, 24–26], different characteristics should be taken into account when designing an MR-based training method. In Table 1, the features of the technologies are specified and assigned to different classifications, which are applied to the use case in the following. The first three categories (task, user, information) play an essential role in designing the training environment. The last two categories (interaction, technology) additionally contribute to the choice of the medium.

Table 1. Specification of general MR-features

Classification	Specification
Task-centered [12, 14, 27]	Procedural vs. declarative knowledge Mainly cognitive vs. physical activity
User-centered [24]	Demographic variables and prerequisites of user with regard to <ul style="list-style-type: none"> • the task (<i>Previous experiences</i>) • the technology (<i>Affinity for technology, Control beliefs while dealing with technology, Use of technology</i>)
Information-centered [25]	Type and dimensionality of the objects <ul style="list-style-type: none"> • Text-based • Images • 2D or 3D visualization
Interaction-centered [26]	Input, control, and feedback <ul style="list-style-type: none"> • Voice/Audio • Visual • Haptic
Technology-centered [23]	<ul style="list-style-type: none"> • Extension of World Knowledge (EWK) • Reproduction Fidelity (RF) • Extent of Presence Metaphor (EPM)

The task of assembling a pneumatic cylinder requires the acquisition of procedural knowledge in order to complete the work steps in the correct order. Dede and colleagues report that VR, for example, can be very effective for learning procedural tasks in which students learn a sequence of steps to complete a task that requires maneuvers in three-dimensional space [12]. With regard to the user, demographic variables such as age, gender, and educational background could play a role with regard to the success of the MR-based training procedure. Even more important, however, is the users' previous knowledge of the task. For example, inexperienced workers may need more or different instructions than experienced workers [28]. The presented use case is primarily aimed at inexperienced temporary workers who have to practice the assembly

without any prior knowledge. The didactic structure of the training should take this into account by offering a modular structure, where steps can be repeated and played in different paces [7]. Furthermore, affinity and enthusiasm for technology may have an influence on the assessment of the system [29].

The objects of the virtual assembly should be displayed as realistically as possible to allow realistic orientation in the training environment. Therefore, true-to-scale 3D representations of the objects were used. Important information was additionally visualized in short, clearly arranged text fields.

In a heuristic evaluation with the MR prototype [30], the lack of possibilities for interaction was particularly criticized. Therefore, the comparison of two different media with different possibilities for interaction should contribute to evaluate the importance of e.g. haptic control. The first medium, which was also used in the preliminary study, is the Microsoft HoloLens. Here, there are possibilities for voice input as well as the possibility to use a clicker for the selection of different steps for interaction. However, the control is primarily visual, by directing the gaze at an object or field and then confirming this by voice or clicker. Since the scenario is projected into a real environment, it can be classified in the Reality-Virtuality spectrum in the middle range between AR and AV.

In contrast, the second medium is intended to offer possibilities for haptic interaction. Therefore, the same training application was implemented in a VR environment, which is related to the right end of the continuum. In the VR environment, it is possible to select steps or to move objects yourself with two controllers. Since the use case requires differentiating between using right and left hand, a feedback function for selecting objects with the right or left controller is built in.

The technical classifications are based on the preliminary work of Milgram and Kishino, to whose publication reference is made for more detailed information [23]. The first medium (AR/AV) uses a partially modelled world in the EWK dimension by virtually projecting the entire machine into the real world. The second medium (ar mixed reality glasses), on the other hand, uses a 360° view of the real environment in which the virtually modelled machine is installed. In terms of RF, both media use 3D animations with high fidelity. The extent of presence in the EPM dimension is relatively well established using Head Mounted Displays (HMDs), although it can be assumed that the VR environment allows an even higher degree of presence.

According to the feature specification with regard to the use case of the cylinder assembly, it becomes evident that the media particularly differ in the interaction possibilities with the work components. The following section presents the deduced hypotheses, from which the study design will be derived later on.

2.3 Hypothesis

The hypotheses to be examined in this study relate on the one hand to the perception and assessment of the respective medium (AR/AV vs. VR). At this point, the evaluation of usability and ergonomics is differentiated. On the other hand, hypotheses on the perception and evaluation of the training process using MR are formulated and tested in the further course of the study. Since there is no evidence for the suitability of

one of the media compared to the other for such an application, the hypotheses are formulated and tested in a non-directional way (Table 2).

Table 2. Hypotheses

H ₁	There is a correlation between technology affinity and the assessment of the perceived system usability
H _{2a}	The system usability of the two technologies AR/AV and VR differs significantly
H _{2b}	The evaluation of ergonomics of the two media AR/AV and VR differs significantly
H ₃	The used media AR/AV and VR differ significantly in their evaluation of the training process

3 Method

3.1 Study Design

The study design consists of an experimental within subject procedure with pre- and post test. This ensures that each participant could test both AR/AV and VR training application. At first, the participants were asked to fill out a questionnaire before they were assigned to a test condition. The order of the test conditions was randomly assigned. Test condition A included the following procedure: (1) Fill out a pre-test questionnaire, (2a) conduct training with AR, (2b) fill out a post-test questionnaire on the use of AR, (3a) conduct training with VR, (3b) fill out a post-test questionnaire on the use of VR, (4) answer the question of which medium would be the best choice. In test condition B the same procedure was chosen, but at first the VR training and subsequently the AR training was performed.

The same scenario was run through with all participants. The training took place in a room outside the production hall in order to avoid disturbing influences. After a brief introduction to the use of the technology, the participants had time to inspect the virtual assembly cell. After the orientation phase, the participants were able to watch a guided exercise scenario. The participants did not have to become active themselves, but had to use the clicker to proceed with the next step. Afterwards, they were able to view the entire assembly process again in animated mode - this was possible either in real time or in a slower version. The last step included the independent execution of the assembly steps. In the AR/AV version, the components had to be selected via eye control and clicker - the animation of those components was then carried out automatically if the selection was correct. In the VR version, the participants had the task of gripping the components themselves with the controllers (a distinction was made here between gripping with the right and left hand) and placing them in the correct position. The duration of use of the devices varied between 8 and 25 min.

The description of the questionnaires used in pre- and post-tests is presented below. Subsequently, an overview of the participants of the study is given.

3.2 Questionnaires

Pre-test. In a pre-test, the participants were asked about their demographic data (age, gender, educational level) and their current job position. In order to investigate the participants' previous knowledge of the task, they were asked whether they had ever carried out pneumatic cylinder assembly on the machine before. In addition, the participants were asked whether they had ever participated in a training procedure using AR/VR.

Subsequently, the participants were asked about their technical affinity with five items (e.g., "My enthusiasm for technology is...") on a six-level scale from "very low" to "very high". To complete the data on the participants, they were also asked which media (e.g. laptop, smartphone, tablets) are available to them, how often they use them and how easy-to-use the respective medium is. In addition, we used the "locus of control" questionnaire (KUT) to assess general control beliefs while dealing with technology [31]. With its 8 items (e.g., "I feel so helpless when dealing with technical devices that I do not even touch them") on a scale of six levels from "not at all" to "absolutely", the German questionnaire achieves a reliability of Cronbach's $\alpha = .85$.

Post-test. The post-test was divided into two parts for each test condition. First, questions were answered to evaluate the training medium. First, the System Usability Scale (SUS; [32]) was used to evaluate the respective technology. A SUS score between 60–80 means that the system is marginally acceptable, scores above 80 indicate good to very good system usability and 100 points indicate an excellent rating system that fully meets the users' expectations. In order to ensure that the answer scale of the entire questionnaire is consistent, a six level scale "do not agree at all" to "fully agree" was used in this study. Accordingly, an adjusted factor was used in the calculation of the overall score in order to ensure the comparability and significance of the SUS score. This scale reached a Cronbach's alpha of $\alpha = .80$ in this study.

In order to be able to cover not only system usability but also ergonomic issues, six self-created items were used (e.g., "The field of vision of the glasses restricted me.") and assessed on the same scale. This scale reached an $\alpha = .68$ and thus has to be treated with caution. Furthermore, the questionnaire contained two open questions: "What difficulties did you encounter in the learning process with the training medium?" and "What functionalities (e.g. speech/gestural control) did you find particularly helpful?"

The second part included the evaluation of the training process. The Nasa Task Load Index (NASA TLX) was used to measure the perceived load during the task. It measures the subjectively perceived demand with a multidimensional scale that distinguishes, for example, between physical and psychological stress [33]. The German summary contains six dimensions, namely mental, physical and temporal stress as well as performance, effort and frustration. The original scale has 20 gradations from "very low" to "very high". Adapted to the German version, we have used a 10-step scale with the poles "little" and "much".

In order to capture the evaluation of the training process in depth, two open questions were asked: "To what extent do you feel prepared for the task by using the system?" and "To what extent would you prefer the system used to other training

methods (please explain)? After both test conditions were carried out, a final question was asked: “Which of the two tested training media would you prefer? Please give reasons for your answer”.

3.3 Participants

A total of nine men and seven women participated in the study ($N = 16$). Four participants were under 20, six of the participants were between 21–25 years old, one participant between 26–30 years old, four participants were between 31–40 years old and one participant was between 41–45 years old. The sample consisted of six company employees, two temporary workers, and the remaining eight were trainees, apprentices or managers. Three participants stated that they had intermediate level of secondary education, three had technical college entrance qualification. Six participants stated that their highest educational level was high school and/or university entrance qualification, the remaining eight already had an academic degree (nBachelor = 2, nMaster = 1, nPhD = 1).

Two participants stated that they already had experience with MR-based training methods. One of the participants had already conducted the MR-based training for the respective assembly, three other participants already knew the assembly, but based on manual training.

3.4 Analysis

The analysis of the collected data was conducted using SPSS. Since we are dealing with a small sample ($N = 16$), the inferential statistical evaluation is based on a non-parametric data level. For group comparisons, the Wilcoxon test for paired samples is used. In the following, an overview of descriptive statistics is given first. In order to comply with the SUS analysis requirements, the items on the system usability and ergonomics scales were adjusted from one to six to zero to five.

4 Results

An overview of the results of the pre-test is given below. First, descriptive statistics for the respective questionnaires in the pre- and post-test are reported. Subsequently the test results of the hypotheses are presented. In the framework of the pre-test, the use of technology, the handling of technology and the affinity of the sample towards technology are presented.

4.1 Pre-test Results

All study participants reported to have access to laptop/PC and smartphone for private or professional purposes. 14 participants work with tablets, while only three participants used Microsoft HoloLens or Oculus Rift/HTC Vive. The ease of use of these technologies was assessed on a scale from one “very difficult” to six “very easy”. The Microsoft HoloLens received the lowest value with a mean value of 4.50 (Min = 4,

Max = 5, SD = .58, n = 3), the smartphone received the highest mean value with $\bar{x} = 5.56$ (Min = 1, Max = 6, SD = 1.3, n = 16).

An overview of the participants' affinity for technology is given in Table 3. The locus of control while dealing with technology in the sample had a mean value of $\bar{x} = 4.44$ (min = 3.13, max = 5.38, SD = 0.74, n = 16).

Table 3. Descriptive statistics for technical affinity.

n = 16	Mean	SD	Min	Max
Enthusiasm for technology	4.75	1.24	2	6
Interest in technology	4.69	1.30	2	6
Technological understanding	4.19	1.17	2	6
Distrust of technology	2.63	0.62	2	4
Technical skills	4.38	1.09	2	6

Hypothesis 1: There is a correlation between technology affinity and the assessment of the perceived system usability

A Spearman's correlation was run to determine the relationship between technical affinity values and SUS Index for AR/AV and VR. No correlation was found, neither for the AR/AV SUS Index ($= -.12$, $p = .67$, $n = 16$), nor for VR's system usability ($= .25$, $p = .35$, $n = 16$).

4.2 Post-test Results

The results of the post-test include an overview of the descriptive statistics for the evaluation of the media and the respective learning process. In each case, the results of the hypothesis tests are presented afterwards.

Hypothesis 2a: The usability of the two media AR/AV and VR differs significantly

Descriptive statistics (mean value, standard deviation, minimum and maximum) of the System Usability Scale are presented in Table 4. The six level scale was adjusted to zero = "do not agree at all" to five = "fully agree". The calculation of the SUS score shows that the AR/AV system has an overall score of 63.63. The SUS score for the VR system is 72.75.

The Wilcoxon test for paired samples shows a significant difference with a median of 35.5 for AR/VR and a median of 37.5 for VR ($z = -2.047$, $p = 0.041$, $n = 16$). Hypothesis 1a can therefore be confirmed.

Hypothesis 2b: The evaluation of ergonomics of the two media AR/AV and VR differs significantly

Descriptive statistics for the six-item scale for ergonomics is shown in Table 5. According to the Wilcoxon test for ergonomics of AR/VR (median = 23.5) and VR (median = 23.5) with $z = -.596$ and $p = .551$, it shows that the central tendencies of the respective test conditions are not different. Hypothesis H1b must therefore be rejected.

Table 4. Descriptive statistics for system usability. (The following items have been translated by the author into English for better comprehensibility.)

n = 16	AR/AV				VR			
	Mean	SD	Min	Max	Mean	SD	Min	Max
I think that I would like to use this system frequently	2.81	1.33	1	5	3.81	0.83	2	5
I found the system unnecessarily complex. (rec)	3.69	0.95	2	5	4.00	0.63	3	5
I thought the system was easy to use	3.13	1.09	1	5	3.81	0.75	2	5
I think that I would need the support of a technical person to be able to use this system. (rec)	3.06	1.18	1	5	3.19	1.11	1	5
I found the various functions in this system were well integrated	3.00	0.89	1	4	3.63	0.50	3	4
I thought there was too much inconsistency in this system. (rec)	2.81	1.22	1	4	3.88	0.81	2	5
I would imagine that most people would learn to use this system very quickly	3.25	1.00	1	5	3.31	0.95	1	5
I found the system very cumbersome to use. (rec)	3.31	1.08	1	5	4.00	0.63	3	5
I felt very confident using the system	3.00	1.21	1	5	3.25	0.86	2	4
I needed to learn a lot of things before I could get going with this system. (rec)	3.75	1.06	1	5	3.50	1.26	1	5

Difficulties Using the Media. Within the framework of the open questions, the participants reported the following difficulties in the training process with the medium AR/AV: The glasses were very heavy and/or unsuitable for wearer of glasses, limited field of vision (5), lack of integration of the hands (2), not realistic, difficulties with orientation (Where does the next step take place?) (4), difficulties with clicker operation.

When using VR, the participants reported the following problems: The field of vision flickers/is blurred (2), dizziness (2), spatial boundaries are missing, it is difficult

Table 5. Descriptive statistics for ergonomics.

n = 16	AR/AV				VR			
	Mean	SD	Min	Max	Mean	SD	Min	Max
The virtual training made me feel uncomfortable. (rec)	4.38	1.20	1	5	3.88	0.96	2	5
Wearing the glasses was pleasant	2.75	1.24	1	5	3.31	0.87	2	5
Working with the glasses was exhausting for me. (rec)	3.69	1.14	1	5	3.63	0.89	2	5
Wearing the glasses had some after-effects on me. (rec)	4.25	0.93	2	5	4.25	0.86	2	5
The field of vision of the glasses restricted me. (rec)	1.94	1.44	0	5	3.38	1.41	0	5
Wearing glasses has exhausted my eyes. (rec)	4.06	1.00	2	5	3.44	1.26	0	5
Sum	3.67	0.72	2.00	4.50	3.81	0.71	2.00	5.17

to immerse oneself in the animation, lack of intuitiveness when operating the controllers (e.g. when grasping), sound feedback is sometimes confusing.

Helpful Functionalities. The following comments were noted as helpful features for AR/AV: Clear audio instructions through the steps as a supplement to the text, pleasant wearing comfort, real environment is still visible (2), free movement in space is possible, cursor control via view, as well as clear color highlighting of the components to be used. The following features were found to be helpful when using VR glasses: Controller (possibility to grasp and move the components) (7), differentiation between right and left hand (2), color highlighting of the components to be used (2), high level of realism by imitating the real environment, and comfortable wearing comfort.

Hypothesis 3: The used media AR/AV and VR differ significantly in their evaluation of the training process

The evaluation of the NASA Task Load Index shows that the perception of the test persons regarding their load during the task does not differ according to medium. Figure 3 shows the mean values of the individual scales for the media in comparison. The Wilcoxon test shows that hypothesis 2 must be rejected with $z = -0.910$ and $p = 0.363$.

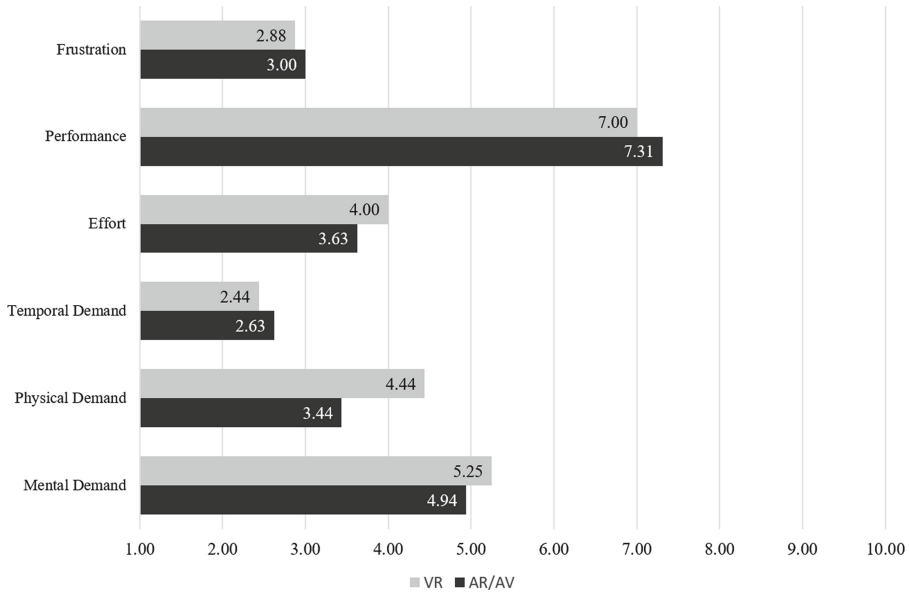


Fig. 3. NASA task load index for AR/AV and VR (n = 16).

Preparation for the Task by the System. When using the AR/AV medium, participants noted that they felt well prepared (6) because they knew the process. Some participants noted that they would repeat the exercise several times before the actual assembly (2). Others said that they felt less prepared because of the lack of orientation (3) and poor transferability to the real workplace. With the VR system, the majority of the test persons felt well to very well prepared for the upcoming task (4). One participant noticed that he felt as safe as if he had received a personal instruction to the system. Two participants also indicated that they would repeat the exercise in advance. One participant states that he feels relatively safe, but that he would not be able to answer adequately if he were asked questions.

Comparison of MR to Other Training Methods. Some participants indicate that AR/AV System is preferable to text-based process instructions (3). It is especially emphasized that the method is suitable for an initial familiarization with the process. Personal learning by another employee is still preferred by two participants. Due to the lack of haptics, the system is merely a supplement to, but not a replacement for, instructions on the physical machine. Participants positively noted that each employee can learn at his/her own pace, the training is carried out in a very standardized manner and no time or money loss is incurred due to cycle time overruns. One participant stated that he prefers a mixed form of both systems.

The VR system is described by the participants as very realistic. The participants state that it is particularly suitable for obtaining an overview and orientation of the workplace. It is preferred over text forms, but the advantage of VR over simple videos is questioned. However, when problems arise, the participants see difficulties.

Choice of Medium. The subjects rated the VR system significantly better and would choose to use it (14 for VR, 2 not quite clear, 0 for AR). The mentioned advantages of VR in this question refer to the fact that orientation is easier, the use of the hands helps to memorize procedures, the field of vision is larger and there is a higher wearing comfort. The test persons describe the experience as closer to reality.

A correlation matrix was created exploratively in order to identify the extent to which there are correlations between the decision for VR and the surveyed parameters system usability, ergonomics, and task load. Table 6 shows that there are significant point-biserial correlations between the decision for VR and system usability ($= .63$, $n = 16$, $p < 0.01$) as well as between VR and the evaluation of ergonomics ($= .65$, $n = 16$, $p < 0.01$). There seems to be no significant correlation between the task load and the decision for VR.

Table 6. Correlation matrix

n = 16	Decision_VR	Ergo_VR	SUS_VR	TLX_VR
Decision_VR		.63**	.65**	.46*
Ergo_VR			.61*	.54*
SUS_VR				.51*
TLX_VR				

**The correlation is significant (bilateral) at the level of 0.01.

*The correlation is significant (bilateral) at the level of 0.05.

5 Conclusion

5.1 Limitations

Before discussing the results of the present study, the limitations of the study are presented briefly. An essential aspect is the small size and representativeness of the sample. Especially in studies that take place outside the laboratory but directly in industry, it is difficult to acquire an adequate number of participants. The problem of the low validity of the data of such small number of participants was countered by a conservative approach to statistical analysis and additionally recorded qualitative statements. Nevertheless, any results based on such a small amount of data should be interpreted with caution. Furthermore, the composition of the sample does not fully reflect the characteristics of typical temporary workers. For this reason, we have made sure that the previous knowledge for the task should be as little as possible in order to deduce statements about inexperienced workers. Another problem is the use of a non-validated scale for the factor ergonomics. This scale should be revised and validated in larger surveys. Furthermore, the scale of system usability was evaluated in this study by a six-level item scale and not, as in the original, by a five-level scale. This could also have an influence on the reliability of the scale.

5.2 Discussion of Results

The results on the tested hypotheses are summarized in Table 7 and are discussed below, taking into account the qualitative results.

Table 7. Decision on hypotheses

Hypothesis	Statement	Decision
H ₁	There is a correlation between technology affinity and the assessment of the perceived system usability	Reject
H _{2a}	The usability of the two media AR/AV and VR differs significantly	Accept
H _{2b}	The evaluation of ergonomics of the two media AR/AV and VR differs significantly	Reject
H ₃	The used media AR/AV and VR differ significantly in their evaluation of the training process	Reject

The pre-test shows that the participants in the study had little or no overall experience with MR. Even though the sample has a relatively high affinity for technology and indicates that they feel confident in using technology, there seems to be no correlation with the subsequent evaluation of the systems' usability. The enthusiasm for technology seems to play a subordinate role in the training procedure for the presented use case. This leads to the assumption that people with less technical affinity could also get along well with the system - and vice versa.

The evaluation of the media was divided into the evaluation of system usability and ergonomics. Regarding system usability, both media had a rather lower score, which is described in the manual as only marginally acceptable. However, this could be due to the fact that the participants in the study were asked to closely examine both systems. The assumption that system usability scores differ can be confirmed. Data show that the VR system is rated significantly better.

Statistically speaking, there are no differences between the systems in the scale of ergonomics. This result is surprising, as the written comments mainly mention comments on ergonomically related difficulties (e.g. the restricted field of vision). This indicates that the items of the scale should be revised. The written comments on the media show that all participants prefer clear guidance through the process steps (e.g. by highlighting the objects in color or using cue arrows). When independently carrying out the assembly steps, the use of controllers in the VR system was found to be helpful. The high degree of reality provided by the 360° photo of the assembly hall is perceived as pleasant. However, some participants stated that they found the simultaneous perception of reality in AR/AV more pleasant and safer.

With regard to the evaluation of the training process through the NASA TLX, the descriptive statistics show only a small recognizable difference in the area of physical demand, which can be explained by the increased movement and interaction in VR during assembly. Overall, the load during the assembly does not seem to differ between

the media. At this point it should be questioned whether the scale used is sufficiently valid for the criteria to be measured.

In the written comments it becomes evident that the participants feel well prepared for the realistic assembly through both media. Most of them note that they want to repeat the exercise before the assembly is carried out on the real system. Compared to other (e.g. text-based) instructions, the participants regard the MR procedures as superior. However, a few number (two) of the participants prefer the possibility to be trained by another person or that MR is combined with training in a real setup. Accordingly, it should be ensured that the test persons are provided with a supervisor who can answer questions if needed. The high standardization and the possibility of individual learning speed was perceived positively.

The final evaluation of the media clearly shows that the participants prefer the VR system. The main reasons are the good orientation in space, the natural field of view and the use of the controllers to interact with the work pieces. The explorative correlation matrix shows that there are correlations between the decision for VR and the system usability of VR and it's evaluation of ergonomics. Although this does not allow any conclusions to be drawn about a cause-and-effect relationship, it nevertheless confirms the importance of placing a strong focus on usability in the development of such systems and of involving users early in the development process.

5.3 Conclusion and Outlook

The findings allow drawing conclusions about the specification and importance of different MR technology features. The underlying features seem to be more important than the clear separation of the devices. Thus, for a task that requires interaction of the user with a real or virtual object, there should be appropriate possibilities for interaction in the system. This can be realized in AR/AV technologies as well as in VR. The following table (Table 8) summarizes the collected findings in preliminary guidelines for classifications, which should be reviewed and extended in further studies.

Table 8. Guideline for developing MR training systems

Classification	Guideline
Task-centered	<ul style="list-style-type: none"> • Task with clear focus on procedural knowledge and physical activity
User-centered	<ul style="list-style-type: none"> • Conducting an introduction to the technology and its control • Modular design of the training contents in order to take into account different levels of prior knowledge • In the case of videos or automatically running animations, possibilities for reducing/increasing the speed should be offered
Information-centered	<ul style="list-style-type: none"> • Color highlighting of the objects to be focused • Use only short texts or notes and provide the possibility for audio instructions • Clear instructions (e.g. arrows) to guide attention to the next step

(continued)

Table 8. (continued)

Classification	Guideline
Interaction-centered	<ul style="list-style-type: none"> • Parts of the task that require specific physical activities, in order to learn motor processes, should be designed as realistically as possible (e.g. gripping, moving objects) • Feedback about correct or wrong steps should include an explanation (e.g., of what has to be changed)
Technology-centered	<ul style="list-style-type: none"> • The more important the context is for the completion of a task, the better it should be modelled (VR) or the task should be performed directly at the machine (e.g. with instructions from AR) • HMDs are particularly suitable for tasks in which the hands are to be used. Other tasks can also be taught using hand-held devices (e.g. smartphones, tablets) • A natural or wide field of view is essential for the presence experience in the virtual or augmented world

The listed points result mainly from the qualitative results of the participants' written comments. Thus, it is evident that the collection of qualitative data is essential when evaluating new systems. In order to be able to integrate the empirical results of the quantitative survey, a further study should be conducted with a larger sample - if necessary also in a laboratory setting - and compared with the available findings. Furthermore, possible relevant user factors (such as openness to new experiences or the current mood) should be surveyed and examined for correlations with the evaluation of the systems.

Acknowledgement. This work is part of the project “ELLI 2 - Excellent Teaching and Learning in Engineering Sciences” and was funded by the Federal Ministry of Education and Research (BMBF), Germany.

References

1. Benešová, A., Hirman, M., Steiner, F., Tupa, J.: Analysis of education requirements for electronics manufacturing within concept industry 4.0. In: 2018 41st International Spring Seminar on Electronics Technology (ISSE), pp. 1–5. IEEE, Piscataway (2018)
2. Chang, M.M.L., Ong, S.K., Nee, A.Y.C.: Approaches and challenges in product disassembly planning for sustainability. *Procedia CIRP* **60**, 506–511 (2017)
3. Kerin, M., Pham, D.T.: A review of emerging industry 4.0 technologies in remanufacturing. *J. Cleaner Prod.* **237**, 117805 (2019)
4. Konings, J., Vanormelingen, S.: The impact of training on productivity and wages: firm-level evidence. *Rev. Econ. Stat.* **97**(2), 485–497 (2015)
5. Kim, M., Park, K.-B., Choi, S.H., Lee, J.Y., Kim, D.Y.: AR/VR-based live manual for user-centric smart factory services. In: Moon, I., Lee, Gyu M., Park, J., Kiritsis, D., von Cieminski, G. (eds.) *APMS 2018. IAICT*, vol. 536, pp. 417–421. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-99707-0_52

6. Velosa, J.D., Cobo, L., Castillo, F., Castillo, C.: Methodological proposal for use of virtual reality VR and augmented reality AR in the formation of professional skills in industrial maintenance and industrial safety. In: Auer, M.E., Zutin, D.G. (eds.) *Online Engineering & Internet of Things*. LNNS, vol. 22, pp. 987–1000. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-64352-6_92
7. Werrlich, S., Daniel, A., Ginger, A., Nguyen, P.A., Notni, G.: Comparing HMD-based and paper-based training. In: 2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 134–142. IEEE (2018)
8. Andaluz, V.H., et al.: Multi-user industrial training and education environment. In: De Paolis, L.T., Bourdot, P. (eds.) *AVR 2018*. LNCS, vol. 10851, pp. 533–546. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-95282-6_38
9. Ghandi, S., Masehian, E.: Review and taxonomies of assembly and disassembly path planning problems and approaches. *Comput. Aided Des.* **67–68**, 58–86 (2015)
10. Choi, S., Jung, K., Noh, S.D.: Virtual reality applications in manufacturing industries: past research, present findings, and future directions. *Concurr. Eng.* **23**(1), 40–63 (2015)
11. Daling, L., Abdelrazeq, A., Sauerborn, C., Hees, F.: A comparative study of augmented reality assistant tools in assembly. In: Ahram, T., Falcão, C. (eds.) *AHFE 2019*. AISC, vol. 972, pp. 755–767. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-19135-1_74
12. Dede, C.J., Jacobson, J., Richards, J.: Introduction: virtual, augmented, and mixed realities in education. In: Liu, D., Dede, C., Huang, R., et al. (eds.) *Virtual, Augmented, and Mixed Realities in Education*, pp. 1–19. Springer, Singapore (2017). https://doi.org/10.1007/978-981-10-5490-7_1
13. Müller, B.C., et al.: Motion tracking applied in assembly for worker training in different locations. *Procedia CIRP* **48**, 460–465 (2016)
14. Guo, Q.: Learning in a mixed reality system in the context of Industrie 4.0. *J. Tech. Educ.* **3**(2), 92–115 (2015)
15. Milgram, P., Colquhoun, H.: A taxonomy of real and virtual world display integration. In: Ohta, Y., Tamura, H. (eds.) *Mixed reality: Merging real and virtual worlds*, pp. 5–30. Springer, Berlin (1999). https://doi.org/10.1007/978-3-319-08234-9_205-1
16. Azuma, R., Bailiot, Y., Behringer, R., Feiner, S., Julier, S., MacIntyre, B.: Recent advances in augmented reality. *IEEE Comput. Graph. Appl.* **21**(6), 34–47 (2001)
17. Chicaiza, E.A., De la Cruz, E.L., Andaluz, V.H.: Augmented reality system for training and assistance in the management of industrial equipment and instruments. In: Bebis, G., et al. (eds.) *ISVC 2018*. LNCS, vol. 11241, pp. 675–686. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-03801-4_59
18. Regenbrecht, H., et al.: An augmented virtuality approach to 3D videoconferencing. In: *Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality, ISMAR 2003*, pp. 290–291. IEEE, Washington (2003)
19. Ternier, S., Klemke, R., Kalz, M., Van Ulzen, P., Specht, M.: AR learn: augmented reality meets augmented virtuality. *J. Univers. Comput. Sci. Technol. Learn. Phys. Virt. Spaces* **18**(15), 2143–2164 (2012)
20. Normand, J.M., Servières, M., Moreau, G.: A new typology of augmented reality applications. In: *Proceedings of the 3rd Augmented Human International Conference, AH 2012*, pp. 1–8. ACM, New York (2012)
21. Paiva Guimarães, M., Dias, D.R.C., Mota, J.H., Gnecco, B.B., Durelli, V.H.S., Trevelin, L. C.: Immersive and interactive virtual reality applications based on 3D web browsers. *Multimedia Tools Appl.* **77**(1), 347–361 (2018)

22. Porrás, A.P., Solís, C.R., Andaluz, V.H., Sánchez, J.S., Naranjo, C.A.: Virtual training system for an industrial pasteurization process. In: De Paolis, L.T., Bourdot, P. (eds.) AVR 2019. LNCS, vol. 11614, pp. 430–441. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-25999-0_35
23. Milgram, P., Kishino, F.: A taxonomy of mixed reality visual displays. *IEICE Trans. Inf. Syst.* **E77-D**(12), 1321–1329 (1994)
24. Lindemann, R.W., Noma, H.: A classification scheme for multi-sensory augmented reality. In: Proceedings of the 2007 ACM Symposium on Virtual Reality Software and Technology, pp. 175–178. ACM, New York (2007)
25. Tönnies, M., Plecher, D.A.: Presentation Principles in Augmented Reality - Classification and Categorization Guidelines Version 1.0. Technical Report (TUM-II111). Technische Universität München (2011)
26. Mackay, W.E.: Augmented reality: linking real and virtual worlds - a new paradigm for interacting with computers. In: Proceedings of AVI 1998, ACM Conference on Advanced Visual Interfaces AVI, pp. 13–21. ACM, New York (2000)
27. Ortiz, J.S., et al.: Teaching-Learning process through VR applied to automotive engineering. In: Proceedings of the 2017 9th International Conference on Education Technology and Computers, ICETC 2017, pp. 36–40. ACM, New York (2017)
28. Radkowski, R., Herrema, J., Oliver, J.: Augmented reality-based manual assembly support with visual features for different degrees of difficulty. *Int. J. Hum.-Comput. Inter.* **31**(5), 337–349 (2015)
29. Karrer, K., Glaser, C., Clemens, C., Bruder, C.: Technikaffinität erfassen—der Fragebogen TA-EG. In: Lichtenstein, A., Stößel, C., Clemens, C. (eds.) *Der Mensch im Mittelpunkt technischer Systeme*. 8. Berliner Werkstatt Mensch-Maschine-Systeme, ZMMS Spektrum, vol. 22, no. 29, pp. 196–201. VDI, Düsseldorf (2009)
30. Nielsen, J.: Enhancing the explanatory power of usability heuristics. In: Adelson, B., Dumais, S., Olson, J. (eds.) Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 1994, pp. 152–158. ACM, New York (1994)
31. Beier, G.: Kontrollüberzeugungen im Umgang mit Technik. *Rep. Psychol.* **24**(9), 684–693 (1999)
32. Brooke, J.: SUS-A quick and dirty usability scale. In: Jordan, P.W., Thomas, B., Weerdmeester, B.A., McLelland, I.L. (eds.) *Usability evaluation in industry*, pp. 189–194. Taylor and Francis, London (1996)
33. Hart, S.G.: NASA-task load index (NASA-TLX): 20 Years Later. In: Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting, pp. 904–908. HFES, Santa Monica (2006)

Publication 3:

Daling, L. M., & Schlittmeier, S. J. (2024). Effects of augmented reality-, virtual reality-, and mixed reality-based training on objective performance measures and subjective evaluations in manual assembly tasks: a scoping review. *Human Factors*, 66(2), 589-626. <https://doi.org/10.1177/00187208221105135>

Reproduced with permission from SAGE.

Effects of Augmented Reality-, Virtual Reality-, and Mixed Reality–Based Training on Objective Performance Measures and Subjective Evaluations in Manual Assembly Tasks: A Scoping Review

Lea M. Daling  and Sabine J. Schlittmeier , RWTH Aachen University, Germany

Objective: The present scoping review aims to transform the diverse field of research on the effects of mixed reality-based training on performance in manual assembly tasks into comprehensive statements about industrial needs for and effects of mixed reality-based training.

Background: Technologies such as augmented and virtual reality, referred to as mixed reality, are seen as promising media for training manual assembly tasks. Nevertheless, current literature shows partly contradictory results, which is due to the diversity of the hardware used, manual assembly tasks as well as methodological approaches to investigate the effects of mixed reality-based training.

Method: Following the methodological approach of a scoping review, we selected 24 articles according to predefined criteria and analyzed them concerning five key aspects: (1) the needs in the industry for mixed reality-based training, (2) the actual use and classification of mixed reality technologies, (3) defined measures for evaluating the outcomes of mixed reality-based training, (4) findings on objectively measured performance and subjective evaluations, as well as (5) identified research gaps.

Results: Regarding the improvement of performance and effectiveness through mixed reality-based training, promising results were found particularly for augmented reality-based training, while virtual reality-based training is mostly—but not consistently—as good as traditional training.

Application: Mixed reality-based training is still not consistently better, but mostly at least as good as traditional training. However, depending on the use case and technology used, the training outcomes in terms of assembly performance and subjective evaluations show promising results of mixed reality-based training.

Keywords: training evaluation, virtual environments, transfer of training, manual materials handling, analysis and evaluation, human performance modeling, immersive environments

Address correspondence to Lea M. Daling, Information Management in Mechanical Engineering, RWTH Aachen University, Dennewartstr. 27, Aachen 52068, Germany; e-mail: lea.daling@ima.rwth-aachen.de

HUMAN FACTORS

2024, Vol. 66(2) 589–626

DOI:10.1177/00187208221105135

Article reuse guidelines: sagepub.com/journals-permissions

Copyright © 2022, Human Factors and Ergonomics Society.

BETWEEN POTENTIALS AND CHALLENGES: MIXED REALITY-BASED TRAINING FOR MANUAL ASSEMBLY TASKS

Augmented (AR) and virtual (VR)-based training is increasingly used in industrial application areas such as product design, production and manufacturing, and health applications (Berg & Vance, 2017; Blattgerste et al., 2017; Horigome et al., 2020). Despite the popularity of AR or VR, also referred to as mixed reality (MR) technologies (Skarbez et al., 2021), the evidence on their effectiveness and efficiency as a training medium is extremely diverse. This holds particularly true for the training of manual assembly tasks in production, although this is considered a highly relevant area for MR-based training (Werrlich, Lorber, et al., 2018). Here, MR-based training is increasingly used to train employees inexperienced in certain assembly tasks, which vary more frequently and more quickly due to the increasing variety and complexity of products compared to the long prevailing mass or flow production (AlGeddawy & ElMaraghy, 2012). However, looking at the recently published empirical studies on the effects of training, it has quickly become apparent that this is—in many respects—a diverse field of research and that the results on the effects of augmented and virtual reality-based training have been very heterogeneous and partly contradictory. We present a scoping review in order to transform the diverse studies on the effects of MR-based training on performance in manual assembly tasks into comprehensive statements about industrial needs, training outcome measures, effects of training on objectively measured performance, users' subjective evaluations, and research gaps in the field.



Preparatory, we will first look at what manual assembly tasks are and what characterizes them. As a next step, the term of mixed reality (MR) is presented as an overarching construct to subsume the different instances of AR and VR technologies. Then, an introduction into the research field of training for manual assembly tasks provides insight into the need and the goal of the current scoping review.

Manual Assembly Tasks

Manual assembly tasks represent a significant part of activities in the production environment, despite the trend toward automation of manual tasks (Kothiyal & Kayis, 1995). Manual assembly activities can be described as the entirety of all operations for the assembly of objects with geometrically determined shape, such as joining (e.g., screwing, nailing, welding, gluing, soldering, clamping), handling (e.g., grab, place, turn over, move, secure), inspecting (e.g., check, control, measure), and adjusting (e.g., setting or auxiliary operations such as caulking, deburring, and (un)packing) (Lotter, 2006). Thus, an assembly process is characterized by activities ranging from providing the material to be assembled to inspection and packaging, whereby the complexity (e.g., number of steps or assembly parts and the difficulty of assembly) of the assembly process of part or whole products can vary greatly (Lotter, 2006).

According to an analysis by Yuviler-Gavish, Krupenia, and Gopher (2013) of what skills underlie assembly tasks, it appeared that while both sensorimotor and cognitive skills are involved, the most important skill required for manual assembly tasks involves procedural skills. According to Koziol and Budding (2012), procedural learning does not only refer to the acquisition of cognitive skills but primarily to motor skills and habits. Specifically, in manual assembly tasks, workers have to be taught on the one hand what actions to perform, in what order, and with which method (cognitive skills). On the other hand, assembly tasks usually consist of a complex sequence of steps and require knowledge of specific procedures and techniques (procedural skills) (Gavish et al., 2015). In contrast to solely cognitive tasks and factual

information (e.g., to select, sequence, and use the correct assembly parts), which can be explicitly retrieved, procedural tasks (e.g., routine sequences of actions such as grasping, turning, or pressing) usually need to be repeated and trained several times before the learning outcome is demonstrated through improved task performance.

One potential goal of industry and production is to keep the work cycle time of assembly tasks as short as possible while at the same time ensuring the quality of products assembled (Wang et al., 2009). Nevertheless, important factors have to be considered in the evaluation of this goal, such as the weight of the parts and complexity of the product, the individual work capacity, health and safety of the workers, and appropriate training and qualification (Kothiyal & Kayis, 1995).

Until recently, training for manual assembly tasks have been primarily conveyed by means of paper-, video-, or trainer-based formats. These methods have been established over years and are still widely used today. However, they show significant shortcomings concerning time-efficiency (e.g., plant managers or learning mentors need to take time to train novices which results in a high expenditure of personnel resources), location independence (e.g., when training is carried out on-site on running machines in the production line), or individual learning requirements (e.g., paper manuals follow the one-size-fits-all approach and have not been adapted to possible prior knowledge) (Hou et al., 2013).

Due to the increasing variety and complexity of products, there has been growing interest in the industry for flexible, effective (i.e., selecting a suitable and successful training method to enhance training outcomes) and efficient (i.e., saving time and financial resources) training methods to provide workers quickly, safely, and reliably with the necessary cognitive and procedural skills (Doolani, Wessels, et al., 2020; Gavish et al., 2011). For this reason, digital training methods such as AR- or VR-based training are increasingly used to train employees inexperienced in a certain assembly task (Guo, 2015; Wang et al., 2016). However, a look into this highly interdisciplinary field of research

quickly reveals that there are very heterogeneous and partly inconsistent results on the impact of AR- and VR-based training on training outcomes (Gavish et al., 2015; Loch et al., 2019; Werrlich, Nguyen et al., 2018). In the following, a characterization of AR and VR technologies as instances of the broader term MR is provided in order to classify and understand different findings on the effectiveness of these training means.

MR-Based Training for Manual Assembly Tasks

AR and VR training systems are designed to prepare workers safely and efficiently for a task without, for example, downgrading the cycle times of machines or causing occupational safety risks through mistakes (Sautter & Daling, 2021). While AR-based training is mostly used to display virtual objects into the real world, for example, by overlaying virtual objects or instructions onto the workspace, training in VR environments enables the user to interact in a computer-generated 3D environment (e.g., using a virtual tool to assemble components), while the real world is (partially) hidden (Milgram & Kishino, 1994). In order to research the effectiveness of these training methods, it is essential to understand to what extent AR- or VR-based training is even comparable.

In an early definition, Milgram and Kishino (1994) grouped AR and less-commonly used augmented virtuality (AV) technologies under the term MR. VR, on the other hand, was excluded from this term as a component of virtual environments (Milgram & Kishino, 1994). Since then, there has been much discussion around the development of taxonomies of AR and VR technologies (Lindeman & Noma, 2007; Mackay, 2000; Normand et al., 2012). In this review, we follow the most recent consideration by Skarbez et al. (2021), who proposed MR as a broad umbrella term for both AR and VR and stated that “mixed reality is broader than previously believed, and, in fact, encompasses conventional virtual reality experiences” (p.1). In their definition, VR experiences are classified as external virtual environments, indicating that only users’ exteroceptive senses, that is, sight, hearing, touch, smell, and taste, are controlled

by the technology. A state in which both exteroceptive and interoceptive senses are stimulated by technology was excluded from the MR-term and has been described as so-called “Matrix-Like” Virtual Environments (Skarbez et al., 2021).

In this course, AR and VR are mainly differentiated according to the extent to which the system is, so to speak, aware of its real environment and can respond to changes in that environment (introduced as extent of world knowledge by Milgram & Kishino, 1994) and their degree of immersion. Immersion can be defined as objective parameters of a system that are, on the one hand, displays (in all sensory modalities) and, on the other hand, tracking capabilities that ensure high fidelity and lead to changes in users’ perception of the environment (Slater, 2004). Skarbez et al. (2021) stated that AR systems generally have low or medium immersion, but a higher level of world knowledge, while external virtual reality systems (i.e., VR) generally have high immersion, with little or no world knowledge. Together, they are classified as instances of MR, which is defined as an environment in which the physical (i.e., real) world and virtual objects and stimuli are presented together within a single percept (Skarbez et al., 2021). The joint consideration of AR and VR as part of MR enables the analysis of different training formats with special consideration of additional features, which are used for, for example, the interaction with tools and assembly parts through controller or haptic devices. Consequently, in this scoping review, we use the term MR for various technologies and hardware of AR to VR as shown in Figure 1.

Although MR as a training medium offers new visualization and learning possibilities, great skepticism has been prevailed in the industry for several years, primarily related to the return on the initial investment in terms of the cost of installing the system hardware and software (Gallagher et al., 2005). However, the costs of technologies are subject to constant change, which is why the benefit of using MR must be made clear independently. Despite the identified shortcomings of traditional training methods, it remains unclear to the industry why to use MR training and what potential outcomes

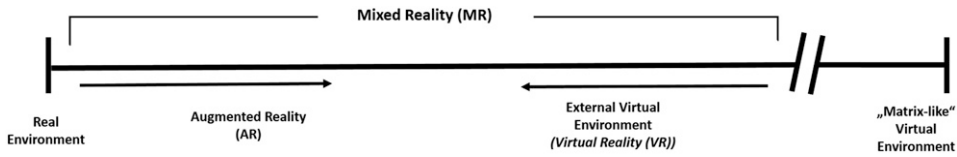


Figure 1. Revision of the Reality-Virtuality Continuum (based on Milgram & Kishino, 1994), where Mixed Reality Was Used as Umbrella Term for AR and VR technologies (based on Skarbez et al., 2021). Note that VR was Considered as Part of External Virtual Environments in Skarbez et al., 2021.

to expect. To increase industry confidence in MR-based training, ongoing research aims to demonstrate the benefits of MR-based training through various performance and accuracy measures.

Two recent reviews have revealed that how successfully MR-based training has been used to improve users' performance varied based on several factors, such as task type and population being trained (Doolani, Wessels, et al., 2020; Kaplan et al., 2021). In their meta-analysis, Kaplan and co-authors (2021) differentiated between cognitive, physical, and mixed tasks and found that MR (described here as XR) is a powerful training medium especially for physical tasks that involved some sort of bodily training, such as aerobic, dancing, or balance activities (Kaplan et al., 2021; Prasertsakul et al., 2018; Rose et al., 2000). Overall, no significant effects on performance measures have been found for cognitive tasks, such as remembering facts or information. In accordance with our previous definition, maintenance or manual assembly tasks have mainly been classified as mixed tasks, meaning that a combination of cognitive and physical tasks was required in training. For those tasks, the picture has been very ambiguous regarding the potential benefits of MR-based training on performance with $d = -.07$. While one study found that compared to no training, training in VR has improved the speed of a maintenance task (Ganier et al., 2014), other studies who have compared MR-based training with conventional training have not been able to support these superior effects of MR (González-Franco et al., 2016; Webel et al., 2013). However, Kaplan et al. (2021) reported that these findings were not consistent due to different MR technologies used, varying tasks and training methods, as well as different

performance measures. Furthermore, they claimed that the sparsity of data made it extremely difficult to perform a meta-analysis. In their review, Kaplan et al. (2021) concluded that further research would need to consider a broader scope of outcome variables to research the benefit of training. Moreover, they considered a wide range of training tasks and populations (e.g., stroke patients, technicians, students, etc.). Thus, an isolated consideration of manual assembly training is indicated to further examine the validity of these results.

The second recent review of Doolani, Wessels, and co-authors (2020) on the use of various MR technologies in manufacturing, safety, education, military, rehabilitation, and medical training already provided insights into the tasks for which MR technologies have been used within various phases of the work and manufacturing process. Here, the overall suitability of MR in all phases of manufacturing processes and first tendencies for the particular suitability of VR in introductory or orientation phases have already become apparent. Although AR has been mentioned to be useful in later phases of inspection or for the use of hand tools or rare machinery, the authors have summarized VR to be superior to AR tools. Overall, the technically oriented review by Doolani, Wessels, et al. (2020) described the shortcomings that still need to be overcome concerning the standardization of hardware and software to investigate MR across different applications. In this context, the authors emphasized the importance of including further interaction modalities in the consideration of MR technologies to gain deeper insights into their effectiveness.

The two reviews have clearly shown that MR-based training have not yet shown a clear

advantage over traditional training. Nevertheless, Kaplan et al. (2021) and Doolani, Wessels, et al. (2020) concluded that across all tasks and fields of application, MR-based training did not show any disadvantage either. At this point, it remains to be stated that MR-based training is at least as good as traditional training with regard to performance measures.

For industry practitioners to decide on whether MR-based training might be more suitable than traditional training for the use case of manual assembly tasks, further aspects need to be taken into account beyond the factors considered in the previous reviews. First of all, higher applicability of results need to be ensured by exclusively evaluating studies that examine MR-based training in the context of manual assembly tasks and comparing it to other formats of traditional training. Furthermore, the consideration of the outcomes of MR-based training is indicated to go beyond the purely quantitative consideration of performance to other relevant qualitative aspects that might play a role in the decision for or against the suitability of a training medium (e.g., other psychological aspects such as immersion, task load, or user experience) (Kaplan et al., 2021). Finally, it is proposed to take into account that a broad range of MR technologies (hardware and software) has been used in prevalent studies, which aggravates the comparability of their impact on user's performance and subjective evaluations. Therefore, an in-depth consideration of the respective MR technology and interaction features is needed to make overarching statements about the effects of different MR-based training formats on training outcomes such as performance.

The aspects mentioned are essential extensions of previous reviews to improve the understanding of the impact of MR-based training compared to traditional training in the use case of manual assembly tasks. The present review addresses these aspects. The objectives and guiding questions of the review resulting from the current state of the literature are presented in the following section.

Purpose of the Scoping Review and Review Questions

The present scoping review aimed to provide a comprehensive understanding for researchers and practitioners in the industry on the impact of MR-based training versus traditional training on user performance in manual assembly tasks. Thus, we transformed different research findings on the impact of MR-based training for manual assembly tasks on performance and subjective evaluations into concise and meaningful conclusions and provided a close link between research outcomes and industry needs.

In order to synthesize the wide range of literature in the field, this review followed the methodology of a scoping review (Arksey & O'Malley, 2005). In the case of particularly heterogeneous research evidence on MR-based training for manual assembly tasks, this type of review is suitable for summarizing and disseminating research findings and to describe in more detail the findings and range of research. Thus, it serves as a precursor to a systematic review, aiming at identifying key characteristics or factors related to the concept of MR-based training for manual assembly tasks (Munn et al., 2018).

To ensure the applicability of the research results for the use cases of industry, the first step of the review was to elicit the needs and requirements on the part of the industry and what benefits are expected from the use of MR in the context of manual assembly. This topic was examined in review question (RQ) 1. Furthermore, the present review extended existing findings from previous reviews by missing in-depth consideration of different MR technologies and their interaction features used in manual assembly training. Here, the focus was on whether systematic differences in training outcomes are depending on technology and features used (RQ 2). Moreover, the evaluation of training outcomes went beyond the purely quantitative consideration of performance and thus included relevant qualitative aspects and subjective evaluations of the users. To achieve this, we categorized and compared how different objective measures and subjective evaluations as training outcomes of MR-based training were

defined and measured (RQ 3). This in turn will help to make future research in the field comparable with regard to dependent variables.

The central question of the scoping review related to the effects of MR-based training concerning the in RQ 3 identified outcome measures. To this end, we analyzed in RQ 4 how different MR-based training formats impact user performance and subjective evaluations compared to traditional training. Finally, current research gaps in the field of MR-based training were analyzed, discussed, and practical implications were derived (RQ 5). Within the scope of this review, all findings from the five review questions listed below were translated into understandable and applicable statements.

RQ 1. What are the industrial needs and expected benefits of using MR-based trainings in assembly tasks?

RQ 2. What kind of MR technologies are currently used for training procedural assembly tasks in the industrial context?

RQ 3. What measures to capture objective performance effects and subjective evaluations are used to assess the outcomes of MR-based training?

RQ 4. What are the effects of using MR-based training compared to traditional training regarding the different outcome measures?

RQ 5. What research gaps are reported by the authors?

METHOD

The method of this scoping review referred to the iterative approach for scoping reviews proposed by Peters and co-authors (2020), which was based on the five stages framework proposed by Arksey and O'Malley (2005) as well as Levac et al. (2010). In the following, inclusion criteria are described and summarized according to the population, concept, and context (PCC) scheme. Then, based on defined inclusion and exclusion criteria, the search strategy and selected studies with *post hoc*

adjustment of the criteria based on terminologies and content used are presented. The extraction of results was subsequently described using the PRISMA Flow diagram adapted for scoping reviews (Moher et al., 2009). The objectives, inclusion criteria, and methods for this scoping review were specified in advance and documented in a protocol (Daling, 2021).

Inclusion Criteria

The research questions derived through the current state of research provided a clear framework for inclusion criteria related to types of participants, concept, context, and types of sources, which are described in the following.

Types of Participants. MR-based learning systems are particularly used when employees who are inexperienced in a new task or temporary workers need to be trained at short notice. However, even experienced employees can be novices when it comes to a task they have not had to complete before. The prerequisite for the inclusion of participants was therefore that they have no previous experience for the task under investigation. Accordingly, the review included articles examining healthy and adult participants who were currently working or being trained for a job, including students or university samples or employees from the industry.

Concept. The scoping review investigated current research on MR-based training for manual assembly tasks, aiming at (1) exploring the needs and expected benefits mentioned in the articles, (2) mapping the technologies and features being used, (3) categorizing relevant outcome measures, (4) analyzing the effects and outcomes of MR-based training in comparison with traditional training as well as (5) identifying relevant research gaps. Thus, the review focused on learning or training systems, which is why on-the-job assistance systems were not included. As a result, only studies were considered in which training of a certain assembly task ranging from simple, more abstract tasks to highly complex assembly tasks was conducted and then actual task performance in performing the respective task was measured. The investigated training had to include one of the core terms of augmented (reality), virtual (reality), or

mixed reality or, if these terms were not specifically mentioned in the description of the technology, meet the presented definition of MR (based on Skarbez et al., 2021). Thus, we included trainings in which physical world and virtual objects or stimuli were presented together within a single percept. Moreover, only articles with a comparison or control group that included another form of training (e.g., paper-based, video-based, trainer-based, or other technologies) were included.

Context. The research object of this review referred exclusively to procedural and cognitive training for industrial manual assembly. More abstract tasks (such as assembling Lego parts or 3D puzzles) were considered as long as the task included at least one typical assembly activity (joining, handling, inspecting, or adjusting). Medical training or other procedural tasks were not considered. Furthermore, the tasks to be trained should not include any further cooperation or collaboration with other humans or machines, or robots.

Types of Sources. From the existing literature, only primary studies were included in the scoping review. Meta-analyses or systematic reviews were excluded. To ensure the quality of the empirical results, no gray literature, blogs, or similar were considered. Since technologies are constantly evolving and some are not comparable with very early versions, only studies published between 2010 and 2020 were included. English- and German-language articles were considered, regardless of the authors' origin or location.

Search Strategy

The literature search covering the topic was conducted on three of the most established scientific search engines "ProQuest," "Web of Science," and "EbscoHost" (see Table 1). First, an initial limited search was conducted as a pilot. After analyzing text words contained in the title and abstract of retrieved papers, a second search using identified keywords narrowed down the results to 1881 on ProQuest (extracted on 27th November 2020), 214 on Web of Science (extracted on 21st November 2020) and 38 on

EbscoHost extracted on 24th November 2020) as shown in the search string listed in Table 1.

All identified papers were imported to the reference management tool EndNote and the duplicates were deleted, resulting in a total of 2043 articles (see Figure 2). After filtering through the titles, the identified papers were reduced to 205 and yet again narrowed down to 138 after a practical screening of the abstracts.

Extraction of Results

An overview of the data extraction process is specified and summarized in Figure 2. A random sample of 25 articles was selected out of the 138 articles and full texts were screened by two reviewers using defined inclusion and eligibility criteria. Within this pilot test, the reviewers agreed that usability studies without additional objective performance indicators, concept papers, or articles with a purely technical focus would have to be excluded to filter findings on the effects and outcomes of MR-based training. In order to include studies with reliable and precise estimates, it was also determined that only studies with sample sizes of $n \geq 10$ should be considered in the final review, ensuring the statistical foundation of the conclusions (Hackshaw, 2008; Aguinis & Harden, 2008).

After assessing the modified eligibility, 24 studies were included in the review. The most frequent reasons for excluding articles were a lack of content fit to the topic (e.g., remote assistance for maintenance in Wang et al., 2019; or developing models for VR teleoperation in Lipton et al., 2018) (49 articles) and an exclusive technical focus, where the analysis presented was not based on human data (e.g., Leu et al., 2013; Xia et al., 2012) (30 articles). Thirteen articles were excluded because important statistical values (e.g., M, SD/SE, p) were not reported, there was a lack of examination of the requirements for statistical testing of small samples, sample sizes were not reported, or because the chosen level of alpha was set to $>.05$. Furthermore, 14 concept papers were excluded, as well as three articles with a purely descriptive statistical analysis, three articles with a sample with less than 10 participants, and two

TABLE 1: Key Search Strings Used on the Databases Proquest, Web of Science, and Ebscohost.

Database	Search date	Search strings
Proquest	27/11/2020	((mainsubject.Exact("virtual reality") OR mainsubject.Exact("augmented reality") OR mainsubject.Exact("mixed reality")) AND mainsubject.Exact("assembly" OR "assembling")) AND mainsubject(training)) AND la.exact("English" OR "German")) NOT (at.exact("Evidence Based Healthcare" OR "Review" OR "Literature Review") AND la.exact("ENG" OR "GER")) NOT subt.exact("rehabilitation" OR "robotics" OR "machine learning" OR "patients" OR "neural networks" OR "neurosciences" OR "brain research" OR "stroke" OR "surgery" OR "cognition & reasoning" OR "brain" OR "older people" OR "cultural heritage" OR "medical imaging" OR "artificial neural networks" OR "walking" OR "proteins" OR "electroencephalography" OR "medicine")) NOT "medi*"
Web of Science	21/11/2020	Timespan=2010–2020 TS=(virtual reality assembly training) OR ts=(VR assembly training) OR ts=(AR assembly training) OR ts=(augmented reality assembly training) OR ts=(mixed reality assembly training) OR ts=(MR assembly training) OR ts=(virtual reality assembly learning) OR ts=(virtual reality assembly teaching) OR ts=(extended reality assembly training) OR Refined by: [excluding] DOCUMENT TYPES: (REVIEW)
Ebscohost	24/11/2020	Advanced search with keywords and field tags virtual reality assembly training OR VR assembly training OR AR assembly training OR augmented reality assembly training OR mixed reality assembly training OR MR assembly training OR virtual reality assembly learning OR virtual reality assembly teaching OR extended reality assembly training NOT TX medic* NOT TI review

articles whose dependent variable did not involve any performance measures as outcome.

A draft charting table using Excel was developed and tested by two reviewers independently screening ten randomly selected papers. The data chart was finally narrowed down to five main categories: Overview, background and need for MR, method and measures, results as well as identified research gaps, which are subsequently described in the following.

Overview. The overview category included the following information: Published year, country of origin, type of material, research design of the study, and whether the paper was published in a peer-reviewed journal. All information was summarized and charted.

Background and Need for MR. In the background and need for MR category, the information regarding what was defined as the need for MR training was collected, and what kind of training task was given in the study. Qualitative content analysis (Mayring, 2014) was conducted to analyze the described needs, requirements, and expected benefits of MR regarding RQ1. The aim of the content analysis was the definition of precise categories that capture the substance of the investigated content. In the first step, approximately half of the selected papers were reread to get an overview of the relevant sections of the papers in preparation for the development of the category system. The focus was on those sections containing descriptions of the general context, specific problems, and the relevance of the topic

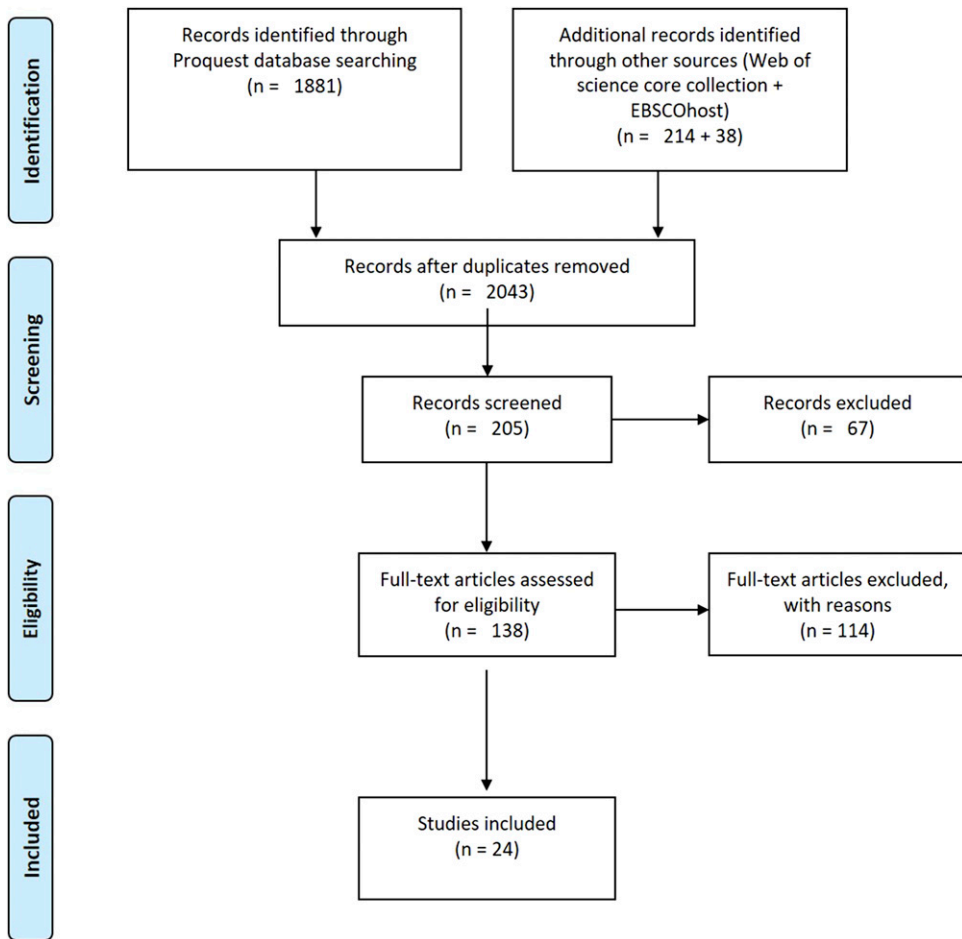


Figure 2. PRISMA flow diagram for scoping reviews.

selected in the article. The central statements from these text sections were paraphrased and tagged with keywords, which were used to classify similar groups and larger thematic areas. After the various content-related aspects of the data were identified in this way, a coding guideline was developed and used to code and subsequently analyze all material.

Method and Measures. In the method and measures category, information about the study design and the MR technologies used in the study was collected. Regarding RQ2, information on MR technologies used was collected, categorized, and sorted along the revised reality-virtuality continuum (based on Skarbez et al., 2021). Furthermore, the experimental data

such as the hypothesis, experimental, and control group specifications as well as objective and subjective measurements were collected. Within the scope of RQ3, all measures used in the reviewed articles regarding objective performance and subjective evaluations were classified and frequencies of the identified categories were calculated.

Results of MR-Based Training. The reported results of MR-based training on objective and subjective measures in comparison to respective control groups were summarized and broken down into short and clear statements. Statistically significant results were marked accordingly. Effects were clustered by technology and counted by frequency. Subsequently, a more in-

depth analysis of the effects was performed, taking into account the task, type of population, and other study characteristics.

Identified Research Gaps. In the identified research gap category (RQ5), important discussion points were summarized and implications for possible future work were derived, following the same approach of the above-mentioned qualitative content analysis (Mayring, 2014).

RESULTS

All results are presented following the above-mentioned categories and review questions. First, an overview of all reviewed articles is given. Subsequently, the identified industrial needs and expected benefits of using MR-based training in assembly tasks are reported according to RQ1. Then, the results of RQ2 on what kind of MR technologies were currently used for training procedural assembly tasks in the industrial context are presented. Afterward, we present the results of RQ3, revealing which measures were used to capture objective performance and subjective evaluations as outcome measures of MR-based training. In the section of RQ4, the results concerning the effects of using MR-based training compared to traditional training regarding the different outcome measures are presented. Finally, the identified research gaps of the reviewed articles are reported in RQ5.

Overview

A summary of the 24 articles included in this scoping review can be seen in Table 2. From the included articles in the review, most were published in 2015 ($n=4$), 2018 ($n=5$), and 2019 ($n=5$). The first authors of these articles were located in 13 different countries at the time of publication. Based on the location of the first authors affiliation, most articles were from Germany ($n=5$), USA ($n=3$), and Australia ($n=3$), followed by Denmark ($n=2$), Canada ($n=2$), UK ($n=2$), and Spain ($n=2$). One article each could be located in Belgium, Saudi Arabia, Brazil, Italy, and New Zealand. The research disciplines of the first authors' affiliation can be

categorized as follows: Seven of the first authors' affiliations were related to Industrial and Mechanical Engineering (Al-Ahmari et al., 2018; Gavish et al., 2015; Hoedt et al., 2017; Roldan et al., 2019; Velaz et al., 2014; Werrlich, Lorber, et al., 2018; Werrlich, Nguyen, et al., 2018), five to Computer Science and Simulation (Doolani, Owens, et al., 2020; deMoura & Sadagic, 2019; González-Franco et al., 2016; Murcia-Lopez & Steed, 2018; Webel et al., 2013), five to Building Science and Construction Research (Cooper et al., 2018; Hou & Wang, 2013; Hou et al., 2013; Hou et al., 2015), and another five to Human Computer Interaction and Human Factors (Carlson et al., 2015; González-Franco et al., 2016; Langley et al., 2016; Oren et al., 2012; Westerfield et al., 2015). Three other articles can be classified as being grounded in Information Systems and Business Development (Koumaditis et al., 2019, 2020; Loch et al., 2019) and one in Environmental Engineering (Kwiatek et al., 2019).

Type of material was divided into Journal Articles (72%) and Conference Proceedings (28%). Seventeen studies were conducted in a lab environment with student or university samples, three articles used non-specified or random samples in their lab-based studies, and five were conducted as a field study with employees from the industry. At this point, it is important to mention that the article of Gavish et al. (2015) contained two independent experiments, which were analyzed separately and thus counted twice. Different types of conventional training were used as control groups in the articles, as specified in Table 2. Some used several groups of comparisons: From the articles included, seven used paper-based training as control group (Hou & Wang, 2013; Hou et al., 2013, 2015; Kwiatek et al., 2019; Murcia-Lopez & Steed, 2018; Roldan et al., 2019). Video-based training was used eight times as control group (Doolani, Owens, et al., 2020; Gavish et al., 2015 (2x); Koumaditis et al., 2020; Loch et al., 2019; Murcia-Lopez & Steed, 2018; Webel et al., 2013; Velaz et al., 2014). Trainer-based training, that is, training with real human instructors, was used within six articles (González-Franco et al., 2016; Hoedt et al., 2017; Koumaditis et al., 2019, 2020; Langley

TABLE 2: Summary of Included Articles.

Author	Title	Year	Country of Origin	Type of Material	Disciplinary Field	Lab or Field Study	N
Doolani, Owens, et al.	vIS: An Immersive Virtual Storytelling System for Vocational Training	2020	USA	Journal article	Computer science	Lab (student sample)	30
Al-Ahmari et al.	Evaluation of 3D printing approach for manual assembly training	2018	Saudi Arabia	Journal article	Industrial engineering	Lab (student sample)	25
Carlson et al.	Virtual Training: Learning Transfer of Assembly Tasks	2015	USA	Journal article	Human computer interaction	Lab (student sample)	63
Cooper et al.	The effects of substitute multisensory feedback on task performance and the sense of presence in a virtual reality environment	2018	Canada	Journal article	Construction science	Lab (random sample)	17
deMoura and Sadagic	The Effects of Stereopsis and Immersion on Bimanual Assembly Tasks in a Virtual Reality System	2019	Brazil	Conference Proceedings	Air force/naval school	Lab (student sample)	68
Gavish et al.	Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks	2015	Italy	Journal article	Industrial engineering	VR: Lab AR: Field (industrial employee sample)	40
Gonzalez-Franco et al.	Immersive Augmented Reality Training for Complex Manufacturing Scenarios	2016	UK	Journal article	Computer science and Simulation	Lab but with real training scenarios (industrial employee sample)	24
Hoedt et al.	The evaluation of an elementary virtual training system for manual assembly	2017	Belgium	Journal article	Industrial Systems Engineering and Product Design	Lab but with a realistic work place and task (sample not specified)	26
Hou and Wang	A study on the benefits of augmented reality in retaining working memory in assembly tasks: A focus on differences in gender	2013	Australia	Journal article	Construction science	Lab (student sample)	28
Hou et al.	Using Animated Augmented Reality to Cognitively Guide Assembly	2013	Australia	Journal article	Construction science	Lab (student sample)	20; 30
Hou et al.	Using Augmented Reality to Facilitate Piping Assembly: An Experiment-Based Evaluation	2015	Australia	Journal article	Construction science	Lab (student sample)	18
Kourmaditis et al.	Effectiveness of Virtual Versus Physical Training: The Case of Assembly Tasks, Trainer's Verbal Assistance, and Task Complexity	2020	Denmark	Journal article	Information Systems and Business Development	Lab (sample not specified)	100

(Continued)

TABLE 2: (Continued)

Author	Title	Year	Country of Origin	Type of Material	Disciplinary Field	Lab or Field Study	N
Koumaditis et al.	Immersive Training: Outcomes from Small Scale AR/VR Pilot-Studies	2019	Denmark	Conference Proceedings	Information Systems and Business Development	Field (industrial employee sample)	10
Kwiatk et al.	Impact of augmented reality and spatial cognition on assembly in construction	2019	Canada	Journal article	Civil and Environmental engineering	Lab (industrial employee vs. student sample)	61
Langley et al.	Establishing the Usability of a Virtual Training System for Assembly Operations within the Automotive Industry	2016	Germany/UK	Journal article	Human factors	Field (industrial employee sample)	30
Loch et al.	Using Real-time Feedback in a Training System for Manual Procedures	2019	Germany	Conference Paper	Automation and Information systems	Lab (student sample)	20
Murcia-Lopez and Steed	A Comparison of Virtual and Physical Training Transfer of Bimanual Assembly Tasks	2018	UK	Journal article	Computer science	Lab (student sample)	60
Oren et al.	Puzzle Assembly Training: Real World vs. Virtual Environment	2012	USA	Conference Proceedings	Human computer interaction	Lab (student sample)	10
Roldan et al.	A training system for Industry 4.0 operators in complex assemblies based on virtual reality and process mining	2019	Spain	Journal article	Automation and Robotics	Lab (student sample)	20
Velaz et al.	The Influence of Interaction Technology on the Learning of Assembly Tasks Using Virtual Reality	2014	Spain	Journal article	Mechanical engineering	Lab (random sample)	57
Webel et al.	An augmented reality training platform for assembly and maintenance skills	2013	Germany	Journal article	Computer science	Field (industrial employee sample)	20
Werrlich, Lorber, et al.	Assembly Training: Comparing the Effects of Head-Mounted Displays and Face-to-Face Training	2018	Germany	Conference Proceedings	Mechanical Engineering	Field (industrial employee sample)	36
Werrlich, Nguyen, et al.	Evaluating the training transfer of Head-Mounted Display based training for assembly tasks	2018	Germany	Conference Proceedings	Mechanical Engineering	Field (industrial employee sample)	30
Westerfeld et al.	Intelligent Augmented Reality Training for Motherboard Assembly	2015	New Zealand	Journal article	Human interface technology	Lab (student sample)	16

Note: Country of origin and disciplinary field was based on the first authors' affiliation. Samples with students and university staff were indicated as *student samples*, samples with real workers or end users were indicated as *industrial employee sample*, other samples were *not specified* or *random samples*.

et al., 2016; Werrlich, Lorber, et al., 2018). Physical objects (such as 3D prints) were used as control group in four experiments (Al-Ahmari et al., 2018; Carlson et al., 2015; Oren et al., 2012; Murcia-Lopez & Steed, 2018). In five studies, other technologies and interaction features were used as control group (Cooper et al., 2018; deMoura & Sadagic, 2019; Werrlich, Nguyen, et al., 2018; Westerfield et al., 2015, Velaz et al., 2014).

In the following, the results are clustered and analyzed according to the five review questions of the present scoping review, starting with the identified needs of the industry of using MR-based training.

RQ1: What are the Industrial Needs and Expected Benefits of Using MR-Based Trainings in Assembly Tasks?

We conducted a qualitative content analysis based on Mayring's approach (2014) using the software MAXQDA (Version 2020) on the 24 selected articles to gain a differentiated perspective on the existing needs in the industry context, the main potentials as well as the expectations of using MR in manual assembly training.

The category system that resulted from the qualitative content analysis comprised various main categories and subcategories as shown in Figure 3. The first block contains different statements on the general context of manual assembly in Industry 4.0. Almost all authors emphasized that the trend toward a high degree of product diversity and customizability increased the complexity of assembly tasks making employee training a key factor in this context. As a central requirement for the training processes, several authors stated that the training should be as (cost-)efficient as possible (e.g., low costs for hardware and developing or adjusting the software) and should take place in a safe environment that can be easily adapted to the frequently changing products in the assembly line (Doolani, Owens, et al., 2020; deMoura & Sadagic, 2019; González-Franco et al., 2016; Koumaditis et al., 2019; Loch et al., 2019; Murcia-Lopez & Steed, 2018; Oren

et al., 2012). Currently, training in the industry was often carried out using paper manuals, which was mentioned to be time-consuming to interpret and could lead to misunderstandings and errors due to ambiguous information (Gavish et al., 2015; Hou et al., 2013; Kwiatek et al., 2019; Westerfield et al., 2015). Taking into account the status quo in manual assembly, it became apparent that the benefit of using MR as training medium was influenced by a reciprocal process of considering the industry needs and unforeseen possibilities offered by technology. The benefits of MR mentioned in the articles were summarized in the second block. Overall, it was emphasized that through the use of MR, information could be presented in a way that was easier to understand (González-Franco et al., 2016; Hou & Wang, 2013; Kwiatek et al., 2019), for example, because it was displayed as a virtual replica directly on the relevant objects, such as in AR (Hou & Wang, 2013; Webel et al., 2013; Werrlich, Lorber, et al., 2018; Werrlich, Nguyen, et al., 2018; Westerfield et al., 2015). In VR, training of assembly processes could be simulated in a safe and engaging way (Doolani, Owens, et al., 2020; deMoura & Sadagic, 2019; González-Franco et al., 2016; Velaz et al., 2014). Overall, MR systems actively guided the workers at individual pace through the assembly process (Hou & Wang, 2013; Webel et al., 2013; Werrlich, Lorber, et al., 2018).

It was found that the concept of productivity in the papers referred to numerous different aspects and was often used non-specifically. Within the framework of the content analysis, three content-related dimensions were identified based on which the improvement potential arising from MR was described using the terms performance, effectiveness, or efficiency, as indicated in Figure 3. The first dimension referred to the direct influence MR has on employee performance, that is, their cognitive and motor ability to perform assembly procedures quickly and without errors as well as their subjective user experience (Cooper et al., 2018; Doolani, Owens, et al., 2020; Hou et al., 2013; Hou et al., 2015; Kwiatek

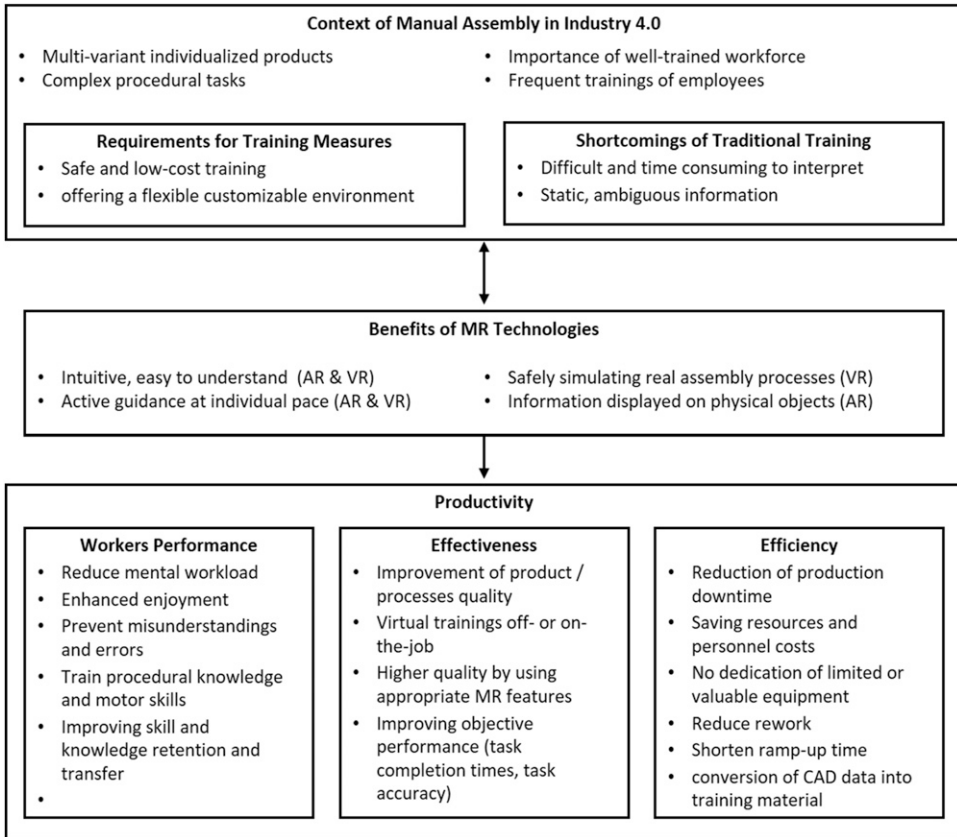


Figure 3. Identified needs and expected benefits of MR-based trainings for manual assembly tasks in the context of Industry 4.0 (RQ1).

et al., 2019; Webel et al., 2013). In the second dimension, which is effectiveness, aspects were summarized which referred to the success of the training outcome itself by using well-designed MR technologies and features for certain tasks, but also to the organization of other tasks and activities relevant for enterprises. In this context, MR did not appear solely as an alternative training medium, but as a component of the general digitalization of work processes (Hoedt et al., 2017; Langley et al., 2016; Oren et al., 2012). The third dimension, which is efficiency, referred to the saving of resources. Traditional on-the-job training on the assembly line was usually accompanied by a loss of productivity. This was characterized by the use of personnel or valuable machinery to allow new employees to

learn assembly through hands-on practice and to develop an understanding of the work processes. Rework during training and machine downtimes were thus in contrast to the highest possible production capacity exploitation. MR, in contrast, provides time-, location-, and trainer-independent training that reduces the time required to practice on the machine itself. Moreover, the overall efficiency could be enhanced using existing CAD models, or production planning and training could be partially carried out in parallel (Hoedt et al., 2017; Hou & Wang, 2013; Oren et al., 2012; Roldan et al., 2019; Velaz et al., 2014). To investigate these identified needs, studies with a wide variety of different MR technologies and features were examined, which are analyzed and classified below.

RQ2: What Kind of MR Technologies is Currently Used for Training Procedural Assembly Tasks in the Industrial Context?

The technologies used in the reviewed articles varied in their hardware, software, and specific functionalities related to their extent of world knowledge, immersion, and fidelity. For the scoping review presented here, we used a consistent categorization of technologies that might have deviated from the authors’ original definition (e.g., if the authors defined Microsoft HoloLens as MR, we categorized it as AR HMD, since MR was used here as umbrella term). In order to form comparable categories, the technical descriptions of the hardware and interaction features used were listed and classified according to the revised reality-virtuality continuum and the presented definition of MR (based on Skarbez et al., 2021). This resulted in the main categories AR-based training, screen-based VR training, and VR head-mounted-display (HMD)–based training (Figure 4). The respective allocations of articles to main- and subcategories are described below. The study of Gavish et al. (2015) is listed twice, since two independent technologies were investigated.

Twelve articles were identified using AR-based formats and thus were allocated to the left side of the continuum, where a high extent of world knowledge and perception of real elements was enabled by the presented MR-based training (i.e., AR-based trainings displaying virtual objects onto the real world). These were assigned to the three subcategories AR projectors (Hou & Wang, 2013; Hou et al., 2013, 2015; Loch et al., 2019), AR handhelds (Gavish et al., 2015; Kwiatek et al., 2019; Webel et al., 2013) and AR head-mounted-displays (HMD) (González-Franco et al., 2016; Koumaditis et al., 2019; Werrlich, Lorber, et al., 2018; Werrlich, Nguyen, et al., 2018; Westerfield et al., 2015). AR projectors were defined as being fixed to the environment (e.g., monitors or fixed projectors above the workstation), whereas AR handhelds were used as a mobile and flexible tool, that is, tablets, connected either to the user or the environment. AR HMD technologies were characterized as displays being permanently attached to the users’ heads while still allowing them a see-through view of the real environment.

Eight articles used screen-based VR training, which was allocated in the middle of the continuum, representing MR solutions between reality and virtuality. These

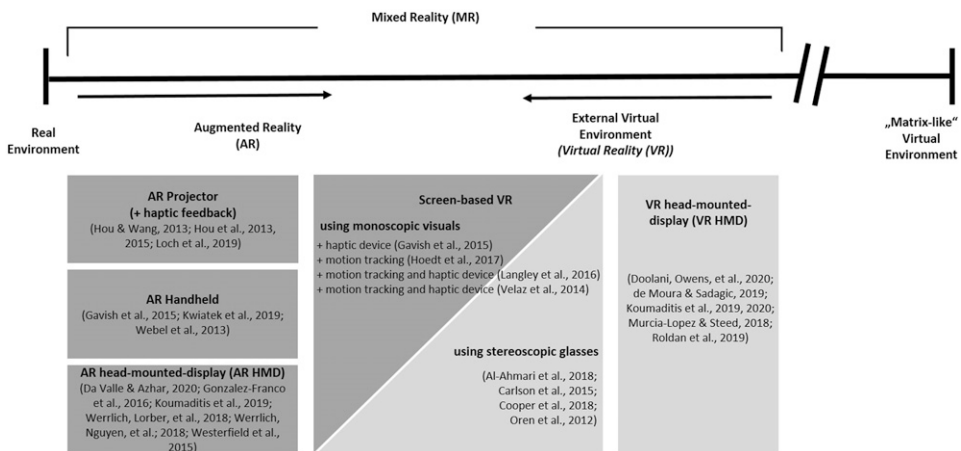


Figure 4. Classification of MR technologies and their interaction features used in the reviewed articles as AR-based training, Screen-based VR training, and VR head-mounted display (HMD)–based training (RQ2) according to the revised reality-virtuality continuum based on Skarbez et al. (2021).

technologies were further differentiated according to their use of monoscopic or stereoscopic visuals. We used the description monoscopic screen-based VR training if the authors indicated that they were presenting 3D-modeled environments on a screen without any use of, for example, stereoscopic glasses. Monoscopic VR can be defined as presenting an image simultaneously to both eyes (Singer et al., 1995), making it less immersive and providing a relatively high extent of world knowledge. Nevertheless, they were mentioned as a relevant form of VR in industry-related assembly training (Gavish et al., 2015; Hoedt et al., 2017; Langley et al., 2016; Velaz et al., 2014). Screen-based VR training using stereoscopic glasses was used in another four studies. Stereoscopic visualization in VR creates an illusion of depth through two two-dimensional images corresponding to the view of a scene from two different angles (Singer et al., 1995). Thus, these technologies were allocated closer to the right side of the revised reality-virtuality continuum, enabling a lower extent of world knowledge and higher immersion (Al-Ahmari et al., 2018; Carlson et al., 2015; Cooper et al., 2018; Oren et al., 2012).

Another five articles used VR HMD-based training, which was allocated close to the right, that is, the external virtual environment (Doolani, Owens, et al., 2020; deMoura & Sadagic, 2019; Koumaditis et al., 2020; Murcia-Lopez & Steed, 2018; Roldan et al., 2019). The category of VR HMD had a low

extent of world knowledge, following the definition that the real environment was hidden through closed goggles and the user was completely immersed in the 3D environment (Milgram & Kishino, 1994).

In the course of this review, the selected publications were analyzed regarding their effects on training outcomes in manual assembly tasks and thereby clustered along with the defined technology categories AR projectors, AR handhelds, AR HMD, screen-based VR, and VR HMD. In the following section, we first present an overview and classification of collected training outcome measures to ensure comparability of effects concerning dependent variables.

RQ3: What Measures to Capture Objective Performance Effects and Subjective Evaluations are Used to Assess the Outcomes of MR-Based Training?

The dependent variables used to measure training outcomes of participants after MR-based training were divided into objective measures and subjective evaluations and evaluated by frequency. In the following, objective measures were defined as being independent of the observer by using stop-watches or error logs. Subjective evaluations included self-assessment questionnaires filled by participants as well as opinions and general impressions of the observer. An overview of all variables that were collected more than once and their frequency is provided in Table 3. At this point, variables that were collected were counted only once per article, even if

TABLE 3: Number of Outcome Measures of MR-Based Trainings Used in the Reviewed Articles (RQ 3).

Objective Measures	Number of Articles (out of 24)	Subjective Measures	Number of Articles (out of 24)
Task completion time (TCT)	20	Perceived usability, Ease of use	10
Task accuracy (TA)	20	Perceived difficulty of the task	6
Training time	10	Perceived Workload	5
Long-term skill retention (TCT)	4	Feeling of Presence	3
Long-term skill retention (TA)	3	Others: Single items on user's subjective assessments	8

TABLE 4: The Effects of MR-Based Trainings on Objective Performance Measures and Subjective Evaluations (RQ4).

Author (Year)	Technology	Assembly Task (J) Joining, (H) Handling, (I) Inspecting, (A) Adjusting	Objective Measures					Subjective Measures					
			Task Completion Time (TCT)	TCT (long-term retention)	Task Accuracy (TA)	TA (long-term skill retention)	Training Time	Usability	Workload	Feeling of Presence	Task Difficulty	Others	
Hou and Wang (2013)	AR Projector (Monitor)	Lego Mindstorm Assembly (J)(H)(I)(A)	AR better than paper*	—	AR better than paper*	—	—	—	—	—	—	—	
Hou et al. (2013)	AR Projector	Lego Mindstorm Assembly (J)(H)(I)(A)	AR better than paper*	—	AR better than paper*	—	—	—	—	—	—	AR system was perceived as positive and easy to use	
Hou et al. (2015)	AR Projector (Monitor)	Piping assembly (J)(H)(I)(A)	AR better than paper*	—	AR better than paper*	—	—	—	—	Workload lower with AR than paper*	—	—	
Loch et al. (2019)	AR Projector + haptic feedback	Product assembly (J)(H)(I)(A)	AR better than video*	—	AR better than video*	—	—	—	—	AR better than video	—	Ease of learning (AR = video)	
Gavish et al. (2015)	AR Handheld	Electronic actuator assembly (J)(H)(I)(A)	AR not different from video	—	AR better than video* (unsolved errors)	—	AR not as good as video*	—	—	AR better than VR	—	Satisfaction with performance (AR > VR)	
Kwiatek et al. (2019)	AR Handheld	Pipe spool assembly (J)(H)(I)(A)	AR better than paper*	—	—	—	—	—	—	—	—	Participants recommend the tool	
Weibel et al. (2013)	AR Handheld	Physical assembly task (J)(H)(I)(A)	AR not different from video	—	AR better than video*	—	—	—	—	—	—	—	
González-Franco et al. (2016)	AR HMD	Aircraft door assembly (J)(H)(I)(A)	—	—	AR not different from trainer	—	AR not as good as trainer*	—	—	—	—	—	Knowledge retention (AR = trainer)
Kouraditis et al. (2019)	AR HMD	Real product assembly task (J)(H)(I)(A)	AR better than trainer*	—	—	—	—	—	—	Workload (effort) lower with AR than trainer*	—	—	—
Wierlich, Lorber, et al. (2018)	AR HMD	Real engine assembly (J)(H)(I)(A)	—	—	AR not as good as trainer*	—	AR not as good as trainer*	—	—	—	—	—	Knowledge retention (AR = trainer)
Wierlich, Nguyen, et al. (2018)	AR HMD+ Quiz	Real engine assembly (J)(H)(I)(A)	AR HMD not as good as conventional AR*	—	AR HMD better than conventional AR*	—	AR HMD not different from conventional AR	—	—	Workload with AR HMD not different from conventional AR	—	—	—
Westerfield et al. (2015)	Intelligent AR HMD	Motherboard assembly (J)(H)(I)(A)	Intelligent AR better than conventional AR*	—	Intelligent AR not different from conventional AR	—	Intelligent AR not different from conventional AR	—	—	—	—	—	No differences in subjective evaluation
Gavish et al. (2015)	Screen-based VR (monoscopic + haptic device)	Electronic actuator assembly (J)(H)(I)(A)	VR not different from video	—	VR not different from video	—	VR not as good as video*	—	—	VR not as good as AR*	—	—	Users would rather recommend AR

(Continued)

TABLE 4: (Continued)

		Objective Measures					Subjective Measures				
Author (Year)	Technology	Task Completion Time (TCT)	TCT (long-term skill retention)	Task Accuracy (TA)	TA (long-term skill retention)	Training Time	Usability	Workload	Feeling of Presence	Task Difficulty	Others
	Assembly Task (J) Joining, (H) Handling, (I) Inspecting, (A) Adjusting										
Heedt et al. (2017)	Screen-based VR (monoscopic + motion tracking)	VR not different from real training	—	VR not different from real training	—	—	—	—	—	—	—
Langley et al. (2016)	Screen-based VR (monoscopic + motion tracking and haptic device)	—	VR not different from trainer + paper	—	VR better than trainer + paper*	—	—	—	—	—	System is easy to use and more enjoyable
Velaz et al. (2014)	Screen-based VR (monoscopic + motion tracking and haptic device)	VR not different from video, mouse, haptic device, and 2D	—	VR not different from video, mouse, haptic device, and 2D	—	VR not as good as video*, mouse*, haptic device*, and 2D*	VR not different from video, mouse, haptic device, and 2D	—	—	VR not different from video, mouse, haptic device, and 2D	Haptic device is perceived more consistent
Al-Ahmari et al. (2018)	Screen-based VR (as control group) (stereoscopic glasses)	VR not as good as 3D prints* but better than paper	—	VR not as good as 3D prints* but not different from paper	—	—	—	VR not different from 3D prints and paper	—	—	—
Carlson et al. (2015)	Screen-based VR (stereoscopic glasses)	VR not as good as physical*	VR not different from physical	—	—	VR not as good as physical	VR not as good as physical	—	VR not different from physical	VR not easier than physical*	Recall strategy in VR: color
Cooper et al. (2018)	Screen-based VR (stereoscopic glasses)	VR with audio cues better than white noise*, VR with tactile* better than no tactile*, VR with visual cue not different from no visual cue	—	—	—	—	—	—	Higher presence correlates with faster TCT; audio, tactile or visual cues enhance feeling of presence	—	User's subjective experience is related to the overall task performance
Oren et al. (2012)	Screen-based VR (stereoscopic glasses + gloves)	VR not different from physical blocks	—	VR not different from physical blocks	—	VR not as good as physical blocks*	VR not as good as physical blocks*	—	—	VR not different from physical blocks	No difference in realism or helpfulness

(Continued)

TABLE 4: (Continued)

		Objective Measures						Subjective Measures			
Author (Year)	Technology	Task Completion Time (TCT)	TCT (long-term skill retention)	Task Accuracy (TA)	TA (long-term skill retention)	Training Time	Usability	Workload	Feeling of Presence	Task Difficulty	Others
Assembly Task (J) Joining, (H) Handling, (I) Inspecting, (A) Adjusting											
Dodani, Owens, et al. (2020)	VR HMD	VR not different from paper and video	VR better than paper* not different from video	VR not different from paper and video	VR better than paper* and video*	VR better than paper* not different from video	VR better than video and paper	—	—	—	—
deMoura and Sadagic (2019)	VR HMD (stereoscopic)	Stereoscopic VR better than monoscopic VR*, stereoscopic, screen-based VR* and monoscopic, screen-based VR*	—	Stereoscopic VR not different from monoscopic VR*, stereoscopic, VR better than stereoscopic and monoscopic, screen-based VR*	—	—	VR better than 3D PC and 2D display	—	—	VR easier than 3D PC* and 2D display*	eye strain observed as symptom for Stereoscopic view, general discomfort for non-stereoscopic
Koumaditis et al. (2020)	VR HMD	VR not as good as video* and video+trainer* VR not different from VR+trainer	—	VR not as good as video*, and video+trainer*, VR not different from VR+trainer	—	—	—	TLX negatively predicts TA and positively predicts TCT	—	—	Trainer's verbal assistance has no positive impact
Murcia-Lopez and Seedorf (2018)	VR HMD	VR better than paper* and paper + video* = paper + video + blocks	VR not different from paper, paper + video and paper + video + blocks	VR not as good as paper*, paper + video* and paper + video +blocks*	VR not different from paper, paper + video + blocks	VR not as good as paper* and paper + video* VR not different from paper + video+ blocks	VR better than paper*	—	—	VR easier than paper* and paper + blocks*	—
Roldan et al. (2019)	VR HMD + controller	—	—	VR better than paper* (easy assembly), VR not different from paper in complex assembly	—	—	—	—	Immersion with VR higher than paper*	VR easier than paper*	—

Note: An asterisk (*) indicates statistically significant results with $p < .05$. A dash (—) indicates that this variable was not reported in the paper cited. The article of Gavish et al. (2015) contained two independent experiments, which were analyzed separately and thus reported twice.

they were collected in two experiments. In [Table 4](#), the results of individual experiments within articles are reported separately.

The most frequently used objective measure was performance, either the time needed for completing the assembly after training (task completion time) and/or the accuracy in assembling a product, that is, the error rate or quality of task processing (task accuracy). Task completion time (TCT) was measured in 20 out of 24 articles ([Doolani, Owens, et al., 2020](#); [Al-Ahmari et al., 2018](#); [Carlson et al., 2015](#); [Cooper et al., 2018](#); [deMoura & Sadagic, 2019](#); [Gavish et al., 2015](#); [Hoedt et al., 2017](#); [Hou & Wang, 2013](#); [Hou et al., 2013, 2015](#); [Koumaditis et al., 2020](#); [Kwiatek et al., 2019](#); [Langley et al., 2016](#); [Loch et al., 2019](#); [Murcia-Lopez & Steed, 2018](#); [Oren et al., 2012](#); [Roldan et al., 2019](#); [Velaz et al., 2014](#); [Webel et al., 2013](#); [Werrlich, Nguyen, et al., 2018](#); [Westerfield et al., 2015](#)). Similarly, task accuracy (TA) was also collected in 20 of the 24 articles ([Doolani, Owens, et al., 2020](#); [Al-Ahmari et al., 2018](#); [deMoura & Sadagic, 2019](#); [Gavish et al., 2015](#); [González-Franco et al., 2016](#); [Hoedt et al., 2017](#); [Hou & Wang, 2013](#); [Hou et al., 2013, 2015](#); [Koumaditis et al., 2020](#); [Langley et al., 2016](#); [Loch et al., 2019](#); [Murcia-Lopez & Steed, 2018](#); [Oren et al., 2012](#); [Roldan et al., 2019](#); [Velaz et al., 2014](#); [Webel et al., 2013](#); [Werrlich, Lorber, et al., 2018](#); [Werrlich, Nguyen, et al., 2018](#); [Westerfield et al., 2015](#)). Additionally, in ten articles statements were made about the training time required with the respective technology to learn how to fulfill the defined assembly task ([Doolani, Owens, et al., 2020](#); [Carlson et al., 2015](#); [Gavish et al., 2015](#); [González-Franco et al., 2016](#); [Murcia-Lopez & Steed, 2018](#); [Oren et al., 2012](#); [Velaz et al., 2014](#); [Werrlich, Lorber, et al., 2018](#); [Werrlich, Nguyen, et al., 2018](#); [Westerfield et al., 2015](#)). To record how well the training content was remembered in the long term, that is, after a certain time, skill retention was assessed as the assembly success regarding TCT after a certain time period had passed in three articles ([Doolani, Owens, et al., 2020](#); [Carlson et al., 2015](#); [Murcia-Lopez & Steed, 2018](#)). Long-term skill retention concerning TA was assessed in two articles ([Doolani, Owens, et al., 2020](#); [Murcia-Lopez & Steed, 2018](#)).

Subjective evaluations were assessed to complement the findings on objective performance and included all information collected through questionnaires and the subjective assessment of users. These variables were mostly related to usability factors. Of the nine articles that captured usability ([Doolani, Owens, et al., 2020](#); [Carlson et al., 2015](#); [deMoura & Sadagic, 2019](#); [Gavish et al., 2015](#); [Loch et al., 2019](#); [Murcia-Lopez & Steed, 2018](#); [Oren et al., 2012](#); [Velaz et al., 2014](#); [Werrlich, Nguyen, et al., 2018](#)), five used the standardized System Usability Scale of Brooke (1996) ([Doolani, Owens, et al., 2020](#); [deMoura & Sadagic, 2019](#); [Gavish et al., 2015](#); [Velaz et al., 2014](#); [Werrlich, Nguyen, et al., 2018](#)) and one used the USE questionnaire of Lund (2001) ([Loch et al., 2019](#)). The difficulty of the task was measured in six articles ([Carlson et al., 2015](#); [deMoura & Sadagic, 2019](#); [Murcia-Lopez & Steed, 2018](#); [Oren et al., 2012](#); [Velaz et al., 2014](#); [Roldan et al., 2019](#)), which was sometimes also operationalized as the difficulty of completing the task while interacting with the device (e.g., [Velaz et al., 2014](#)). The workload was assessed in five articles through the NASA TLX of Hart & Staveland (1988) ([Al-Ahmari et al., 2018](#); [Hou et al., 2015](#); [Koumaditis et al., 2019, 2020](#); [Werrlich, Nguyen, et al., 2018](#)). In three articles, the feeling of presence was assessed ([Carlson et al., 2015](#); [Cooper et al., 2018](#); [Roldan et al., 2019](#)). One single paper recorded the subjective assessment or satisfaction of the user's performance ([Gavish et al., 2015](#)). Additionally, in seven articles various single items were used to assess user's perceptions (e.g., stress, frustration, seriousness) ([deMoura & Sadagic, 2019](#); [Gavish et al., 2015](#); [Hou et al., 2013](#); [Langley et al., 2016](#); [Oren et al., 2012](#); [Velaz et al., 2014](#); [Westerfield et al., 2015](#)).

The effects of the investigated MR-based training formats regarding the defined performance measures TCT and TA as well as the subjective evaluations are summarized in [Table 4](#) and reported in the following in comparison to traditional training.

RQ4: What are the Effects of Using MR-Based Training Compared to

Traditional Training Regarding the Different Outcome Measures?

The effects and outcomes related to MR-based training compared to traditional training were analyzed in terms of the objectively measured performance outcomes (TCT, TA) and subjective evaluations described previously. Looking at the frequencies exclusively, AR-based training ($n = 12$) led to statistically significantly better performance than traditional training in 7 of 10 studies in which TCT was collected as the dependent variable (Hou & Wang, 2013; Hou et al., 2013, 2015; Koumaditis et al., 2019; Kwirotek et al., 2019; Loch et al., 2019; Westerfield et al., 2015), the same in two studies (Gavish et al., 2015; Webel et al., 2013), and statistically significantly worse in one (Werrlich, Nguyen, et al., 2018). In terms of TA, which was measured in 10 AR-based studies, the AR condition performed statistically significantly better in eight studies (Gavish et al., 2015; Hou & Wang, 2013; Hou et al., 2013, 2015; Loch et al., 2019; Webel et al., 2013; Werrlich, Lorber, et al., 2018; Werrlich, Nguyen, et al., 2018) and equally well twice (González-Franco et al., 2016; Westerfield et al., 2015). None of the articles tested long-term skill retention after AR-based training.

Looking at the different VR-based training methods ($n = 13$), we observed that concerning short-term, that is, immediate effects after training on TCT, statistically significantly better performance was achieved in one study using VR HMD compared to paper- and video-based training (Murcia-Lopez & Steed, 2018) and in another when compared to screen-based VR training (deMoura & Sadagic, 2019). In five articles, VR-based training led to equally good results compared to traditional training (Doolani, Owens, et al., 2020; Gavish et al., 2015; Hoedt et al., 2017; Oren et al., 2012; Velaz et al., 2014). Statistically significantly worse results compared to traditional training regarding TCT were found in three articles (Al-Ahmari et al., 2018; Carlson et al., 2015; Koumaditis et al., 2020). Moreover, Cooper et al. (2018) reported positive effects of additional audio and haptic cues on TCT. Regarding TA, VR-based training led to equally good

results as traditional training in five studies (Doolani, Owens, et al., 2020; Gavish et al., 2015; Hoedt et al., 2017; Oren et al., 2012; Velaz et al., 2014). In three studies, VR-based training was statistically significantly worse in TA than traditional training (Al-Ahmari et al., 2018; Koumaditis et al., 2020; Murcia-Lopez & Steed, 2018). Roldan et al. (2019) found different results for easy and complex assembly tasks, and deMoura & Sadagic, 2019 showed again that VR HMD was statistically significantly better than screen-based VR. With regard to long-term skill retention, four studies showed that VR-based training led to the same or even statistically significantly better results than traditional training, even when the immediate effects were not as good as traditional training (Carlson et al., 2015; Doolani, Owens, et al., 2020; Langley et al., 2016; Murcia-Lopez & Steed, 2018).

Beyond the overview presented above, an in-depth look into the study context (i.e., lab vs. field study), task type, and training specifications provided further insights into the effects of MR-based training. In each section, we first summarized the effects of the respective MR technologies AR projectors, AR handhelds, AR HMD, screen-based VR, and VR HMD as training media in comparison to other training formats. Subsequently, we analyzed how the effects related to objective performance measures and subjective evaluations varied depending on task type and training specifications. Main findings on the effects of MR-based training and statistically significant group differences are summarized in Table 4, where the specifications of the assembly task are indicated by stating whether joining, handling, inspecting, and/or adjusting activities were involved.

AR PROJECTORS

AR projectors as MR technologies were located on the revised reality-virtuality continuum on the left side, that is, close to the real environment, indicating a high extent of world knowledge. In total, four articles focused on AR projectors as MR technology to train assembly tasks (Hou et al., 2013; Hou & Wang, 2013; Hou et al., 2015; Loch et al., 2019). Overall, the evaluation and analysis showed that the

projector-based AR training led to statistically significant better results in terms of TCT and TA when compared to paper-based manuals or video-based training. All assembly tasks included joining, handling, inspecting, and adjusting activities. Subjective evaluation of AR projectors showed that AR projectors were perceived as easy to use and statistically significantly reduced workload compared to paper-based training. The individual findings of the studies about objective measures such as TCT, TA, and training time as well as results of subjective evaluations are summarized below.

Effects of AR Projector-based Training on Objective Performance Measures

Hou and co-authors (2013) compared an animated AR projector system with a paper-based manual system in two different experiments ($n = 20$ and $n = 30$) to assess performance. In a lab-based Lego® assembly, the authors found that participants made statistically significantly fewer mistakes (TA) and took statistically significantly less time (TCT) to assemble when being trained with the AR-based projection monitor. An improvement in the learning curve was illustrated by the fact that AR-trained participants were able to remember more assembly instructions from the previous training task than those trained with the paper manual. Hou and Wang (2013) further investigated the gender-specific effects of the use of AR (here used on a monitor) on performance ($n = 28$). The experimental group used AR, while the control group was trained using a paper manual with 3D elements. Both groups consisted of seven males and seven females. In comparison to the paper-manual control group, both male and female participants showed statistically significantly better performance in terms of TCT and TA. A statistically significant gender difference was only found concerning the control group, caused by the fact that manual-based training was more effective for males. In another study, Hou et al. (2015) investigated how AR affected performance and workload in a lab-based construction piping scenario ($n = 18$). Compared to training

with isometric drawings, the assembly novices trained with AR performed statistically significantly better, since they made half the number of errors (TA) and took only half the time to complete the task (TCT) and rework activities. The superior results of AR projectors were also confirmed by the study of Loch et al. (2019), who showed that using AR-based projection on a workbench with physical objects was statistically significantly better in terms of TCT and TA ($n = 20$) compared to a video-based system. In their study, they used a pick-to-light principle that enabled direct feedback for correct or wrong steps in the assembly of a circular flange onto a baseplate.

Effects of AR Projector-based Training on Subjective Evaluations

Overall, AR projectors were rated better in terms of ease of use, usefulness, and workload reduction in the subjective evaluation. Hou and co-authors (2013) showed that the paper-based training was perceived as complicated and cumbersome. However, a statistically significant difference was only reflected in the perceived workload, which was significantly higher in almost all scales of the NASA TLX (Hart & Staveland, 1988) for the paper manual group than in the group guided with AR. No difference was reported for the subscale *subjectively perceived performance* of the NASA TLX. Hou and Wang (2013) did not include subjective evaluations in their study. In the piping assembly task (Hou et al., 2015), assembly novices reported perceiving the AR system as easier to use and navigate, while experiencing statistically significantly lower workload (NASA TLX; Hart & Staveland, 1988) during the task. Loch et al. (2019) descriptively analyzed the USE questionnaire (Lund, 2001), containing 30 items on usefulness, ease of use, ease of learning, and satisfaction. The authors found that satisfaction and usefulness were rated higher when using the AR projector system compared to the video system. However, statistically significant group differences were not reported. Ease of learning was perceived equally positive for both systems.

AR HANDHELDS

Similar to AR projectors, AR handhelds (such as tablets) were used to augment a real training environment with virtual objects. Three of the 24 reviewed studies used AR handheld systems for assembly training and showed that this kind of AR-based training led to statistically significantly better results in TCT in comparison to paper-based isometric drawings (Kwiatk et al., 2019). Equal performance in TCT was shown in comparison with video-based training (Gavish et al., 2015; Webel et al., 2013). In these studies, statistically significantly better performance was shown in TA using AR handhelds compared with video. In particular, AR led to fewer unsolved errors, i.e., errors that were not corrected by the user (Gavish et al., 2015). Overall, all applications were tested in field settings or with employees from the production context using assembly tasks with joining, handling, inspecting, and adjusting activities. Subjective evaluations revealed that AR was perceived to have high usability and that AR-based 3D visualization was assessed as helpful in the assembly process. The following section describes these findings in detail.

Effects of AR Handheld-based Training on Objective Performance Measures

Gavish and co-authors (2015) presented two independent experiments, with one experiment relating to the comparison of an AR system with a video-based control group ($n = 20$) and another examining a VR system compared to a video-based control group ($n = 20$). In this section, the results regarding the AR system compared to the control group are reported. Later on, the second study of their article will be reported under VR systems. The tested AR training system consisted of a tablet PC with a touchscreen that was used to work directly on the machine. In terms of training time, the AR-based training took statistically significantly longer than the control group, which was trained with video. The AR group showed statistically significantly fewer unsolved errors, meaning that participants trained with AR identified and corrected errors more often. Otherwise, there were no differences

between the groups in terms of TCT and the number of errors solved. Comparable results were also shown in the study of Webel and co-authors (2013), who studied a multimodal AR tablet system compared to video-based training for the assembly of an electro-mechanical actuator with experienced workers ($n = 20$). While TCT and the number of errors solved were not different from the video control group, the number of unsolved errors (TA) was statistically significantly lower in the AR group compared to the control group. Furthermore, Webel and co-authors (2013) reported that training with AR took slightly longer (851.9 s) than with video-based training (682.0 s).

While these two studies showed that using AR handhelds for training led to comparable results as video-based training, but statistically significantly reduced unsolved errors, the study by Kwiatek and co-authors (2019) revealed that using AR handhelds led to better results when compared to paper-based training. Here, participants were trained to assemble a complex pipe spool using conventional isometric drawings or AR-based guidance on a tablet. All participants were previously classified into two groups (engineers vs. pipefitters) and into high, medium, and low spatial skills. Across both groups, the use of AR on a tablet statistically significantly reduced assembly time (TCT) and rework time. The authors showed that those whose cognitive abilities were considered to be low benefited more from the AR application than the other participants.

Effects of AR Handheld-based Training on Subjective Evaluations

Gavish and co-authors (2015) reported no comparison between AR and control group in the subjective evaluations, but between both experiments on AR and VR—although both systems were evaluated independently. The authors showed that AR was rated with statistically significantly higher usability and better transfer of training compared to VR. Participant feedback in the study by Kwiatek et al. (2019) revealed that 3D design and visualization were perceived as helpful to facilitate the assembly

and rework process. It needs to be noted that no standardized questionnaire was used in these studies.

AR HMDS

While AR projectors and handhelds allowed the user to look away from virtual augmentations, the latter were inevitably integrated into the user's field of view when using AR HMDS. In the following, we analyzed the effects of using AR as a head-mounted device as a training medium. The effects of AR HMDS were investigated within five articles and mostly conducted in real training settings for all assembly activities joining, handling, inspecting, and adjusting (González-Franco et al., 2016; Koumaditis et al., 2019; Werrlich, Lorber, et al., 2018; Werrlich, Nguyen, et al., 2018; Westerfield et al., 2015). These studies revealed that in comparison to trainer-based formats, training with AR HMD required statistically significantly more time. However, using AR HMDS in assembly training led to statistically significantly higher performance than trainer-based training in terms of TCT, but to the same or even statistically significantly worse accuracy (TA). Additional features in AR (such as quizzes) led to a statistically significant increase in TA. The subjective evaluations showed that AR HMDS were rated as requiring statistically significantly less effort than trainer-based training. All results are further described in the following.

Effects of AR HMD-based Training on Objective Performance Measures

Three studies reported the effects of using AR HMDS for assembly training compared to trainer-based formats. First, the study of González-Franco and co-authors (2016) showed that participants in AR HMD training (using see-through Oculus Rift) achieved the same performance in the assembly of an aircraft maintenance door as participants in traditional face-to-face training ($n = 24$). Here, the authors assessed both factual knowledge in a knowledge retention multiple-choice test and procedural knowledge in a knowledge interpretation test,

where participants were asked to perform the assembly step by step. There was no difference in the scores of knowledge retention and interpretation between the conditions. The latter was treated as the accuracy of the task (TA) in our analysis. Training times were statistically significantly higher in the AR condition compared to a trainer-based format. Second, a comparison between AR HMD (Microsoft HoloLens) and trainer-based training was conducted by Werrlich, Lorber, et al. (2018) in an assembly of a real engine ($n = 36$). Here, training time was statistically significantly higher in the AR HMD group, too. Although participants in the AR HMD group made 10% fewer picking mistakes, 5% fewer assembly order mistakes, and caused 60% less rework, overall assembly quality and quality per tact were statistically significantly higher in the trainer condition. Similar to González-Franco et al. (2016), Werrlich, Lorber, et al. (2018) decided to use a questionnaire capturing factual knowledge acquisition of the performed assembly task besides measuring TA as an assembly performance indicator. The results showed that participants who trained with AR retained factual knowledge equally well as participants who were instructed by a trainer. As a third study comparing AR HMD and trainer-based training, Koumaditis et al. (2019) used a turning table workstation, where a six-part product assembly product composed of supports, mechanical elements, and electronic components was assembled by handling and connecting components without the help of other tools. They showed that the group trained with AR HMD performed statistically significantly better than the group trained by a trainer in terms of TCT ($n = 10$).

Two other studies showed that using AR HMDS with additional features such as intelligent tutors or quizzes in training led to better results than AR HMD-based training without such features. Werrlich, Nguyen, et al. (2018) compared two AR HMD modalities ($n = 30$), both covering different skill levels from beginner to expert. In one of their AR HMD training, an additional quiz determining participants' procedural knowledge was implemented before conducting the final assembly of an

engine. Here, participants had to select the correct sequence of assembly steps and received immediate feedback from the system. Compared to the AR HMD-based training without the quiz, participants in the quiz condition took statistically significantly longer to complete the final assembly task (TCT) but committed statistically significantly fewer sequence errors (TA). [Westerfield and co-authors \(2015\)](#) integrated an intelligent tutor with feedback functionalities into the AR HMD system and tested its effects against an AR HMD system without immediate feedback ($n = 16$). Participants were trained to conduct a motherboard assembly, consisting of identifying and installing five motherboard components: memory, processor, graphics card, TV tuner card, and heatsink. [Westerfield et al. \(2015\)](#) found that the group receiving AR-based training with the intelligent tutor was statistically significantly faster in the final assembly task in terms of TCT compared to the group being trained by the conventional AR HMD system without feedback, but both groups made comparable numbers of errors (TA).

Effects of AR HMD-based Training on Subjective Evaluations

Overall, it appeared that the use of AR HMD for assembly training was perceived as positive and, in some cases, even helped to reduce workload. [Werrlich, Lorber, et al. \(2018\)](#) evaluated the usability of the tested system (Microsoft HoloLens), which was assessed with a mean system usability (SUS) score of 73.5, indicating acceptable to good usability. SUS Scores are ranging from 0 to 100 ([Brooke, 1996](#)), where 68 is considered as average usability score ([Sauro, 2011](#)). However, the authors discussed that making it more user-friendly would enhance its potential. [Koumaditis and co-authors \(2019\)](#) assessed the NASA TLX and indicated that the participants using AR HMD reported requiring statistically significantly less effort to accomplish the task than the participants guided by a trainer. No differences were found for the other scales of the NASA TLX. Testing the AR HMD system with an additional quiz to ensure training transfer revealed that participants' perceived workload did not differ from participants' evaluations who

were trained with the conventional AR HMD method without the quiz ([Werrlich, Nguyen, et al., 2018](#)). System Usability was rated equally excellent for the AR HMD system without (SUS = 90.50) and with the quiz (SUS = 91.83). The results of subjective evaluations in the study of [Westerfield et al. \(2015\)](#) supported previously reported findings. Here, participants were asked to fill out a questionnaire and indicate whether, for example, they perceived the tutor as effective, and whether they felt physically or mentally stressed during the training process. The authors reported that both AR HMD with an intelligent tutor and conventional AR HMD systems were positively evaluated and showed no differences in these subjective evaluations (e.g., perceived stress, effectiveness, frustration).

SCREEN-BASED VR

The category of screen-based VR technologies was characterized by the fact that although virtual, that is, 3D modeled environments were presented, and they were displayed on a screen or monitor. All articles included the four assembly activities joining, handling, inspecting, and adjusting ([Al-Ahmari et al., 2018](#); [Carlson et al., 2015](#); [Cooper et al., 2018](#); [Gavish et al., 2015](#); [Hoedt et al., 2017](#); [Langley et al., 2016](#); [Oren et al., 2012](#); [Velaz et al., 2014](#)). The superordinate category screen-based VR was located in the middle of the revised reality-virtuality continuum (based on [Skarbez et al., 2021](#)), while the subcategory monoscopic VR (i.e., 3D environments that are presented without stereoscopic visuals) provided a higher extent of world knowledge and could thus be located closer to the real environment. Screen-based VR with stereoscopic glasses, on the other hand, enabled depth perception through the addition of, for example, shutter glasses and thus a more intensive immersion in the 3D environment. However, both variants did not exclude users' perceptions of the environment. Results indicated that monoscopic screen-based VR showed comparable effects on TCT as trainer-based training and video-based training and resulted either in the same ([Gavish et al., 2015](#); [Hoedt et al., 2017](#); [Velaz et al., 2014](#)) or statistically significantly higher task accuracy

(Langley et al., 2016). Studies, where screen-based VR was used in combination with stereoscopic glasses, showed that participants who were trained with physical objects performed assembly tasks statistically significantly faster and more accurately than participants trained with screen-based VR and stereoscopic glasses (Al-Ahmari et al., 2018; Carlson et al., 2015). However, with more experience in VR and the use of additional features and cues, results seemed to improve (Cooper et al., 2018; Oren et al., 2012).

In the following, we will first review the effectiveness of monoscopic screen-based VR training with regard to objective performance measures and subjective evaluations. Second, studies using screen-based VR with stereoscopic visuals will be reviewed. Subsequently, the effects of the latter on subjective evaluations will be examined.

Effects of Monoscopic Screen-Based VR Training on Objective Performance Measures

Four out of eight articles on screen-based VR used monoscopic VR (Gavish et al., 2015; Hoedt et al., 2017; Langley et al., 2016; Velaz et al., 2014). These studies will be described in more detail in the following. Gavish and co-authors (2015) used a VR platform consisting of a screen displaying a 3D graphical scene and a haptic device where trainees were able to manipulate the tools and components of the virtual scene to train the assembly of an electronic actuator of a motorized modulating valve ($n = 20$), while the control group was instructed via video. Performance was measured as TCT and number of solved and unsolved errors (TA), where no differences between video and non-HMD VR conditions occurred. However, participants took statistically significantly longer to train with the VR system than with the video system. Similar effects were reported by Hoedt and co-authors (2017), who examined a screen-based VR-trained group vs. a non-trained group ($n = 26$) during an assembly of a medium complex [®]MECCANO sub assembly. The screen-based VR training was realized using a tablet, a screen, and a Kinect V2 [®]Microsoft. During the training period,

participants' hands were tracked in real-time while assembling the virtual objects. At the first measurement time point (T1), the VR-trained participants showed statistically significantly better scores in terms of TCT. At the second measurement point (T2), the control group was considered "trained" and again compared to the first measurement point of the VR group. We only included T2 in our analysis to ensure comparability of the results with regard to trained control groups. Even though the non-trained control group took slightly longer to complete the T1 assembly and thus the training, there were no differences between screen-based VR and the trained control group from T2 onwards. The overall learning rate was found to be equal, although the absolute difference in assembly times was 27% faster in the control group. As the third study in the field of monoscopic screen-based VR, Langley and co-authors (2016) compared their screen-based VR training system to a trainer-based learning environment and showed that in the long-term, their monoscopic VR training resulted in achieving comparable TCT but statistically significantly higher TA. More specifically, they investigated the effectiveness of a virtual training environment on screen using a Wii controller compared to a trainer plus paper-based training procedure for a car door assembly task. After one week, participants ($n = 30$) were asked to perform the final assembly task without guidance. There were no differences in long-term skill retention concerning TCT, but there was a statistically significant difference for long-term skill retention concerning overall and trainer-corrected errors (TA), indicating that participants who were trained with the screen-based VR system made fewer errors one week after training. No difference for TA was found concerning self-corrected errors. Finally, Velaz et al. (2014) investigated different interaction modes in screen-based VR and compared them with video-based training. Four out of five groups trained to assemble an electrohydraulic valve in VR with either a computer mouse, a haptic device, or two configurations of a markerless motion capture system (with 2D or 3D tracking of hands). The screen-based VR + 3D markerless motion capture system was referred to as experimental group here. The fifth group, that is, control group, was

trained with video. Overall, no difference between the training methods was shown regarding TCT and TA. In terms of training time, the screen-based VR + 3D markerless motion capture system took statistically significantly more time than video training as well as training with other interaction devices such as using a mouse, a haptic device, or 2D. In the following, the effects of monoscopic screen-based VR on subjective evaluations will be examined.

Effects of Monoscopic Screen-Based VR Training Training on Subjective Evaluations

Gavish and co-authors (2015) asked participants whether they thought the training with VR rapidly enhanced their skill level, which was answered with “yes” by all participants. In their study, the authors further examined how the presented VR system was subjectively evaluated in comparison to their AR handheld system presented earlier. In a direct comparison, the compared AR handheld system was rated statistically significantly higher than the screen-based VR system in terms of satisfaction with performance, usability, and willingness to recommend the system. The evaluation of different VR interaction types in the study of Velaz and co-authors (2014) revealed that usability and interaction difficulty was perceived the same for all tested interaction types (computer mouse, haptic device, 3D, and 2D) and the video control group. Langley and co-authors (2016) assessed the subjective evaluations of their virtual system in an additional evaluation study by observing and interviewing participants and analyzing the results with theme-based content analysis. Three main usability issues were identified accordingly: (1) Participants stated that extra instruction would be required concerning the Wii controller and relating to its use, (2) the instructions for discarding the screw were insufficient, and (3) that instructions for moving the whole body to change the view area were inadequate. In addition, the participants indicated in their comments that the system was easy to use and that it was more enjoyable than trainer-based training.

Effects of Screen-Based VR Training with Stereoscopic Glasses on Objective Performance Measures

The other four articles used screen-based VR in combination with stereoscopic shutter glasses (Al-Ahmari et al., 2018; Carlson et al., 2015; Cooper et al., 2018; Oren et al., 2012). Al-Ahmari et al. (2018), for example, compared a screen-based VR system with stereoscopic shutter glasses, conventional drawings, and 3D prints at three different scales to train the assembly of a mid bearing ($n = 25$). In total, five groups were tested with each group consisting of five participants. The authors reported that both screen-based VR and drawings performed statistically significantly worse regarding time (TCT) and accuracy (TA) than 3D prints. The study result of Carlson et al. (2015) supported the finding that training with physical objects led to faster task completion than screen-based VR training, but further revealed that using colors in VR had a positive effect on remembering training content better in the long term. In their study, Carlson and co-authors (2015) used a burr puzzle assembly to investigate the influence of color on skill retention success comparing a screen-based VR environment with stereoscopic glasses and a data glove with training using physical objects ($n = 63$). The first run of burr puzzle assembly showed that physically trained participants were statistically significantly faster in TCT than virtually trained participants—regardless of whether they trained with wood or colored objects first. At the point of repeated measurement (T2), the authors showed that the screen-based VR group that trained with color first no longer differed from the physical group regarding TCT. This was explained by the fact that participants reported using remembering the color as their recall strategy. Another comparison of training with physical objects versus stereoscopic screen-based VR was done by Oren et al. (2012), who investigated the assembly of a 3D wooden burr puzzle in a controlled lab experiment with two conditions: On the one hand, training with a stereoscopic head-tracked virtual system with haptic devices and data gloves (VR condition) and on the other hand training with real physical blocks ($n = 10$). The participants in the VR condition took three times

longer to train than the control group. During the test phase, this effect reversed and the virtually trained participants were four times faster in completing the physical task (TCT) than the control group. Nevertheless, this effect was not statistically significant.

Cooper and co-authors (2018) captured the effects of multisensory cues in a screen-based VR environment with wireless shutter glasses and gloves. Participants ($n = 17$) performed a wheel change using a pneumatic tool and received tactile (vibration feedback), audio (task-appropriate sounds), or visual (virtual hands of the participant turning yellow when in contact with virtual objects) cues depending on the condition. While there was no difference between multimodal, bimodal, or unimodal feedback in terms of TCT, audio cues statistically significantly improved TCT compared to white noise, and tactile cues statistically significantly improved TCT compared to the condition where tactile cues were absent. There was no difference regarding TCT between screen-based VR training with and without visual cues.

Effects of Screen-Based VR Training with Stereoscopic Glasses on Subjective Evaluations

Compared to physical 3D prints and paper, screen-based VR with stereoscopic glasses was perceived to cause equal workload in the study of Al-Ahmari and co-authors (2018). Carlson and co-authors (2015) indicated in their study that physically trained participants rated the difficulty and ease of use of the training environment as statistically significantly easier than those who were virtually trained, but both training environments were rated as equally realistic. Oren and co-authors (2012) showed that their virtual system was rated as statistically significantly less easy to use than the physical blocks, whereas task difficulty, realism, and helpfulness of both conditions were not differently assessed. Examination of the multisensory cues such as audio cues or haptic devices showed that additional cues were associated with an increased feeling of presence and improved performance. Focusing on presence as a subjective measure, Cooper and co-authors (2018) revealed that audio, tactile, or visual cues enhanced the feeling of presence. The

latter was assessed with a seven-item questionnaire on immersion and involvement, that is, the sense of being present in the VR environment. Moreover, higher presence correlated with faster TCT and users' subjective experience was proven to be related to the overall task performance, that is, an increased sense of presence during VR was associated with faster task performance.

VR HMD

With respect to the revised reality-virtuality continuum, VR HMDs were located close to the right, that is, more in the direction of the virtual environment, and thus contained lower levels of world knowledge, but enabled higher levels of immersion (based on Skarbez et al., 2021). The effects of hiding the real environment by using closed goggles on training success will be analyzed in the following. VR HMD was used as a form of MR-based training in five articles (Doolani, Owens, et al., 2020; deMoura & Sadagic, 2019; Koumaditis et al., 2020; Murcia-Lopez & Steed, 2018; Roldan et al., 2019). Overall, VR HMD were often used for more abstract tasks like 3D puzzles, which did not cover all, but at least one of the four assembly activities. In the following, the evaluation of the objective performance measures of VR HMD is divided into the comparison with paper-based training, the comparison with different immersion levels, and the comparison with trainer-based use cases. The finding that initial shortcomings of VR HMD-based training could be reversed when the task was repeated after a certain time was not only found in objective performance measures but was also supported by subjective evaluations.

Effects of VR HMD-Based Training on Objective Performance Measures

The following four studies compared VR HMD training to paper-based training, showing that VR HMD mostly led to faster TCT during the final assembly. Moreover, they demonstrated that initial shortcomings of VR HMD training in terms of TA could be reversed in long-term retention tests. Doolani, Owens, et al. (2020) investigated a vocational immersive storytelling system (VR HMD) and compared it to a paper-

based and video-based training for the use of a mechanical micrometer, where handling, inspecting, and adjusting activities were required ($n = 30$). In terms of TCT, there was no difference between the conditions. In a long-term skill retention test, TCT was statistically significantly faster in the VR group compared to paper, but no difference was observed comparing VR and video-based training. TA was measured in the training session and again after seven days (TA). While showing no differences during the first session in TA, the VR HMD system led to statistically significantly higher accuracy compared to video and paper at the second measurement point. Training time was the same for VR and video groups, but VR HMD training was statistically significantly faster than paper-based training. Murcia-Lopez and Steed (2018) supported the observed superiority of VR HMD training over paper-based training. In their study, the authors investigated different physical (paper, paper and video, paper and physical blocks, paper and video plus physical blocks) and virtual conditions (VR plus virtual blocks, VR plus virtual blocks and animations) in a complex 3D burr puzzle task, consisting of handling, inspecting, and adjusting activities ($n = 60$). The overarching best condition turned out to be the combination of paper, video, and blocks and was used as a reference condition. The VR condition performed as well as the reference condition in terms of TCT, but statistically significantly better than paper or paper plus video. However, concerning TA, participants using VR HMD as a training system performed statistically significantly worse at the first measurement point. No more differences between the conditions were shown regarding TCT and TA in the long-term skill retention test. Regarding training time, participants in the VR condition took statistically significantly longer than those in the paper or paper and video condition, but there was no difference to the reference group. Roldan et al. (2019) further showed that prior experience in VR and the simplicity of a task contributed to the success of VR HMD training. The authors tested a process mining-based VR system that was built to ensure the transferability of training content. Twenty participants conducted four different assemblies

including handling, joining, inspecting, and adjusting activities with four difficulty levels. Two assemblies each were taught with paper, two with the immersive VR HMD. It was found that participants in the VR HMD condition achieved statistically significantly higher scores (a combination of TCT and TA) in the easiest assembly, with no difference in the other assemblies. The authors reported that participants with prior experience in VR scored higher after VR HMD training, while the others scored higher after the physical training conditions. However, this difference was not statistically significant.

In contrast to the above-mentioned studies, the following study did not compare VR HMD to traditional paper-based training but rather focused on investigating the influence of different immersion levels, that is, VR HMD versus screen-based VR. deMoura and Sadagic (2019) compared four different VR conditions in a between-subjects design, namely: stereoscopic VR HMD, monoscopic VR HMD, stereoscopic screen-based VR, and monoscopic screen-based VR with $n = 68$. The task of assembling a 3D-printed toy helicopter consisted of joining, handling, inspecting, and adjusting activities. Participants who were trained with an immersive stereoscopic display (Oculus Rift) were statistically significantly faster in task completion (TCT) than participants being trained with monoscopic VR HMD, stereoscopic screen-based VR, or monoscopic screen-based VR. No differences were found between stereoscopic and monoscopic VR HMD concerning TA, but performance with stereoscopic VR HMD was statistically significantly more accurate than with both screen-based VR conditions.

One of the evaluated studies using VR HMD focused on the influence of trainer assistance in VR-based and physical training. In this study by Koumaditis et al. (2020), participants were trained to assemble a complex 3D cube in a physical or VR HMD training with or without trainer assistance ($n = 100$). All assembly activities from joining to adjusting were included. TCT was statistically significantly shorter in the physical condition than in the VR condition, but the presence of the trainer did not affect performance in either case. This training effect pattern was also found with respect to TA.

Effects of VR HMD-Based Training on Subjective Evaluations

The results regarding objective performance measures were supported by the subjective evaluations. It turned out that more experience in VR led to less mental demand and effort. Individual studies also reported that the usability of VR HMD was perceived higher than the usability of paper manuals.

Doolani, Owens, et al. (2020) showed that their storytelling VR system was perceived to be of higher usability than video and paper-based training. Murcia-Lopez and Steed (2018) showed that the VR HMD condition with animated instructions was associated with statistically significantly perceived lower task difficulty rather than conditions that only contained paper or paper plus physical blocks. Moreover, the system usability score of the VR HMD condition was statistically significantly higher with SUS = 76, while paper-based training achieved a SUS of 62, considering that scores between 60 and 80 indicated acceptable to good usability. Roldan and co-authors (2019) reported that the immersive VR HMD system was assessed statistically significantly better than the paper manual in terms of perceived mental demand, subjective performance, perception, learning, and result (i.e., the subjective impression about the results of the test). deMoura and Sadagic (2019) showed that the VR HMD with stereoscopic view was perceived as easiest to use, level of realism as highest and difficulty while interacting as the lowest compared to non-stereoscopic VR HMD, 3D environment on a screen or 2D display. Regarding simulator sickness, eye strain was observed as a symptom for stereoscopic view and general discomfort for non-stereoscopic. Koumaditis and co-authors (2020) found that task load positively predicted TCT, indicating that higher task load led to longer completion times, but at the same time negatively predicted TA. This result was found in both of their two experiments on AR and VR training systems.

SUMMARY OF RQ4

In the following, we will summarize the main results of RQ4 on the effects of using MR-based training compared to traditional training regarding objective performance measures and subjective evaluations. Overall, it can be stated that the results differed according to the technology used and the technology or training method being compared to. Eight of 12 investigated AR training systems (including all three AR technology types projectors, handhelds, and HMD) showed systematic advantages over traditional training (especially paper-based training) concerning objective performance measures such as TCT and TA. In only one of the 12 studies, an enhanced AR HMD system showed a worse result in the objective performance measure TCT—here, however, not in comparison to traditional training, but to an AR training without the additional feature of a knowledge quiz, which could have been the decisive factor. The subjective evaluations showed that AR-based training was consistently rated as providing high usability and kept resulting in a perceived reduction of task load.

More variety was apparent in screen-based VR and VR HMD-based trainings in terms of both performance measures and subjective evaluations. In most cases, screen-based VR or VR HMD led to comparable performance as traditional training. To start with the former, screen-based VR achieved comparable results in TCT and TA when compared to traditional training, that is, paper- or video-based training, regardless of whether monoscopic or stereoscopic visuals were used. Compared to training with physical objects, screen-based VR achieved statistically significantly worse results in two of three cases. Presenting additional cues (such as tactile feedback or task-related sounds) during training enhanced the performance of participants. Considering VR HMD-based training, it was found to perform better regarding TCT than paper-based training formats, which was especially evident in easier or less complex assembly tasks. Regarding TA, VR

HMDs were either comparable or not as good as traditional training. Both screen-based VR and VR HMD-based training often did not result in the same performance as traditional training immediately after training but achieved comparable or better results in long-term skill retention, which was shown across several studies. Concerning subjective evaluations, VR was shown to be rated with higher usability than paper- and video-based training but was perceived as more difficult to use than physical objects. At the same time, a higher feeling of presence was positively correlated with increased performance. With the results of different MR-based training on objective performance measures and subjective evaluations presented, needs for further research were subsequently derived within the reviewed articles and presented below.

RQ5: WHAT RESEARCH GAPS ARE REPORTED BY THE AUTHORS?

The authors of the reviewed articles identified certain research gaps, which were analyzed by means of a qualitative content analysis. The suggested individual thematic fields were grouped into categories related to changes in study design, suggested modifications of independent and dependent as well as investigation of individual differences.

Concerning the overall study design and setup, two articles emphasized that the tested systems should be transferred and evaluated in more realistic use cases (Hou et al., 2015; Roldan et al., 2019). Moreover, it was discussed that larger and more diverse samples are needed, investigating how different individuals or groups (e.g., novices vs. professionals) would profit from MR systems (Hou et al., 2013).

Suggested changes in independent variables are referring to different training modalities, technology features as well as task-related changes. Regarding the training modalities, it was mentioned that a comparison of various training platforms and presentation formats is needed to deepen the confidence in the implementation of MR applications in practice (Kwiatk et al., 2019; Webel et al., 2013;

Werrlich, Nguyen, et al., 2018). Comparing AR and VR systems in the same study with comparable features was considered an important next step (Gavish et al., 2015; Roldan et al., 2019). Moreover, research on the effect of different guidance cues or instructions (e.g., color) (Al-Ahmari et al., 2018; Murcia-Lopez & Steed, 2018) and trainer impact (Koumaditis et al., 2020) on objective and subjectively perceived assembly performance (Cooper et al., 2018) was mentioned. Furthermore, enabling training on-demand using intelligent systems (Doolani, Owens, et al., 2020) was considered an important next step in the reviewed articles. The need to explore further technology features was also expressed. Here, it was mainly referred to as the implementation of different interaction modalities, such as vibrotactile and auditory feedback (Loch et al., 2019; deMoura & Sadagic, 2019; Velaz et al., 2014). Task-related changes were suggested to test the applications in longer and more complicated tasks (Westerfield et al., 2015), expecting AR and VR systems to have a significant advantage over traditional training because of their potential to focus on enhancing the cognitive understanding of the task (Gavish et al., 2015).

Regarding dependent variables, it was stressed that using biophysical measurements (such as arousal and stress levels) in addition to the subjective evaluations could facilitate triangulating data and provide a clearer set of findings (Koumaditis et al., 2019). Investigating how to prevent the effect of forgetting in order to enhance long-term skill retention was also mentioned (Hoedt et al., 2017). Two studies claimed that investigating correlations between objective performance and theories of self-representation, first-person interaction, and the feeling of presence would be of relevance for further research (Cooper et al., 2018; González-Franco et al., 2016).

Finally, the reviewed studies suggested investigating further potentially influencing individual differences, that is, the personal motivation in using new technologies (Werrlich, Lorber, et al., 2018) or individual recall strategies (Carlson et al., 2015).

DISCUSSION

The present scoping review aimed to transform the broad and diverse literature of MR-based training in manual assembly tasks into comprehensive statements on the effects of different formats of MR-based training on objective performance measures and subjective evaluations. In this review, MR was used as an umbrella term for AR-based, screen-based VR, and VR HMD-based training. Through the analysis of 24 selected articles on MR-based assembly training, the review identified industry needs for MR-based training, classified currently used and researched MR technologies, and provided an overview of a set of outcome measures that have been employed in the reviewed studies. Concerning the effects of MR-based training on objective performance and subjective evaluations, results indicated promising results for the use of AR projectors, AR handhelds, and AR HMD-based training. In contrast, neither screen-based VR nor VR HMD-based training did show consistently better results in direct comparison to traditional training. In the following, the main findings on the effects of MR-based training are discussed in relation to the identified industry needs that have to be considered in the industry when it comes to choosing an appropriate training method for manual assembly: Enhancing *worker performance*, *effectiveness*, and *efficiency*. After the discussion of the limitations of this scoping review, we will conclude by assessing the results and providing an outlook on what needs to be considered in future research.

Main Findings and Statements on MR-Based Training for Manual Assembly

One of three key industry needs that was mentioned in the reviewed articles was to improve worker performance, aiming to reducing errors and enhancing the satisfaction of workers in the assembly process. In this context, MR-based training was expected to reduce cognitive load and avoid misunderstandings (e.g., by direct visualization of instructions on the component). Within the reviewed studies, we found that MR-based training led to several positive

aspects with regard to worker performance. AR-based training, for example, was found to reduce workload compared to paper- and trainer-based training, which could be explained by the fact that the virtual objects and instructions were directly displayed in the users' view. Compared to video-based training, AR-based training was found to be perceived with higher usability. Especially for projector-based AR and AR HMD, this advantage could be associated with hands-free interaction using the system, making bimanual coordination easier. This is also an interesting implication for tablet-based AR, where this advantage could also be exploited through a fixation or mobile tablet arm. VR-based training, on the one hand, was shown to be perceived with comparable or higher usability than paper- and video-based training, which could be due to the fact that the VR environment makes it possible to simulate the completion of the task in an environment that is close to everyday work, for example, through storytelling or appropriate visualization of the environment. On the other hand, VR was rated with lower usability compared to training with physical objects. This suggests that the actual grasping and feeling of, for example, individual components should still be an essential part of training. From this, it can be deduced that the potential of VR could be increased in future training by viewing and touching the real components as well as the final product before the VR exercise starts. However, using additional audio, tactile, or visual cues in VR-based training positively influenced users' experience with regard to the degree of involvement in the task, that is, presence, which in turn led to higher performance than, for example, VR-based training without additional cues. This was especially evident for task-related cues, which supported the participants in focusing on the respective task. Thus, cues that highlight and corroborate task-related actions might enhance the natural interaction with the system and the task to be trained. Overall, more experience with the technology led to better results. Thus, especially with regard to the subjective factors usability and workload, MR-based training might be a powerful tool. However, it became apparent that subjective evaluation measures

were operationalized in very different ways. Moreover, these studies did not always use existing standardized and validated questionnaires (such as John Brooke's System Usability Scale, [Brooke, 1996](#), or the NASA Task Load Index, [Hart & Staveland, 1988](#)), so that comparability of results across studies cannot be ensured. Overall, it can be stated that subjective perceptions are a core concept of user performance that should be further evaluated by, that is, including usability studies.

While these results on user performance were mostly related to the subjective evaluations of the MR-based training by the user, the second identified need of the industry, namely effectiveness, was related to objectively measured training success and thus assembly performance. Effectiveness was mostly assessed within the reviewed studies as training time during training and the time of completing a task (TCT), as well as number of errors (task accuracy, TA) during the final assembly. Overall, all AR-based training formats led to consistently better assembly performance with regard to TCT and TA than paper-based training. Compared to video-based training, AR-based training led to equal or better assembly performance. Compared to traditional training, AR-based training performed worse in task accuracy only once when compared to a trainer-based format. Additional features like intelligent tutors further improved the results of AR-based training. Thus, with its additional benefits in terms of subjective variables such as reduced workload, AR-based training promises to be a reliable medium for improving user performance and effectiveness. Screen-based VR or VR HMD-based training formats were mostly as good as traditional training. Compared to paper-based training, VR-based training led to equal or better results regarding TCT, but to mixed results regarding TA. Compared to video-based training or training with physical objects, no clear advantage of VR-based training occurred. Interestingly, VR-based training was found to be better when less complex assembly tasks were conducted. The still mixed results of VR-based training so far give reason to reflect more precisely if and when VR should be used as a substitute for traditional training, e.g., to increase safety or location

independence of the training. However, besides the current findings on TCT and TA, the time required for handovers and interfaces between different production areas and workplaces, and the ease of this workflow, might be also relevant when evaluating effectiveness. When assessing task accuracy, it should also be considered whether some errors should perhaps be weighted more heavily than others because they result in rework at later points in time. Similarly, a shorter training time, which was sometimes used as a main criterion for effectiveness, should be critically reflected. It is to be questioned whether a shorter training time is really beneficial if longer and more intensive training could in turn increase the overall assembly performance. Moreover, most assembly performance measures were assessed only at one point in time. At this point of time, however, very little is known about the effects of MR-based training on long-term knowledge and skill retention, which is an important factor when it comes to the application in industry. No AR-related studies included long-term skill retention tests. With regard to VR, only four studies evaluated long-term skill retention, but these studies showed promising effects of VR-based training in the long-term. From this, we conclude that the majority of reviewed studies have not yet exploited the full potential of MR technologies as a training medium. Future studies should therefore include the evaluation of long-term skill retention.

Efficiency, as a third industry need emerging from the reviewed articles, builds on the two other factors of performance and effectiveness. Efficiency was often mentioned as industry need in the context of the reduction and savings of financial, personnel, and time resources through the use of MR. In the reviewed studies, however, cost- or time-efficiency was not explicitly addressed or evaluated in most studies. In some articles, it was mentioned that it still took a lot of effort to program MR technologies in such a way that they can be used for the respective use cases and connected to existing interfaces and technologies. Some studies claimed to have used more cost-effective solutions by using existing materials such as a usual screen and monoscopic VR without providing details on costs or effort. Some authors discussed in the research outlook

that the workflow for creating MR-based training could be made more efficient by, for example, using existing CAD models. However, since no evaluable information on cost- or time-efficiency was provided beyond these more general aspects, these aspects were not considered within this review. This could be an important extension for further reviews.

Limitations of this Review

The five research questions examined in this scoping review have led to first conclusions about the effects of MR-based training on objective performance and subjective evaluations, which should be classified under consideration of the limitations of this review. First of all, the findings discussed within this review only represent a restricted sample of the overall undertaken research activities in the field of MR training in manual assembly tasks. In this context, it should be mentioned that a considerable number of articles had to be excluded since the results reported could not be reliably evaluated with the information and justifications given. Particularly in industry-oriented research, studies are conducted with relatively few participants ($n < 10$), often due to resource constraints. What was surprising to us, however, was the amount of studies which had to be excluded because they reported and interpreted results of parametric tests (e.g., ANOVA) applied to groups with, for example, $n < 10$ without testing or reporting test requirements. It is therefore urgent to create an understanding across disciplines as to which statistical procedures are suitable in order to obtain robust and valid results even with small sample sizes. Moreover, even most of the included studies that were able to report statistically significant effects of AR- and VR-based training did not report effect sizes that describe the magnitude of the demonstrated effects. For this reason, the report of effect sizes was not made a premise for inclusion of a study in this scoping review. However, it would be highly desirable if it became standard practice for studies to report effect sizes in the future, so that this statistical measure can be captured in future reviews and considered when drawing conclusions.

Second, in the present scoping review, the classification of different MR technologies into the categories AR-based training, screen-based VR training, and VR HMD-based training was built based on the given and sometimes only superficially description of the technologies in the reviewed articles. Thus, a precise assignment was sometimes difficult and might have deviated from their original classification in some cases. On the one hand, this clearly shows that MR represents a continuum of partially overlapping technologies. On the other hand, it also shows the need for a more differentiated description of the hardware and software used in the studies to avoid misunderstandings and misinterpretations.

Third, the diversity of research methods and, for example, objective performance measures prevalent in the literature made the comparability of results challenging. For instance, the designation of the objective outcome measures varied greatly. Accuracy consisted, for example, in the measurement of correct steps or of different error types, such as corrected errors or uncorrected errors, which were not further differentiated in the context of this review. To facilitate the comparability of studies in this area in the future, we proposed a uniform use of terminology describing objective outcome measures, such as Task Completion Time (TCT) and Task Accuracy (TA). This will allow the research community to quickly and clearly identify which measures were used and would facilitate comparing of results across different research articles.

Finally, it must be said that although the importance of subjective evaluations was emphasized, only studies that also collected objective performance measures were included in this review. Thus, pure usability studies were excluded, which might limit the validity of information about subjective perceptions of the technologies. Future reviews could therefore provide additional insights by focusing particularly on outcome variables such as user satisfaction, technology acceptance, usability, individual recall strategies, and perceived task load.

In addition to this, it would of course be highly desirable if instead of the present scoping review a systematic review could be conducted

in the field of MR-based training for manual assembly tasks in the future. Since at this stage, however, it must be stated that there is not enough data yet, further conclusive experimental studies on the efficacy of MR in this context have to be performed first.

CONCLUSION AND OUTLOOK

The present scoping review of 24 articles brings new insights into the effects of MR-based training on objective performance measures and subjective evaluations in the context of manual assembly tasks. Regarding the effects on objective performance measures, we found that AR projectors, AR handhelds, and AR HMD were shown to improve objective performance measures such as task completion time (TCT) and task accuracy (TA) compared to paper- and video-based training, while screen-based VR and VR HMD-based training mostly led to comparable assembly performance as traditional training. However, when long-term skill retention of the final assembly task was tested, it was shown that participants who trained with VR were often able to compensate for previously equal or worse results compared to traditionally trained participants. Since the effects of MR-based training on long-term knowledge and skill retention were rarely considered in the reviewed studies, these should definitely be investigated in future research. Concerning subjective evaluations, AR-based training formats were overall reported to be easy to use and to decrease perceived mental demand and task load. VR-based formats were mostly evaluated with equal or higher usability scores compared to paper- or video-based training but showed shortcomings compared to training with physical objects. However, when the feeling of presence increased, for example, by combining VR technologies with additional task-related interaction features, participants' performance increased, too.

All in all, MR-based training can be seen as a promising alternative to traditional formats, especially with regard to the identified industry needs for *improved worker performance* and *increased effectiveness*. However,

traditional training might be still more *efficient* when it comes to investing time or human resources in the development and implementation of the training. In future research, the results of experimental studies should be made comparable by using consistent objective performance measures such as TCT and TA together with subjective evaluations such as system usability, technology acceptance, and task load in order to increase the industry's confidence in these methods. For this purpose, it is also essential for researchers to clearly define and classify the MR technologies used as well as their interaction possibilities, ensuring the interpretability of research results. In future reviews, the consideration of usability studies would allow to give more weight and significance to subjective evaluations. Investigating the efficiency of MR-based training formats requires including data and information on cost- and time-efficiency of using MR technologies in training workflows. Overall, the dedicated consideration of these industry needs can contribute to bridging the gap between research and practice in the field of MR-based training in manual assembly tasks.

KEY POINTS

- The current literature on MR-based training in manual assembly provides inconsistent and hardly comparable results due to the diversity of MR technologies, assembly tasks, and research methods.
- In manual assembly tasks, AR-based training formats lead consistently to better results regarding objective performance measures, that is, task completion time and task accuracy, as well as regarding subjective evaluations of usability and perceived task load than traditional, that is, paper or video-based training.
- Less consistent, VR-based training formats lead to comparable results as traditional training but increase assembly performance in the long term.
- Future research has to provide more comparable results on industry-related use cases that can be transferred to the industry.

ORCID iDs

Lea M. Daling  <https://orcid.org/0000-0002-0324-3935>

Sabine J. Schlittmeier  <https://orcid.org/0000-0001-9051-4547>

REFERENCES

- Aguinis, H., & Harden, E. E. (2008). Sample size rules of thumb evaluating three common practices. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences* (1st ed., pp. 267–284). Routledge. <https://doi.org/10.4324/9780203867266>
- Al-Ahmari, A., Ameen, W., Abidi, M. H., & Mian, S. H. (2018). Evaluation of 3D printing approach for manual assembly training. *International Journal of Industrial Ergonomics*, *66*, 57–62. <https://doi.org/10.1016/j.ergon.2018.02.004>
- AlGeddawy, T., & ElMaraghy, H. (2012). Product variety management in design and manufacturing: Challenges and strategies. In H. A. ElMaraghy (Ed.), *Enabling manufacturing competitiveness and economic sustainability* (pp. 518–523). Springer. https://doi.org/10.1007/978-3-642-23860-4_85
- Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, *8*, 19–32. <https://doi.org/10.1080/1364557032000119616>
- Berg, L. P., & Vance, J. M. (2017). Industry use of virtual reality in product design and manufacturing: A survey. *Virtual reality*, *21*, 1–17. <https://doi.org/10.1007/s10055-016-0293-9>
- Blattgerste, J., Streng, B., Renner, P., Pfeiffer, T., & Essig, K. (2017). Comparing conventional and augmented reality instructions for manual assembly tasks. In *Proceedings of the 10th international conference on pervasive technologies related to assistive environments* (pp. 75–82). Association for Computing Machinery. <https://doi.org/10.1145/3056540.3056547>
- Brooke, J. (1996). SUS-A quick and dirty usability scale. In *Usability evaluation in industry* (pp. 189–194). Taylor & Francis. <https://doi.org/10.1201/9781498710411-35>
- Carlson, P., Peters, A., Gilbert, S. B., Vance, J. M., & Luse, A. (2015). Virtual training: Learning transfer of assembly tasks. *IEEE Transactions on Visualization and Computer Graphics*, *21*, 770–782. <https://doi.org/10.1109/tvcg.2015.2393871>
- Cooper, N., Milella, F., Pinto, C., Cant, I., White, M., & Meyer, G. (2018). The effects of substitute multisensory feedback on task performance and the sense of presence in a virtual reality environment. *PLoS One*, *13*, Article e0191846. <http://doi.org/10.1371/journal.pone.0191846>
- Daling, L. M. (2021). *Impact of augmented, virtual and mixed reality-based trainings on performance in manual assembly: A scoping review protocol*. OSF. <https://doi.org/10.17605/OSF.IO/QJSDE>
- deMoura, D. Y., & Sadagic, A. (2019). The effects of stereopsis and immersion on bimanual assembly tasks in a virtual reality system. In *2019 IEEE conference on virtual reality and 3D user interfaces (VR)* (pp. 286–294). IEEE. <https://doi.org/10.1109/VR.2019.8798112>
- Doolani, S., Owens, L., Wessels, C., & Makedon, F. (2020a). vIS: An immersive virtual storytelling system for vocational training. *Applied Sciences*, *10*, 8143. <http://doi.org/10.3390/app10228143>
- Doolani, S., Wessels, C., Kanal, V., Sevastopoulos, C., Jaiswal, A., Nambiappan, H., & Makedon, F. (2020b). A review of extended reality (xr) technologies for manufacturing training. *Technologies*, *8*, 77. <https://doi.org/10.3390/technologies8040077>
- Gallagher, A. G., Ritter, E. M., Champion, H., Higgins, G., Fried, M. P., Moses, G., Smith, C. D., & Satava, R. M. (2005). Virtual reality simulation for the operating room: Proficiency-based training as a paradigm shift in surgical skills training. *Annals of surgery*, *241*, 364–372. <https://doi.org/10.1097/01.sla.0000151982.85062.80>
- Ganier, F., Hoareau, C., & Tisseau, J. (2014). Evaluation of procedural learning transfer from a virtual environment to a real situation: A case study on tank maintenance training. *Ergonomics*, *57*, 828–843. <https://doi.org/10.1080/00140139.2014.899628>
- Gavish, N., Gutierrez, T., Webel, S., Rodriguez, J., Peveri, M., Bockholt, U., & Tecchia, F. (2015). Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks. *Interactive Learning Environments*, *23*, 778–798. <https://doi.org/10.1080/10494820.2013.815221>
- Gavish, N., Gutierrez Seco, T., Webel, S., Rodriguez, J., Peveri, M., & Bockholt, U. (2011). Transfer of skills evaluation for assembly and maintenance training. *BIO Web of Conferences*, *1*, 00028. <https://doi.org/10.1051/bioconf/20110100028>
- González-Franco, M., Cermeron, J., Li, K., Pizarro, R., Thorn, J., Hannah, P., Hutabarat, W., Tiwari, A., & Bermell-Garcia, P. (2016). Immersive augmented reality training for complex manufacturing scenarios. *ArXiv, abs/1602.01944*.
- Guo, Q. (2015). Learning in a mixed reality system in the context of, Industrie 4.0. *Journal of Technical Education*, *3*, 92–115. <https://doi.org/10.48513/joted.v3i2.60>
- Hackshaw, A. (2008). Small studies: Strengths and limitations. *The European Respiratory Journal*, *32*, 1141–1143. <https://doi.org/10.1183/09031936.00136408>
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology*, *52*, 139–183. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
- Hoedt, S., Claeys, A., Van Landeghem, H., & Cottyn, J. (2017). The evaluation of an elementary virtual training system for manual assembly. *International Journal of Production Research*, *55*, 7496–7508. <https://doi.org/10.1080/00207543.2017.1374572>
- Horigome, T., Kurokawa, S., Sawada, K., Kudo, S., Shiga, K., Mimura, M., & Kishimoto, T. (2020). Virtual reality exposure therapy for social anxiety disorder: A systematic review and meta-analysis. *Psychological Medicine*, *50*, 2487–2497. <https://doi.org/10.1017/S0033291720003785>
- Hou, L., & Wang, X. Y. (2013). A study on the benefits of augmented reality in retaining working memory in assembly tasks: A focus on differences in gender. *Automation in Construction*, *32*, 38–45. <https://doi.org/10.1016/j.autcon.2012.12.007>
- Hou, L., Wang, X. Y., Bernold, L., & Love, P. E. D. (2013). Using animated augmented reality to cognitively guide assembly. *Journal of Computing in Civil Engineering*, *27*, 439–451. [https://doi.org/10.1061/\(asce\)cp.1943-5487.0000184](https://doi.org/10.1061/(asce)cp.1943-5487.0000184)
- Hou, L., Wang, X. Y., & Truijens, M. (2015). Using augmented reality to facilitate piping assembly: An experiment-based evaluation. *Journal of Computing in Civil Engineering*, *29*, Article 05014007. [https://doi.org/10.1061/\(asce\)cp.1943-5487.0000344](https://doi.org/10.1061/(asce)cp.1943-5487.0000344)
- Kaplan, A. D., Cruitt, J., Endsley, M., Beers, S. M., Sawyer, B. D., & Hancock, P. A. (2021). The effects of virtual reality, augmented reality, and mixed reality as training enhancement methods: A

- meta-analysis. *Human Factors*, 63, 706–726. <https://doi.org/10.1177/0018720820904229>
- Kothiyal, K. P., & Kayis, B. (1995). Workplace design for manual assembly tasks: Effect of spatial arrangement on work-cycle time. *International Journal of Occupational Safety and Ergonomics*, 1, 136–143. <https://doi.org/10.1080/10803548.1995.11076310>
- Koumaditis, K., Chinello, F., Mitkidis, P., & Karg, S. (2020). Effectiveness of virtual versus physical training: The case of assembly tasks, trainer's verbal assistance, and task complexity. *IEEE Computer Graphics and Applications*, 40, 41–56. <https://doi.org/10.1109/mcg.2020.3006330>
- Koumaditis, K., Venckute, S., Jensen, F. S., & Chinello, F. (2019). Immersive training: Outcomes from small scale AR/VR Pilot-Studies. In *2019 IEEE conference on virtual reality and 3D user interfaces (VR)* (pp. 1–5). IEEE. <https://doi.org/10.1109/VR44988.2019.9044162>
- Kozioł, L. F., & Budding, D. E. (2012). Procedural learning. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (pp. 2694–2696). Springer. https://doi.org/10.1007/978-1-4419-1428-6_670
- Kwiatk, C., Sharif, M., Li, S., Haas, C., & Walbridge, S. (2019). Impact of augmented reality and spatial cognition on assembly in construction. *Automation in Construction*, 108, Article 102935. <https://doi.org/10.1016/j.autcon.2019.102935>
- Langley, A., Lawson, G., Hermawati, S., D'Cruz, M., Apold, J., Arlt, F., & Mura, K. (2016). Establishing the usability of a virtual training system for assembly operations within the automotive industry. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 26, 667–679. <https://doi.org/10.1002/hfm.20406>
- Leu, M. C., ElMaraghy, H. A., Nee, A. Y., Ong, S. K., Lanzetta, M., Putz, M., Zhu, W., & Bernard, A. (2013). CAD model based virtual assembly simulation, planning and training. *CIRP Annals*, 62, 799–822. <https://doi.org/10.1016/j.cirp.2013.05.005>
- Levac, D., Colquhoun, H., & O'Brien, K. K. (2010). Scoping studies: Advancing the methodology. *Implementation science*, 5, 1–9. <https://doi.org/10.1186/1748-5908-5-69>
- Lindeman, R. W., & Noma, H. (2007). A classification scheme for multi-sensory augmented reality. In S. N. Spencer (Ed.), *Proceedings of the 2007 ACM symposium on virtual reality software and technology* (pp. 175–178). Association for Computing Machinery. <https://doi.org/10.1145/1315184.1315216>
- Lipton, J. I., Fay, A. J., & Rus, D. (2018). Baxter's homunculus: Virtual reality spaces for teleoperation in manufacturing. *IEEE Robotics and Automation Letters*, 3, 179–186. <https://doi.org/10.1109/LRA.2017.2737046>
- Loch, F., Ziegler, U., & Vogel-Heuser, B. (2019). Using real-time feedback in a training system for manual procedures. *IFAC PapersOnLine*, 52, 241–246. <https://doi.org/10.1016/j.ifacol.2019.12.089>
- Lotter, B. (2006). Einführung. In B. Lotter & H. P. Wiendahl (Eds.), *Montage in der industriellen produktion: Ein handbuch für die praxis* (pp. 1–8). Springer. https://doi.org/10.1007/3-540-36669-5_1
- Lund, A. M. (2001). Measuring usability with the USE questionnaire. *Usability Interface*, 8, 3–6. www.stcsig.org/usability/newsletter/index.html
- Mackay, W. E. (2000). Augmented reality: Dangerous liaisons or the best of both worlds? In *Proceedings of DARE 2000 on designing augmented reality environments* (pp. 170–171). Association for Computing Machinery. <https://doi.org/10.1145/354666.354697>
- Mayring, P. (2014). *Qualitative content analysis: theoretical foundation, basic procedures and software solution*. Klagenfurt. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-395173>
- Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, 77, 1321–1329
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA, Group (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS medicine*, 6, Article e1000097. <https://doi.org/10.1371/journal.pmed.1000097>
- Munn, Z., Peters, M. D., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC medical research methodology*, 18, 1–7. <https://doi.org/10.1186/s12874-018-0611-x>
- Murcia-Lopez, M., & Steed, A. (2018). A comparison of virtual and physical training transfer of bimanual assembly tasks. *IEEE Transactions on Visualization and Computer Graphics*, 24, 1574–1583. <https://doi.org/10.1109/tvcg.2018.2793638>
- Normand, J. M., Servières, M., & Moreau, G. (2012). A new typology of augmented reality applications. In *Proceedings of the 3rd augmented human international conference* (pp. 1–8). Association for Computing Machinery. <https://doi.org/10.1145/2160125.2160143>
- Oren, M., Carlson, P., Gilbert, S., & Vance, J. M. (2012). Puzzle assembly training: Real world vs. virtual environment. In S. Coquillart, S. Feiner, & K. Kiyokawa (Eds.), *2012 IEEE virtual reality workshops* (pp. 27–30). IEEE. <https://doi.org/10.1109/VR.2012.6180873>
- Peters, M. D. J., Godfrey, C., McInerney, P., Munn, Z., Tricco, A. C., & Khalil, H. (2020). Chapter 11: Scoping reviews. In E. Aromataris & Z. Munn (Eds.), *JBI manual for evidence synthesis*. JBI. <https://doi.org/10.46658/JBIMES-20-12>
- Prasertsakul, T., Kaimuk, P., Chinjenpradit, W., Limroongreungrat, W., & Charoensuk, W. (2018). The effect of virtual reality-based balance training on motor learning and postural control in healthy adults: A randomized preliminary study. *Biomedical engineering online*, 17, 1–17. <https://doi.org/10.1186/s12938-018-0550-0>
- Roldan, J. J., Crespo, E., Martin-Barrio, A., Pena-Tapia, E., & Barrientos, A. (2019). A training system for industry 4.0 operators in complex assemblies based on virtual reality and process mining. *Robotics and Computer-Integrated Manufacturing*, 59, 305–316. <https://doi.org/10.1016/j.rcim.2019.05.004>
- Rose, F. D., Attree, E. A., Brooks, B. M., Parslow, D. M., Penn, P. R., & Ambihapahan, N. (2000). Training in virtual environments: Transfer to real world tasks and equivalence to real task training. *Ergonomics*, 43, 494–511. <https://doi.org/10.1080/001401300184378>
- Sauro, J. (2011). *A practical guide to the system usability scale: Background, benchmarks, & best practices*. Measuring Usability
- Sautter, B., & Daling, L. (2021). Mixed Reality Supported Learning for Industrial on-the-job Training. *Proceedings of the Conference on Learning Factories*. <http://doi.org/10.2139/ssrn.3864189>
- Singer, M. J., Ehrlich, J. A., & Cinq-Mars, S. (1995). *Task performance in virtual environments: Stereoscopic versus*

- monoscopic displays and head-coupling* (Technical Report 1034). US Army Research Institute for the Behavioral and Social Sciences
- Skarbez, R., Smith, M., & Whitton, M. C. (2021). Revisiting Milgram and Kishino's reality-virtuality Continuum. *Frontiers in Virtual Reality*, 2, Article 647997. <https://doi.org/10.3389/frvir.2021.647997>
- Slater, M., Danieleto, S., Pooley, M., Cheng Teh, L., Gidley-Baird, A., & Barden, J. A. (2004). A note on presence terminology. *Presence Connect*, 3, 1–5. http://www0.cs.ucl.ac.uk/research/vr/Projects/Presencia/ConsortiumPublications/ucl_cs_papers/presence-terminology.htm
- Velaz, Y., Arce, J. R., Gutierrez, T., Lozano-Rodero, A., & Suescun, A. (2014). The influence of interaction technology on the learning of assembly tasks using virtual reality. *Journal of Computing and Information Science in Engineering*, 14, Article 041007. <https://doi.org/10.1115/1.4028588>
- Wang, P., Bai, X., Billingham, M., Zhang, S., Han, D., Lv, H., He, W., Yan, Y., Zhang, X., & Min, H. (2019). An MR remote collaborative platform based on 3D CAD models for training in industry. In *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)* (pp. 91–92). IEEE. <https://doi.org/10.1109/ISMAR-Adjunct.2019.00038>
- Wang, X., Ong, S. K., & Nee, A. Y. C. (2016). A comprehensive survey of augmented reality assembly research. *Advances in Manufacturing*, 4, 1–22. <https://doi.org/10.1007/s40436-015-0131-4>
- Wang, L., Keshavarzmanesh, S., Feng, H. Y., & Buchal, R. O. (2009). Assembly process planning and its future in collaborative manufacturing: A review. *The International Journal of Advanced Manufacturing Technology*, 41, 132–144. <https://doi.org/10.1007/s00170-008-1458-9>
- Webel, S., Bockholt, U., Engelke, T., Gavish, N., Olbrich, M., & Preusche, C. (2013). An augmented reality training platform for assembly and maintenance skills. *Robotics and Autonomous Systems*, 61(4), 398–403. <https://doi.org/10.1016/j.robot.2012.09.013>
- Werrlich, S., Lorber, C., Nguyen, P. A., Yanez, C. E. F., & Notni, G. (2018a). Assembly training: comparing the effects of head-mounted displays and face-to-face training. In J. Y. C. Chen & G. Fragomeni (Eds.), *Virtual, augmented and mixed reality: Interaction, navigation, visualization, embodiment, and simulation, VAMR 2018, Pt I* (Vol. 10909, pp. 462–476). Springer. https://doi.org/10.1007/978-3-319-91581-4_35
- Werrlich, S., Nguyen, P.-A., & Notni, G. (2018b). Evaluating the training transfer of Head-Mounted Display based training for assembly tasks. In *PETRA '18: Proceedings of the 11th pervasive technologies related to assistive environments conference* (pp. 297–302). Association for Computing Machinery. <https://doi.org/10.1145/3197768.3201564>
- Westerfield, G., Mitrovic, A., & Billingham, M. (2015). Intelligent Augmented Reality Training for Motherboard Assembly. *International Journal of Artificial Intelligence in Education*, 25, 157–172. <https://doi.org/10.1007/s40593-014-0032-x>
- Yuviler-Gavish, N., Krupenia, S., & Gopher, D. (2013). Task analysis for developing maintenance and assembly VR training simulators. *Ergonomics in Design: The Quarterly of Human Factors Applications*, 21, 12–19. <https://doi.org/10.1177/1064804612463214>
- Xia, P., Lopes, A. M., Restivo, M. T., & Yao, Y. X. (2012). A new type haptics-based virtual environment system for assembly training of complex products. *The International Journal of Advanced Manufacturing Technology*, 58, 379–396. <https://doi.org/10.1007/s00170-011-3381-8>
- Lea M. Daling, M.Sc. is a research associate at the Institute of Information Management in Mechanical Engineering as well as an external PhD student in the teaching and research area of Work and Engineering Psychology at the RWTH Aachen University. She completed her master's in social, organizational, and work psychology at the Freie Universität Berlin in 2017.
- Sabine J. Schlittmeier is head of the teaching and research area of Work and Engineering Psychology at RWTH Aachen University. She was previously Professor of Psychology with a focus on Auditory Cognition at RWTH Aachen University (2018–2020; funded by the HEAD-Genuit Foundation) and Professor of Applied Psychology at HSD Hochschule Döpfer in Cologne (2015–2018).

Date received: December 6, 2021

Date accepted: May 12, 2022

Publication 4:

Daling, L. M., Tenbrock, M., Isenhardt, I., & Schlittmeier, S. J. (2023). Assemble it like this!—Is AR-or VR-based training an effective alternative to video-based training in manual assembly?. *Applied Ergonomics*, 110,104021. <https://doi.org/10.1016/j.apergo.2023.104021>

Reproduced with permission from Elsevier.



Assemble it like this! – Is AR- or VR-based training an effective alternative to video-based training in manual assembly?

Lea M. Daling^{a,b,*}, Marisa Tenbrock^{a,b}, Ingrid Isenhardt^a, Sabine J. Schlittmeier^b

^a WZL Chair of Production Metrology and Quality Management & Institute for Information Management in Mechanical Engineering, RWTH Aachen University, Dennewartstr. 27, 52068, Aachen, Germany

^b Work and Engineering Psychology, Institute of Psychology, RWTH Aachen University, Jägerstr. 17-19, 52066, Aachen, Germany

ARTICLE INFO

Keywords:

Virtual reality
Augmented reality
Training and long-term retention

ABSTRACT

AR- and VR-based training is increasingly being used in the industry to train workers safely and effectively for new tasks. In this study, we investigated and compared the effects of AR-, VR- and video-based training on short- and long-term objective performance measures and subjective evaluations in a manual assembly task. Our results showed that there was no difference between AR-, VR- and video-based training concerning the objective performance measures task completion time and error count. However, in the subjective evaluations VR-based training showed a significantly higher perceived task load and a lower usability rating than the AR- and video-based training regimes. An exploratory analysis additionally revealed partially better results for AR than for VR after adjusting the data for the age of the participants. Future research should further investigate the advantage of AR- and video-based methods over VR when the age and technology experience of participants are taken into account.

1. Introduction: Assemble it like this! – Is AR- or VR-based training an effective alternative to video-based training in manual assembly?

In recent years, new technologies have been introduced to all areas of production to enhance the effectiveness, efficiency, and flexibility of the overall production workflow (Eversberg et al., 2021; Kim et al., 2018; Miqueo et al., 2020; Xu et al., 2021). This also changed the qualification requirements for workers to be prepared for new tasks effectively (e.g., reducing errors), efficiently (e.g., saving time resources), and flexibly (e.g., being able to adapt to new requirements) (Joshi et al., 2021; Kagermann et al., 2013; Sautter and Daling, 2021). In the field of manual assembly tasks in particular, the increasing individualization and complexity of products, shorter product life cycles, and staff turnover all led to a higher demand for appropriate training and assistance technologies (AlGeddawy and ElMaraghy, 2012; Daling and Schlittmeier, 2022; Werrlich et al., 2018). In this context, digital technologies such as augmented reality (AR) and virtual reality (VR) have been introduced as promising media for the training and qualification of

workers (Heinz et al., 2019; Joshi et al., 2021; Lawson et al., 2016; Wang et al., 2016, 2019) providing them with safe, hazard-free, as well as flexible, location- and time-independent training (Kaplan et al., 2021).

Current findings on the effectiveness of AR and VR technologies for training manual assembly tasks already revealed promising results for their immediate effects on objective performance measures compared to conventional training (Hou and Wang, 2013; Koumaditis et al., 2019; Murcia-López and Steed, 2018; Roldan et al., 2019). To date, however, there is still too little knowledge about the long-term effects of training with AR or VR on skill and knowledge retention, as concluded by Daling & Schlittmeier in a recent scoping review (2022). In addition, this review discussed the need for a more holistic consideration of objective measures and subjective evaluations, since most studies provided results on objective measures of performance, but rarely also systematically examined subjective assessments. Lastly, the review noted that AR- and VR-based training have only been studied separately, so currently it is not possible to state whether differences exist between AR and VR when it comes to training manual assembly tasks. The present study addresses these research gaps by evaluating and comparing AR- and VR-based

Abbreviations: VR, Virtual Reality; AR, Augmented Reality.

* Corresponding author. Chair of Production Metrology and Quality Management & Institute for Information Management in Mechanical Engineering, RWTH Aachen University, Dennewartstr. 27, 52068, Aachen, Germany.

E-mail addresses: lea.daling@ima.rwth-aachen.de (L.M. Daling), marisa.tenbrock@rwth-aachen.de (M. Tenbrock), ingrid.isenhardt@ima.rwth-aachen.de (I. Isenhardt), sabine.schlittmeier@psych.rwth-aachen.de (S.J. Schlittmeier).

<https://doi.org/10.1016/j.apergo.2023.104021>

Received 25 July 2022; Received in revised form 15 March 2023; Accepted 17 March 2023

Available online 30 March 2023

0003-6870/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

training in a manual assembly task regarding short- and long-term effects on objective performance measures as well as subjective evaluations.

1.1. Training of manual assembly tasks using AR and VR technologies

Training of manual assembly tasks imparts to the trainee both: factual knowledge of the components and their manner and sequence of assembly; and procedural skills that include, for example, joining, handling, inspecting, and adjusting (Lotter, 2006). Traditionally, workers have acquired the knowledge and skills during training performed by trainers or more experienced colleagues. Additionally, paper-based instructions – and later also videos – have been used as a supplement or replacement for face-to-face training (Werrlich et al., 2018). Within the framework of Industry 4.0, the need for more flexible methods such as AR- and VR-based training has become apparent, as conventional training formats showed limitations in terms of adaptability and efficiency: for example, if personnel resources had to be made available to teach new variants more frequently, or if the cycle time of machines had to be reduced during on-site training (Funk and Schmidt, 2021).

AR and VR are defined as different instances on a spectrum between reality and virtuality (Milgram and Kishino, 1994). Both systems have in common that, to some extent, both real-world and virtual-world objects and stimuli are presented together within a single percept (Skarbez et al., 2021). In AR-based assembly training, virtual objects or instructions are usually projected onto the real workstation in combination with physical components. VR-based assembly training simulates the real workspace in a 3D modeled environment and includes the handling of virtual objects using controllers (Doolani et al., 2020; Gavish et al., 2015; Kaplan et al., 2021; Milgram and Kishino, 1994). Although AR and VR both make it possible to circumvent the above-mentioned limitations of conventional training in terms of reduced personal efficiency and adaptability, it is to be asked whether the training success using these technologies actually compares to that of conventional training. Ideally this question should be asked before implementation.

1.2. Evaluating training success of AR- and VR-based training

Put simply, training is successful if the trainee actually learns the knowledge and skills that need to be trained. Success indicators are often taken to be the so-called objective measures, such as measures of performance. However, subjective experience can also be used for the evaluation of training measures. One example is the perceived task load during training. Evaluation of subjective experience tends to be used more often when the recording of performance measures is difficult (cp. Daling and Schlittmeier, 2022). Moreover, subjective evaluations can be an important predictive indicator for the acceptance and use of a system (Longo, 2018; Venkatesh et al., 2008) because they provide insights into in-depth information about opinions, attitudes, behaviors, satisfaction and preferences related to the system under investigation (DeLone and McLean, 2002). In fact, we believe that successful training is characterized by both aspects: objectifiable effectiveness and positive subjective evaluations. So AR and VR technologies should not fall behind conventional training in either of these evaluation dimensions – better yet, they should perform better. In the following, we will take a closer look at the current empirical evidence of AR- and VR-based training on objective performance measures and subjective evaluations in manual assembly tasks.

1.2.1. Objective performance measures

When evaluating training success after AR- and VR-based training, a first reference is usually to examine whether the training was effective in preparing users for the new task. Common objective performance measures in manual assembly training are task completion time (TCT) and the number of errors (e.g., Ariansyah et al., 2021; Drouot et al.,

2022; Murcia-López and Steed, 2018; Werrlich et al., 2018). Moreover, in addition to the short-term effects of training on performance, long-term skill retention, i.e., remembering the assembly task over several days and weeks, is another important indicator of objective performance in an increasingly flexible production environment (Doolani et al., 2020).

When looking at the short-term effects, many studies concluded that AR-based training was either comparable or even superior to paper- or video-based training (e.g., DaValle and Azhar, 2020; Hou and Wang, 2013; Koumaditis et al., 2019; Kwiatek et al., 2019; Loch et al., 2019; Westerfield et al., 2015). At the same time, recent studies on VR-based training revealed similarly successful results when using VR compared to video- or paper-based training (e.g., Doolani et al., 2020; Koumaditis et al., 2020; Murcia-López and Steed, 2018; Roldan et al., 2019). Only two studies attempted to compare AR- and VR-based training. Gabajová et al. (2019) investigated the training times of an industrial plug assembly and found that while both AR and VR training formats reduced training time compared to the paper-based method, there was no difference between those two formats. A very recent study comparing task completion times after AR-, VR- or video-based training in a simple industrial maintenance task showed initial results suggesting that with increasing task difficulty the AR group showed an advantage compared to the video and VR group (Liu et al., 2022).

Long-term effects of AR- and VR-based training, however, have barely been considered in current research (see the review by Daling and Schlittmeier, 2022). Only Gavish et al. (2011) examined knowledge retention in a procedural task a few hours after AR-based training. Here, the AR group showed fewer uncorrected errors, but took the same amount of time to complete the task as a group trained with video. A few initial studies on the long-term effects of VR-based training found that VR was significantly worse in the immediate, i.e. short-term, comparison, but that these differences were no longer noticeable when long-term skill retention was assessed, e.g. after two weeks, compared to training with physical blocks, paper manuals, or video based-training (Doolani et al., 2020; Murcia-López and Steed, 2018). Based on these results, it can be assumed that AR- and VR-based training is at least as good or even better in direct comparison to traditional training in both the short- and long-term, although long-term effects have not yet been sufficiently investigated.

1.2.2. Subjective evaluations

In addition to objective performance measures, subjective evaluations provide information about the perspective of users, which is essential for a holistic evaluation of a system. The subjective evaluation of users can be a decisive factor in determining whether a system is actually used and accepted (Jang et al., 2021). Another factor is the frequent lack of any possibility to measure objective measures outside of laboratory settings since individual employee performance evaluations are not allowed in many companies. In current performance studies on AR and VR, subjective evaluations often played a subordinate role (González-Franco et al., 2016; Hoedt et al., 2017; Hou and Wang, 2013). In some cases, they were recorded instead of objective measures in usability studies (Kostaras and Xenos, 2012; Li et al., 2019; Webel et al., 2013). Among the studies that included subjective evaluations, most findings were collected based on self-report variables such as subjective task load indices and system usability scores, which were used in addition to some non-validated questionnaires (Al-Ahmari et al., 2018; Doolani et al., 2020; Carlson et al., 2015; Hou et al., 2015; Koumaditis et al., 2019; Murcia-López and Steed, 2018; Oren et al., 2012; Wang et al., 2019).

When using AR-based training, an increase in objective performance was often accompanied by positive impacts on subjective evaluations. Hou et al. (2015) showed that using AR led to a reduced task load in addition to outperforming paper training with regard to completion times and error rates. Koumaditis et al. (2019) discussed how the reduction in perceived task load using AR was due to the provision of

contingent, in-situ information through temporal and spatial anchoring on relevant objects, because the AR-based training could direct the trainee's attention precisely to where it was needed at the time it was needed (Hou et al., 2013, 2015; Koumaditis et al., 2019). Concerning perceived system usability, AR-based training was found to be rated higher than video-based training (Loch et al., 2019) but lower than paper manuals (Wang et al., 2019).

In terms of subjective evaluations of VR-based training, Al-Ahmari et al. (2018) found in their study that while VR-based training led to worse results in objective performance measures than training with physical objects, there was no difference in task load between the training groups. When looking at system usability, results differed: Murcia-López and Steed (2018) found that VR resulted in better performance than paper-based training while being rated with higher usability scores whereas Oren et al. (2012) showed that VR-based training led to equally good results in objective measures compared to physical blocks but was rated with lower usability. In a direct comparison, Gavish et al. (2015) found that AR-based training was rated statistically significantly better than VR in terms of satisfaction with performance, usability, and willingness to recommend the system. However, the authors used different assembly use cases for AR and VR, which diminishes the comparability of these findings. Liu et al. (2022) found that no group difference in cognitive load was found between AR-, VR- and video-based groups during easier maintenance tasks. However, during difficult tasks the AR group showed a lower workload than the VR group, while VR and video showed no difference. All in all, the results on subjective evaluations indicate positive effects from AR- and VR-based training on user experience, although the results for VR-based training are more heterogeneous than those for AR-based training.

1.3. Aim of the present study and hypotheses

At this point, we can conclude that existing research already provided promising results on the short-term effects of AR- and VR-based training for manual assembly tasks on objective performance measures. However, important questions about the long-term effects of AR- and VR-based training have not been sufficiently studied yet. In addition, in classical performance studies, the analysis of subjective evaluations has played a subordinate role. Until now, no studies exist that take a joint and holistic view of the short- and long-term effects of these training formats in consideration of both objective performance measures and subjective evaluations. Lastly, it has become apparent that although both AR- and VR-based systems have each been compared with conventional methods, no direct comparison of training success after using AR versus VR has yet been conducted.

The present study aimed to address these research gaps by comparing AR-based training and VR-based training with conventional video training concerning (a) short-term and long-term effects on objective performance measures and (b) subjective evaluations within one study. Participants followed an assembly training in either AR, VR, or video formats for the manual assembly of a LEGO® MINDSTORMS® EV3 robot, which consisted of ten steps. The structure and content of the training were identical in AR, VR, and video. Within the scope of this paper, two research questions will be addressed:

(RQ1). Do AR-, VR- and video-based trainings show differences in their impact on short- and long-term objective performance?

(RQ2). Do AR-, VR- and video-based trainings differ in how they are subjectively evaluated by users?

To answer these questions, we first looked at the short- and long-term effects of AR and VR-based training on objective performance measures in comparison with video-based training. TCT and error count were used as outcome variables in an assembly task that was performed twice: immediately after training (T1) and two weeks after training (T2). According to Schmidt and Bjork (1992), a retention phase after the initial

training should be long enough to ensure that any temporary effects of the independent variable have been dissipated. The two-week interval is the current standard in the literature for long-term retention of industry training (Carlson et al., 2015; Murcia-López and Steed, 2018) and realistic in occupational settings: within Industry 4.0, which is characterized by mass customization and product variant variety, it is common that some variants are only assembled from time to time. For both short-term and long-term effects (i.e. on both T1 and T2), we expected that both AR- and VR-based training would lead to better objective performance measures (lower TCT and fewer errors) in comparison to video-based training.

Second, the subjective evaluations of the training served as an additional essential characteristic for training success. Thus, perceived task load (Hart and Staveland, 1988) and experienced usability (Brooke, 1996) were explored with regard to differences between the AR-, VR- and video-based training. Both task load and system usability referred to the users' experience during training. We expected lower perceived task load and higher system usability in both the AR and the VR groups compared to the video group.

After the analysis on these a priori hypotheses and as the collected data allowed, the authors performed another post hoc exploratory analysis of the data to gain deeper insights into the effectiveness and potential differences of AR-, VR- and video-based training. This analysis provided insights into differences in training duration as well as the importance of user-related variables such as age and prior experience of participants.

2. Materials and methods

To answer the research questions regarding the short- and long-term effects of AR-, VR- and video-based training in terms of objective performance measures and subjective evaluations, we used a 3×2 repeated-measures experimental design. Participants were divided into three groups and trained to learn the procedural assembly process of a LEGO® MINDSTORMS® EV3 robot using either AR-, VR- or video-based instructions. The AR and video groups were trained at a physical workplace consisting of an assembly cell and industrial storage systems. The VR group was trained in a 3D-modeled virtual workplace, which was designed to resemble the physical workplace (see Fig. 1). The training slides of the instruction protocol were identical for all groups, while the training medium differed (a video of each training and the instruction protocol can be found in the supplementary data, Appendix A.1–4). On the virtual and physical desks, open storage boxes contained the 14 components needed for assembly (tires, server motors, EV3 stone, steel ball, frame, cables, axles, module carriers, various connecting pins, mounting rings, spacer pins and angled connectors). Participants were instructed to imagine being an inexperienced worker in a factory during a production peak, and to memorize the correct sequential order of the ten-step assembly process. Training success was evaluated both immediately after training (T1) and two weeks later (T2). In the following, all methods and materials used in the study will be described.

2.1. Participants

In total, $N = 103$ participants took part in the study and were randomly allocated into three experimental groups as follows: both the video group and AR group had 34 participants, while the VR group consisted of 35 participants. However, due to technical problems with the VR application, seven participants in the VR group were excluded from the analysis, so that this group finally entered data analysis with 28 participants. Of the 96 participants included in the analyses, 51 were male (53.1%) and 45 were female (46.9%). The mean age was 30.19 years, with a standard deviation of 11.46 years, a minimum age of 18 years, and a maximum age of 73 years.

Participants were recruited via word of mouth, mailing lists, social networks, email, and posters. Participation was voluntary and

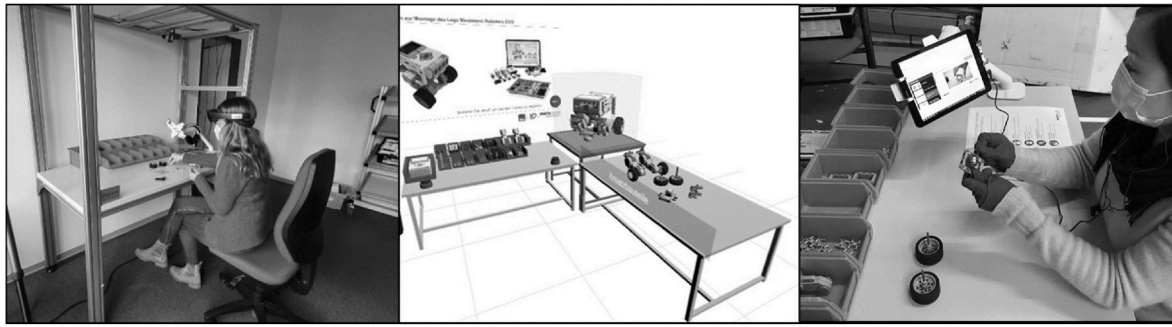


Fig. 1. Physical workplace of the AR group (left), 3D modeled workspace of the VR group (middle), and physical workplace of the video group (right).

participants did not receive any payment or reward. The inclusion criterion was a minimum age of 18 years and informed consent was provided before the experiment. An a priori power analysis for a repeated measures ANOVA with within-between interaction using G*Power (Faul et al., 2007) determined a required total sample size of $N = 81$ for a power of 0.8 and a medium effect (Cohen's f) of 0.25 (Cohen, 1988) with a significance level of 5%. The study protocol was approved by the ethics committee of RWTH Aachen University (Ethics approval number 2021_02_FB7_RWTH AACHEN).

In the video group, 26.5% (nine participants) reported no prior experience with video tutorials, and 73.5% (25 participants) indicated having used video tutorials before. In the VR group, 40.7% (eleven participants) reported having no experience using VR, and 59.3% (16 participants) reported having used VR before. In the AR group, 62.5% (20 participants) indicated having no experience using AR, and 37.6% (12 participants) reported prior experience. Across all groups, 53.1% (51 participants) reported having student status, 49% (47 participants) reported being employed (multiple selections were possible), and 4.2% (four participants) were retired.

2.2. Materials

2.2.1. Technical equipment

Participants in the AR-based training used the Microsoft HoloLens 1 HMD. The AR HMD was operated with either speech commands, hand gestures or the clicker, depending on the participants' preference. The AR-based training was programmed in Unity Version 2019.4.3f1, using the Microsoft Mixed Reality Toolkit. Participants in the AR group used physical LEGO® MINDSTORMS® EV3 assembly components. The instructional slides were projected from the AR onto the white wall at the participants' eye level. A slowly rotating holographic 3D animation, showing each step in detail, was projected into the environment to their right (a video of the AR training can be found in the supplementary data, Appendix A.1). The training was holographically guided and performed with physical components.

In the VR-based training, participants were trained using an Oculus Rift VR HMD and controllers which simulated their hand movements operating the system (Auto Hand® - VR Physics Interaction, Unity). The training was programmed in Unity version 2019.4.3f1 and simulated the assembly process with virtually simulated LEGO® MINDSTORMS® EV3 3D assembly components. The instructional slides were projected at participants' eye level in the virtual environment. Detailed, moving 3D animations of the assembly steps were projected onto a virtual table on the right side of the participants (a video of the VR training can be found in the supplementary data, Appendix A.2). The training was guided and performed using virtual components.

Participants in the video-based training used a Samsung tablet (10.4 inches) and a touch pen to operate on the screen. Similar to the AR group, participants used physical LEGO® MINDSTORMS® EV3 assembly components. Instructional slides were presented on the tablet screen, which was fixed with a mounted tablet arm. The steps were shown as

video clips of real hands performing the task. The tablet presenting the slides and videos was placed to the participants' right side (a video of the video training can be found in the supplementary data, Appendix A.3). The training was guided with video instructions and performed using physical components.

2.2.2. Questionnaires

Task load was measured using the German version of the Raw NASA-Task Load Index (NASA-TLX) questionnaire (Hart and Staveland, 1988; Hart, 2006). The NASA-TLX is a self-report instrument measuring perceived task load on six dimensions: mental demand, physical demand, temporal demand, performance demand, effort, and frustration. Each of the six 1-item scales was presented with a horizontal line with two poles ranging from "low" (= 0) to "high" (= 100).

Usability was measured using the German version of the System Usability Scale (SUS; Brooke, 1996; Gao et al., 2020). SUS items were rated on a five-point Likert scale from "strongly disagree" to "strongly agree". Examples of items are "I found the various functions in this system were well integrated" and "I found the system unnecessarily complex".

2.3. Procedure

The study design included three groups that were trained using either AR, VR, or video. The experiment consisted of two single sessions for all participants. T1 consisted of the training and the assembly: the procedure took 75–100 min. T2 took place two weeks later, with participants performing the assembly again with no training provided: the procedure took 15–35 min. A procedure flowchart is presented in Fig. 2.

2.3.1. First experimental session

Upon arrival at T1, participants gave informed consent and then proceeded with a demographic questionnaire querying gender, age, current occupational status, industry branch, experience using LEGO® Technic, and experience using either VR, AR, or video tutorials. Before starting the training procedure, participants in all groups were shown the physical target object, i.e., the assembled LEGO® MINDSTORMS® EV3 robot. Then, participants in all groups received an introduction to the medium used and a tutorial to familiarize themselves with the technology. They were advised that they were allowed to take as much time as they needed during the training phase to memorize the process, as they had to repeat the assembly afterward without assistance. In all groups, participants were able to set the pace of the training themselves. The training consisted of two parts. The first part consisted of highly detailed instructions on slides and audio tracks being presented to guide the participants step by step through the assembly (for the detailed steps, see the instruction protocol in Appendix A.4). In the AR group, holographic 3D animations of the single assembly steps were presented in addition to the slides and holographic buttons were used to skip back and forth. In the VR group, virtual 3D animations were presented, and virtual buttons were used to navigate through the training content. In

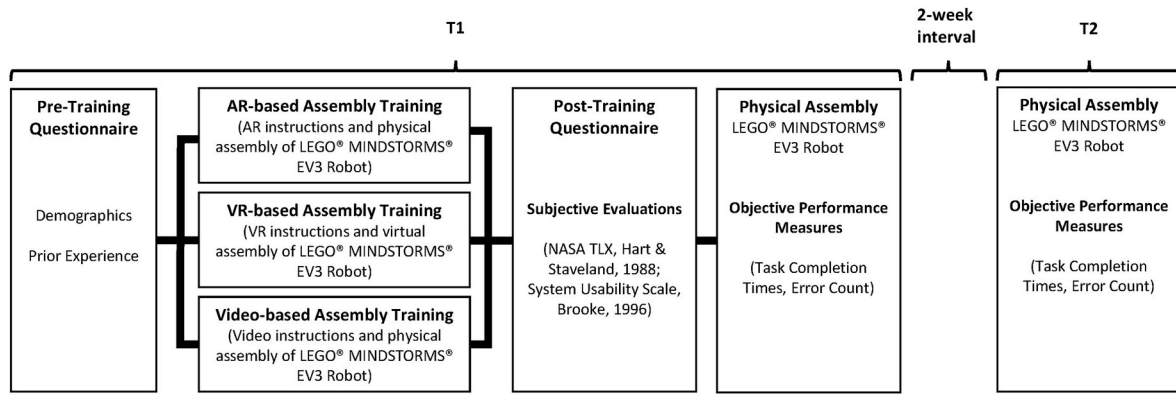


Fig. 2. Procedure Flowcharts for the first (T1) and second (T2) experimental sessions.

the video group, videos of each step were presented, and participants could press pause and play with the touch pen (see Appendix A.1-3). The second part was a shorter version of the training, consisting of the instructional slides only, without the audio track, animations, or videos. Altogether, each participant assembled the robot two times during the training.

After the training phase, participants filled in the NASA-TLX and the SUS. Then they performed the assembly. This took place at the physical workstation for all groups. Participants were instructed to assemble the robot as quickly and as accurately as possible. If needed, participants were allowed to use a paper manual once per step. However, they were instructed that for each time they used the manual, one error point would be added. During the assembly, a video camera recording participants' hands was placed to their left to help count the assembly time (i.e. TCT) and errors. The assembly was deemed to be completed when the LEGO® MINDSTORMS® EV3 robot was put into operation.

2.3.2. Second experimental session

The second experimental session took place two weeks after the first to measure long-term effects of training. For T2, participants were asked to perform the assembly as quickly and accurately as possible from their memory. The paper-based manual was at the participants' disposal with the same condition that each use would lead to one error point being added. Assembly completion time (TCT) and errors were counted with the help of a video camera recording as in T1.

3. Results

3.1. Results on a priori hypotheses

In the following, the results regarding the a priori hypotheses are reported. First, the results of the short- and long-term effects of AR-, VR- and video-based training on objective performance measures TCT and error count are reported, followed by the results of the subjective

evaluations of AR-, VR- and video-based training..

3.1.1. Short- and long-term effects of AR-, VR- and video-based training on objective performance measures

Concerning both objective performance measures TCT and error count, we expected that AR- and VR-based training would be more effective than video-based training in the short- (T1) and the long-term (T2). Descriptive results of TCT and error count can be seen in Table 1. An outlier analysis was conducted for all variables on the group level. Z scores were inspected for outliers above three standard deviations, resulting in the pairwise exclusion of identified cases from the analysis.

Task completion times. TCT describes the time needed by the participants to assemble the robot and put it into operation, which was measured during the assemblies in T1 and T2. TCTs are reported in minutes, seconds were transformed into decimal seconds. Results on

Table 2

Descriptive results of the subjective evaluations NASA-task load index (NASA-TLX) and system usability score (SUS).

Training	Subjective Evaluations	N	Min	Max	M	SE	95%-CI
AR	NASA TLX	34	17.17	57.17	35.75	-1.85	[31.98, 39.51]
	SUS	34	62.5	100.0	81.76	1.78	[78.14, 85.39]
VR	NASA TLX	28	27.83	73.67	52.29	2.03	[48.13, 56.45]
	SUS	28	25.0	85.0	61.07	3.2	[54.52, 67.63]
Video	NASA TLX	34	16.5	64.5	40.25	2.11	[35.95, 44.55]
	SUS	34	62.5	100.0	85.96	1.58	[82.76, 89.76]

Note. Due to an outlier analysis, identified cases were excluded pairwise. SE = standard error; CI = confidence interval.

Table 1

Descriptive results of the objective performance measures TCT and error count for T1 and T2.

Training	Performance Measures	Point of Measurement											
		T1						T2					
		N	Min	Max	M	SE	95%-CI	N	Min	Max	M	SE	95%-CI
AR	TCT	33	3.52	8.47	5.1	.24	[4.62, 5.58]	33	6.01	15.51	9.79	.48	[8.80, 10.79]
	Error count	34	0	10	3.29	.45	[2.37, 4.21]	34	1	16	7.06	.68	[2.86, 4.55]
VR	TCT	27	4	9.95	5.74	.29	[5.13, 6.36]	27	5.27	15.73	9.29	.52	[5.67, 8.45]
	Error count	27	0	9	3.7	.41	[2.86, 4.55]	27	1	15	9.15	.61	[7.88, 10.41]
Video	TCT	33	3.07	11.43	5.13	.28	[4.55, 5.70]	33	5.49	15.97	10.02	.51	[8.98, 11.05]
	Error count	33	0	8	3.55	.36	[2.81, 4.28]	33	1	16	7.59	.62	[6.33, 8.85]

Note. Task completion times (TCTs) are reported in minutes, seconds were transformed to decimal seconds. Subsequent to an outlier analysis, identified cases were excluded pairwise. SE = standard error; CI = confidence interval.

short- and long-term effects of AR- and VR-based training on TCT in comparison with the video group can be seen in Fig. 3.

A 3×2 ANOVA with the between-subjects factor *training group* (AR, VR, video) and the within-subject factor *time of measurement* (T1, T2) revealed that the main effect of *time of measurement* was significant, $F(1, 90) = 262.99, p \leq .001, \eta_p^2 = .75$. This main effect is important because TCT significantly increased from T1 to T2, regardless of the group. On T1, overall mean TCT was $M = 5.32$ min, $SE = .16$, 95%-CI [5.01, 5.64], and on T2 overall mean TCT was $M = 9.70$ min, $SE = .29$, 95%-CI [9.12, 10.29], with a mean difference of $M_{Diff} = 4.38$ min, $SE = .27$, 95%-CI [3.84, 4.92] between T1 and T2. The main effect of *training group* was not significant, $F(2, 90) = 2.52, p = .964$, indicating that there was no difference in TCT between the three training groups. Moreover, the interaction between *time of measurement* and *training group* was not significant, $F(2, 90) = 2.52, p = .111$, indicating that training methods did not affect TCT for T1 differently than for T2.

Error count. During the assembly, a video camera recorded participants' hands to count errors afterward. The videos of T1 and T2 were analyzed, coded, and summed up based on a pre-defined error protocol. For each of the ten steps of the assembly procedure, the severity of the error was defined. For severe errors, two errors were counted, for minor errors, one error was counted. Severe errors were counted based on errors which hindered further progression of the assembly task, for example mounting the cables to the wrong socket, which made it impossible to put the robot into operation. Minor errors were counted, for example, when participants were mixing up right and left, or mixing up the order of sequence. When participants realized and resolved an error, the error did not count. Using the manual was counted with one error point. The minimum error rate was zero, while a maximum of 68 errors were possible.

A 3×2 ANOVA with the between-subjects factor *training group* (AR, VR, video) and the within-subject factor *time of measurement* (T1, T2) revealed that the main effect for *time of measurement* was significant, $F(1, 91) = 144.95, p \leq .001, \eta_p^2 = .61$ (see Fig. 4). This indicated that the error count increased for all participants, regardless of the group, with an overall mean of $M = 3.51$ errors, $SE = .24$, 95%-CI [3.04, 3.99] on T1 and an overall mean of $M = 7.93$ errors, $SE = 0.38$, 95%-CI [7.19, 8.68] on T2. Across all groups, there was a mean difference of $M_{Diff} = 4.42$ errors between T1 and T2, $SE = .37$, 95%-CI [3.69, 5.147]. The main effect of *training group* was not significant, $F(2, 91) = 1.975, p = .145$, indicating that there was no difference in error count between the training groups. Furthermore, the interaction between *time of*

measurement and *training group* was not significant $F(2, 91) = 1.88, p = .159$, indicating that training methods did not affect error count for T1 differently than for T2.

3.1.2. Results on subjective evaluations of AR-, VR- and video-based training

Concerning subjective evaluations, we expected that AR- and VR-based training would be rated with less task load and higher system usability than video training. Descriptive results of NASA-TLX and SUS can be seen in Table 2. An outlier analysis was conducted for all variables on the group level. Z scores were inspected for outliers above three standard deviations, resulting in the pairwise exclusion of identified cases from the analysis.

NASA-TLX. The overall NASA-TLX task load estimate was calculated as the unweighted average of the scores of each subscale. Although reliability was questionable for the overall NASA-TLX score (Cronbach's $\alpha = .66$), a one-way ANOVA with the between-subjects factor *training group* (AR, VR, video) revealed a statistically significant difference in *task load* between the training groups, $F(2, 93) = 17.06, p \leq .001, \eta_p^2 = .27$. Tukey's post hoc analysis revealed that there was no difference between AR and the video group with a mean difference of $M_{Diff} = -4.49, SE = 2.75$, 95%-CI [-11.05, 2.06], $p = .317$, but participants in the VR group reported a statistically significantly higher task load compared to participants in the video group with a mean difference of $M_{Diff} = 12.04, SE = 2.89$, 95%-CI [5.14, 18.94], $p \leq .001$ (see Fig. 5). Moreover, there was a statistically significant difference between VR and AR with a mean difference of $M_{Diff} = 16.54, SE = 2.89$, 95%-CI [9.64, 23.44], $p \leq .001$.

System usability. Reliability of the German SUS was acceptable for positively formulated items (Cronbach's $\alpha = .78$), while questionable reliability was found for the inverse items (Cronbach's $\alpha = 0.69$), which was probably caused by the fact that some of the inverted items also included more neutral statements such as "I needed to learn a lot of things before I could get going with this system". To compute the SUS score, ratings were transformed to a range from 0 to 100. A score of 60–80 indicates acceptable usability, a score above 80 indicates good to very good usability, and a score of 100 indicates excellent usability (Brooke, 1996). SUS scores were rated with good usability in the video and AR group and with acceptable usability in the VR group (see Table 2). Since Levene's Test showed that variances were not equal ($p < .001$), Welch's ANOVA was used to analyze group differences (see Fig. 6). System usability as indicated by the SUS differed significantly

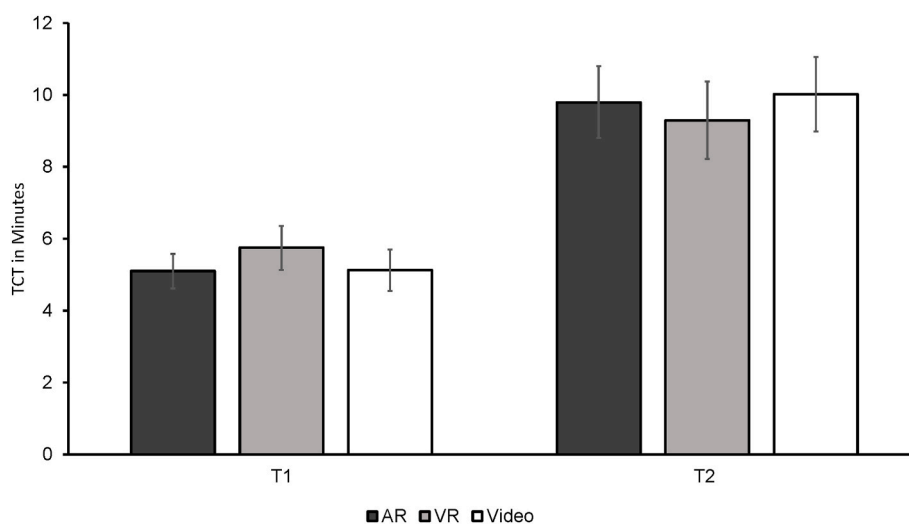


Fig. 3. Short- and long-term effects of AR-,VR- and video-based training on task completion time (TCT).

Note. Due to identified outliers, a total of 33 participants from the AR group, 27 participants from the VR group and 33 participants from the video group were included in the analysis. Means with 95%-confidence interval are depicted.

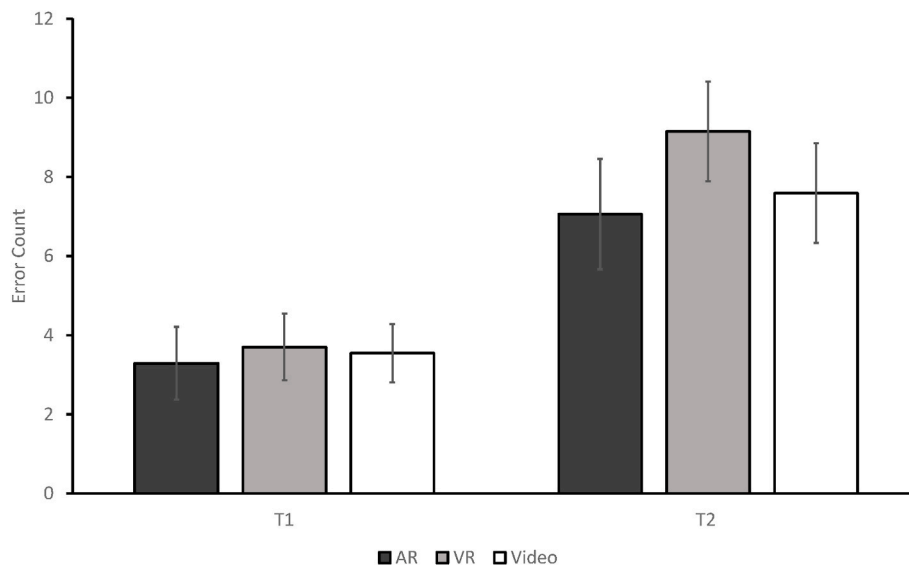


Fig. 4. Short- and long-term effects of AR-,VR- and video-based training on error count.

Note. Due to identified outliers, a total of 34 participants from the AR group, 27 participants from the VR group, and 33 participants from the video group were included in the analysis. Means with 95%-confidence interval are depicted.

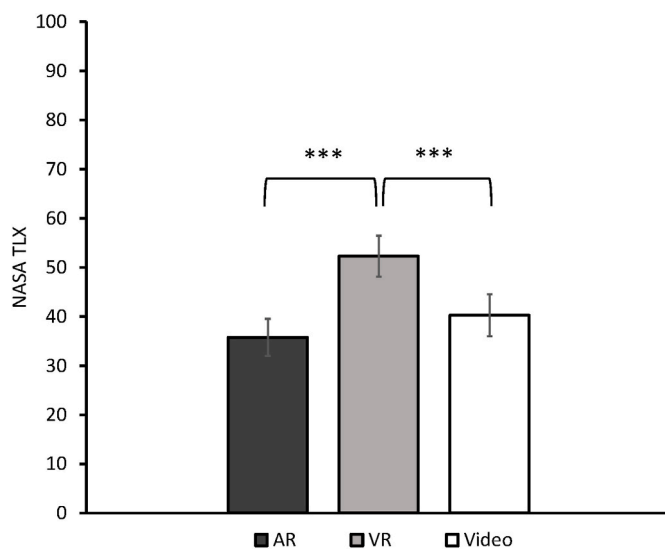


Fig. 5. Perceived task load (NASA-TLX; Hart, 2006) during AR-,VR- and video-based training.

Note. 34 participants from the AR group, 28 participants from the VR group, and 34 participants from the video group were included in the analysis. Means with 95%-confidence interval are depicted. *** $p < .001$.

between the three groups, Welch's $F(2, 55.7) = 24.17, p \leq .001, \eta_p^2 = .43$. Games-Howell comparisons indicated that the SUS of the VR group differed significantly from the SUS of both the AR group with a mean difference of $M_{Diff} = -20.7, SE = 3.66, 95\%-CI [-29.57, -11.81], p \leq .001$ and the video group with a mean difference of $M_{Diff} = -24.88, SE = 3.56, 95\%-CI [-33.55, -16.21], p \leq .001$. No difference was found between AR and video groups with a mean difference of $M_{Diff} = -4.19, SE = 2.37, 95\%-CI [-9.88, 1.51], p = .190$.

Summarizing the analysis regarding the a priori hypotheses, it can be stated that there was no difference between the AR, VR, and video groups in their effect on short- and long-term objective performance measures, but TCT and error count significantly increased from T1 to T2 in all groups (RQ1). Concerning subjective evaluations, it can be stated that the VR group reported significantly higher task load and lower

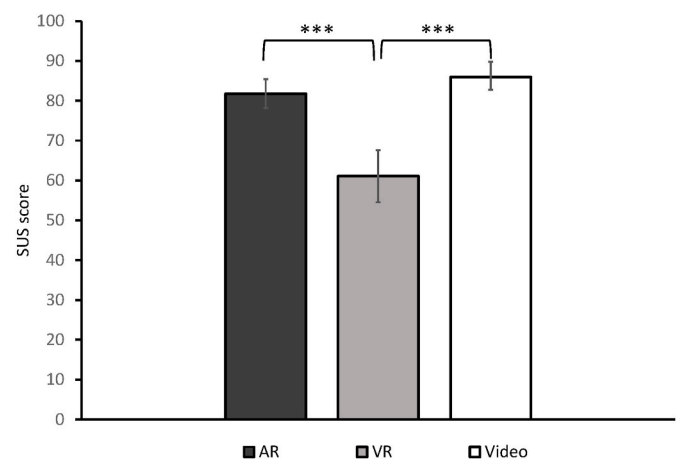


Fig. 6. Perceived system usability (SUS; Brooke, 1996) of AR-, VR- and video-based training.

Note. 34 participants from the AR group, 28 participants from the VR group, and 34 participants from the video group were included in the analysis. Means with 95%-confidence interval are depicted. *** $p < .001$.

system usability than the AR and video groups (RQ2). An overview of these results is presented in Table 3. The main findings are further supplemented by conducting exploratory analyses, which are presented in Section 3.2.

3.2. Results of exploratory analyses

Following the data analysis to decide on our a priori research hypotheses, we conducted post hoc exploratory analyses of the dataset to obtain input for future studies. Here, the focus was on the training duration of the three training methods as well as potential influences by user-related variables. Even though no hypotheses were associated a priori and the study design and sample selection were not aimed at these aspects, we decided to publish the findings during the present study since they provide deeper insights into possible differences between the three training methods (AR-, VR- and video-based).

Table 3
Overview and summary of results on a priori hypotheses.

Measures			Training Method		
			Video	AR	VR
T1(after training)	Subjective evaluations	NASA-TLX	not different from each other lower than VR*	lower than VR*	higher than AR and video*
		System Usability	not different from each other higher than VR*	higher than VR*	lower than AR and video*
T1	Objective performance	TCT	not different from each other		
T2	Objective performance	Error Count	not different from each other		
		TCT	not different from each other		
		Error Count	not different from each other		

Note. An asterisk (*) indicates statistically significant results with $p < .05$.

3.2.1. Exploratory analysis of group differences in training duration

A one-way ANOVA was conducted to assess the effects of training groups on training duration. The training duration indicated how long the participants used that group's technology to learn the assembly procedure on T1. Training durations differed significantly between groups, $F(2, 92) = 39.01, p < .001, \eta_p^2 = .46$. The training duration was lowest in the AR group, $M = 22.94$ min, $SE = .93$, intermediate in the video group, $M = 24.39$ min, $SE = .77$, and highest in the VR group, $M = 34.49$ min, $SE = 1.24$. Tukey's post hoc analysis revealed that differences between video- and VR-based training were significant, with a mean difference of $M_{Diff} = -10.09$ min, $SE = 1.41$, 95%-CI [-13.46, -6.73], $p \leq .001$, as well as between VR and AR, with a mean difference of $M_{Diff} = 11.55$ min, $SE = 1.40$, 95% CI [8.21, 14.89] $p \leq .001$. There was no difference between video and AR ($M_{Diff} = 1.46$ min, $SE = 1.34$, 95% CI [-1.74, 4.66], $p = .524$).

3.2.2. Exploratory analysis of the influence of age of participants on objective performance measures

Age as a user-related factor did not significantly differ between groups, Welch's $F(2, 59.28) = 1.5, p = .232$. A one-way ANCOVA was conducted to determine a statistically significant difference between training groups (AR, VR, video) on performance (TCT, error count) while controlling for age. Age was identified as a significant covariate for TCT on T1, $F(1,89) = 12.89, p = .001, \eta_p^2 = .13$, TCT on T2, $F(1,90) = 4.52, p = .036, \eta_p^2 = .05$, and for error count on T2, $F(1,91) = 6.53, p = .012, \eta_p^2 = .07$, but not for error count on T1, $F(1,89) = 3.82, p = .054$. Results indicated that with an increase in age, TCT and error count increased ($b = .049, 95\%-CI [.02, .07]$ for TCT on T1, $b = .057, 95\%-CI [.00, .11]$ for TCT on T2, $b = .082, 95\%-CI [.02, .15]$ for error count on T2). Although no group difference could be demonstrated in the previous analyses, there was a significant effect of training group on TCT on T1 after adjusting for age, $F(2, 89) = 6.36, p < .05, \eta_p^2 = .66$. Bonferroni-corrected post hoc analysis revealed that, after adjusting for age, the VR group took statistically significantly longer (i.e. a higher TCT) than the AR group to complete the assembly with a mean difference of $M_{Diff} = .89$ s, $SE = .37$, 95%-CI [.16, 1.64], $p < .05$. At the same time, no difference was found between VR and the video group ($M_{Diff} = .69$ s, $SE = .37$, 95%-CI [-.03, 1.42], $p = .062$) or between AR and the video group ($M_{Diff} = -.21$ s, $SE = .35$, 95%-CI [-.91, .49], $p = .056$) for TCT on T1. Moreover, adjusting for age resulted in a significant effect of training groups on error count on T2, $F(2, 91) = 3.52, p < .05, \eta_p^2 = .07$. Bonferroni-corrected post hoc analysis revealed that, after adjusting for age, the VR group made statistically significantly more errors than the AR group with a mean difference of $M_{Diff} = 2.38$ errors, $SE = .92$, 95%-CI [-4.620, -1.15], $p < .05$. No difference was found between VR and the video group ($M_{Diff} = 1.66$ errors, $SE = .90$, 95%-CI [-.54, 3.86], $p = .206$) as well as between AR and the video group ($M_{Diff} = -.73$ errors, $SE = .86$, 95%-CI [-2.83, 1.38], $p = 1.00$) for error count on T2. However, after adjusting for age, no statistically significant difference in training groups was found for TCT on T2, $F(2, 90) = .78, p = .460$, and error count on T1, $F(2, 89) = .43, p =$

.650.

3.2.3. Exploratory analysis of the influence of prior experience on objective performance measures

Prior experience with the respective technology did not differ between groups, $F(2, 90) = 2.01, p = .140$. A one-way ANCOVA was conducted to determine a statistically significant difference between training groups (AR, VR, video) on performance (TCT, error count) after controlling for prior experience. The ANCOVA revealed that for TCT on T1, prior experience was a significant covariate, $F(1,87) = 4.52, p = .036, \eta_p^2 = .05$, indicating that when participants have prior experience, their TCTs were lower in the short-term, i.e., they finished the assembly faster than participants with less prior experience ($b = -.51, 95\%-CI [-.99, -.03]$). However, results showed no difference between the training groups in TCT on T1 after controlling for prior experience, $F(2,87) = 1.59, p = .209$, indicating that prior experience had no influence on whether the training groups differed in their assembly times directly after training. Prior experience was not a significant covariate for TCT on T2, $F(1,88) = 2.31, p = .133$, error count on T1 $F(1,87) = 1.51, p = .222$, and error count on T2, $F(1,89) = 1.1, p = .299$.

An overview of the results of the exploratory data analysis are summarized in Table 4.

4. Discussion

This study aimed to assess and compare AR- and VR-based training in a complex manual assembly task in terms of its effects on objective performance measures and subjective evaluations compared to video-based training. The assembly task was performed two times with a two-week interval so that the training methods could be compared regarding short-term and long-term training effects and outcomes. Assembly errors and task completion time (TCT) of the assembly task (objective performance) were used for comparison as well as perceived task load and perceived system usability (subjective evaluations).

4.1. Differences between AR-, VR- and video-based training in their impact on short- and long-term objective performance

When testing our a priori hypothesis, we found that there was no difference between the AR-, VR- and video-based training groups concerning the objective performance measures TCT and error count in the short- and long-term. However, all groups showed a significant but similar decrease in performance over time. Exploratory data analysis further refined the results and after adjusting the data for age of the participants, partially better results for AR than for VR were revealed. These findings are explained and discussed in more detail below.

The decrease in performance over time could be either due to the high complexity of the assembly steps or because participants were not explicitly informed that they were supposed to perform the assembly again from memory after two weeks. Further research should explore which factors contribute to performance decline and how the effect of

Table 4
Overview and summary of the exploratory analysis.

Measures		Training Method		
		Video	AR	VR
T1 (Training)	Training duration		not different from each other faster than VR*	slower than AR and video*
T1	Objective performance when controlled for age	TCT	not different from AR and VR	slower than AR*
		Error Count	–	–
	Objective performance when controlled for prior experience	TCT	not different from each other	–
		Error Count	–	–
T2	Objective performance when controlled for age	TCT	–	–
		Error Count	not different from AR and VR	not different from video less errors than VR*
	Objective performance when controlled for prior experience	TCT	–	–
		Error Count	–	–

Note. An asterisk (*) indicates statistically significant results with $p < .05$. Empty rows (–) indicate that no effect of the covariate could be found.

forgetting can be minimized in all training methods. Doolani, Owens, and co-authors (2020) discussed that, for example, the use of storytelling could be a helpful approach to remembering processes more easily over time.

Another factor influencing short- and long-term objective performance could be related to user-related variables. Overall, in our study a higher age contributed to longer assembly times in the short-term and more errors in the long-term. This is a known effect, as older adults are more likely to perform worse than younger adults on procedural assembly tasks due to age-related declines in working memory (Abubakar and Wang, 2019; Morrell and Park, 1993). After adjusting for the effect of age, the AR group's performance was significantly faster in the short-term and the group showed significantly fewer errors in the long-term than the VR group. AR-based training might thus be particularly helpful in enhancing the time effectiveness of assembly in the short-term, which supports the results of previous studies (Hou and Wang, 2013; Koumaditis et al., 2019; Kwiatek et al., 2019; Loch et al., 2019) as well as highly recent results found by Liu et al. (2022), who showed that for training maintenance tasks, there was an advantage of AR over VR and video-based training as task difficulty increased. Moreover, these results support the previous finding that AR has particular potential to contribute to training success in the long-term by reducing error rates and thus contributing to the overall efficiency, effectiveness, and flexibility of production (Gavish et al., 2011). The finding that the performance after VR-based training showed worse results than AR could be justified by a higher need for the transfer of the virtual training environment to the real assembly use case. However, after adjusting for age, no difference in task completion times on T1 and T2 and error count on T2 could be found between VR-based and video-based training nor between AR-based and video-based training. Accordingly, both AR- and VR-based training did not differ significantly from the more conventional video-based method. In their first attempt to test AR and VR against a video control group, Gavish et al. (2015) came to similar conclusions. In their study, VR did not show any advantage over the video-based training of a relatively simple six-step actuator assembly task. AR, in contrast, showed no advantage concerning assembly times, but led to fewer unsolved errors than video-based training in the short-term. In our study, no differentiation of errors into solved vs. unsolved errors has been realized, which could have contributed to a better understanding of the effects of AR-vs. VR-based training on error count in assembly tasks (Werrlich et al., 2018; Werrlich et al., 2018). Both Gavish et al. (2015) and Liu et al. (2022) further discussed that the similarity of the training methods might be related to relatively easy assembly tasks and hypothesized that AR and VR might have a significant advantage over conventional methods when a more demanding

task is used. This, however, cannot be supported by our study results. Moreover, the authors discussed that participants who have more experience in using AR and VR technologies might be able to use these technologies more efficiently (Gavish et al., 2015). In our study, we found that prior experience of participants indeed appeared to play an important role: across all groups, more experience corresponded with shorter TCT in the short-term. Comparable findings on experience and performance were found by Schwarz and co-authors (2020) as well as Abubakar and Wang (2019), who identified experience as the most significant human factor affecting individual performance. Whether increasing prior experience can be expected to have a positive effect on training success for all training methods should be clarified in the context of further research.

At this point, it can be concluded from the results of the exploratory analysis of the covariates age and prior experience that older persons may need more support in using these technologies and that prior experience may contribute to higher short-term success of the training. However, since this analysis was exploratory and neither study design nor sample selection were aimed at these aspects, interpreting these results should be treated with caution. Further research is needed to uncover specific differences between and advantages of the training methods.

4.2. Differences between AR-, VR- and video-based training in their subjective evaluation by users

In addition to the results on objective performance, our study showed that the subjective evaluations of AR-, VR-, and video-based training, namely task load and usability, also provide exciting insights into the different perceptions of the training methods by the users. For instance, we found significant differences between the groups in terms of perceived task load during training. Contrary to our expectation that AR would generate less task load than video training by allowing instructions to be directly displayed on the real components and workspace (Hou et al., 2015; Koumaditis et al., 2019; Liu et al., 2022), AR- and video-based training resulted in a similar perceived task load. This could have been caused by the fact that the video training was also presented directly at the workplace and thus the transfer of what was seen to the actual assembly task was easier than in a scenario where the video is shown outside the workplace. Moreover, task load was only measured by using subjective ratings, which could limit the expressiveness of the results. In a recent study of Drouot et al. (2022), ocular and behavioral data indicated that AR led to increased mental workload in comparison to computerized instructions. Thus, the inclusion of additional objective data sources to assess task load in AR and VR

training is recommended for future studies. Interestingly, the VR-based training resulted in a significantly higher task load than the other two training systems, which could be because rotating and handling virtual components might require more effort than working with real hands and components (Schwarz et al., 2020). At this stage, we can therefore conclude that for this type of complex assembly task, AR and video-based training are particularly suitable for providing assembly instructions with a comparatively low workload.

Similar findings were found concerning perceived usability, revealing that AR- and video-based training were perceived to have the same degree of *good* usability, while the VR-based training was rated with significantly lower, but still *acceptable* usability. These results confirm the current data regarding good to very good usability of AR-based training (Daling and Schlittmeier, 2020; Loch et al., 2019) as well as a less consistent and sometimes even worse assessment of VR-based training (Daling and Schlittmeier, 2020; Oren et al., 2012). Most likely, participants in the VR group found it more difficult or cumbersome to handle and merge virtual objects since the object was not manipulated directly, but via controllers representing virtual hands (da Silva Marinho et al., 2022; Gavish et al., 2015; Schwarz et al., 2020). Additional interviewing of participants after training about the reasons for their perceptions, e.g., via a retrospective think-aloud session, would be useful for future studies to further elucidate the reasons for low SUS in the VR condition. The analysis of training duration revealed that participants in the VR group took the longest to complete the training, which might reflect difficulties in dealing with VR. The disadvantages of VR compared to AR and video in terms of subjective evaluations and longer training durations should be addressed and investigated in further studies, e.g., through technological improvements of interaction features.

5. Conclusion

In this study, we investigated the short- and long-term effects of AR-, VR- and video-based training on objective performance measures as well as their subjective evaluation in a manual assembly task. Concerning objective performance, our analysis showed that there was no difference between AR-, VR- and video-based training in terms of their short- and long-term effects on task completion time and task accuracy. However, when controlling for age, the AR group's performance was revealed to have faster task-completion times in the short-term and fewer errors in the long-term compared to the VR group. In addition, across the groups more prior experience corresponded with better short-term results. Future research should thus investigate the use of these technologies at different age and expertise levels to better determine whether one of the technologies is more suitable for certain user groups. In all training groups there was a significant decline in performance over time. This raises the need for an investigation of factors influencing long-term skill retention.

Furthermore, we could show that subjective evaluations such as task load and usability provide important insights into how users perceive the technologies differently. In our study, AR- and video-based training were better evaluated in terms of task load and system usability than VR-based training. These results might play a significant role in the acceptance of AR-, VR- and video-based training in the industry. For training novices in complex assembly tasks, both AR and video seem to be particularly suitable, while the usability and intuitive use of VR in training still needs to be improved. Future research should focus on the extent to which VR can further improve its usability and on its impact on perceived task load. In addition, the possible impact of subjective evaluations of the training methods on objective performance measures should be explored in further studies.

Taking into account the effects on both objective performance and subjective evaluations, AR-based training in particular can be considered an effective alternative to video-based training to ensure short- and long-term training success in manual assembly tasks. In the future,

advanced features of AR technologies could even increase their potential and enable completely new training possibilities beyond the restrictions of conventional video-based training. While VR showed no difference to the video-based method with regard to objective performance measures, first indications could be found that VR performed worse compared to AR. In addition, VR showed a clear disadvantage compared to AR and video in terms of subjective evaluations. This might be compensated in the future by task load reduction, increased usability and improved confidence of users in handling VR, but currently AR seems to be the more suitable technology for manual assembly tasks.

In future research, associating objective performance measures and subjective evaluations should be an integral part of validating the conclusions found here comparing AR-, VR- and other conventional training in terms of their short- and long-term effects.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.apergo.2023.104021>.

References

- Abubakar, M.I., Wang, Q., 2019. Key human factors and their effects on human centered assembly performance. *Int. J. Ind. Ergon.* 69, 48–57. <https://doi.org/10.1016/j.ergon.2018.09.009>.
- Al-Ahmari, A., Ameen, W., Abidi, M.H., Mian, S.H., 2018. Evaluation of 3D printing approach for manual assembly training. *Int. J. Ind. Ergon.* 66, 57–62. <https://doi.org/10.1016/j.ergon.2018.02.004>.
- AlGeddawy, T., ElMaraghy, H.A., 2012. Product variety management in design and manufacturing: challenges and strategies. In: ElMaraghy, H.A. (Ed.), *Enabling Manufacturing Competitiveness and Economic Sustainability*. Springer, pp. 518–523. https://doi.org/10.1007/978-3-642-23860-4_85.
- Ariansyah, D., Erkoyuncu, J.A., Eimontaite, I., Johnson, T., Oostveen, A., Fletcher, S.R., Sharples, S.C., 2021. A head mounted augmented reality design practice for maintenance assembly: toward meeting perceptual and cognitive needs of AR users. *Appl. Ergon.* 98, 103597. <https://doi.org/10.1016/j.apergo.2021.103597>.
- Brooke, J., 1996. SUS-A quick and dirty usability scale. *Usability Eval. Ind.* 189 (194), 4–7. <https://doi.org/10.1201/9781498710411-35>.
- Carlson, P., Peters, A., Gilbert, S.B., Vance, J.M., Luse, A., 2015. Virtual training: learning transfer of assembly tasks. *IEEE Trans. Visual. Comput. Graph.* 21 (6), 770–782. <https://doi.org/10.1109/tvcg.2015.2393871>.
- Cohen, J., 1988. *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates, New Jersey.
- Daling, L.M., Schlittmeier, S.J., 2022. Effects of augmented reality-, virtual reality-, and mixed reality-based training on objective performance measures and subjective evaluations in manual assembly tasks: a scoping review. *Hum. Factors* 0 (0). <https://doi.org/10.1177/00187208221105135>.
- da Silva Marinho, A., Tertton, U., Jones, C.M., 2022. Cybersickness and postural stability of first time VR users playing VR videogames. *Appl. Ergon.* 101, 103698. <https://doi.org/10.1016/j.apergo.2022.103698>.
- DaValle, A., Azhar, S., 2020. An investigation of mixed reality technology for onsite construction assembly. *MATEC Web Conf.* 312 (2), 06001. <https://doi.org/10.1051/mateconf/202031206001>.
- DeLone, W.H., McLean, E.R., 2002. Information systems success revisited. In: *Proceedings of the 35th Annual Hawaii International Conference on System Sciences*. IEEE, pp. 2966–2976. <https://doi.org/10.1109/HICSS.2002.994345>.
- Doolani, S., Owens, L., Wessels, C., Makedon, F., 2020a. vIS: an immersive virtual storytelling system for vocational training. *Appl. Sci.* 10 (22), 8143. <https://doi.org/10.3390/app10228143>.
- Doolani, S., Wessels, C., Kanal, V., Sevastopoulos, C., Jaiswal, A., Nambiappan, H., Makedon, F., 2020b. A review of extended reality (xr) technologies for manufacturing training. *Technologies* 8 (4), 77. <https://doi.org/10.3390/technologies8040077>.
- Drouot, M., Le Bigot, N., Bricard, E., De Bougrenet, J.L., Nourrit, V., 2022. Augmented reality on industrial assembly line: impact on effectiveness and mental workload. *Appl. Ergon.* 103, 103793. <https://doi.org/10.1016/j.apergo.2022.103793>.
- Eversberg, L., Grosenick, P., Meusel, M., Lambrecht, J., 2021. An industrial assistance system with manual assembly step recognition in virtual reality. In: *Proceedings of 2021 International Conference on Applied Artificial Intelligence (ICAPAI)*. IEEE, pp. 1–6. <https://doi.org/10.1109/ICAPAI49758.2021.9462061>.

- Faul, F., Erdfelder, E., Lang, A.-G., Buchner, A., 2007. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. <https://doi.org/10.3758/BF03193146>.
- Funk, J., Schmidt, L., 2021. Evaluation of an augmented reality instruction for a complex assembly task: comparison of a smartphone-based augmented reality instruction with a conventional paper instruction for the teach-in phase in manual assembly. *com* 20 (1), 63–72. <https://doi.org/10.1515/icom-2021-0005>.
- Gabajová, G., Furmannová, B., Medvecká, I., Grznár, P., Krajčovič, M., Furmann, R., 2019. Virtual training application by use of augmented and virtual reality under university technology enhanced learning in Slovakia. *Sustainability* 11 (23), 6677. <https://doi.org/10.3390/su11236677>.
- Gao, M., Kortum, P., Oswald, F., 2020. Multi-Language toolkit for the system usability scale. *Int. J. Hum. Comput. Interact.* 36 (3), 1883–1901. <https://doi.org/10.1080/10447318.2020.1801173>.
- Gavish, N., Gutierrez Seco, T., Webel, S., Rodriguez, J., Peveri, M., Bockholt, U., 2011. Transfer of skills evaluation for assembly and maintenance training. *BIO Web Conf.* 1, 00028 <https://doi.org/10.1051/bioconf/20110100028>.
- Gavish, N., Gutierrez, T., Webel, S., Rodriguez, J., Peveri, M., Bockholt, U., Tecchia, F., 2015. Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks. *Interact. Learn. Environ.* 23 (6), 778–798. <https://doi.org/10.1080/10494820.2013.815221>.
- González-Franco, M., Cermeron, J., Li, K., Pizarro, R., Thorn, J., Hannah, P., Hutabarat, W., Tiwari, A., Bermell-Garcia, P., 2016. *Immersive Augmented Reality Training for Complex Manufacturing Scenarios*. ArXiv abs/1602.01944.
- Hart, S.G., 2006. NASA-task load index (NASA-TLX); 20 years later. In: *Proceedings Of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50. Sage publications, Los Angeles, pp. 904–908. <https://doi.org/10.1177/154193120605000909>, 9.
- Hart, S.G., Staveland, L.E., 1988. Development of NASA-TLX (task load index): results of empirical and theoretical research. *Adv. Psychol.* 52, 139–183. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9).
- Heinz, M., Büttner, S., Röcker, C., 2019. Exploring training modes for industrial augmented reality learning. In: *Proceedings of the 12th ACM International Conference on Pervasive Technologies Related to Assistive Environments*. Association for Computing Machinery, New York, pp. 398–401. <https://doi.org/10.1145/3316782.3322753>.
- Hoedt, S., Claeys, A., Van Landeghem, H., Cottyn, J., 2017. The evaluation of an elementary virtual training system for manual assembly. *Int. J. Prod. Res.* 55 (24), 7496–7508. <https://doi.org/10.1080/00207543.2017.1374572>.
- Hou, L., Wang, X.Y., 2013. A study on the benefits of augmented reality in retaining working memory in assembly tasks: a focus on differences in gender. *Autom. ConStruct.* 32, 38–45. <https://doi.org/10.1016/j.autcon.2012.12.007>.
- Hou, L., Wang, X.Y., Bernold, L., Love, P.E.D., 2013. Using animated augmented reality to cognitively guide assembly. *J. Comput. Civ. Eng.* 27 (5), 439–451. [https://doi.org/10.1061/\(asce\)cp.1943-5487.0000184](https://doi.org/10.1061/(asce)cp.1943-5487.0000184).
- Hou, L., Wang, X.Y., Truijens, M., 2015. Using augmented reality to facilitate piping assembly: an experiment-based evaluation. *J. Comput. Civ. Eng.* 29 (1), 05014007 [https://doi.org/10.1061/\(asce\)cp.1943-5487.0000344](https://doi.org/10.1061/(asce)cp.1943-5487.0000344).
- Jang, J., Ko, Y., Shin, W.S., Han, I., 2021. Augmented reality and virtual reality for learning: an examination using an extended technology acceptance model. *IEEE Access* 9, 6798–6809. <https://doi.org/10.1109/ACCESS.2020.3048708>.
- Joshi, S., Hamilton, M., Warren, R., Faucett, D., Tian, W., Wang, Y., Ma, J., 2021. Implementing Virtual Reality technology for safety training in the precast/prestressed concrete industry. *Appl. Ergon.* 90, 103286 <https://doi.org/10.1016/j.apergo.2020.103286>.
- Kagermann, H., Hellbig, J., Hellingner, A., Wahlster, W., 2013. Recommendations for implementing the strategic initiative INDUSTRIE 4.0: securing the future of German manufacturing industry; Final report of the Industrie 4.0 Working Group. *Forschungsunion*. <https://gateway-bayern.de/BV041628398>.
- Kaplan, A.D., Cruik, J., Endsley, M., Beers, S.M., Sawyer, B.D., Hancock, P.A., 2021. The effects of virtual reality, augmented reality, and mixed reality as training enhancement methods: a meta-analysis. *Hum. Factors* 63 (4), 706–726. <https://doi.org/10.1177/0018720820904229>.
- Kim, M., Park, K., Choi, S.H., Lee, J.Y., Kim, D., 2018. AR/VR-Based live manual for user-centric smart factory services. In: Moon, I., Lee, G., Park, J., Kiritsis, D., von Cieminski, G. (Eds.), *IFIP International Conference On Advances In Production Management Systems (APMS)*, vol. 536. Springer, Cham, pp. 417–421. https://doi.org/10.1007/978-3-319-99707-0_52.
- Kostaras, N., Xenos, M., 2012. Usability evaluation of augmented reality systems. *Intell. Decis. Technol.* 6 (2), 139–149. <https://doi.org/10.3233/IDT-2012-0130>.
- Koumaditis, K., Chinello, F., Mitkidis, P., Karg, S., 2020. Effectiveness of virtual versus physical training: the case of assembly tasks, trainer's verbal assistance, and task complexity. *IEEE Comput. Graph. Appl.* 40 (5), 41–56. <https://doi.org/10.1109/mcg.2020.3006330>.
- Koumaditis, K., Venckute, S., Jensen, F.S., Chinello, F., 2019. Immersive training: outcomes from small scale AR/VR pilot-studies. In: *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, pp. 1–5. <https://doi.org/10.1109/VR44988.2019.9044162>.
- Kwiatek, C., Sharif, M., Li, S., Haas, C., Walbridge, S., 2019. Impact of augmented reality and spatial cognition on assembly in construction. *Autom. ConStruct.* 108, 102935 <https://doi.org/10.1016/j.autcon.2019.102935>. Article.
- Lawson, G., Salantri, D., Waterfield, B., 2016. Future directions for the development of virtual reality within an automotive manufacturer. *Appl. Ergon.* 53 (Pt B), 323–330. <https://doi.org/10.1016/j.apergo.2015.06.024>.
- Li, Z., Wang, J., Yan, Z., Wang, X., Anwar, M.S., 2019. An interactive virtual training system for assembly and disassembly based on precedence constraints. In: *InComputer Graphics International Conference 2019*. Springer, Cham, pp. 81–93. https://doi.org/10.1007/978-3-030-22514-8_7.
- Liu, X.-W., Li, C.-Y., Dang, S., Wang, W., Qu, J., Chen, T., Wang, Q.-L., 2022. Research on training effectiveness of professional maintenance personnel based on virtual reality and augmented reality technology. *Sustainability* 14 (21), 14351. <https://doi.org/10.3390/su142114351>. MDPI AG.
- Loch, F., Ziegler, U., Vogel-Heuser, B., 2019. Using real-time feedback in a training system for manual procedures. *IFAC-PapersOnLine* 52 (19), 241–246. <https://doi.org/10.1016/j.ifacol.2019.12.089>.
- Longo, L., 2018. Experienced mental workload, perception of usability, their interaction and impact on task performance. *PLoS One* 13 (8), e0199661. <https://doi.org/10.1371/journal.pone.0199661>.
- Lotter, B., 2006. Einführung. In: Lotter, B., Wiendahl, H.P. (Eds.), *Montage in der industriellen Produktion: Ein Handbuch für die Praxis*. Springer, Berlin, Heidelberg, pp. 1–8. https://doi.org/10.1007/3-540-36669-5_1.
- Milgram, P., Kishino, F., 1994. A taxonomy of mixed reality visual displays. *IEICE Trans. Info Syst.* 77 (12), 1321–1329.
- Miqueo, A., Torralba, M., Yagüe-Fabra, J.A., 2020. Lean manual assembly 4.0: a systematic review. *Appl. Sci.* 10 (23), 8555. <https://doi.org/10.3390/app10238555>.
- Morrell, R.W., Park, D.C., 1993. The effects of age, illustrations, and task variables on the performance of procedural assembly tasks. *Psychol. Aging* 8 (3), 389–399. <https://doi.org/10.1037/0882-7974.8.3.389>.
- Murcia-López, M., Steed, A., 2018. A comparison of virtual and physical training transfer of bimanual assembly tasks. *IEEE Trans. Visual. Comput. Graph.* 24 (4), 1574–1583. <https://doi.org/10.1109/tvcg.2018.2793638>.
- Oren, M., Carlson, P., Gilbert, S., Vance, J.M., 2012. Puzzle assembly training: real world vs. Virtual environment. In: Coquillart, S., Feiner, S., Kiyokawa, K. (Eds.), *2012 IEEE Virtual Reality Workshops*. IEEE, pp. 27–30. <https://doi.org/10.1109/VR.2012.6180873>.
- Roldan, J.J., Crespo, E., Martín-Barrio, A., Pena-Tapia, E., Barrientos, A., 2019. A training system for industry 4.0 operators in complex assemblies based on virtual reality and process mining. *Robot. Comput. Integrated Manuf.* 59, 305–316. <https://doi.org/10.1016/j.rcim.2019.05.004>.
- Sautter, B., Daling, L., 2021. Mixed reality supported learning for industrial on-the-job training. In: *Proceedings of the Conference on Learning Factories*. <https://doi.org/10.2139/ssrn.3864189>.
- Schmidt, R.A., Bjork, R.A., 1992. New conceptualizations of practice: common principles in three paradigms suggest new concepts for training. *Psychol. Sci.* 3 (4), 207–218. <https://doi.org/10.1111/j.1467-9280.1992.tb00029.x>.
- Schwarz, S., Regal, G., Kempf, M., Schatz, R., 2020. Learning success in immersive virtual reality training environments: practical evidence from automotive assembly. In: *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences*. Shaping Society, pp. 1–11. <https://doi.org/10.1145/3419249.3420182>.
- Skarbez, R., Smith, M., Whitton, M.C., 2021. Revisiting Milgram and Kishino's reality-virtuality continuum. *Front. Virtual Real.* 2, 647997 <https://doi.org/10.3389/frvir.2021.647997>. Article.
- Venkatesh, V., Brown, S.A., Maruping, L.M., Bala, H., 2008. Predicting different conceptualizations of system use: the competing roles of behavioral intention, facilitating conditions, and behavioral expectation. *MIS Q.* 483–502. <https://doi.org/10.2307/25148853>.
- Wang, X., Ong, S.K., Nee, A.Y.C., 2016. A comprehensive survey of augmented reality assembly research. *Adv. Manuf.* 4, 1–22. <https://doi.org/10.1007/s40436-015-0131-4>.
- Wang, C., Tsai, N., Lu, J., Wang, M.J., 2019. Usability evaluation of an instructional application based on Google Glass for mobile phone disassembly tasks. *Appl. Ergon.* 77, 58–69. <https://doi.org/10.1016/j.apergo.2019.01.007>.
- Webel, S., Bockholt, U., Engelke, T., Gavish, N., Olbrich, M., Preusche, C., 2013. An augmented reality training platform for assembly and maintenance skills. *Robot. Autom. Syst.* 61 (4), 398–403. <https://doi.org/10.1016/j.robot.2012.09.013>.
- Werrlich, S., Lorber, C., Nguyen, P.A., Yanez, C.E.F., Notni, G., 2018a. Assembly training: comparing the effects of head-mounted displays and face-to-face training. In: Chen, J.Y.C., Fragomeni, G. (Eds.), *Virtual, Augmented and Mixed Reality: Interaction, Navigation, Visualization, Embodiment, and Simulation*, Vamr 2018, Pt I, vol. 10909. Springer, Cham, pp. 462–476. https://doi.org/10.1007/978-3-319-91581-4_35.
- Werrlich, S., Nguyen, P.-A., Notni, G., 2018b. Evaluating the training transfer of Head-Mounted Display based training for assembly tasks. In: *PETRA '18: Proceedings Of the 17th Pervasive Technologies Related To Assistive Environments Conference*, pp. 297–302. <https://doi.org/10.1145/3197768.3201564>.
- Westerfield, G., Mitrovic, A., Billingham, M., 2015. Intelligent augmented reality training for motherboard assembly. *Int. J. Artif. Intell. Educ.* 25 (1), 157–172. <https://doi.org/10.1007/s40593-014-0032-x>.
- Xu, X., Lu, Y., Vogel-Heuser, B., Wang, L., 2021. Industry 4.0 and Industry 5.0-Inception, conception and perception. *J. Manuf. Syst.* 61, 530–535. <https://doi.org/10.1016/j.jmsy.2021.10.006>.
- Lea M. Daling, M.Sc. completed her master's in social, organizational, and work psychology at the Freie Universität Berlin in 2017. Since then, she is a research associate at the Chair of Production Metrology and Quality Management & Institute for Information Management in Mechanical Engineering (WZL-MQ/IMA) as well as an external Ph.D. student in the teaching and research area of Work and Engineering Psychology at the RWTH Aachen University.

Marisa Tenbrock, M.Sc. completed her master's degree in psychology at RWTH Aachen University in 2021. During her studies, she focussed on work and engineering psychology and worked as a research assistant at the Chair of Production Metrology and Quality Management & Institute for Information Management in Mechanical Engineering. She is now a consultant at a strategy consulting firm.

Prof. Dr. Phil. Ingrid Isenhardt is Academic Director at WZL-MQ/IMA (Chair of Production Metrology and Quality Management & Institute for Information Management in Mechanical Engineering Information Management in Mechanical Engineering) as an associate (apl.) professor at the Faculty of Mechanical Engineering with a second position at the

Faculty of Philosophy of RWTH Aachen University. Her research focus is on socio-technical aspects of digital transformation, sustainable production & value creation, as well as information and data management in mechanical engineering.

Prof. Dr. Sabine J. Schlittmeier is head of the teaching and research area of Work and Engineering Psychology at RWTH Aachen University. She was previously a Professor of Psychology with a focus on Auditory Cognition at RWTH Aachen University (2018–2020; funded by the HEAD-Genuit Foundation) and a Professor of Applied Psychology at HSD Hochschule Döpfer in Cologne (2015–2018).