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

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

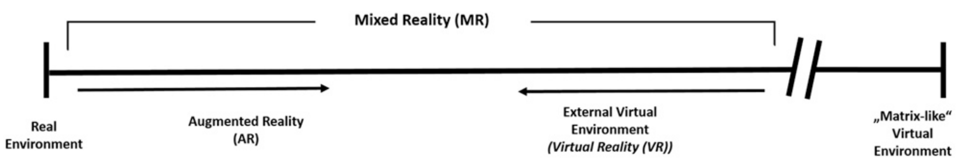

Figure 1. Revision of the Reality-Virtuality Continuum (based on Milgram & Kishino, 1994), where Mixed Reality Was Used as Umbrella Term for AR and VR technologies (based on Skarbez et al., 2021). Note that VR was Considered as Part of External Virtual Environments in Skarbez et al., 2021.

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

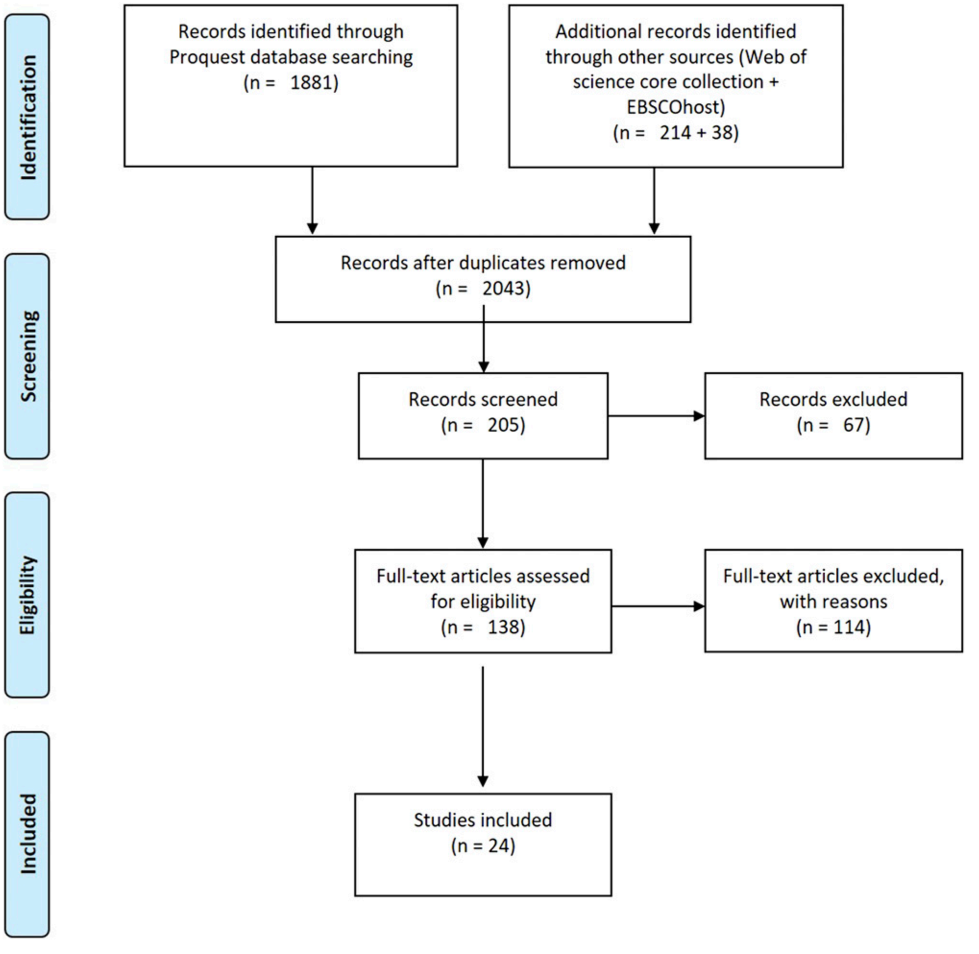

Figure 2. PRISMA flow diagram for scoping reviews.

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
Koumaditis 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
Kwiatek 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
Westerfield 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

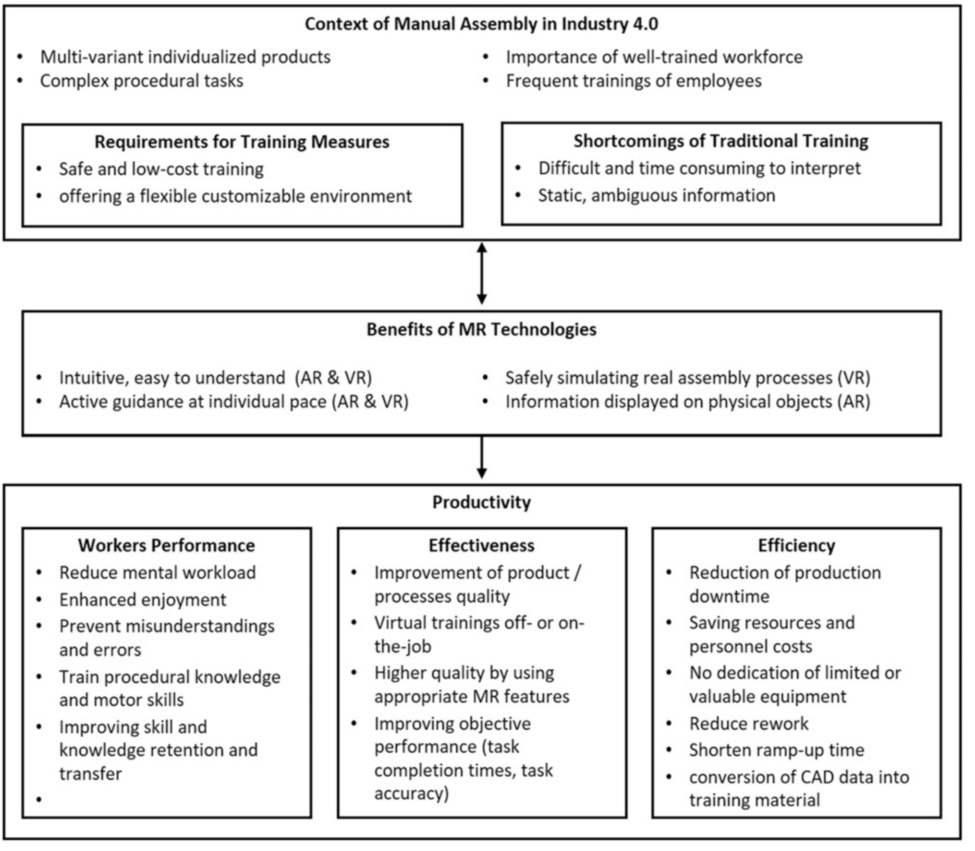

Figure 3. Identified needs and expected benefits of MR-based trainings for manual assembly tasks in the context of Industry 4.0 (RQ1).

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

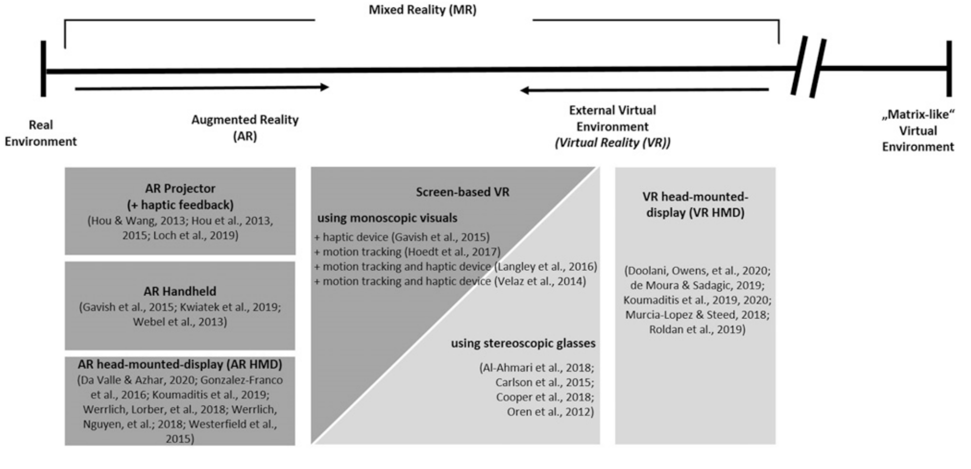

Figure 4. Classification of MR technologies and their interaction features used in the reviewed articles as AR-based training, Screen-based VR training, and VR head-mounted display (HMD)–based training (RQ2) according to the revised reality-virtuality continuum based on Skarbez et al. (2021).

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).

			Objective Measures					Subjective Measures				
Author (Year)	Technology	Assembly Task (J) Joining, (H) Handling, (I) Inspecting, (A) Adjusting	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
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
Webel 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)
Koumaditis et al. (2019)	AR HMD	Real product assembly task (J) (H) (I) (A)	AR better than trainer*	—	—	—	—	—	Workload (effort) lower with AR than trainer*	—	—	—
Werrlich, 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)
Werrlich, 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	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	Assembly Task (J) Joining, (H) Handling, (I) Inspecting, (A) Adjusting	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
Hoedt et al. (2017)	Screen-based VR (monoscopic + motion tracking)	sub assembly (J) (H) (I) (A)	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)	assembly of a car door (J) (H) (I) (A)	—	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)	electrohydaulic valve assembly (J) (H) (I) (A)	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)	assemble a mid bearing (J) (H) (I) (A)	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)	assembling six-piece burr puzzles (H) (I)	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)	Wheel change (H) (I) (A)	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)	Burr puzzle with blocks (H) (I)	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	Assembly Task (J) Joining, (H) Handling, (I) Inspecting, (A) Adjusting	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
Doolani, Owens, et al. (2020)	VR HMD	Using a mechanical diameter (H) (I) (A)	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)	assembling a toy helicopter (J) (H) (I) (A)	Stereoscopic VR better than monoscopic VR*, stereosc. screen-based VR* and monosc. screen-based VR *	—	Stereoscopic VR not different from monoscopic VR*, stereoscopic, VR better than stereosc. and monosc. screen-based VR*	—	—	VR better than VR non-ster., 3D PC and 2D display	—	—	VR easier than 3D PC* and 2D display*	eye strain observerd as symptom for Stereoscopic view, general discomfort for non-stereoscopic
Koumaditis et al. (2020)	VR HMD	Complex cube assembly (J) (H) (I) (A)	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 Steed (2018)	VR HMD	Burr puzzle (H) (I) (A)	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 and 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	Assembly of blocks (J) (H) (I) (A)	—	—	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; Kwiatek 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 (Kwiatek 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. pipe-fitters) 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 (Kwiatek 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 to 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., Strenge, 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/JQSDE

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

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

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

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*, 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
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
Kwiatek, 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.1000097s
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., & Ambihaipahan, 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., Danieletto, 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., Billinghurst, 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., & Billinghurst, 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