Applied Ergonomics 110 (2023) 104021

Contents lists available at ScienceDirect

Applied Ergonomics

journal homepage: www.elsevier.com/locate/apergo

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

Lea M. Daling [a,b,*], Marisa Tenbrock [a,b], Ingrid Isenhardt [a], Sabine J. Schlittmeier [b]

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

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

ARTICLE INFO

Keywords:
Virtual reality
Augmented reality
Training and long-term retention

ABSTRACT

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

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

In recent years, new technologies have been introduced to all areas of production to enhance the effectiveness, efficiency, and flexibility of the overall production workflow (Eversberg et al., 2021; Kim et al., 2018; Miqueo et al., 2020; Xu et al., 2021). This also changed the qualification requirements for workers to be prepared for new tasks effectively (e.g., reducing errors), efficiently (e.g., saving time resources), and flexibly (e.g., being able to adapt to new requirements) (Joshi et al., 2021; Kagermann et al., 2013; Sautter and Daling, 2021). In the field of manual assembly tasks in particular, the increasing individualization and complexity of products, shorter product life cycles, and staff turnover all led to a higher demand for appropriate training and assistance technologies (AlGeddawy and ElMaraghy, 2012; Daling and Schlittmeier, 2022; Werrlich et al., 2018). In this context, digital technologies such as augmented reality (AR) and virtual reality (VR) have been introduced as promising media for the training and qualification of workers (Heinz et al., 2019; Joshi et al., 2021; Lawson et al., 2016; Wang et al., 2016, 2019) providing them with safe, hazard-free, as well as flexible, location- and time-independent training (Kaplan et al., 2021).

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

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

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

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

https://doi.org/10.1016/j.apergo.2023.104021
Received 25 July 2022; Received in revised form 15 March 2023; Accepted 17 March 2023
Available online 30 March 2023

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

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

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

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

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

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

1.2.1. Objective performance measures

When evaluating training success after AR- and VR-based training, a first reference is usually to examine whether the training was effective in preparing users for the new task. Common objective performance measures in manual assembly training are task completion time (TCT) and the number of errors (e.g., Ariansyah et al., 2021; Drouot et al., 2022; Murcia-López and Steed, 2018; Werrlich et al., 2018). Moreover, in addition to the short-term effects of training on performance, long-term skill retention, i.e., remembering the assembly task over several days and weeks, is another important indicator of objective performance in an increasingly flexible production environment (Doolani et al., 2020).

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

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

1.2.2. Subjective evaluations

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

When using AR-based training, an increase in objective performance was often accompanied by positive impacts on subjective evaluations. Hou et al. (2015) showed that using AR led to a reduced task load in addition to outperforming paper training with regard to completion times and error rates. Koumaditis et al. (2019) discussed how the reduction in perceived task load using AR was due to the provision of contingent, in-situ information through temporal and spatial anchoring on relevant objects, because the AR-based training could direct the trainee's attention precisely to where it was needed at the time it was needed (Hou et al., 2013, 2015; Koumaditis et al., 2019). Concerning perceived system usability, AR-based training was found to be rated higher than video-based training (Loch et al., 2019) but lower than paper manuals (Wang et al., 2019).

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

1.3. Aim of the present study and hypotheses

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

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

(RQ1). Do AR-, VR- and video-based trainings show differences in their impact on short- and long-term objective performance?
(RQ2). Do AR-, VR- and video-based trainings differ in how they are subjectively evaluated by users?

To answer these questions, we first looked at the short- and long-term effects of AR and VR-based training on objective performance measures in comparison with video-based training. TCT and error count were used as outcome variables in an assembly task that was performed twice: immediately after training (T1) and two weeks after training (T2). According to Schmidt and Bjork (1992), a retention phase after the initial training should be long enough to ensure that any temporary effects of the independent variable have been dissipated. The two-week interval is the current standard in the literature for long-term retention of industry training (Carlson et al., 2015; Murcia-López and Steed, 2018) and realistic in occupational settings: within Industry 4.0, which is characterized by mass customization and product variant variety, it is common that some variants are only assembled from time to time. For both short-term and long-term effects (i.e. on both T1 and T2), we expected that both AR- and VR-based training would lead to better objective performance measures (lower TCT and fewer errors) in comparison to video-based training.

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

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

2. Materials and methods

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

2.1. Participants

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

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

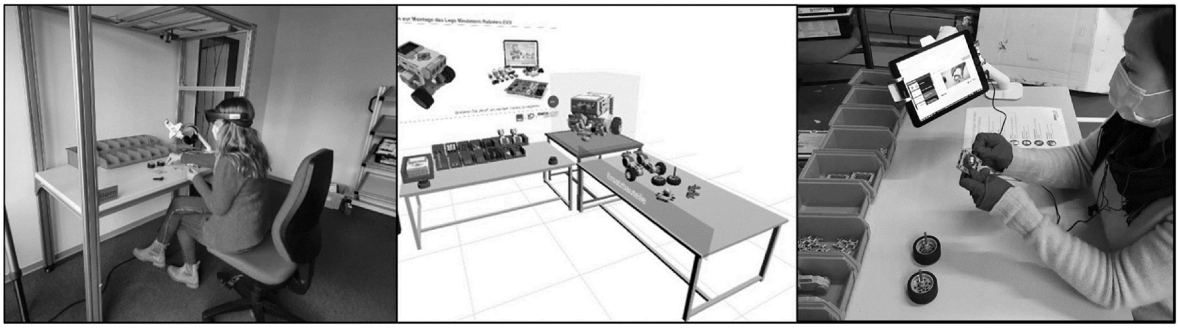

Fig. 1. Physical workplace of the AR group (left), 3D modeled workspace of the VR group (middle), and physical workplace of the video group (right).

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

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

2.2. Materials

2.2.1. Technical equipment

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

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

Participants in the video-based training used a Samsung tablet (10.4 inches) and a touch pen to operate on the screen. Similar to the AR group, participants used physical LEGO® MINDSTORMS® EV3 assembly components. Instructional slides were presented on the tablet screen, which was fixed with a mounted tablet arm. The steps were shown as video clips of real hands performing the task. The tablet presenting the slides and videos was placed to the participants' right side (a video of the video training can be found in the supplementary data, Appendix A.3). The training was guided with video instructions and performed using physical components.

2.2.2. Questionnaires

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

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

2.3. Procedure

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

2.3.1. First experimental session

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

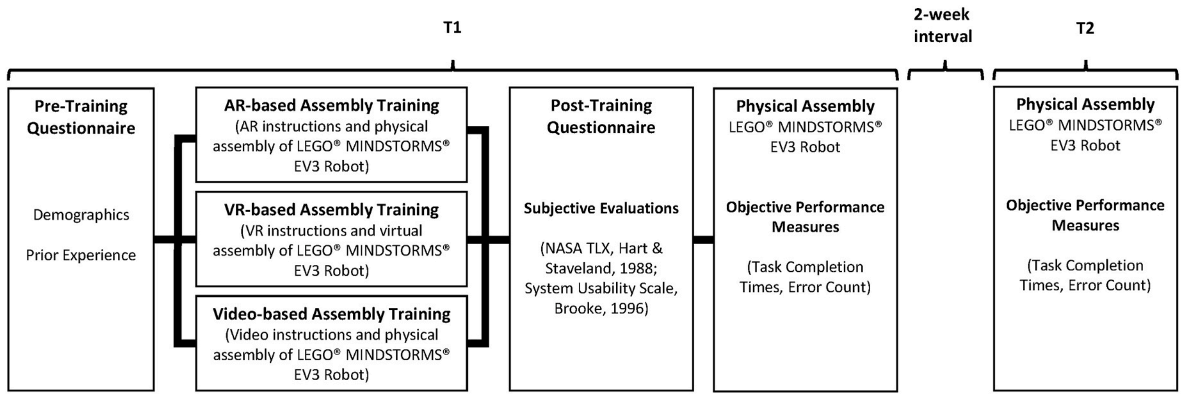

Fig. 2. Procedure Flowcharts for the first (T1) and second (T2) experimental sessions.

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

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

2.3.2. Second experimental session

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

3. Results

3.1. Results on a priori hypotheses

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

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

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

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

Table 2
Descriptive results of the subjective evaluations NASA-task load index (NASA-TLX) and system usability score (SUS).

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

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

Table 1
Descriptive results of the objective performance measures TCT and error count for T1 and T2.

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

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

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

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

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

A 3 × 2 ANOVA with the between-subjects factor *training group* (AR, VR, video) and the within-subject factor *time of measurement* (T1, T2) revealed that the main effect for *time of measurement* was significant, $F(1, 91) = 144.95$, $p \leq .001$, $\eta_p^2 = .61$ (see Fig. 4). This indicated that the error count increased for all participants, regardless of the group, with an overall mean of $M = 3.51$ errors, $SE = .24$, 95%-CI [3.04, 3.99] on T1 and an overall mean of $M = 7.93$ errors, $SE = 0.38$, 95%-CI [7.19, 8.68] on T2. Across all groups, there was a mean difference of $M_{\text{Diff}} = 4.42$ errors between T1 and T2, $SE = .37$, 95%-CI [3.69, 5.147]. The main effect of *training group* was not significant, $F(2, 91) = 1.975$, $p = .145$, indicating that there was no difference in error count between the training groups. Furthermore, the interaction between *time of measurement* and *training group* was not significant $F(2, 91) = 1.88$, $p = .159$, indicating that training methods did not affect error count for T1 differently than for T2.

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

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

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

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

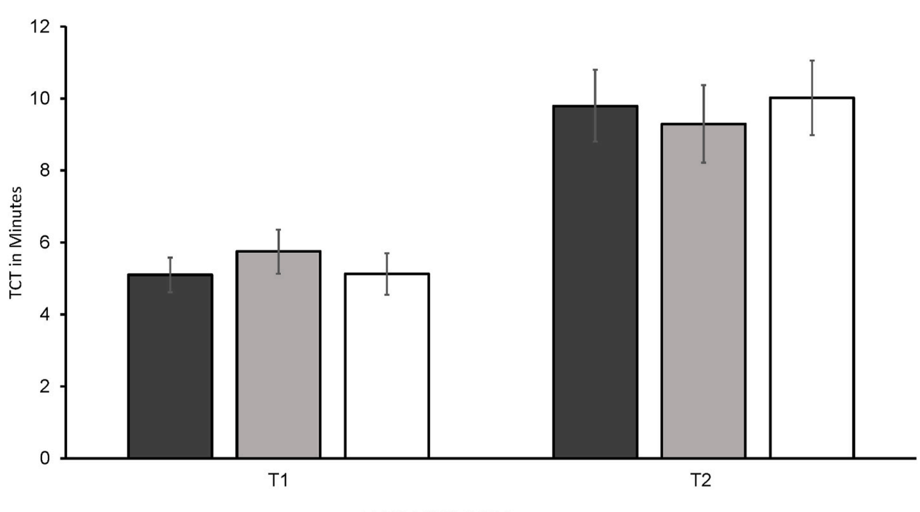

Fig. 3. Short- and long-term effects of AR-,VR- and video-based training on task completion time (TCT).
Note. Due to identified outliers, a total of 33 participants from the AR group, 27 participants from the VR group and 33 participants from the video group were included in the analysis. Means with 95%-confidence interval are depicted.

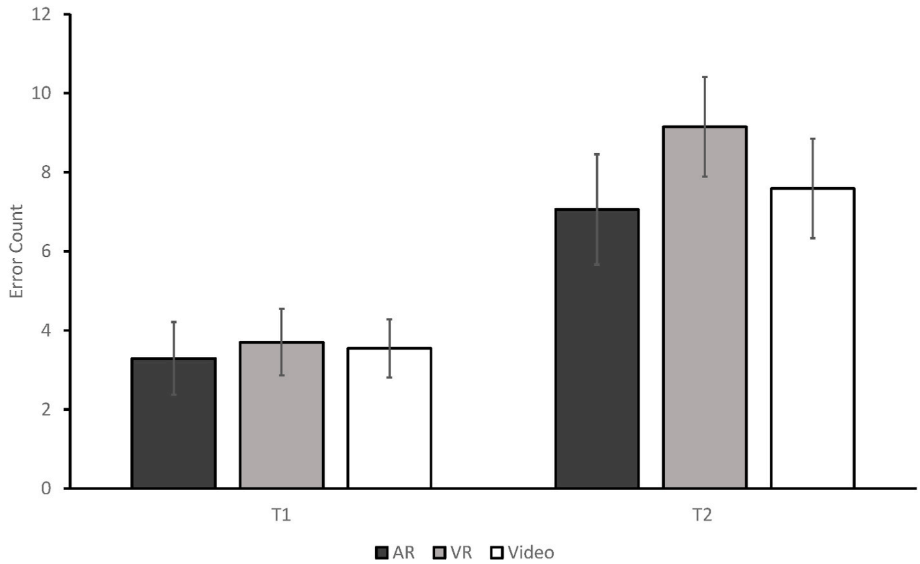

Fig. 4. Short- and long-term effects of AR-,VR- and video-based training on error count.
Note. Due to identified outliers, a total of 34 participants from the AR group, 27 participants from the VR group, and 33 participants from the video group were included in the analysis. Means with 95%-confidence interval are depicted.

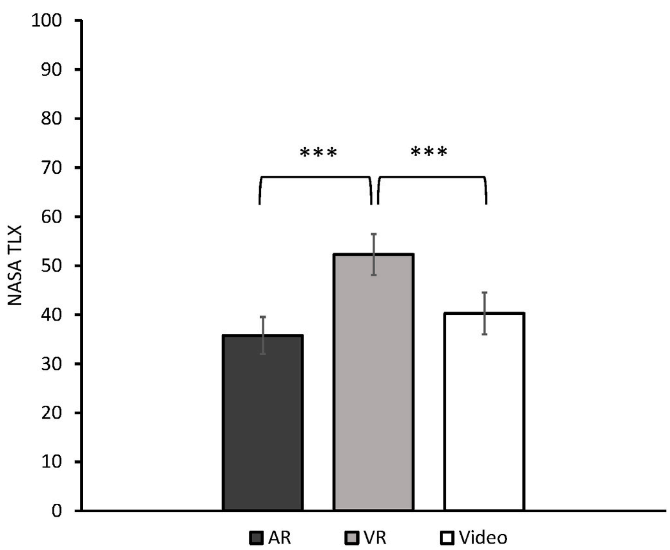

Fig. 5. Perceived task load (NASA-TLX; Hart, 2006) during AR-,VR- and video-based training.
Note. 34 participants from the AR group, 28 participants from the VR group, and 34 participants from the video group were included in the analysis. Means with 95%-confidence interval are depicted. ***$p < .001$.

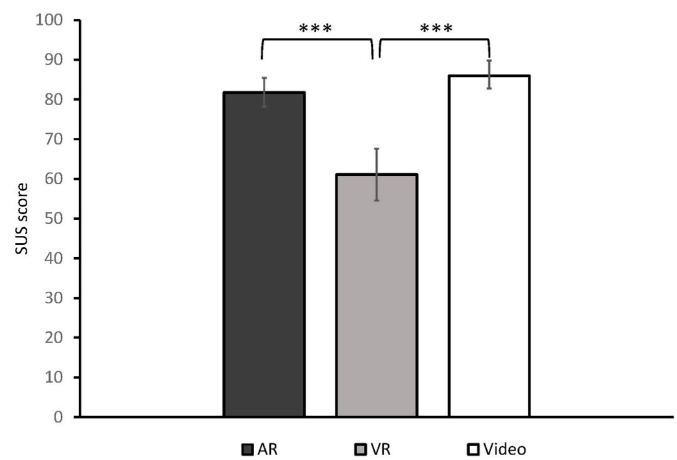

Fig. 6. Perceived system usability (SUS; Brooke, 1996) of AR-, VR- and video-based training.
Note. 34 participants from the AR group, 28 participants from the VR group, and 34 participants from the video group were included in the analysis. Means with 95%-confidence interval are depicted. ***$p < .001$.

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

Summarizing the analysis regarding the a priori hypotheses, it can be stated that there was no difference between the AR, VR, and video groups in their effect on short- and long-term objective performance measures, but TCT and error count significantly increased from T1 to T2 in all groups (RQ1). Concerning subjective evaluations, it can be stated that the VR group reported significantly higher task load and lower system usability than the AR and video groups (RQ2). An overview of these results is presented in Table 3. The main findings are further supplemented by conducting exploratory analyses, which are presented in Section 3.2.

3.2. Results of exploratory analyses

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

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

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

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

3.2.1. Exploratory analysis of group differences in training duration

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

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

Age as a user-related factor did not significantly differ between groups, Welch's $F(2, 59.28) = 1.5$, $p = .232$. A one-way ANCOVA was conducted to determine a statistically significant difference between *training groups* (AR, VR, video) on *performance (TCT, error count)* while controlling for *age*. Age was identified as a significant covariate for TCT on T1, $F(1,89) = 12.89$, $p = .001$, $\eta_p^2 = .13$, TCT on T2, $F(1,90) = 4.52$, $p = .036$, $\eta_p^2 = .05$, and for error count on T2, $F(1,91) = 6.53$, $p = .012$, $\eta_p^2 = .07$, but not for error count on T1, $F(1,89) = 3.82$, $p = .054$. Results indicated that with an increase in age, TCT and error count increased ($b = .049$, 95%-CI [.02,.07] for TCT on T1, $b = .057$, 95%-CI [.00,.11] for TCT on T2, $b = .082$, 95%-CI [.02,.15] for error count on T2). Although no group difference could be demonstrated in the previous analyses, there was a significant effect of *training group* on *TCT on T1* after adjusting for *age*, $F(2, 89) = 6.36$, $p < .05$, $\eta_p^2 = .66$. Bonferroni-corrected post hoc analysis revealed that, after adjusting for *age*, the VR group took statistically significantly longer (i.e. a higher TCT) than the AR group to complete the assembly with a mean difference of $M_{\text{Diff}} = .89$ s, $SE = .37$, 95%-CI [.16, 1.64], $p < .05$. At the same time, no difference was found between VR and the video group ($M_{\text{Diff}} = .69$ s, $SE = .37$, 95%-CI [-.03, 1.42], $p = .062$) or between AR and the video group ($M_{\text{Diff}} = -0.21$ s, $SE = .35$, 95%-CI [-.91, .49], $p = .056$) for TCT on T1. Moreover, adjusting for *age* resulted in a significant effect of *training groups* on *error count* on T2, $F(2, 91) = 3.52$, $p < .05$, $\eta_p^2 = .07$. Bonferroni-corrected post hoc analysis revealed that, after adjusting for *age*, the VR group made statistically significantly more errors than the AR group with a mean difference of $M_{\text{Diff}} = 2.38$ errors, $SE = .92$, 95%-CI [−4.620, −.15], $p < .05$. No difference was found between VR and the video group ($M_{\text{Diff}} = 1.66$ errors, $SE = .90$, 95%-CI [-.54, 3.86], $p = .206$) as well as between AR and the video group ($M_{\text{Diff}} = -.73$ errors, $SE = .86$, 95%-CI [−2.83, 1.38], $p = 1.00$) for error count on T2. However, after adjusting for *age*, no statistically significant difference in *training groups* was found for *TCT on T2*, $F(2, 90) = .78$, $p = .460$, and *error count on T1*, $F(2, 89) = .43$, $p = .650$.

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

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

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

4. Discussion

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

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

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

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

Table 4
Overview and summary of the exploratory analysis.

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

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

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

Another factor influencing short- and long-term objective performance could be related to user-related variables. Overall, in our study a higher age contributed to longer assembly times in the short-term and more errors in the long-term. This is a known effect, as older adults are more likely to perform worse than younger adults on procedural assembly tasks due to age-related declines in working memory (Abubakar and Wang, 2019; Morrell and Park, 1993). After adjusting for the effect of age, the AR group's performance was significantly faster in the short-term and the group showed significantly fewer errors in the long-term than the VR group. AR-based training might thus be particularly helpful in enhancing the time effectiveness of assembly in the short-term, which supports the results of previous studies (Hou and Wang, 2013; Koumaditis et al., 2019; Kwiatek et al., 2019; Loch et al., 2019) as well as highly recent results found by Liu et al. (2022), who showed that for training maintenance tasks, there was an advantage of AR over VR and video-based training as task difficulty increased. Moreover, these results support the previous finding that AR has particular potential to contribute to training success in the long-term by reducing error rates and thus contributing to the overall efficiency, effectiveness, and flexibility of production (Gavish et al., 2011). The finding that the performance after VR-based training showed worse results than AR could be justified by a higher need for the transfer of the virtual training environment to the real assembly use case. However, after adjusting for age, no difference in task completion times on T1 and T2 and error count on T2 could be found between VR-based and video-based training nor between AR-based and video-based training. Accordingly, both AR- and VR-based training did not differ significantly from the more conventional video-based method. In their first attempt to test AR and VR against a video control group, Gavish et al. (2015) came to similar conclusions. In their study, VR did not show any advantage over the video-based training of a relatively simple six-step actuator assembly task. AR, in contrast, showed no advantage concerning assembly times, but led to fewer unsolved errors than video-based training in the short-term. In our study, no differentiation of errors into solved vs. unsolved errors has been realized, which could have contributed to a better understanding of the effects of AR-vs. VR-based training on error count in assembly tasks (Werrlich et al., 2018; Werrlich et al., 2018). Both Gavish et al. (2015) and Liu et al. (2022) further discussed that the similarity of the training methods might be related to relatively easy assembly tasks and hypothesized that AR and VR might have a significant advantage over conventional methods when a more demanding task is used. This, however, cannot be supported by our study results. Moreover, the authors discussed that participants who have more experience in using AR and VR technologies might be able to use these technologies more efficiently (Gavish et al., 2015). In our study, we found that prior experience of participants indeed appeared to play an important role: across all groups, more experience corresponded with shorter TCT in the short-term. Comparable findings on experience and performance were found by Schwarz and co-authors (2020) as well as Abubakar and Wang (2019), who identified experience as the most significant human factor affecting individual performance. Whether increasing prior experience can be expected to have a positive effect on training success for all training methods should be clarified in the context of further research.

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

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

In addition to the results on objective performance, our study showed that the subjective evaluations of AR-, VR-, and video-based training, namely task load and usability, also provide exciting insights into the different perceptions of the training methods by the users. For instance, we found significant differences between the groups in terms of perceived task load during training. Contrary to our expectation that AR would generate less task load than video training by allowing instructions to be directly displayed on the real components and workspace (Hou et al., 2015; Koumaditis et al., 2019; Liu et al., 2022), AR- and video-based training resulted in a similar perceived task load. This could have been caused by the fact that the video training was also presented directly at the workplace and thus the transfer of what was seen to the actual assembly task was easier than in a scenario where the video is shown outside the workplace. Moreover, task load was only measured by using subjective ratings, which could limit the expressiveness of the results. In a recent study of Drouot et al. (2022), ocular and behavioral data indicated that AR led to increased mental workload in comparison to computerized instructions. Thus, the inclusion of additional objective data sources to assess task load in AR and VR training is recommended for future studies. Interestingly, the VR-based training resulted in a significantly higher task load than the other two training systems, which could be because rotating and handling virtual components might require more effort than working with real hands and components (Schwarz et al., 2020). At this stage, we can therefore conclude that for this type of complex assembly task, AR and video-based training are particularly suitable for providing assembly instructions with a comparatively low workload.

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

5. Conclusion

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

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

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

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

Declaration of competing interest

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

Appendix A. Supplementary data

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

References

Abubakar, M.I., Wang, Q., 2019. Key human factors and their effects on human centered assembly performance. Int. J. Ind. Ergon. 69, 48–57. https://doi.org/10.1016/j.ergon.2018.09.009.

Al-Ahmari, A., Ameen, W., Abidi, M.H., Mian, S.H., 2018. Evaluation of 3D printing approach for manual assembly training. Int. J. Ind. Ergon. 66, 57–62. https://doi.org/10.1016/j.ergon.2018.02.004.

AlGeddawy, T., ElMaraghy, H.A., 2012. Product variety management in design and manufacturing: challenges and strategies. In: ElMaraghy, H.A. (Ed.), Enabling Manufacturing Competitiveness and Economic Sustainability. Springer, pp. 518–523. https://doi.org/10.1007/978-3-642-23860-4_85.

Ariansyah, D., Erkoyuncu, J.A., Eimontaite, I., Johnson, T., Oostveen, A., Fletcher, S.R., Sharples, S.C., 2021. A head mounted augmented reality design practice for maintenance assembly: toward meeting perceptual and cognitive needs of AR users. Appl. Ergon. 98, 103597 https://doi.org/10.1016/j.apergo.2021.103597.

Brooke, J., 1996. SUS-A quick and dirty usability scale. Usability Eval. Ind. 189 (194), 4–7. https://doi.org/10.1201/9781498710411-35.

Carlson, P., Peters, A., Gilbert, S.B., Vance, J.M., Luse, A., 2015. Virtual training: learning transfer of assembly tasks. IEEE Trans. Visual. Comput. Graph. 21 (6), 770–782. https://doi.org/10.1109/tvcg.2015.2393871.

Cohen, J., 1988. Statistical Power Analysis for the Behavioral Sciences. Lawrence Erlbaum Associates, New Jersey.

Daling, L.M., Schlittmeier, S.J., 2022. Effects of augmented reality-, virtual reality-, and mixed reality–based training on objective performance measures and subjective evaluations in manual assembly tasks: a scoping review. Hum. Factors 0 (0). https://doi.org/10.1177/00187208221105135.

da Silva Marinho, A., Terton, U., Jones, C.M., 2022. Cybersickness and postural stability of first time VR users playing VR videogames. Appl. Ergon. 101, 103698 https://doi.org/10.1016/j.apergo.2022.103698.

DaValle, A., Azhar, S., 2020. An investigation of mixed reality technology for onsite construction assembly. MATEC Web Conf. 312 (2), 06001 https://doi.org/10.1051/matecconf/202031206001.

DeLone, W.H., McLean, E.R., 2002. Information systems success revisited. In: Proceedings of the 35th Annual Hawaii International Conference on System Sciences. IEEE, pp. 2966–2976. https://doi.org/10.1109/HICSS.2002.994345.

Doolani, S., Owens, L., Wessels, C., Makedon, F., 2020a. vIS: an immersive virtual storytelling system for vocational training. Appl. Sci. 10 (22), 8143. https://doi.org/10.3390/app10228143.

Doolani, S., Wessels, C., Kanal, V., Sevastopoulos, C., Jaiswal, A., Nambiappan, H., Makedon, F., 2020b. A review of extended reality (xr) technologies for manufacturing training. Technologies 8 (4), 77. https://doi.org/10.3390/technologies8040077.

Drouot, M., Le Bigot, N., Bricard, E., De Bougrenet, J.L., Nourrit, V., 2022. Augmented reality on industrial assembly line: impact on effectiveness and mental workload. Appl. Ergon. 103, 103793 https://doi.org/10.1016/j.apergo.2022.103793.

Eversberg, L., Grosenick, P., Meusel, M., Lambrecht, J., 2021. An industrial assistance system with manual assembly step recognition in virtual reality. In: Proceedings of 2021 International Conference on Applied Artificial Intelligence (ICAPAI). IEEE, pp. 1–6. https://doi.org/10.1109/ICAPAI49758.2021.9462061.

Faul, F., Erdfelder, E., Lang, A.-G., Buchner, A., 2007. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav. Res. Methods 39, 175–191. https://doi.org/10.3758/BF03193146.

Funk, J., Schmidt, L., 2021. Evaluation of an augmented reality instruction for a complex assembly task: comparison of a smartphone-based augmented reality instruction with a conventional paper instruction for the teach-in phase in manual assembly. com 20 (1), 63–72. https://doi.org/10.1515/icom-2021-0005.

Gabajová, G., Furmannová, B., Medvecká, I., Grznár, P., Krajčovič, M., Furmann, R., 2019. Virtual training application by use of augmented and virtual reality under university technology enhanced learning in Slovakia. Sustainability 11 (23), 6677. https://doi.org/10.3390/su11236677.

Gao, M., Kortum, P., Oswald, F., 2020. Multi-Language toolkit for the system usability scale. Int. J. Hum. Comput. Interact. 36 (3), 1883–1901. https://doi.org/10.1080/10447318.2020.1801173.

Gavish, N., Gutierrez Seco, T., Webel, S., Rodriguez, J., Peveri, M., Bockholt, U., 2011. Transfer of skills evaluation for assembly and maintenance training. BIO Web Conf. 1, 00028 https://doi.org/10.1051/bioconf/20110100028.

Gavish, N., Gutierrez, T., Webel, S., Rodriguez, J., Peveri, M., Bockholt, U., Tecchia, F., 2015. Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks. Interact. Learn. Environ. 23 (6), 778–798. https://doi.org/10.1080/10494820.2013.815221.

González-Franco, M., Cermeron, J., Li, K., Pizarro, R., Thorn, J., Hannah, P., Hutabarat, W., Tiwari, A., Bermell-Garcia, P., 2016. Immersive Augmented Reality Training for Complex Manufacturing Scenarios. ArXiv abs/1602.01944.

Hart, S.G., 2006. NASA-task load index (NASA-TLX); 20 years later. In: Proceedings Of the Human Factors and Ergonomics Society Annual Meeting, vol. 50. Sage publications, Los Angeles, pp. 904–908. https://doi.org/10.1177/154193120605000909, 9.

Hart, S.G., Staveland, L.E., 1988. Development of NASA-TLX (task load index): results of empirical and theoretical research. Adv. Psychol. 52, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9.

Heinz, M., Büttner, S., Röcker, C., 2019. Exploring training modes for industrial augmented reality learning. In: Proceedings of the 12th ACM International Conference on Pervasive Technologies Related to Assistive Environments. Association for Computing Machinery, New York, pp. 398–401. https://doi.org/10.1145/3316782.3322753.

Hoedt, S., Claeys, A., Van Landeghem, H., Cottyn, J., 2017. The evaluation of an elementary virtual training system for manual assembly. Int. J. Prod. Res. 55 (24), 7496–7508. https://doi.org/10.1080/00207543.2017.1374572.

Hou, L., Wang, X.Y., 2013. A study on the benefits of augmented reality in retaining working memory in assembly tasks: a focus on differences in gender. Autom. ConStruct. 32, 38–45. https://doi.org/10.1016/j.autcon.2012.12.007.

Hou, L., Wang, X.Y., Bernold, L., Love, P.E.D., 2013. Using animated augmented reality to cognitively guide assembly. J. Comput. Civ. Eng. 27 (5), 439–451. https://doi.org/10.1061/(asce)cp.1943-5487.0000184.

Hou, L., Wang, X.Y., Truijens, M., 2015. Using augmented reality to facilitate piping assembly: an experiment-based evaluation. J. Comput. Civ. Eng. 29 (1), 05014007 https://doi.org/10.1061/(asce)cp.1943-5487.0000344.

Jang, J., Ko, Y., Shin, W.S., Han, I., 2021. Augmented reality and virtual reality for learning: an examination using an extended technology acceptance model. IEEE Access 9, 6798–6809. https://doi.org/10.1109/ACCESS.2020.3048708.

Joshi, S., Hamilton, M., Warren, R., Faucett, D., Tian, W., Wang, Y., Ma, J., 2021. Implementing Virtual Reality technology for safety training in the precast/prestressed concrete industry. Appl. Ergon. 90, 103286 https://doi.org/10.1016/j.apergo.2020.103286.

Kagermann, H., Helbig, J., Hellinger, A., Wahlster, W., 2013. Recommendations for implementing the strategic initiative INDUSTRIE 4.0: securing the future of German manufacturing industry; Final report of the Industrie 4.0 Working Group. Forschungsunion. https://gateway-bayern.de/BV041628398.

Kaplan, A.D., Cruit, J., Endsley, M., Beers, S.M., Sawyer, B.D., Hancock, P.A., 2021. The effects of virtual reality, augmented reality, and mixed reality as training enhancement methods: a meta-analysis. Hum. Factors 63 (4), 706–726. https://doi.org/10.1177/0018720820904229.

Kim, M., Park, K., Choi, S.H., Lee, J.Y., Kim, D., 2018. AR/VR-Based live manual for user-centric smart factory services. In: Moon, I., Lee, G., Park, J., Kiritsis, D., von Cieminski, G. (Eds.), *IFIP International Conference On Advances In Production Management Systems* (*APMS*), vol. 536. Springer, Cham, pp. 417–421. https://doi.org/10.1007/978-3-319-99707-0_52.

Kostaras, N., Xenos, M., 2012. Usability evaluation of augmented reality systems. Intell. Decis. Technol. 6 (2), 139–149. https://doi.org/10.3233/IDT-2012-0130.

Koumaditis, K., Chinello, F., Mitkidis, P., Karg, S., 2020. Effectiveness of virtual versus physical training: the case of assembly tasks, trainer's verbal assistance, and task complexity. IEEE Comput. Graph. Appl. 40 (5), 41–56. https://doi.org/10.1109/mcg.2020.3006330.

Koumaditis, K., Venckute, S., Jensen, F.S., Chinello, F., 2019. Immersive training: outcomes from small scale AR/VR pilot-studies. In: 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). IEEE, pp. 1–5. https://doi.org/10.1109/VR44988.2019.9044162.

Kwiatek, C., Sharif, M., Li, S., Haas, C., Walbridge, S., 2019. Impact of augmented reality and spatial cognition on assembly in construction. Autom. ConStruct. 108, 102935 https://doi.org/10.1016/j.autcon.2019.102935. Article.

Lawson, G., Salanitri, D., Waterfield, B., 2016. Future directions for the development of virtual reality within an automotive manufacturer. Appl. Ergon. 53 (Pt B), 323–330. https://doi.org/10.1016/j.apergo.2015.06.024.

Li, Z., Wang, J., Yan, Z., Wang, X., Anwar, M.S., 2019. An interactive virtual training system for assembly and disassembly based on precedence constraints. In: In*Computer Graphics International Conference* 2019. Springer, Cham, pp. 81–93. https://doi.org/10.1007/978-3-030-22514-8_7.

Liu, X.-W., Li, C.-Y., Dang, S., Wang, W., Qu, J., Chen, T., Wang, Q.-L., 2022. Research on training effectiveness of professional maintenance personnel based on virtual reality and augmented reality technology. Sustainability 14 (21), 14351. https://doi.org/10.3390/su142114351. MDPI AG.

Loch, F., Ziegler, U., Vogel-Heuser, B., 2019. Using real-time feedback in a training system for manual procedures. IFAC-PapersOnLine 52 (19), 241–246. https://doi.org/10.1016/j.ifacol.2019.12.089.

Longo, L., 2018. Experienced mental workload, perception of usability, their interaction and impact on task performance. PLoS One 13 (8), e0199661. https://doi.org/10.1371/journal.pone.0199661.

Lotter, B., 2006. Einführung. In: Lotter, B., Wiendahl, H.P. (Eds.), Montage in der industriellen Produktion: Ein Handbuch für die Praxis. Springer, Berlin, Heidelberg, pp. 1–8. https://doi.org/10.1007/3-540-36669-5_1.

Milgram, P., Kishino, F., 1994. A taxonomy of mixed reality visual displays. IEICE Trans. Info Syst. 77 (12), 1321–1329.

Miqueo, A., Torralba, M., Yagüe-Fabra, J.A., 2020. Lean manual assembly 4.0: a systematic review. Appl. Sci. 10 (23), 8555. https://doi.org/10.3390/app10238555.

Morrell, R.W., Park, D.C., 1993. The effects of age, illustrations, and task variables on the performance of procedural assembly tasks. Psychol. Aging 8 (3), 389–399. https://doi.org/10.1037/0882-7974.8.3.389.

Murcia-López, M., Steed, A., 2018. A comparison of virtual and physical training transfer of bimanual assembly tasks. IEEE Trans. Visual. Comput. Graph. 24 (4), 1574–1583. https://doi.org/10.1109/tvcg.2018.2793638.

Oren, M., Carlson, P., Gilbert, S., Vance, J.M., 2012. Puzzle assembly training: real world vs. Virtual environment. In: Coquillart, S., Feiner, S., Kiyokawa, K. (Eds.), 2012 IEEE Virtual Reality Workshops. IEEE, pp. 27–30. https://doi.org/10.1109/VR.2012.6180873.

Roldan, J.J., Crespo, E., 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. Robot. Comput. Integrated Manuf. 59, 305–316. https://doi.org/10.1016/j.rcim.2019.05.004.

Sautter, B., Daling, L., 2021. Mixed reality supported learning for industrial on-the-job training. In: Proceedings of the Conference on Learning Factories. https://doi.org/10.2139/ssrn.3864189.

Schmidt, R.A., Bjork, R.A., 1992. New conceptualizations of practice: common principles in three paradigms suggest new concepts for training. Psychol. Sci. 3 (4), 207–218. https://doi.org/10.1111/j.1467-9280.1992.tb00029.x.

Schwarz, S., Regal, G., Kempf, M., Schatz, R., 2020. Learning success in immersive virtual reality training environments: practical evidence from automotive assembly. In: Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences. Shaping Society, pp. 1–11. https://doi.org/10.1145/3419249.3420182.

Skarbez, R., Smith, M., Whitton, M.C., 2021. Revisiting Milgram and Kishino's reality-virtuality continuum. Front. Virtual Real. 2, 647997 https://doi.org/10.3389/frvir.2021.647997. Article.

Venkatesh, V., Brown, S.A., Maruping, L.M., Bala, H., 2008. Predicting different conceptualizations of system use: the competing roles of behavioral intention, facilitating conditions, and behavioral expectation. MIS Q. 483–502. https://doi.org/10.2307/25148853.

Wang, X., Ong, S.K., Nee, A.Y.C., 2016. A comprehensive survey of augmented reality assembly research. Adv. Manuf. 4, 1–22. https://doi.org/10.1007/s40436-015-0131-4.

Wang, C., Tsai, N., Lu, J., Wang, M.J., 2019. Usability evaluation of an instructional application based on Google Glass for mobile phone disassembly tasks. Appl. Ergon. 77, 58–69. https://doi.org/10.1016/j.apergo.2019.01.007.

Webel, S., Bockholt, U., Engelke, T., Gavish, N., Olbrich, M., Preusche, C., 2013. An augmented reality training platform for assembly and maintenance skills. Robot. Autonom. Syst. 61 (4), 398–403. https://doi.org/10.1016/j.robot.2012.09.013.

Werrlich, S., Lorber, C., Nguyen, P.A., Yanez, C.E.F., Notni, G., 2018a. Assembly training: comparing the effects of head-mounted displays and face-to-face training. In: Chen, J.Y.C., Fragomeni, G. (Eds.), Virtual, Augmented and Mixed Reality: Interaction, Navigation, Visualization, Embodiment, and Simulation, Vamr 2018, Pt I, vol. 10909. Springer, Cham, pp. 462–476. https://doi.org/10.1007/978-3-319-91581-4_35.

Werrlich, S., Nguyen, P.-A., Notni, G., 2018b. Evaluating the training transfer of Head-Mounted Display based training for assembly tasks. In: PETRA '18: *Proceedings Of the 11th Pervasive Technologies Related To Assistive Environments Conference*, pp. 297–302. https://doi.org/10.1145/3197768.3201564.

Westerfield, G., Mitrovic, A., Billinghurst, M., 2015. Intelligent augmented reality training for motherboard assembly. Int. J. Artif. Intell. Educ. 25 (1), 157–172. https://doi.org/10.1007/s40593-014-0032-x.

Xu, X., Lu, Y., Vogel-Heuser, B., Wang, L., 2021. Industry 4.0 and Industry 5.0-Inception, conception and perception. J. Manuf. Syst. 61, 530–535. https://doi.org/10.1016/j.jmsy.2021.10.006.

Lea M. Daling, M.Sc. completed her master's in social, organizational, and work psychology at the Freie Universität Berlin in 2017. Since then, she is a research associate at the Chair of Production Metrology and Quality Management & Institute for Information Management in Mechanical Engineering (WZL-MQ/IMA) as well as an external Ph.D. student in the teaching and research area of Work and Engineering Psychology at the RWTH Aachen University.

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

Prof. Dr. Phil. Ingrid Isenhardt is Academic Director at WZL-MQ/IMA (Chair of Production Metrology and Quality Management & Institute for Information Management in Mechanical Engineering Information Management in Mechanical Engineering) as an associate (apl.) professor at the Faculty of Mechanical Engineering with a second position at the Faculty of Philosophy of RWTH Aachen University. Her research focus is on socio-technical aspects of digital transformation, sustainable production & value creation, as well as information and data management in mechanical engineering.

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