Does One-Size Training Fit All? Evaluating Adaptive Learning for VR Assembly Training

ARNO VERSTRAETE, Hasselt University - Flanders Make, Belgium
EVA GEURTS, Hasselt University - Flanders Make, Belgium
MAARTEN WIJNANTS, Hasselt University - Flanders Make, Belgium

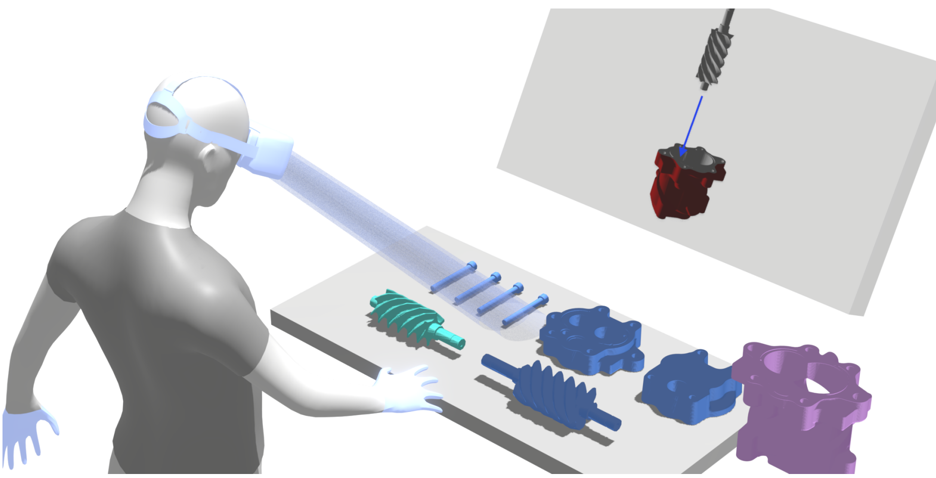


Fig. 1. Concept illustration of our adaptive learning approach in VR showing the sources of adaptation by utilizing hand tracking, eye tracking, and contextual data for real-time performance monitoring during assembly training. The system detects whether the user interacts with the correct objects, highlighted in teal, or wrong objects, highlighted in blue. It also checks if the objects are placed at the correct target location, indicated by a purple highlight. The system provides immediate feedback and guidance to enhance learning outcomes through adaptive support mechanisms, such as showing instructions.

Virtual Reality (VR) is gaining popularity and is increasingly adopted across various industries for its potential to deliver immersive and effective skill development. However, we observe that VR training often follows a one-size-fits-all approach. Trainings typically do not adapt to to individual skill levels, which is particularly important in industrial assembly, where user profiles and expertise levels vary widely. To address this, we applied the concept of adaptive learning to VR assembly training, enabling the system to dynamically provide assistance levels when users struggle and gradually reduce support as their proficiency increases. This paper investigates the learning performance and subjective impact of two types of such adaptive approaches and a non-adaptive variant in a VR user study with 36 participants. The results show that adaptive training

Authors' Contact Information: Arno Verstraete, Hasselt University - Flanders Make, Digital Future Lab, Diepenbeek, Belgium, arno.verstraete@uhasselt.be; Eva Geurts, Hasselt University - Flanders Make, Digital Future Lab, Diepenbeek, Belgium, eva.geurts@uhasselt.be; Maarten Wijnants, Hasselt University - Flanders Make, Digital Future Lab, Diepenbeek, Belgium, maarten.wijnants@uhasselt.be.

ACM 2573-0142/2025/6-ARTEICS010
https://doi.org/10.1145/3734187

significantly enhances user experience and reduces perceived workload. At the same time, adaptive VR learning is found to have a positive impact on learning performance (quantified as a reduced number of assembly mistakes after training). In summary, our findings underscore the potential of applying adaptive learning approaches in VR. To guide future research, we propose guidelines to support the practical adoption of adaptive learning in VR training in manufacturing and beyond.

CCS Concepts: • **Human-centered computing** → **Interactive systems and tools**; **Empirical studies in HCI**; **Virtual reality**; • **Applied computing** → *Computer-assisted instruction*; **Interactive learning environments**.

Additional Key Words and Phrases: Virtual Reality, Adaptive learning, Manufacturing, Industrial applications

ACM Reference Format:
Arno Verstraete, Eva Geurts, and Maarten Wijnants. 2025. Does One-Size Training Fit All? Evaluating Adaptive Learning for VR Assembly Training. *Proc. ACM Hum.-Comput. Interact.* 9, 4, Article EICS010 (June 2025), 26 pages. https://doi.org/10.1145/3734187

1 Introduction

In the domain of industrial operations, training is crucial for mastering assembly processes, ensuring compliance with safety protocols, and effectively integrating digital workflows. From an engineering standpoint, training methods must balance efficiency, scalability, and safety while adapting to the rapid technological advancements reshaping the industry. Traditionally, on-the-job training—where employees acquire skills through hands-on experience under expert supervision—has been the standard approach for transferring operational knowledge [18]. While this method enables direct interaction with experienced instructors and blends theoretical learning with practical application, it presents significant challenges. In particular, it is costly, difficult to scale, inconsistent, and lacks reusability, all while introducing safety risks due to the potential for accidents [18, 22]. Additionally, user profiles and expertise levels vary greatly which makes it even more challenging to provide the optimal training to each individual.

The adoption of digital technologies in industrial training has created new opportunities for enhanced learning [30]. For example, digital work instructions, accessible through various devices, provide a dynamic platform for training, allowing operators to access information on-demand and at their own pace while working [31]. These digital resources enhance efficiency and facilitate consistent training and support but lack personalization [18]. On the other hand, Virtual Reality (VR) stands out as a transformative tool for industrial training, offering highly immersive, interactive, and safe learning environments. VR technology enables the simulation of complex scenarios without the real-world consequences that might arise in traditional training. As VR evolves rapidly, it is being adopted across various sectors, offering a viable alternative—or complement—to these traditional training methods [44]. By simulating assembly processes in a virtual environment, VR training enables operators to develop their skills without the limitations of physical equipment, providing an interactive and immersive learning experience that improves engagement and retention [26]. However, most traditional VR training solutions follow a "one-size-fits-all" approach, providing fixed instructional support that does not adapt to individual skill levels or learning progress [33, 40]. This de facto approach overlooks the diverse backgrounds, experiences, and learning speeds of trainees, limiting its effectiveness in industrial settings [32].

To address these shortcomings, we present an approach that integrates *adaptive learning* in VR training environments, which is defined as a data-driven training method that dynamically adjusts instructional content, feedback, and difficulty based on real-time user performance and prior training history. Leveraging built-in VR capabilities such as hand and eye tracking, our system continuously monitors user interactions to provide personalized guidance, adapting support levels and corrective feedback to match each learner's needs. This shift from static, uniform instruction to an adaptive, individualized learning experience holds the potential to make VR training more responsive, efficient, and tailored to the diverse workforce in industrial operations.

This paper presents our contribution to comparing the impact of adaptive VR training with traditional VR training in the context of manual assembly in the manufacturing industry. Section 2 first presents a review of prior art in the domain of (adaptive) learning and Extended Reality (XR) training solutions. Section 3 describes how we apply adaptive learning to VR assembly training by using dynamic real-time support, feedback mechanisms, and adaptive learning algorithms. We evaluated our approach in an experimental study (N=36) comparing three learning approaches in VR: automated adaptive training (without user control), semi-automated adaptive training (with user control), and traditional, generic training. We describe our evaluation methodology in Section 4 and the evaluation results in Section 5. We conclude this paper with recommendations for adaptive VR learning in assembly contexts and how they can be applied to a broader application domain in Section 6. Generally speaking, our findings inform future research on designing and engineering VR training systems by identifying key factors that influence the effectiveness of (adaptive) learning, and by exploring how real-time adaptation can enhance learning outcomes in various industrial and educational contexts.

2 Related work

We will first provide an overview of existing digital adaptive learning strategies and their relevance to our work. Additionally, we will present the state-of-the-art in XR training solutions and how they have compared different training approaches and strategies, along with the learning metrics used to evaluate their effectiveness.

2.1 Adaptive learning strategies

Taylor et al. [39] define personalized and adaptive learning as a method that is *customized for, or adapted to, the individual learners*. In this approach, each learner follows a unique learning path at their own pace, based on prior knowledge and learning speed. The adoption of adaptive learning has grown in recent years, largely driven by the increasing popularity of distance learning [15]. Traditionally, adaptive learning has been implemented through technologies such as adaptive e-learning or Learning Management Systems (LMS) [6, 42]. By simulating the personalized guidance of human tutors, adaptive learning serves also as a foundation for more advanced AI-driven solutions [13, 28, 35]. While both adaptive learning systems and traditional LMS incorporate key components for personalization, they differ in their approach to monitoring learner progress. Adaptive systems employ real-time analysis, continuously tracking learner performance to make immediate adjustments—an approach aligned with the scaffolding teaching technique, which provides timely support and guidance [46]. In contrast, traditional LMS typically assess learners at fixed intervals rather than dynamically adjusting content in real time [37] which is in line with what we usually encounter in XR solutions as well.

Another essential component of adaptive learning is the use of algorithms that modify instructional content based on the learner's unique needs [11]. These algorithms consider various factors, including learning styles, preferences, and past performance, enabling dynamic adjustments to the learning path. Aligned with Cognitive Load Theory, this approach optimizes learning efficiency and engagement by ensuring that cognitive demands remain manageable, allowing learners to focus on essential information without being overwhelmed [41]. Research on digital and adaptive learning also highlights the importance of personalized approaches in modern learning environments. Digital learning platforms often struggle with information overload, where excessive content can overwhelm learners, reducing comprehension and increasing cognitive fatigue [1]. This challenge is particularly evident in virtual learning environments that lack personalized learning paths, as they present identical content to all users, leading to redundancy and inefficiency.

Adaptive learning technologies offer a solution by dynamically tailoring content to individual learners. By analyzing real-time performance, these systems can filter out irrelevant information and emphasize key concepts, ensuring a more efficient and engaging learning experience. Studies indicate that integrating adaptive elements into digital learning environments can enhance motivation, engagement, and perceived competence [9, 34].

2.2 Comparing various training modalities: insights from AR, VR, and traditional methods

The use of XR technologies—including Augmented Reality (AR), Virtual Reality (VR)- and video-based instruction—continues to reshape training methodologies in industrial settings. Several review studies have examined the effectiveness of these approaches. Cazeri et al. [5] specifically investigated XR applications within Industry 4.0, while broader analyses by Kaplan et al. [21], Daling et al. [7], and Doolani et al. [10] explored the impact of XR-based training across multiple industrial domains. A common limitation highlighted in these reviews is the lack of standardized comparison measures, limited assessments of skill transfer to real-world applications, and insufficient evaluations of long-term knowledge retention.

Adaptive training systems aim to personalize learning experiences by dynamically adjusting instructional content based on user progress. Huang et al. [18] developed an adaptive AR tutoring system that modified instruction levels in real-time. Their study showed that adaptive guidance was more effective and preferred over static methods, but its evaluation was limited to a VR environment, leaving open questions about its real-world applicability. Similarly, Studer et al. [38] explored an open-ended, personalized XR-based training approach for drilling tasks, integrating safety instructions, setup guidance, and real-time feedback on errors and execution quality. Work like that of Neges et al. [29] also investigates the feasibility of adding haptic feedback to these virtual environments, which can be helpful in training in certain use cases.

The effectiveness of XR-based training compared to traditional instruction methods has been widely debated. Liu et al. [22] examined training effectiveness across VR, AR, video-based, and hands-on training modalities for maintenance tasks. Their findings indicated that while traditional training methods were superior for simple tasks, AR provided the most benefits for complex, multi-level procedures. Similarly, Havard et al. [16] compared AR tablet-based instructions to conventional PDF manuals, finding that while AR facilitated object identification, experienced operators often relied on textual guidance, highlighting the need for adaptable training approaches.

Daling et al. [8] investigated how prior experience influences learning outcomes in AR, VR, and video-based training for assembly tasks. Their study found no significant performance differences between the modalities, suggesting that a trainee's familiarity with technology plays a more critical role than the training medium itself. This aligns with Murcia et al. [27], who examined VR-based versus physical training for bimanual assembly. Their results showed that VR training with animated guidance could be as effective as physical practice when assembling complex objects. However, retention tests conducted two weeks later revealed performance declines across all conditions, underscoring the need for further research on long-term learning outcomes in XR training.

2.3 Learning metrics in XR

Since we focus on comparing different learning strategies in VR, it is essential to define the metrics that assess learning performance. Literature highlights the importance of both objective and subjective measures in XR training. Objective metrics typically include task execution time, error count, and error severity, while subjective metrics assess workload, motivation, and perceived competence. A well-known subjective measure for task load evaluation is NASA-TLX [14], which has been widely used in training research. For instance, Ji et al. [19] employed NASA-TLX to analyze workload in VR pilot training and examined its correlation with EEG data.

Studies also emphasize the psychological aspects of training. Dhiman et al. [9] investigated the impact of multimedia AR instructions on motivation and engagement in craft learning. Their evaluation included both objective metrics (task execution quality) and subjective measures (perceived competence and intrinsic motivation) based on Self-Determination Theory.

Existing XR training studies are not consistent in the metrics that are used for assessing learning impact. Some focus solely on objective measures, while others incorporate a mix of objective and subjective assessments. The tutoring system of Huang et al. [18] measures total training duration, number of repetitions, level of support, and error count to track learning progression. Additionally, they evaluated subjective aspects, including user perception of accuracy, understanding, memorization, and confidence. Similarly, Chen et al. [6] analyzed step completion time, error types, and error criticalness, where more severe errors required additional corrective steps.

Daling et al. [8] combined NASA-TLX for cognitive load, SUS for usability, and task execution metrics in their evaluation of VR and AR-based assembly training. Likewise, Ulmer et al. [40] explored the effects of gamification in VR training by assessing task execution time, error count, and NASA-TLX scores to examine motivation.

While Kaplan et al. [21] did not focus on specific training metrics, their literature review highlighted pre- and post-training assessments as key methods for evaluating XR training effectiveness. These assessments use a range of performance metrics to compare skill acquisition before and after training, providing insights into short-term learning outcomes.

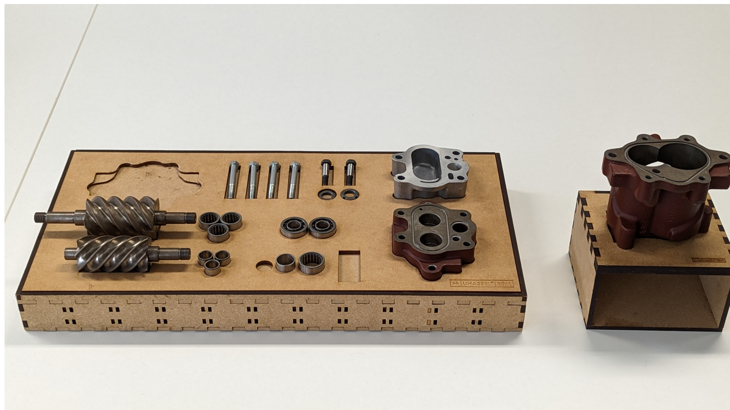
However, many studies lack long-term evaluations or fail to assess skill transfer from virtual to real-world tasks. This is particularly crucial in VR training, where the absence of physical feedback can impact learning retention and real-world applicability [20, 45]. Given that solutions such as AdapTutAR [18] specifically investigated adaptive learning, it is especially relevant to examine whether adaptive guidance enhances skill transfer beyond the virtual environment. As Mayer [23] emphasizes in the context of multimedia learning, integrating both long-term assessment and skill transfer evaluation is essential for a comprehensive understanding of training effectiveness.

3 Adaptive learning in VR: an assembly use case

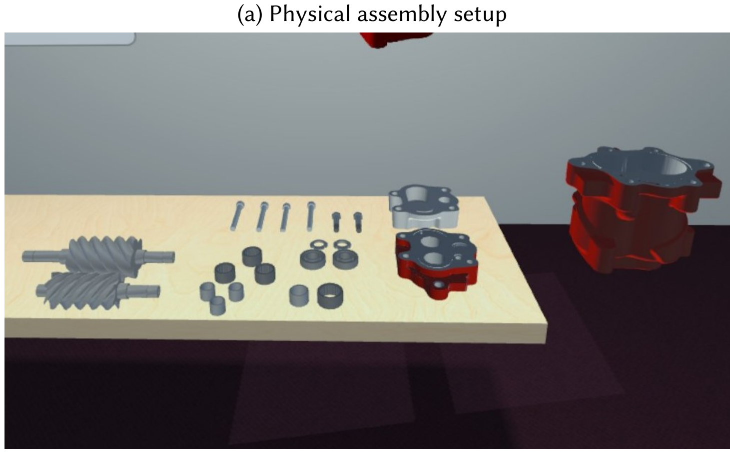
Table 1. 14 tasks required to assemble the compressor.

ID	Part	Task	# Parts
0	Inner bearing small	Place in the left hole of main housing	1
1	Inner bearing large	Place in the right hole of main housing	1
2	Bearing	Place in left hole of main housing	1
3	Bearing large	Place in right hole of main housing	1
4	Rotor female	Slide in left hole of main housing	1
5	Rotor male	Slide in right hole of main housing	1
6	Top lid	Place on top of main housing	1
7	Inner bearing small	Place on each of the rotors	2
8	Bearing	Place on each of the rotors	2
9	Small bolts	Insert in left and right screw-hole of top lid	2
10	Ball bearing	Place on top of the rotors	2
11	Fine screw	Screw on top of ball bearings	2
12	Bearing sealing	Place on top of top lid	1
13	Large bolt	Screw in holes of bearing sealing	4

In this section, we describe the application of three different learning strategies in VR assembly training: two adaptive variants and one non-adaptive variant. The exploration of personalized



(a) Physical assembly setup



(b) VR replica

Fig. 2. Physical and virtual assembly setup of the compressor.

adaptive VR training and its underlying concept is described in [12]. This paper builds on that by validating the proposed concept with user tests, testing three different learning strategies, and providing guidelines based on those results. This section provides a practical perspective on how the three learning strategies are applied within a specific training scenario. This practical use case will be used in our experiment to compare the effectiveness of the three learning strategies (see Section 4 and 5).

3.1 Use case for VR training

We performed several pilot studies with different realistic assembly setups to identify a use case that is challenging enough to require training but not too complex that additional instructions, such as paper-based guidance, become necessary. These pilot studies showed that this compressor assembly case was best suited for several important reasons. Firstly, it allows for effective learning within a single VR session, avoiding the need for extended learning paths that could span several days or weeks. This is crucial for our evaluation, as it helps limit variables that might impact the learning experience and thus keep the study manageable. Also, the complexity of the compressor assembly is such that it can be thoroughly executed without reliance on additional instructions, e.g. paper-based instructions during physical execution. However, it cannot be carried out effectively without any prior training or initial instructions. By eliminating the need for additional paper instructions, we ensure that the VR training is the only available learning tool, thus providing a more immersive and uninterrupted learning experience. This approach helps us assess the effectiveness of different learning strategies in VR training without the interference of external learning aids.

Figure 2 shows the physical version of the parts needed to assemble the compressor and the virtual replica of it in the VR training environment. The compressor assembly contains 14 steps and 22 components (see Table 1). For each step, one type of part must be picked and assembled in the right position. A step can involve multiple parts of the same kind, resulting in (almost identical) substeps that can be performed in any order. For example, for the final task, four large bolts have to be screwed into the assembly. The order in which these four substeps are performed does not matter in the virtual or the physical assembly.

The VR compressor training provides four support systems to help the trainee in executing tasks during training:

Object highlighting This support system highlights parts that have not been assembled yet in green if they are needed to complete the current task. Some parts are unique to the assembly, while other parts can have several duplicates. In case there are duplicates, this system will highlight all parts that are of the same type, even if the task only requires one object.

Wrong object highlighting This system will highlight parts in red if they are the incorrect parts for the current task when the user picks them up. This way, the user is informed of their error when they pick up an incorrect object.

Target highlighting A hologram appears in the shape of the to-be-assembled part at the place where it needs to be placed in the assembly. This informs the user of the desired position for the current part.

Instructions Instructional images are displayed on the wall in front of the trainee. These instructions show the user what actions need to be performed to complete the current task.

3.2 Learning strategies in VR

As explained earlier, the main focus of this paper is to investigate the impact of different learning strategies on learning performance and the user experience when applied in VR assembly training. We propose three different learning strategies: Automated adaptive learning, Semi-automated adaptive learning and Non-adaptive learning. The first two incorporate the push-versus-pull mechanism investigated in prior research [4, 25], while the third follows a non-adaptive approach:

Automated adaptive learning In this push-based adaptive learning variant, the system automatically provides additional instructional support in a gradual way whenever the trainee exceeds predefined time thresholds for a training task. When the trainee has completed the training and starts a new iteration, it remembers what support systems were active the last time the trainee completed a particular task and will activate all those systems again, except for one. This naturally reduces the amount of support while the trainee becomes more proficient.

Semi-automated adaptive learning In this pull-based variant, the trainee can request additional support on their own initiative by pressing a help button in VR, allowing them to control when and how much assistance they receive during training. Just like in automated adaptive learning, some support is removed for each task after successfully completing an iteration of the training.

Non-adaptive learning This non-adaptive variant always provides the trainee with all instructional information available, which is in line with traditional VR training.

For the particular case of compressor assembly training (see Section 3.1), it is important that the trainee learns the sequence of the assembly steps, enabling them to efficiently recall and apply them in a physical assembly setting. Huang et al. [18] introduced relevant terminology to explain important concepts in the context of our adaptive learning approach: *Sources of adaptation* and *Targets of adaptation*. Sources of adaptation are factors that can be tracked during the training (e.g., task execution speed) and can hence be used to trigger the adaptive learning system. Targets of adaptation refer to the support systems that instruct the learner, which can be altered by the adaptive learning system. The four support systems described in Section 3.1 are our targets of adaptation, they can all be used to support the trainee when needed. Our sources of adaptation are the trainee's performance during each task, contextual data, hand tracking, and eye tracking data. Our adaptive learning algorithm determines what particular aspect of the task (contextual data) the user is struggling with based on performance (user data), and it will activate a support system accordingly.

The algorithm determines what type of error the user is making, and thus which accompanying support system should be activated, by checking if the user falls into a *failure state*. For instance, when we recall the example of holding the wrong object for executing the current training task, the algorithm can detect an *Object Failure*. The algorithm uses the sources of adaptation (the trainee's performance during each task, contextual data, hand tracking, and eye tracking data) as input. Each *failure state* is linked to support systems that can be activated or deactivated during the training. Table 2 shows these links for the VR training we are discussing here. The algorithm can also give a *default failure* as the outcome. This only happens in situations where the algorithm was not able to make a clear decision between the three other *failure states*. In this case, a default hierarchy is followed, which is shown in Figure 3. Figure 4 shows the decision-making process of the algorithm. It is triggered by either a time threshold if automated adaptive learning is used or manually by the user if semi-automated adaptive learning is used. The T_x values represent the duration for which the user has looked at object x, which is derived from eye tracking data. These durations are considered for each task and each iteration of the training separately. Additionally, the time thresholds that trigger these failure states can be set on an individual task level. Hand-tracking data is used for decisions that involve knowledge about whether the user is holding correct or incorrect objects. The conceptual idea behind this algorithm is that it tries to decide what part of the current task is not clear to the user by looking at which parts the user has picked up and which parts the user is looking at the most. For example, if the user is only holding incorrect parts, we can assume that the user does not know what the correct object is, and thus, we should trigger an object failure. However, if the user is not holding any objects at all, we rely on the eye tracking data to make a decision. If the user is looking a lot at the table with parts, we assume the user does not know which part to pick, and an object failure is triggered. But if the user is mainly looking at the location where the part has to be assembled (target), the user might be unsure where the asked part should be placed, and we trigger a target failure.

3.2.1 *Automated adaptive learning.*

We will explain how this system works practically with the automated adaptive learning variant by using an example scenario of a trainee using this VR training. Assume the user is performing the training for the first time. At step 9 of the training (see Table 1), the user needs to assemble two small bolts, each on one side of the top lid of the assembly. The training system has no history for the user in this training because it is the first time they are performing it. In that case, only the *instructions* support system is active at the start of the task. If the user has trouble picking the correct bolts and grabs, for example, one of the large bolts,

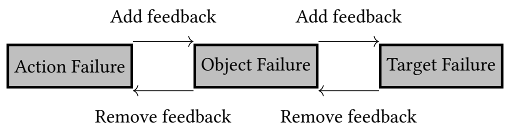


Fig. 3. Default hierarchy for failure states. When the Default Failure is reached (see also Figure 4), support systems are incrementally added from left to right. This means that when the Default Failure is reached, we check if a support system from the Action Failure can be activated. If all of these systems are already active, we check if a system from Object Failure can be activated, and lastly from Target Failure. Removing support happens in the same manner but in reverse order.

Table 2. Assignment of support systems to failure states for compressor case.

	Support system hierarchy	
Object Failure	Wrong object highlighting	Object highlighting
Target Failure	Target highlighting	
Action Failure	Instructions	

which are meant for step 13 instead, the adaptive learning system should help the user identify the correct object. The automated adaptive learning system will trigger after 10 seconds and will identify whether the user is struggling with picking the correct object (object failure). In that case, *wrong object highlighting* will be activated, or *object highlighting* if the prior was already active. Imagine that the user now puts down the large bolt and picks up a small bolt but tries to place it in the wrong hole in the top lid. The system will trigger again 10 seconds after the previous time the threshold was crossed. The adaptive learning algorithm will decide that the user is struggling with finding the correct location to insert the bolt (target failure) and will activate *target highlighting*, which will highlight the target location by visualizing a hologram.

The algorithm not only adds more information when needed but also removes information when the user no longer needs it. This happens every time the user starts a new iteration of the training. For each task, the system will remember which support systems were active the last time the user completed that task and will activate all those systems again, except for one. This follows the reverse hierarchy of Figure 3 and is based on the *Cognitive Load Theory*, stating that cognitive demands should remain manageable without overwhelming the trainee [41]. So in the case of our task 9 example, if the user had *instructions*, *wrong object highlighting* and *target highlighting* active when they completed the task previously, the next training iteration will start with all those systems already active except for *target highlighting*. This system will naturally cause the user to have more support in tasks they struggle with and less support in tasks they are proficient in, creating a personalized and adaptive environment. The more the user trains, the more proficient they will become in each task, causing support to steadily reduce over time. Ideally, in the end, the user will have learned the assembly procedure by heart, meaning that they can perform the assembly without any support system active.

Within the VR environment, there is also a panel that indicates the number of support systems that are currently active (see Figure 5). When a new system is activated, it is added to this virtual display and highlighted with accompanying audible feedback. This display not only makes the user aware of whether a particular system is active but also allows them to visually evaluate their performance during the training. Users can notice that they perform well when the display shows fewer active systems.

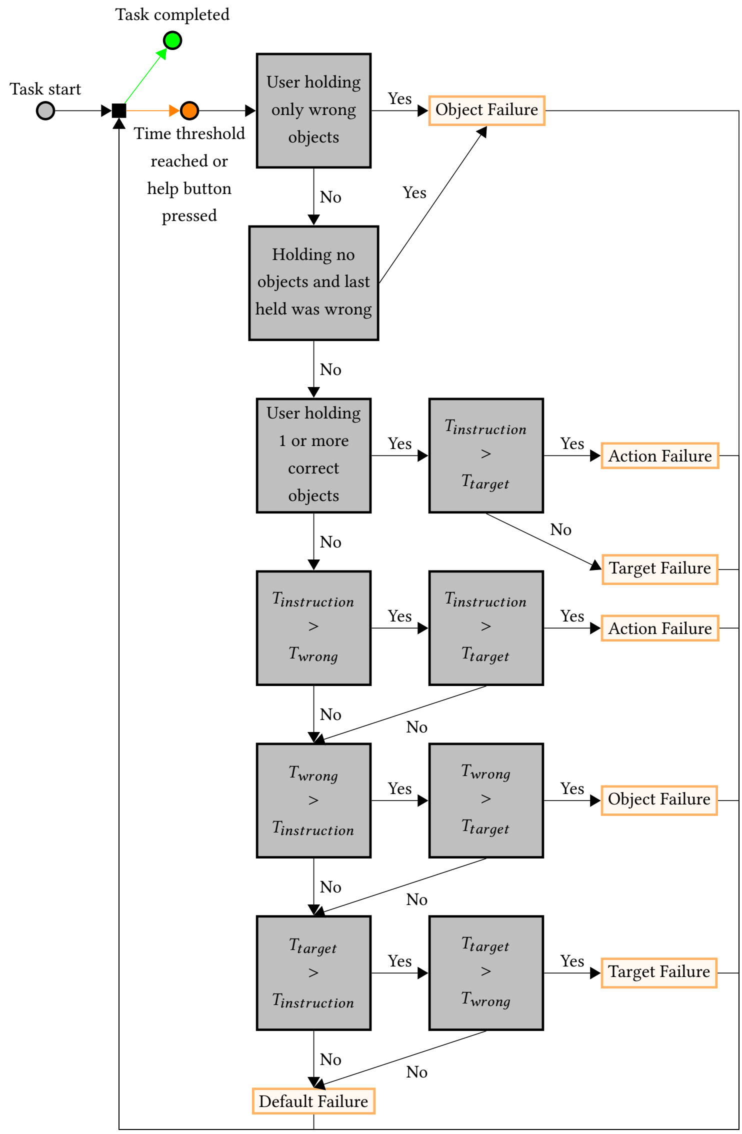


Fig. 4. Support is adaptively activated during the training depending on the failure state outcome of the algorithm. Firstly, the objects the user is holding are taken into consideration. If these do not trigger a failure state, the eye tracking data will be compared. $T_{instructions}$, T_{target} and T_{wrong} refer to the time the user has been looking at elements instructions, the target location where the current part has to be placed, and the wrong parts for the current task respectively during the current task.

3.2.2 Semi-automated adaptive learning. Semi-automated adaptive learning works the same as automated adaptive learning (see Section 3.2.1) except for how support is activated. It also starts with only *instructions* active when the system has no historical training data of a user. It also allows for the addition of support during tasks and the removal of support between iterations. The key difference is that the semi-automated approach is a pull variant, which means that the trainee

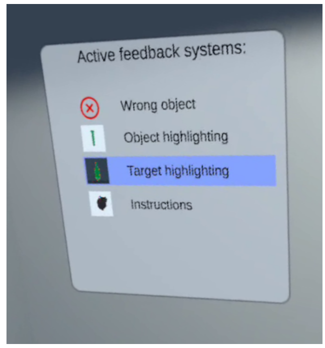


Fig. 5. This is what the panel looks like when all four systems are active. The blue outline indicates that target highlighting was the most recently activated system.

is in control of deciding when additional support is needed for a particular training task. When semi-automated adaptive learning is used, the system remains passive until the user decides to press the help button. However, once triggered, it is still the algorithm that determines what type of support is additionally activated. So, in the example of task 9, if the user struggles with picking the correct object and presses the help button, the algorithm is triggered and will result in an object failure. The outcome will be the activation of *wrong object highlighting* or, if this support system is already active, *object highlighting*. If the user continues to struggle and presses the help button again, the algorithm will progressively enable additional support systems until all systems are active.

3.2.3 Non-adaptive learning. In the non-adaptive learning variant, all four support systems are always active. So, in the example of task 9, if the user is picking the wrong bolt, the bolt will automatically be highlighted in red. The correct bolts will also already be highlighted in green, the hologram at the target location will be visible and the instructions will be shown. Feedback is also not reduced between training iterations. This variant is in line with traditional VR training which often relies on these highlighting and coloring techniques [43].

4 Methodology

To evaluate the impact of the three different learning strategies, we conducted an experimental study (N=36) using the proposed VR assembly training. For clarity, we will refer to the three conditions as *AC* (automated adaptive learning condition), *SC* (semi-automated adaptive learning condition), and *NC* (non-adaptive learning condition). A between-subject design was employed to minimize learning effects, resulting in 12 participants per condition. The study received approval from our university's ethical committee. The sample size of the study was based on the guidelines provided by [3].

4.1 Apparatus

The VR training was developed for the Meta Quest Pro, using the Oculus VR (OVR) integration for Unity. The headset's eye and hand-tracking capabilities were used for the VR training. The OVR API provides a location and rotation for both eyes. To determine where the user was looking, we used ray casts from both eyes, which were then combined to estimate which particular object the user was looking at. Hand-tracking was used for natural interactions in the VR application; no controllers were used.

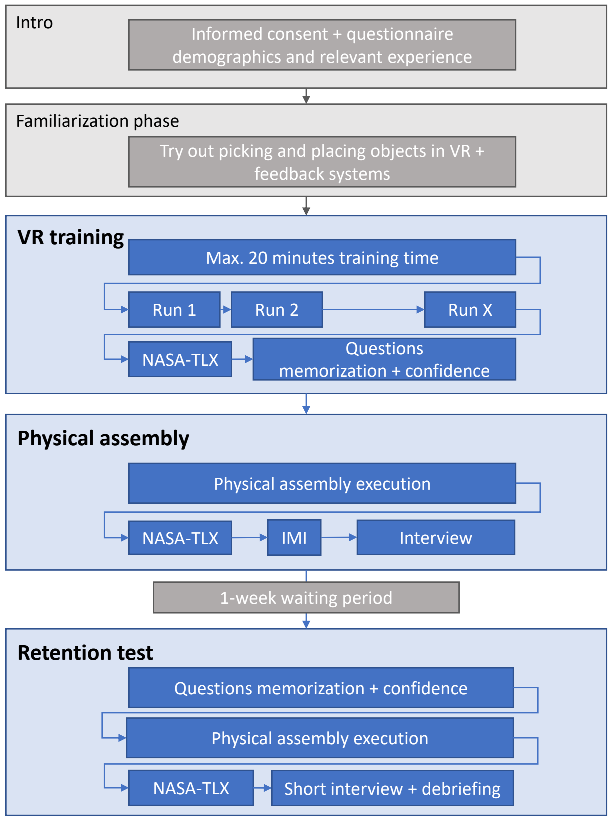


Fig. 6. Overview of the different steps in the user study

4.2 Protocol

Based on the work of Mayer [23] regarding the importance of performing retention tests and the physical transfer of skills, our study consists of three core parts:

- Virtual Reality training,
- physical assembly performance (= reproduction test),
- physical assembly execution after one week (= retention test)

Figure 6 provides an overview of the methodological approach.

4.2.1 Virtual Reality training. In the first part of the study, each user received an introduction to the study along with an informed consent document. Subsequently, they filled out the first part of the questionnaire about demographics, including age, dominant hand, experience with VR and experience with assembly tasks. There was a familiarization phase in the VR environment in which they learned how the assembly interactions work with some objects unrelated to the training (placing a coffee mug, stacking cubes, and inserting a key into a locker) but requiring similar actions in picking and placing objects. The participants also learned about the four support systems in this phase related to the real-time performance tracking support. In the case of the automated adaptive variant, the researcher manually triggered the four systems one at a time for the user. In the case of the semi-automated variant, the users did this themselves with the same help button they would later use in the actual training. In the non-adaptive version, all support systems were already active during the familiarization phase, and the researcher only explained their function. Once participants became sufficiently familiar, they moved to the actual VR training environment. Each participant received a maximum of 20 minutes of training time, an amount determined based on former pilot studies to estimate an appropriate duration. Participants could run through the virtual assembly training multiple times within the provided time frame. After each completed training run, they could press a restart button to start again. They were instructed to train until they believed they knew the assembly well enough to perform it physically without any assistance or instructions (or until their time ran out). We chose to allow freedom in training time in order to investigate whether a certain learning strategy would lead to more efficient training and, thus, less training time. After the virtual training, the participants filled out a NASA-TLX [14] questionnaire regarding their training experience and related perceived workload and also questions about how confident participants felt in their ability to perform the physical assembly and how well they memorized the training steps. During the training, we collected real-time hand and eye tracking data to assess the duration of interaction with objects. This data was used to inform decisions about when to add or remove support systems, ensuring that the training experience remained dynamic and personalized based on the operator's engagement and performance.

4.2.2 Reproduction test. In the second part of the study, the participant had to assemble the physical components without instructions or assistance (see Figure 2a). They were not informed to perform the assembly in exactly the same order as they learned in the virtual training. They were told they could ask the researcher for help if they got stuck. The participant's actions were recorded with a top-down video of the table and assembly setup. Afterward, participants filled out a NASA-TLX and an IMI (Intrinsic Motivation Inventory) questionnaire [24] related to the physical assembly experience. From the IMI questionnaire, we selected three relevant categories for our experiment, i.e. Interest/Enjoyment, Perceived Competence and Effort/Importance, inspired by the work of Dhiman et al. [9]. Lastly, we conducted a semi-structured interview where the participants were asked how they experienced the training session, what improvements they suggested, what features or support were missing, etc.

4.2.3 Retention test. The final part consisted of a retention test that took place one week after completing the training. In this retention test, the participant started with a small questionnaire about their confidence in remembering the assembly sequence. Subsequently, they performed the physical assembly again without any instructions. They could request assistance if needed, and the assembly execution was recorded on video. Lastly, they filled out a NASA-TLX questionnaire and answered some final questions about the training in a short semi-structured interview.

4.2.4 Post-analysis. After the experiment, we invited two experts to classify the errors made in the physical assembly setup. Both experts knew the assembly setup and sequence of execution. They both individually classified the errors observed in the videos from the actual assembly execution of all participants in the reproduction and retention test. Afterward, they collaboratively checked all observed errors to come to an agreement. The following types of errors were classified: immediately corrected errors, later corrected errors (and the associated backtracking depth), non-essential sequence errors, non-corrected assembly errors, and errors where help from a facilitator was needed. The *immediately corrected errors* are errors corrected by the participant before proceeding to the next step. *Later corrected errors* correspond to situations where the participant makes a mistake and executes subsequent assembly steps before realizing their mistake, necessitating them to undo a number of steps (= backtracking depth) to correct their initial mistake. *Non-essential sequence errors* refer to mistakes that do not impact the physical assembly result but do not follow the trained execution sequence. We consider two main groups based on these error categories: minor and severe errors. The minor errors are the *non-essential sequence errors* and the *immediately corrected errors*. The severe assembly errors are the *later corrected errors*, *non-corrected assembly errors*, and *errors that need help from the facilitator*. We kept track of the backtracking depth of each later corrected error, i.e. how many subsequent steps were executed before the mistake was observed.

4.3 Hypotheses

Prior research has shown that adaptive learning in AR leads to improved efficiency [18]. Therefore,
H1: Adaptive learning in VR assembly leads to more efficient training, thus a shorter training duration.
The adaptive learning variant offers more personalized training based on performance. Therefore,
H2: Adaptive learning in VR assembly leads to more effective training:
H2a: faster physical assembly execution,
H2b: reduction of the number of errors in physical assembly execution,
H2c: improved long-term retention of physical assembly knowledge.
As proven by the research of Dhiman et al. [9], applying the Self-Determination Theory (often present in adaptive learning) leads to higher perceived competence. Hence,
H3: Adaptive learning in VR assembly leads to higher perceived competence and success.
Adaptive learning in AR shows to be preferred by users [18]. Hence,
H4: Adaptive learning in VR assembly improves user enjoyment.
Since our adaptive learning approach leads to a reduction of instructions or support compared to a non-adaptive variant,
H5: Adaptive learning in VR assembly leads to an increased level of perceived mental load during training.
Given the personal aspect of adaptive learning leading potentially to more effective training:
H6: Adaptive learning in VR assembly lowers the perceived workload in physical assembly execution after training.
Our adaptive learning approach activates and deactivates support based on real-time performance. Hence,
H7: Adaptive learning in VR assembly lowers the amount of active support and information available.
Although the last hypothesis might sound straightforward, we would like to investigate this to validate whether the adaptive approach leads to a significantly lower amount of information available in the VR environment without negatively affecting the learning experience.

5 Results

We utilized the Shapiro-Wilk test to determine if the data conforms to a normal distribution. If the data is normally distributed and the homogeneity of variances is confirmed using Levene's test, we proceed with ANOVA. If these conditions are unmet, we apply the Kruskal-Wallis test to assess significance. For post-hoc analysis, we employ the Pairwise Wilcoxon rank sum test with Holm-Bonferroni correction and compute the individual effect sizes.

Table 3 provides an overview of the participants' demographic data for each condition and the relevant experience. All participants were students or researchers at the university. We provided the familiarization phase mentioned earlier to optimally reduce the impact of prior VR experience,

Table 3. Overview table of the participants' demographics and relevant experience

	Age		Gender		Dominant hand		Experience with VR				Experience with assembly			
	18-24	25-34	Male	Female	Left	Right	None	Limited	Average	A lot	None	Limited	Average	A lot
AC	12	0	9	3	0	12	5	4	2	1	2	7	3	0
SC	9	3	9	3	2	10	1	7	3	1	2	3	6	1
NC	11	1	10	2	0	12	4	6	2	0	2	7	2	1

Table 4. Results for the average total training time (in seconds) and the number of training runs per condition

	AC		**SC**		**NC**	
	Training time	Nr. of runs	Training time	Nr. of runs	Training time	Nr. of runs
Mean	670.988	4.083	795.523	4.917	629.502	5.083
Std. Dev.	106.188	1.165	197.300	1.881	271.151	2.109

allowing participants to learn the required interactions and continue once they felt comfortable. None of the participants had prior experience with the assembly use case.

5.1 Training efficiency (H1)

Since participants could voluntarily inform the facilitator when they felt ready for the physical assembly execution, we analyzed the training time for each condition. None of the participants required the full 20-minute duration, as all indicated they were ready to proceed before the time expired. We noticed that SC took the longest (M=795.5 seconds), followed by AC (M=671.0 sec) and NC (M=629.5 sec). However, no significant statistical differences were found (ANOVA F(2,33)=2.172, p=0.130, η^2=0.116). Therefore, we cannot accept hypothesis **H1** (training time). Regarding the number of training sessions (iterations), AC showed the least (M=4.1), followed by SA (M=4.9) and NC (M=5.1). Similarly, no significant statistical difference was found (ANOVA F(2,33)=1.106, p=0.343, η^2=0.063). Table 4 shows an overview of these training metrics.

5.2 Training effectiveness (H2)

Regarding execution time, participants in AC performed the physical assembly setup fastest in the reproduction test (M=157 sec), followed by SC (M=164 sec), and NC showed the slowest performance (M=191 sec). However, no statistically significant differences were found (F(2,33)=1.145, p=0.331, η^2=0.065). Therefore, we cannot support hypothesis **H2a** (execution time).

Next to execution time, we also analyzed the performance errors during the physical assembly based on video recordings (see Section 4.2.4). We initially found a significant difference in severe errors for the three conditions with a Kruskal-Wallis test (H(2) = 4.014, p = 0.028, η^2=0.196). However, pairwise comparisons showed that AC and SC are short in significance (p=0.059 and p=0.059, respectively) compared to NC. Therefore, we cannot support hypothesis **H2b**. However, we want to emphasize the more positive trend in the adaptive learning conditions since we noticed that two participants made severe errors in AC and SC compared to seven in NC. For the minor errors, we did not encounter a statistically significant difference (F(2,33)=0.363, p=0.699, η^2=0.022) among the conditions.

In the retention test, one week after the training session, participants were asked to complete the physical assembly execution again. We investigated the same aspects as mentioned for the reproduction test. No significant results were found for the number of errors. However, we observed that in the AC variant, four participants made severe errors, while for SC and NC respectively seven and six participants made severe errors. Regarding execution time, we noticed very similar averages: M=188 sec for AC, M=191 sec for SC, and M=180 sec for NC, indicating that the type of learning approach had a limited impact on long-term retention knowledge. Therefore, we cannot accept hypothesis **H2c**.

5.3 Perceived competence and success (H3)

We investigated three different categories of the IMI questionnaire, one of which is Perceived Competence (6 questions). This category gives an indication of whether someone perceives to feel competent and successful in performing a certain activity. The questions related to this category did not reveal any significant differences across the three conditions (F(2,33)=0.218, p=0.0805, η^2=0.013). Given the result, we cannot support hypothesis **H3**. However, on average, we notice the most positive scores for AC in Perceived Competence (see Table 5). Looking at the questions we asked before the reproduction and retention tests regarding memorization and confidence, we did not encounter any statistically significant results.

5.4 User experience and enjoyment (H4)

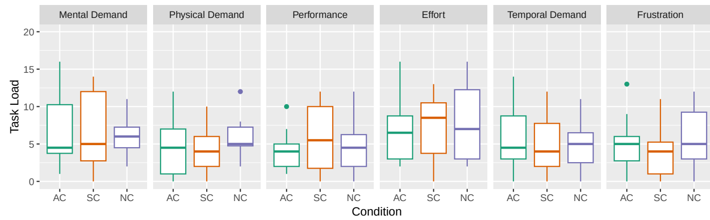
Another relevant category in the IMI questionnaire is Interest/Enjoyment. No statistically significant effect was found within this category among the three conditions (F(2,33)=2.678, p=0.084, η^2=0.140). However, on average, we notice a positive trend in scores for AC and SC (see Table 5). Looking at individual questions within this category, we found that for the question "I thought this was a boring activity", the scores were significantly affected by the training modality (H(2)=9.111, p=0.011, η^2=0.262). Pairwise comparisons showed that AC and SC are significantly less boring (p=0.023 and p=0.023, respectively) than NC. There were no significant differences between the two adaptive strategies (p=0.914). Given the positive trends in the category Interest/Enjoyment and the significant outcome regarding boredom, we accept hypothesis **H4** (user enjoyment).

Table 5. Descriptive statistics IMI questionnaire

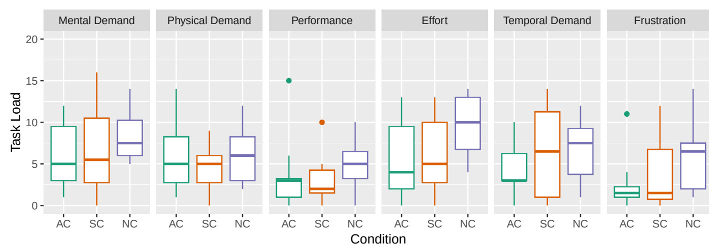
	Interest/Enjoyment			Perceived competence			Effort/Importance		
	AC	SC	NC	AC	SC	NC	AC	SC	NC
Mean	5.940	6.071	5.381	5.403	5.250	5.208	4.367	5.133	4.617
Std. Dev.	0.811	0.671	0.836	0.709	0.917	0.620	1.209	1.177	0.944

5.5 Mental effort and perceived workload (H5 & H6)

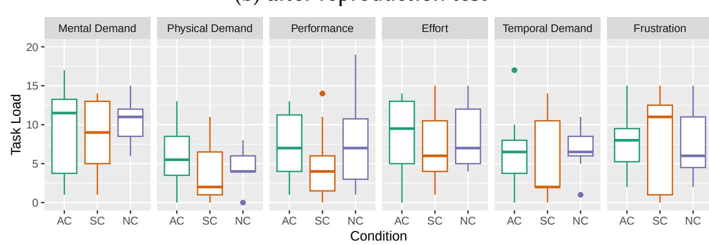
As mentioned in Section 4, we carried out a NASA-TLX questionnaire after the virtual training session, the reproduction test, and the retention test. No statistically significant findings were found in the NASA-TLX after the VR training and after the retention test (see Figures 7a and 7c). Regarding perceived mental demand during training (**H5**), we observe in Figure 7a that this score is slightly lower in AC and SC compared to NC after training. Since the results for mental demand do not show any significant differences, we cannot support hypothesis **H5**. However, we found a significant result for effort after the reproduction test. Effort scores were significantly affected by the training modality H(2)=7.490, p=0.024, η^2=0.184). Pairwise comparisons showed that AC and SC significantly reduced effort scores (p=0.042 and p=0.047, respectively) compared to NC. There were no significant differences between the two adaptive strategies (p=0.846). Given these findings, we accept hypothesis **H6**. Looking at Figure 7b, we observe that all NASA-TLX parameters score better on average for both adaptive learning conditions than for the non-adaptive variant after physical assembly execution right after the training. Although not statistically significant, we like



(a) after VR training



(b) after reproduction test



(c) after retention test

Fig. 7. Boxplots for NASA-TLX questionnaires

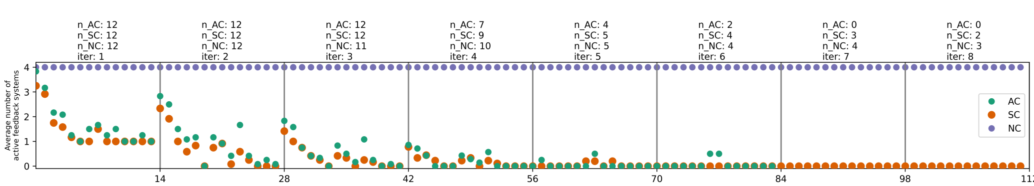


Fig. 8. Average amount of active support systems per task. (Please be aware that AC and SC data often overlap in iterations 4 to 6.)

to emphasize the findings for Frustration after completing the reproduction test (see Figure 7b). Reported averages for AC are 2.4, 4.1 for SC, and 6.1 for NC, indicating that SC and NC caused more frustration. Based on these aforementioned outcomes, hypothesis **H6** (perceived workload) is supported.

In the retention test, one week after the training session, participants were asked to complete the physical assembly execution again. We investigated the same aspects as mentioned for the reproduction test. The NASA-TLX showed no significant differences (see Figure 7c).

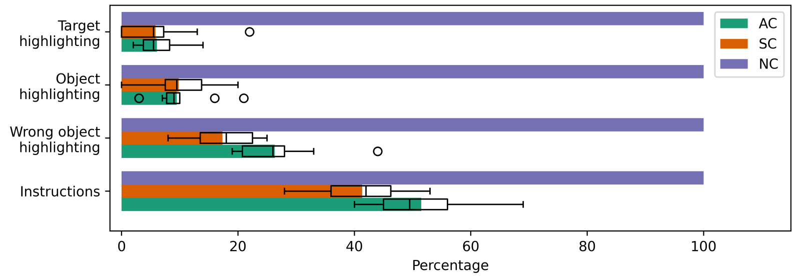


Fig. 9. Boxplots for the amount of time (in %) each support system was active for each variant on average.

5.6 Variations in support (H7)

We were also interested in the amount of support participants received in the different conditions (see Figure 8). The first training iteration had the most active support systems. The first tasks also generally had more active support systems than later tasks. These tasks required the participants to pick bearings that all look very similar but have small differences in size. The participants tended to pick the wrong bearings in these tasks in the first iterations, which explains the increase in active support systems due to increased complexity. The more iterations the participants performed, the less support systems were active as they gained knowledge throughout the runs. The final iteration of a training usually had no or very little active support systems left. This might be a more objective indication that participants are ready to start the physical assembly. Also, we can see that both AC and SC follow a very similar pattern, although there is a significant difference in the number of average support systems active per condition (F(2,33)=1517.396, p<.001, η^2=0.989). Post-hoc comparisons showed that SC has a significantly lower amount of support systems active than AC and NC (p=0.033 and p<0.001), and AC has a significantly lower amount than NC (p=0.003). Therefore, we can accept hypothesis **H7** (lower amount of supportive information). It is important to note that even when all four support systems are deactivated, users still experience some "support" that is unavoidable when interacting in VR. For example, participants of the SC variant tended to use snapping of objects as feedback. If they kept trying and the object would not snap into place, they inferred they were holding the wrong object without "Wrong object highlighting" being active. This could explain why SC has less active support systems than AC. Figure 9 shows an overview of active support systems' average time (in percentage) per condition. Note that these averages' order is the same as the default hierarchy (see Figure 3). It is logical that the more likely the algorithm is to remove a certain support system, the less time it will be active compared to other support systems. An interesting side effect we observed while analyzing and visualizing the number of active support systems (in Figure 8) is that it provides a more intuitive representation of learning progress for each task and each training iteration—something that is not as easily achieved with the non-adaptive variant.

5.7 Qualitative outcomes

Participants' responses (AC01-AC12, SC01-SC12, NC01-NC12) to the semi-structured interviews (conducted after both the reproduction and retention test, see Figure 6) were coded and then processed using thematic analysis [2]. The findings that emerged from this analysis are at least partly determined by the pre-defined questions that were asked during the interview.

Concerning hypothesis **H4** on the positive influence of adaptive learning on training enjoyment, 17 out of the 36 participants (AC:8, SC:7, NC:2) indicated they had enjoyed the training experience.

These participants referred to the VR training using words like “fun”, “amusing”, “nice” or “enjoyable”. Given the limited positive feedback from participants in the NC condition, this subjective data seems to corroborate the validity of hypothesis **H4**.

In the reproduction test interview, participants were asked which step(s) of the physical assembly they found most challenging. 19 participants responded that they struggled most with placing the interior bearings (constituting steps 0-3 of the assembly, see Table 1). They mainly reported that it was due to the difficulty of correctly inserting the bearings in the assembly housing (which requires the user’s hand to move quite deep into the housing). One particularly interesting comment was that “I can’t stick my hand through the housing, which made it difficult to insert the bearings” [AC09]. This remark indicates that the participant exploited the lack of haptic feedback between their hand and the assembly housing in the VR training, which allowed the participant to push the bearings through the housing to reach their target location instead of inserting them from the cavity at the top of the housing (as it must happen in the physical assembly). Next to the interior bearings, the two most challenging tasks were the assembly of the rotors (steps 4 and 5) (mentioned by 12 participants) and the assembly of the fine screws (mentioned by nine participants). The most reported problem with the rotors was that participants were unaware that the male rotor needed to be inserted with a twisting motion to make it fit with the female rotor. For the fine screws, during the reproduction test, participants often did not realize that these were, in fact, screws and hence needed to be screwed. These statements are a testament to an important limitation of the studied VR training software: components snap into place when they reach their target location in VR, whereas physical assembly often requires manual skill execution like screwing. All three VR training conditions shared the same snapping behavior, causing comments on this limitation to emerge from all three participant groups in roughly equal proportions.

The subjective data also sheds light on the perceived usefulness of the four support mechanisms. In particular, the data confirms that participants valued wrong object highlighting over correct object highlighting since eight participants explicitly mentioned the added value of wrong object highlighting versus only three participants for correct object highlighting. Only two participants made positive references to the learning benefits of target highlighting; the instructions support mechanism even received only one such positive remark. This might seem contradictory with the objective finding that instructions were the support system that was active the most during training. The subjective data hence contextualizes that this was not due to the perceived usefulness of this support mechanism but rather because it is the support mechanism that is added first and deactivated last during default system behavior (see Figure 3).

Regarding the object highlighting used in some of our support systems, one unexpected disadvantage that three participants mentioned was that it makes it harder to recognize the involved component during physical assembly: “I had trouble recognizing components because they were colored green all the time [during VR training]” [NC8], or “due to the object highlighting, I did not know the real color of the components” [NC12]. This disadvantage impacted the participants in the NC condition the most since the component highlighting mechanisms were active at all times in this condition.

Finally, the interview data underpins the positive influence of adaptive support systems during VR training on learning performance. Four out of 12 participants in the NC condition expressed the desire to learn with limited support mechanisms enabled or even with all support mechanisms disabled without knowing the existence of the AC or SC variant. In particular, both NC11 and NC12 stated that they wanted a training scenario where no component highlighting was active. At the same time, NC9 and NC10 asked for “no support” learning opportunities (e.g., “near the end [of the training session], I wanted to try to complete the training without support” [NC10]). Closely related to this, 12 participants (AC:2, SC:8, NC:2) mentioned that they deliberately adjusted their behavior to optimize their learning process. Both AC participants mentioned that they sometimes deliberately neglected the active support systems to optimize memorization (e.g., "at the end of the training session, I did not watch the active support anymore" [AC10]). SC07 tried to "use as little support as possible to improve the learning process", while all seven other participants stated to only manually request additional support when needed (e.g., "I only used the help button when I was really stuck" [SC10]). In the NC condition, the most interesting learning optimization approach was exhibited by NC08, who "always thought one step ahead [in the assembly sequence] to circumvent the aid of the always active support systems". Finally, nine participants (AC:3, SC:3, NC:3) drew positive relationships between adaptive support and learning performance (cf. hypothesis **H1** and **H2**). The following statement from SC04 nicely summarizes this: " I found it very useful that I had to explicitly ask for additional help because it allowed me to first think for myself [to come to a solution]. If the help were to be present all the time, I would be less forced to think, which would limit my learning".

6 Discussion

In this section, we first examine the generalizability of our approach to ensure that the adaptive strategies can be effectively applied across various VR training scenarios. Based on our findings and encountered limitations, we then propose guidelines to enhance the design of adaptive learning in VR training and recommendations for VR training evaluation. At the same time, we outline future directions for research to further improve and validate adaptive (VR) training approaches.

6.1 Generalizability

6.1.1 Extensibility to various training contexts. We have investigated the impact of different learning strategies in the context of assembly sequence training. The adaptive learning variants relied on the concept of enabling and disabling support systems dynamically. While the presented support systems provide a way to offer multi-leveled support tailored to the presented use case, they are not the only possible approach. In training, where e.g. craftsmanship is more important, the support systems can be complemented or replaced with support systems that demonstrate the required action or skill, e.g. a ghost avatar [18], but still rely on the adaptive learning principles we proposed and investigated.

6.1.2 Task- versus competence-level adaptivity. Although we provided more granular personalized support compared to Huang et al. [18] by focusing on an individual task level, a more holistic approach can be achieved by expanding this adaptivity to the competence level. This way, tasks are grouped based on their contribution to developing the same underlying competence. Upscaling adaptivity in this way could exponentially enhance learning effectiveness, as adjustments would no longer be confined to isolated tasks but instead inform training across multiple tasks and sessions. This shift would enable a more integrated management of failure states, allowing for adaptive support mechanisms that dynamically adjust across multiple tasks within a competence domain rather than in isolation.

While our findings led us to reject hypothesis **H1** regarding training efficiency, this does not imply that the adaptive approach lacks potential in this regard. Rather, our results suggest that achieving meaningful efficiency gains may require a broader, more long-term implementation. Given the constraints of a single and relatively short training session, it was challenging to observe substantial efficiency improvements. However, a holistic, competence-level approach could take important steps toward optimizing training efficiency by providing more sustained and strategically timed adaptations. Additionally, reducing use-case specificity could foster greater continuity in learning, enabling learners to build and transfer skills more effectively across various contexts.

Investigating the learning impact of competence-level support adaptivity and its implications for long-term skill acquisition represents an important avenue for future research.

6.2 Impact of adaptive training

Our study demonstrates that both AC and SC significantly outperform NC regarding perceived workload and user enjoyment during the physical assembly phase (supporting hypotheses **H4** and **H6**). Specifically, AC and SC reduced the perceived effort and seemed to positively impact the number of severe errors in physical assembly execution, aligning with hypothesis **H2b**. Furthermore, our results corroborate that adaptively reducing the set of support mechanisms, in a way that is personalized to the user's competencies, does not reduce physical assembly performance while at the same time significantly enhancing user enjoyment during training (supporting hypothesis **H4**). This suggests that adaptive training systems can sustain and even enhance training performance compared to non-adaptive training, in which the trainee continuously receives the maximal amount of support. The results also show that although significantly lower support is available in the adaptive conditions, it does not impact the perceived mental demand to perform during the training (supporting hypothesis **H5** and **H7**).

Our study also assessed the transfer of VR training knowledge to physical performance and long-term memory retention (contrary to e.g. [18]). This, unfortunately, did not reveal significant differences in retention among training conditions (i.e., hypothesis **H2c** was discarded, see Section 5.2). This finding may be impacted by the fact that physical assembly tasks were incorporated in our methodological approach immediately after VR training took place (see Figure 6). The immediate physical execution likely had a learning effect that we did not take into account upfront, which might have influenced the retention test results by potentially masking the differences between the training conditions. A limited number of participants actually hinted at this during the semi-structured interviews. Future research must continue to explore the impact of combining (adaptive) VR training with physical execution to better understand their respective benefits and limitations, investigate how their interplay affects long-term learning outcomes, and assess the impact on operators within assembly environments.

6.3 Limitations & lessons learned

One limitation of our current study is that we only assessed knowledge retention after one week, based on a single (VR) training session. This approach does not account for the effects of repeated or spaced training sessions over an extended period, which are critical for understanding the dynamics of long-term learning and forgetting. Research has shown that knowledge retention tends to decay over time, and without reinforcement, individuals may forget much of what they have learned [17]. Future research should explore longer-term training strategies, albeit in combination with micro-learnings [36], that incorporate multiple sessions spread across a certain timespan, allowing for the application of principles related to learning and forgetting. Nevertheless, our adaptive approach remains relevant for long-term training scenarios since it is not bound to a specific timeframe. The real-time support systems we developed can be consistently applied across multiple sessions, continuously adapting to the learner's needs and reinforcing learning regardless of the duration or frequency of training.

Secondly, our empirical study was conducted on the use of a compressor assembly. The involved compressor is relatively simple, both in terms of the different types of components involved and the short assembly sequence (see Section 4.1). It is our hypothesis that the learning benefits and impact of adaptive (VR) training will be even more pronounced as the involved product use case or assembly procedure becomes more elaborate or complex. Future research is therefore needed to explore the application of adaptive training across a broader range of scenarios to confirm these benefits and understand their potential impact on objective performance in real-world settings.

This study is used to test the potential benefits of a training system that allows for detailed personalization in terms of support systems without having too many influencing factors from a real-world setting, e.g. distraction, varying assembly knowledge. The participants in our study had mostly no to limited VR experience, which is in line with what we can expect within industrial companies. We deemed it sufficient to first test this concept of personalized VR training within a lab context with participants who might not fully resemble real-world users. A user test with a representative user group would be an ideal follow-up study now that the potential benefits are established. This would also allow for a more complex training case with longer training times.

The experimental study also uncovered an unexpected issue with how we implemented object highlighting in our VR training. though it was reported by only three participants. The employed color changes (e.g., green/red shading) assisted trainees in locating objects in the VR training environment, but at the same time, they made it more difficult for trainees to recall the objects' actual appearance during physical assembly (see Section 5.7). To address this, we recommend exploring alternative highlighting methods, such as outlines, spotlights, or animations, which may improve object recognition without negatively impacting memory retention [43].

Finally, our approach currently focuses mostly on adding support adaptively. Removing support happens according to a predefined hierarchy. Future work should investigate if removing support more adaptively is possible using real-time performance metrics and if doing so positively impacts training performance. One possible direction is to explore whether iterative task completion times can provide insights into training progress, which in turn could inform the adaptive run-time reduction of support systems. Also, we found certain participants in our experimental study to intentionally ignore support in later training iterations, suggesting that the support removal process may need to be faster for subsets of trainees. Future studies should investigate the timing and customization of support reduction to optimize the learning experience. Moreover, as seen in recent literature [13], there is growing potential for AI-driven adaptive learning systems. AI could play a role in dynamically determining when to provide or remove support based on real-time performance and potentially move beyond predefined rules to offer more flexible, performance-driven support tailored to individual learning progress.

6.4 Design guidelines for adaptive training

Based on our user experiment and resulting findings, we propose the following guidelines for designing adaptive training systems:

- **Adapt support based on real-time training performance.** Our study utilized an adaptive system, operating either fully automatically (AC) or semi-automatically (SC), that tailored support to the trainee's specific needs, in contrast to predefined support levels commonly seen in literature [18, 40] The positive outcomes for AC and SC underscore the value of adapting support to the trainee's performance in real time. The training system must not only enable additional support systems adaptively, but also the deactivation of such support systems should happen adaptively based on, e.g., run-time training performance. Additionally, the division of support systems into categories has shown to have a positive impact on the adaptive training system. In our test case, we divided them into "Object Failure", "Target Failure", and "Action Failure". This way, active support systems support the trainee specifically for either picking, placing, or understanding the requested task or interaction, while in other use cases, this might be related to another type of mistake, e.g., wiring electrical cables or following a safety protocol.

Facilitate training personalization. Personalized learning can cater to trainees who exhibit heterogeneous profiles, expertise levels and learning preferences [39]. To enhance training personalization, we recommend affording a combination of automated and semi-automated adaptive support mechanisms, which will grant trainees the flexibility to determine the extent of active support they receive. Such an approach accommodates varying user preferences for challenge and support during training, which we observed in our study: participants in SC tended to use less support than in AC, indicating a need for adaptive support mechanisms that cater to individual user preferences. By combining both strategies, the system can still determine which aspects of adaptive support it controls and which it does not, depending on the training situation. For example, in a critical task where safety measures are crucial, the system might push support, whereas in less critical tasks, the system might allow the user to decide when to increase or decrease support. A second approach is to apply the push (automated) mechanism more with novice trainees when support is essential for the learning progress and rely on the pull (semi-automated) mechanism when trainees get more skilled and experienced.

6.5 Recommendations for the evaluation of VR training

Through our review of related work and the evaluation of our own VR training, we identified several important aspects of the evaluation that we believe should receive more attention in future research:

Account for objective as well as subjective measures. Many training solutions predominantly emphasize objective performance metrics [7, 21]. In our study, both objective measures (such as the number of severe errors) and subjective measures (such as user enjoyment and perceived workload) showed improvements when applying adaptive training (i.e., AC and SC compared to NC). Subjective metrics must hence not be forgotten during the design of adaptive training systems; their relative importance compared to objective metrics must depend on the specific training needs and goals.

The importance of reproduction tests and retention tests. Our experiment further validated the importance of skill transfer and retention testing in accurately assessing the learning impact of VR assembly training, aligning with the broader multimedia learning principles recommended by Mayer et al. [23]. In contrast to prior work [18], which focused solely on evaluations within VR, our findings highlight the necessity of assessing learning outcomes in a physical assembly context to ensure effective skill transfer. This is further supported by the differences we observed in both objective measures (e.g., number of errors) and subjective measures (e.g., perceived performance in NASA-TLX) across different evaluation moments, reinforcing the need for reproduction and retention testing to fully capture the learning impact.

7 Conclusion

In this paper, we have investigated two variants of VR training systems based on adaptive learning. They adjust support to the user's performance either automatically or semi-automatically, providing more assistance when the user struggles and removing it as proficiency increases. We examined the learning effectiveness and perceived impact of these two adaptive variants compared to a non-adaptive version in a user study with 36 participants (12 for each variant). The results show that both adaptive variants are rated better regarding perceived workload and user experience, significantly reducing perceived effort during the physical assembly after the VR training. The adaptive variants were also found to be significantly less boring. The findings also hint that adaptive training may result in fewer severe errors during physical execution and the ability to correct errors immediately before having an impact on the subsequent assembly steps, although additional research is required to solidify these findings. Overall, our results establish the potential benefits of adopting adaptive VR learning approaches for industrial assembly training in terms of training performance as well as learning experience. A set of guidelines is proposed to facilitate and stimulate the practical adoption of adaptive VR training in the manufacturing domain but also beyond. Continued research on adaptive support mechanisms, as well as the interplay between VR training and physical training, remains essential for further optimizing adaptive VR training solutions.

Acknowledgments

This research was supported by Flanders Make, the strategic research centre for the manufacturing industry, in the project SKILLEDWORKFORCE. In addition, we would like to thank the people who participated in our study.

References

[1] Patrick Albus, Andrea Vogt, and Tina Seufert. 2021. Signaling in virtual reality influences learning outcome and cognitive load. *Computers & Education* 166 (2021), 104154. https://doi.org/10.1016/j.compedu.2021.104154

[2] V. Braun and V. Clarke. 2021. *Thematic Analysis: A Practical Guide.* SAGE Publications. https://books.google.be/books?id=mToqEAAAQBAJ

[3] Kelly Caine. 2016. Local standards for sample size at CHI. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 981–992.

[4] Jessica R Cauchard, Jeremy Frey, Octavia Zahrt, Krister Johnson, Alia Crum, and James A Landay. 2019. The positive impact of push vs pull progress feedback: a 6-week activity tracking study in the wild. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–23. https://doi.org/10.1145/3351234

[5] Gustavo Tietz Cazeri, Luis Antonio de Santa-Eulalia, Milena Pavan Serafim, and Rosley Anholon. 2022. Training for Industry 4.0: a systematic literature review and directions for future research. *Brazilian Journal of Operations & Production Management* 19, 3 (2022), 1–19. https://doi.org/10.14488/BJOPM.2022.007

[6] You-Shyang Chen, Chun-Hui Wu, and Ta-g Chen. 2017. An Adaptive e-Learning System for Enhancing Learning Performance: Based on Dynamic Scaffolding Theory. *EURASIA Journal of Mathematics, Science and Technology Education* 14, 3 (Dec. 2017). https://doi.org/10.12973/ejmste/81061

[7] Lea M Daling and Sabine J Schlittmeier. 2024. Effects of augmented reality-, virtual reality-, and mixed reality–based training on objective performance measures and subjective evaluations in manual assembly tasks: a scoping review. *Human factors* 66, 2 (2024), 589–626. https://doi.org/10.1177/00187208221105135

[8] Lea M. Daling, Marisa Tenbrock, Ingrid Isenhardt, and Sabine J. Schlittmeier. 2023. Assemble it like this! – Is AR- or VR-based training an effective alternative to video-based training in manual assembly? *Applied Ergonomics* 110 (jul 2023), 104021. https://doi.org/10.1016/j.apergo.2023.104021

[9] Hitesh Dhiman, Gustavo Alberto Rovelo Ruiz, Raf Ramakers, Danny Leen, and Carsten Röcker. 2024. Designing Instructions using Self-Determination Theory to Improve Motivation and Engagement for Learning Craft. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 673, 16 pages. https://doi.org/10.1145/3613904.3642136

[10] Sanika Doolani, Callen Wessels, Varun Kanal, Christos Sevastopoulos, Ashish Jaiswal, Harish Nambiappan, and Fillia Makedon. 2020. A Review of Extended Reality (XR) Technologies for Manufacturing Training. *Technologies* 8, 4 (Dec. 2020), 77. https://doi.org/10.3390/technologies8040077

[11] Zigang Ge, Miaoyin Xi, and Yanfei Li. 2019. A literature review of the adaptive algorithms adopted in adaptive learning systems. In *2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP)*. IEEE, 254–258.

[12] Eva Geurts, Arno Verstraete, and Maarten Wijnants. 2025. Empowering Adaptive Learning in VR Assembly Training Using Real-time Performance Tracking. In *MIM 2025: 11th IFAC Conference on Manufacturing Modelling, Management and Control.*

[13] Ilie Gligorea, Marius Cioca, Romana Oancea, Andra-Teodora Gorski, Hortensia Gorski, and Paul Tudorache. 2023. Adaptive Learning Using Artificial Intelligence in e-Learning: A Literature Review. *Education Sciences* 13, 12 (2023). https://doi.org/10.3390/educsci13121216

[14] Sandra G Hart. 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 Sage CA: Los Angeles, CA, 904–908. https://doi.org/10.1177/154193120605000909

[15] Aziz Hasanov, Teemu H. Laine, Jongik Kim, and Tae-Sun Chung. 2024. Lightlore: An Adaptation Framework for Design and Development of xAPI-Based Adaptive Context-Aware Learning Environments. *Electronics* 13, 13 (2024). https://doi.org/10.3390/electronics13132498
[16] Vincent Havard, David Baudry, Benoit Jeanne, Anne Louis, and Xavier Savatier. 2021. A use case study comparing augmented reality (AR) and electronic document-based maintenance instructions considering tasks complexity and operator competency level. *Virtual Reality* 25, 4 (Feb. 2021), 999–1014. https://doi.org/10.1007/s10055-020-00493-z
[17] Steven Hoedt, Arno Claeys, El-Houssaine Aghezzaf, and Johannes Cottyn. 2020. Real Time Implementation of Learning-Forgetting Models for Cycle Time Predictions of Manual Assembly Tasks after a Break. *Sustainability* 12, 14 (2020). https://doi.org/10.3390/su12145543
[18] Gaoping Huang, Xun Qian, Tianyi Wang, Fagun Patel, Maitreya Sreeram, Yuanzhi Cao, Karthik Ramani, and Alexander J. Quinn. 2021. AdapTutAR: An Adaptive Tutoring System for Machine Tasks in Augmented Reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM. https://doi.org/10.1145/3411764.3445283
[19] Li Ji, Chen Zhang, Haiwei Li, Ningning Zhang, Changhao Guo, Yong Zhang, and Xiaoyu Tang. 2023. A comprehensive experimental framework based on analysis of the pilot's EEG and NASA-TLX questionnaire in a VR environment. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine* 237, 7 (2023), 869–878.
[20] Julia M Juliano and Sook-Lei Liew. 2020. Transfer of motor skill between virtual reality viewed using a head-mounted display and conventional screen environments. *Journal of neuroengineering and rehabilitation* 17 (2020), 1–13.
[21] Alexandra D Kaplan, Jessica Cruit, Mica Endsley, Suzanne M Beers, Ben D Sawyer, and Peter A Hancock. 2021. The effects of virtual reality, augmented reality, and mixed reality as training enhancement methods: A meta-analysis. *Human factors* 63, 4 (2021), 706–726. https://doi.org/10.1177/0018720820904229
[22] Xiao-Wei Liu, Cheng-Yu Li, Sina Dang, Wei Wang, Jue Qu, Tong Chen, and Qing-Li Wang. 2022. Research on training effectiveness of professional maintenance personnel based on virtual reality and augmented reality technology. *Sustainability* 14, 21 (2022), 14351. https://doi.org/10.3390/su142114351
[23] Richard E Mayer. 2011. Applying the science of learning to multimedia instruction. In *Psychology of learning and motivation*. Vol. 55. Elsevier, 77–108. https://doi.org/10.1016/B978-0-12-387691-1.00003-X
[24] Edward McAuley, Terry Duncan, and Vance V Tammen. 1989. Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis. *Research quarterly for exercise and sport* 60, 1 (1989), 48–58. https://doi.org/10.1080/02701367.1989.10607413
[25] Brian Mok and Wendy Ju. 2014. The push vs pull of information between autonomous cars and human drivers. In *Adjunct Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. 1–5. https://doi.org/10.1145/2667239.2667287
[26] Roxana Moreno and Richard E Mayer. 2002. Learning science in virtual reality multimedia environments: Role of methods and media. *Journal of educational psychology* 94, 3 (2002), 598. https://doi.org/10.1037/0022-0663.94.3.598
[27] Maria Murcia-Lopez and Anthony Steed. 2018. A comparison of virtual and physical training transfer of bimanual assembly tasks. *IEEE transactions on visualization and computer graphics* 24, 4 (2018), 1574–1583. https://doi.org/10.1109/TVCG.2018.2793638
[28] Mir Murtaza, Yamna Ahmed, Jawwad Ahmed Shamsi, Fahad Sherwani, and Mariam Usman. 2022. AI-Based Personalized E-Learning Systems: Issues, Challenges, and Solutions. *IEEE Access* 10 (2022), 81323–81342. https://doi.org/10.1109/ACCESS.2022.3193938
[29] Matthias Neges, Stefan Adwernat, and Michael Abramovici. 2018. Augmented Virtuality for maintenance training simulation under various stress conditions. *Procedia Manufacturing* 19 (2018), 171–178. https://doi.org/10.1016/j.promfg.2018.01.024 Proceedings of the 6th International Conference in Through-life Engineering Services, University of Bremen, 7th and 8th November 2017.
[30] Fabrizio Palmas, David Labode, David A. Plecher, and Gudrun Klinker. 2019. Comparison of a Gamified and Non-Gamified Virtual Reality Training Assembly Task. In *2019 11th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games)*. IEEE. https://doi.org/10.1109/vs-games.2019.8864583
[31] Sebastian Pimminger, Werner Kurschl, Lisa Panholzer, and Johannes Schönböck. 2021. Exploring the learnability of assembly tasks using digital work instructions in a smart factory. *Procedia CIRP* 104 (2021), 696–701. https://doi.org/10.1016/j.procir.2021.11.117
[32] Simon Rafaqat, Sahil Rafaqat, Saoul Rafaqat, and Dawood Rafaqat. 2022. The impact of workforce diversity on organizational performance: a review. *Journal of Economics and Behavioral Studies* 14, 2 (J) (2022), 39–50.
[33] Juan Jesús Roldán, Elena Crespo, Andrés Martín-Barrio, Elena Peña-Tapia, and Antonio Barrientos. 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 (2019), 305–316. https://doi.org/10.1016/j.rcim.2019.05.004
[34] Nailya R. Salikhova, Martin F. Lynch, and Albina B. Salikhova. 2020. Psychological Aspects of Digital Learning: A Self-Determination Theory Perspective. *Contemporary Educational Technology* 12, 2 (Oct. 2020), ep280. https://doi.org/10.30935/cedtech/8584

[35] Sein Minn. 2022. AI-assisted knowledge assessment techniques for adaptive learning environments. *Computers and Education: Artificial Intelligence* 3 (2022), 100050. https://doi.org/10.1016/j.caeai.2022.100050

[36] Mrigank S Shail. 2019. Using micro-learning on mobile applications to increase knowledge retention and work performance: a review of literature. *Cureus* 11, 8 (2019).

[37] Joshua K Strakos, Matthew A Douglas, Blaine McCormick, and Michael Wright. 2023. A learning management system-based approach to assess learning outcomes in operations management courses. *The International Journal of Management Education* 21, 2 (2023), 100802.

[38] Kachina Studer, Hing Lie, Zhen Zhao, Ben Thomson, Dishita G Turakhia, and John Liu. 2024. An Open-ended System in Virtual Reality for Training Machining Skills. In *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems (CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 392, 5 pages. https://doi.org/10.1145/3613905.3648666

[39] Deborah L. Taylor, Michelle Yeung, and A. Z. Bashet. 2021. *Personalized and Adaptive Learning*. Springer International Publishing, Cham, 17–34. https://doi.org/10.1007/978-3-030-58948-6_2

[40] Jessica Ulmer, Sebastian Braun, Chi-Tsun Cheng, Steve Dowey, and Jörg Wollert. 2022. Gamification of virtual reality assembly training: Effects of a combined point and level system on motivation and training results. *International Journal of Human-Computer Studies* 165 (sep 2022), 102854. https://doi.org/10.1016/j.ijhcs.2022.102854

[41] Jeroen JG Van Merrienboer and John Sweller. 2005. Cognitive load theory and complex learning: Recent developments and future directions. *Educational psychology review* 17 (2005), 147–177.

[42] Paulette Vincent-Ruz and Nathan R. B. Boase. 2022. Activating discipline specific thinking with adaptive learning: A digital tool to enhance learning in chemistry. *PLOS ONE* 17, 11 (11 2022), 1–21. https://doi.org/10.1371/journal.pone.0276086

[43] Spyros Vosinakis and Panayiotis Koutsabasis. 2017. Evaluation of visual feedback techniques for virtual grasping with bare hands using Leap Motion and Oculus Rift. *Virtual Reality* 22, 1 (May 2017), 47–62. https://doi.org/10.1007/s10055-017-0313-4

[44] Josef Wolfartsberger, Robert Zimmermann, Gabriele Obermeier, and Daniel Niedermayr. 2023. Analyzing the potential of virtual reality-supported training for industrial assembly tasks. *Computers in Industry* 147 (2023), 103838. https://doi.org/10.1016/j.compind.2022.103838

[45] Peng Yang and Zhi Liu. 2022. The influence of immersive virtual reality (IVR) on skill transfer of learners: The moderating effects of learning engagement. *International Journal of Emerging Technologies in Learning (iJET)* 17, 10 (2022), 62–73.

[46] Jue Yu, Hoisoo Kim, Xiaoli Zheng, Zihan Li, and Xiangxiang Zhu. 2024. Effects of scaffolding and inner speech on learning motivation, flexible thinking and academic achievement in the technology-enhanced learning environment. *Learning and Motivation* 86 (2024), 101982.