



# Effects of Level of Immersion on Virtual Training Transfer of Bimanual Assembly Tasks

Songjia Shen<sup>1\*</sup>, Hsiang-Ting Chen<sup>2</sup>, William Raffe<sup>1</sup> and Tuck Wah Leong<sup>1</sup>

<sup>1</sup>School of Computer Science, University of Technology Sydney, Sydney, NSW, Australia, <sup>2</sup>School of Computer Science, The University of Adelaide, Adelaide, SA, Australia

The availability of consumer-facing virtual reality (VR) headsets makes virtual training an attractive alternative to expensive traditional training. Recent works showed that virtually trained workers perform bimanual assembly tasks equally well as ones trained with traditional methods. This paper presents a study that investigated how levels of immersion affect learning transfer between virtual and physical bimanual gearbox assembly tasks. The study used a within-subject design and examined three different virtual training systems i.e., VR training with direct 3D inputs (HTC VIVE Pro), VR training without 3D inputs (Google Cardboard), and passive video-based training. 23 participants were recruited. The training effectiveness was measured by participant's performance of assembling 3D-printed copies of the gearboxes in two different timings: immediately after and 2 weeks after the training. The result showed that participants preferred immersive VR training. Surprisingly, despite being less favourable, the subjects' performance of video-based training were similar to training on HTC VIVE Pro. However, video training led to a significant performance decrease in the retention test session 2 weeks after the training.

## OPEN ACCESS

### Edited by:

Michael Madary,  
University of the Pacific, United States

### Reviewed by:

Henrique Galvan Debarba,  
IT University of Copenhagen, Denmark  
Mark Billingham,  
University of South Australia, Australia

### \*Correspondence:

Songjia Shen  
songjia.shen@student.uts.edu.au

### Specialty section:

This article was submitted to  
Virtual Reality and Human Behaviour,  
a section of the journal  
Frontiers in Virtual Reality

**Received:** 21 August 2020

**Accepted:** 03 May 2021

**Published:** 20 May 2021

### Citation:

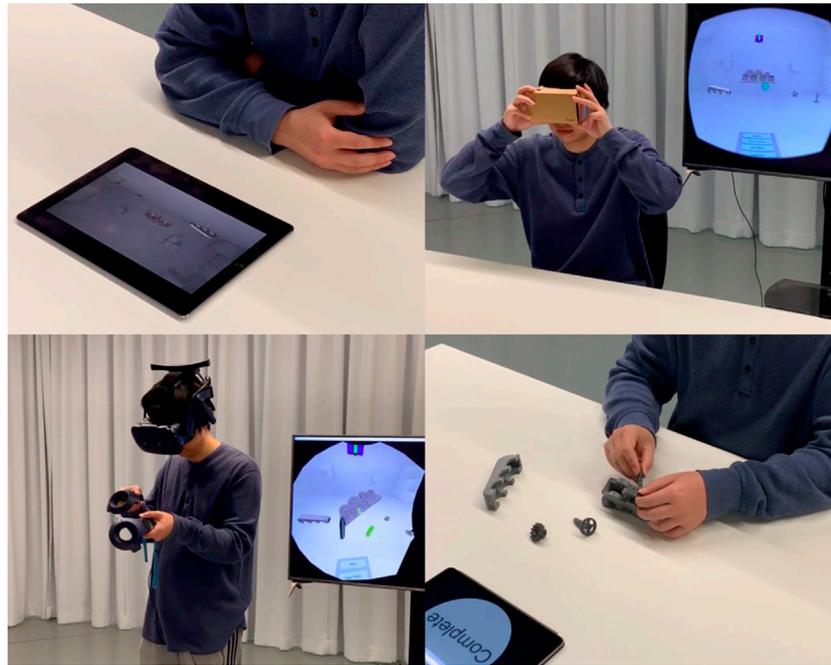
Shen S, Chen H-T, Raffe W and  
Leong TW (2021) Effects of Level of  
Immersion on Virtual Training Transfer  
of Bimanual Assembly Tasks.  
Front. Virtual Real. 2:597487.  
doi: 10.3389/frvir.2021.597487

**Keywords:** assistive systems, head-mounted display, virtual reality, learning transfer, assembly, training

## 1 INTRODUCTION

Assembly workers on production lines require constant training to maintain productivity and to ensure that they continue to work safely. Traditional training is often done at the existing assembly line in the factory. This has several disadvantages, including a reduction of productivity due to a need to suspend the assembly line, a potential waste of expensive materials, and the potential for the trainee to be exposed to physical danger during the training. In this context, virtual reality (VR) training is a good proposition for the manufacturing industry as it leads to minimum disruptions of the production environment and a potentially safer training process. Indeed, several companies in the automotive industry, such as Audi, Ford, and Toyota, have already piloted virtual reality training as part of their employee training programs (Jiang, 2011; Berg and Vance, 2017).

Decades of VR research has provided evidence of the effectiveness of virtual reality training in a wide variety of manufacturing and assembly tasks (Bailenson et al., 2008; Berg and Vance, 2017). Recent works (Oren et al., 2012; Carlson et al., 2015; Murcia-Lopez and Steed, 2018) compared the effectiveness of virtual training and physical training for bimanual assembly tasks of solving 3D burr puzzles. The result showed no significant performance difference between the virtual training and physical training conditions, in terms of assembly times and success rates. However, how the level of immersion i.e., the objective level of sensory fidelity (Bowman and McMahan, 2007), of the virtual training affects the learning transfer from the virtual training to physical assembly tasks is still an



**FIGURE 1** | A user learning assembly tasks in *Video*, *Mobile VR*, and *PC VR* conditions and assembling a 3D-printed gearbox on the table.

open question. Previous experimental protocols focused on the comparison between virtual and physical training. Participants only experienced the virtual training with identical or similar level of fidelity e.g., using the same virtual reality headset but with different types of training instructions and materials.

This paper presents a study that examined the effects of level of immersion on virtual training transfer of bimanual assembly tasks. Participants learned the assembly procedure of functional gearboxes using *Video* (a pre-recorded video of animated assembly process shown on a tablet), *Mobile VR* (Google Cardboard), and *PC VR* (HTC VIVE Pro), as shown in **Figure 1**. During the *Video* condition, the participant passively watched looping videos. The *Mobile VR* provided a stationary semi-immersive experience, where the participant can freely look around the virtual world but can only interact with gearbox pieces with fuse button and default Cardboard Reticle. Lastly, the *PC VR* provided a higher level of immersion, where the participants manipulated the virtual gearbox pieces with 3D-tracked VIVE controllers on both hands. These three levels of immersion were chosen because they represent the most commonly used virtual training systems with significant differences in input modalities and cost differences.

The effectiveness of virtual training was measured in two post-training test sessions which happened immediately after the training (*immediate test*) and 2 weeks later (*retention test*). In both tests, the participants were instructed to physically assemble the 3D-printed version of the gearboxes. The measurements include subjective questionnaire feedback, task completion time, and number of assembly errors.

This study tests the following hypotheses:

- H1: participants would prefer virtual training with a higher level of immersion as the learning experience is more engaging.
- H2: virtual training with a higher level of immersion would yield better assembly performance in both the immediate and retention test sessions, as the immersive virtual training was more memorable and could assist memory recall.
- H3: virtual training with direct 3D-input techniques would yield the best assembly performance, as the hand movements would improve the memory and retrieval of the assembly procedure.

This paper contributes to the understanding of the benefit of immersion (or lack of) toward virtual learning transfer of bimanual assembly tasks. The study is the first to investigate the effectiveness of three representative virtual training platforms i.e., *Video*, *Mobile VR*, and *PC VR*, with a practical gearbox assembly task. We hope the findings presented in this paper will lead to the design of new virtual training systems and provide insights for companies that are interested in deploying virtual training for their employees.

## 2 RELATED WORK

Previous works proposed and evaluated virtual reality training systems in a wide range of domains, such as medical (Stanney et al., 1998; Ruthenbeck and Reynolds, 2015), manufacturing (Boud et al., 1999; Wang et al., 2016), and military (Adams et al., 2001; Bhagat et al., 2016). Here, we highlight some research that focused on the virtual training of assembly and procedural tasks.

Although many of them have showed the training benefits, the effectiveness of virtual environments remains inconclusive due to the use of different reference conditions.

(Adams et al., 2001) compared the learning effectiveness among video, virtual training without haptic feedback, and virtual training with haptic feedback for assembling a LEGO biplane model. They reported that virtual training with haptics had significantly better performance than video. (Gavish et al., 2015) evaluated video, AR and VR training platforms using an electronic actuator assembly task and found AR training resulted in fewer unresolved errors than video watching with physical practice while there were no significant difference in performance between VR training and video watching only. Hall and Horwitz (2001) compared training of device operating between a VR interface (using head-mounted display and PINCH gloves) and a 2D interface (using conventional computer monitor and mouse) and found no significant difference in performance between the groups. Sowndararajan et al. (2008) compared an immersive projection display with a laptop display for memorizing procedures of different complexity. The result showed that higher immersion resulted in better training outcomes only for the more complex procedure. de Moura and Sadagic (2019) evaluated the effects of various combinations of stereopsis and immersive display on assembly training with image-based instructions and reported the immersive stereopsis group outperformed the others.

Instead of using 2D-based training as a control condition, other studies compared virtual training with physical training in real environments. Gonzalez-Franco et al. (2017) studied collaborative training of assembling aircraft doors in a mixed reality setup and a conventional face-to-face scenario and found no significant difference in performance. Funk et al. (2017) conducted an 11-days study in an industrial assembly workplace and found that augmented reality was helpful for untrained workers. Hoedt et al. (2017) evaluated assembly training in a mixed setup with hand tracking enabled and found it led to similar performance as the physical training. Besides, some research also examined effects of different types of instructions on assembly tasks (Yuviler-Gavish et al., 2011; Rodríguez et al., 2012).

Despite the mixed findings, researchers have created a range of applications for assembly and procedural training, considering the various learning affordances enabled by virtual environments (Dalgarno and Lee, 2010). Ritter et al. (2001) developed a VR training application for anatomy education (composing bones and muscles of a 3D virtual foot) and mechanical engineering (assembling a car engine). Gerbaud et al. (2008) created a VR authoring platform based on two previously developed modules (Mollet and Arnaldi, 2006; Mollet et al., 2007), which could support defining object behaviors and operation sequences programmatically for various procedural tasks. Gorecky et al. (2017) designed and implemented a virtual training system for automotive manufacturing, focusing on automation of training content generation. There are also various attempts to integrate computer-aid design models and virtual assembly (Leu et al., 2013).

Some recent works used puzzle pieces as the training target. Such abstract assembly tasks appear to provide better control over the task difficulty and reduces the potential confounders of pre-acquired domain knowledge. Shuralyov and Stuerzlinger (2011) implemented a mouse based system for assembling 3D puzzles on a desktop monitor and recruited two groups of participants (novices and experts) based on their time spent on computers per day and games per week. The result showed that the expert group assembled the puzzle significantly faster and both groups spent more time on rotations. Carlson et al. (2015), based on their previous work (Oren et al., 2012), compared the training effectiveness of a 6-piece burr puzzle assembly task between physical training and virtual training (with motion controls and haptic feedback). The result showed that the participants receiving physical training outperformed the ones receiving virtual training in the test right after the training. However, for the groups with color cues, virtual and physical training became equally effective in the retention test, 2 weeks after the training. Murcia-Lopez and Steed (2018) used a similar experimental protocol as Carlson et al. (2015) and rigorously examined six conditions that mixed physical and virtual training elements, using three puzzles with different complexity levels. The result showed that the virtual training could perform equally well as the physical training although all training methods led to poor performance in the retention session.

Complementing these previous studies (Oren et al., 2012; Carlson et al., 2015; Murcia-Lopez and Steed, 2018), our experiment used a similar experimental protocol (Murcia-Lopez and Steed, 2018) while using functional 3D-printed gearboxes as our assembly training targets, which better resembled actual assembly in production environments and still kept good control of task difficulty and certain level of abstraction. In addition, our experiment specifically examined the effects of different levels of immersion during the virtual training. In line with previous research (Young et al., 2014; Papachristos et al., 2017) on cost-differentiated VR systems, the experiment aimed to understand the benefits (or the limitations) of high-end VR system in virtual training. While the previous works compared the training effectiveness of various virtual environments with physical environments, it has not been examined in detail how different levels of immersion affect training transfer of assembly tasks. Since virtual systems could require significantly different training costs and times, selecting appropriate immersion levels for tasks could be crucial to deploy virtual training in production environments.

### 3 EXPERIMENTAL DESIGN

Our experiment used a within-subject design where participants learned the gearbox assembly procedure in three different conditions of virtual training:

- *Video*: The participant passively watched a pre-recorded animation using VLC on a 12-inch tablet in a sitting pose. The participant could start video play by tapping the video thumbnail.

**TABLE 1 |** The experimental conditions of low, middle and high immersion.

	Video	Mobile VR	PC VR
Resolution	1,920 × 1,080	2,220 × 1,080	2,280 × 1,600
Refresh rate	60 Hz	60 Hz	90 Hz
Field of view	—	80°	110°
User input method	—	Cardboard reticle	3D inputs
Head movement tracking	—	3 DoF	6 DoF
Level of immersion	Low	Mid	High

- *Mobile VR*: The participant wore a low-cost Google Cardboard viewer powered by a smart phone (Samsung Galaxy S8). The participant could manipulate the virtual pieces with the fuse button and the display reticle.
- *PC VR*: The participant wore a wireless HTC VIVE Pro headset and held one 3D-tracked controller on each hand. The participant could grasp and assemble the virtual gearbox pieces using both hands and could freely move around the virtual gearbox. The participant completed the training in a standing pose.

We concur with Dalgarno and Lee (2010) on the view that the representational fidelity and the interactivity types afforded by the virtual environment lead to the degree of immersion. These three different virtual training conditions represent three different levels of immersion: Low, Mid, High. **Table 1**

summarizes the major differences in their display specifications and freedom to interact with the virtual environments.

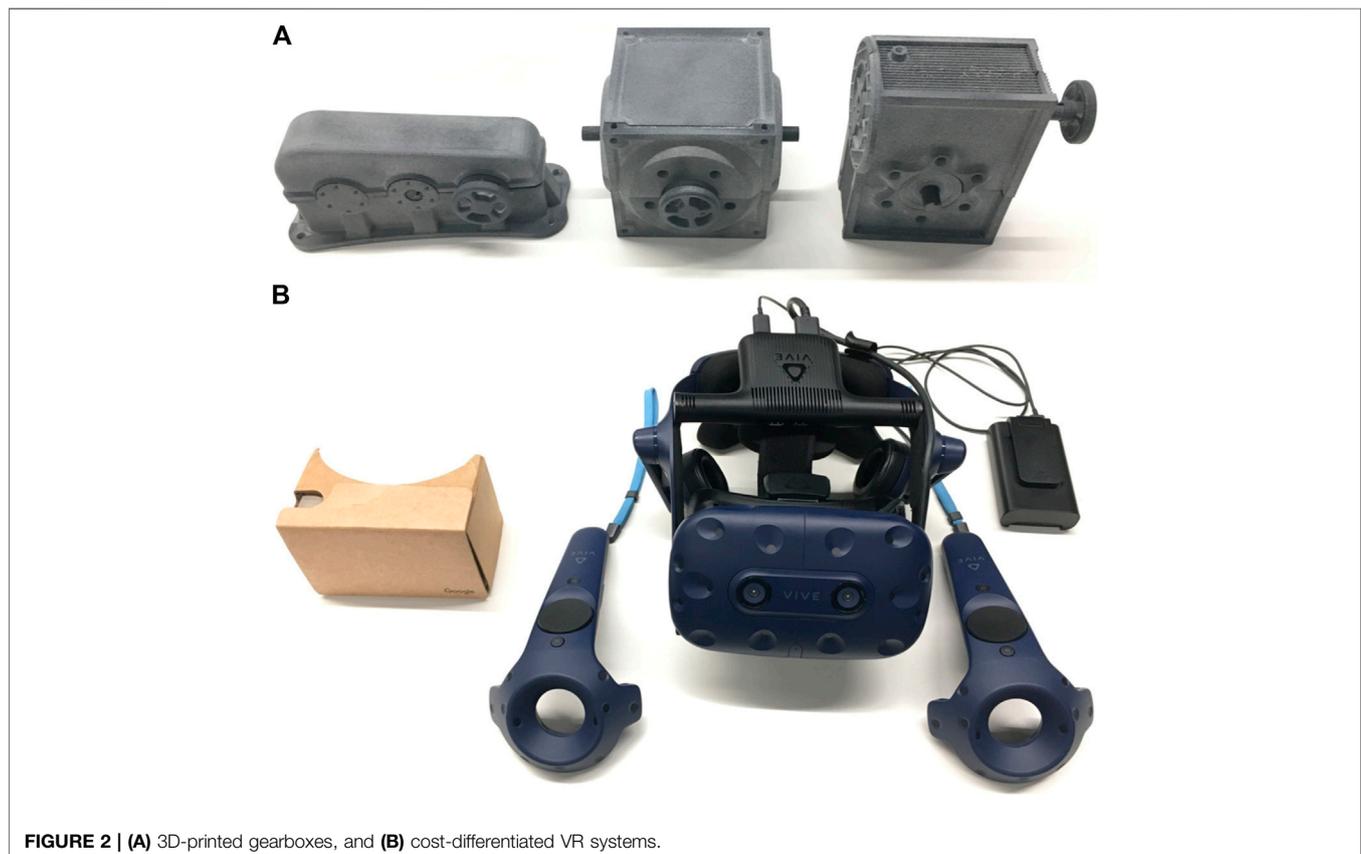
The rendering quality of *PC VR* and *Mobile VR* conditions were similar because of the low number of polygons on 3D assets and virtual environment and the use of plain lighting. The animations used in the *Video* condition were rendered and recorded on PC using the same scene.

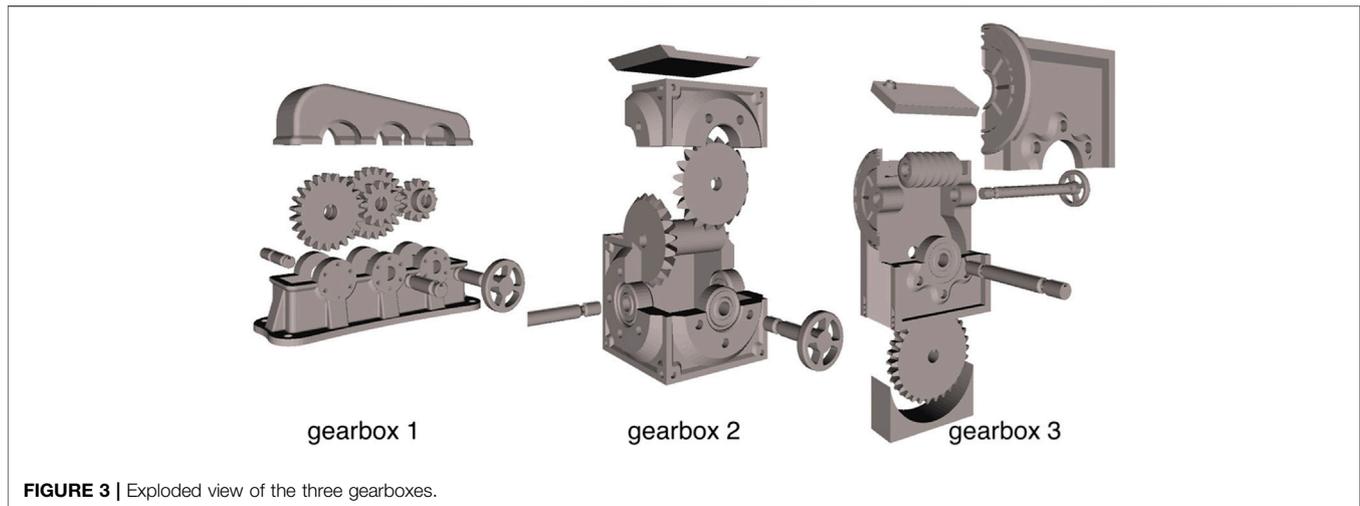
Note that unlike previous works (Oren et al., 2012; Carlson et al., 2015; Murcia-Lopez and Steed, 2018) that used a between-subject experiment design to mitigate the learning effects. Our within-subject experiment design let the participant assemble different gearboxes for each training condition. The pairing of gearboxes and training conditions was based on a 3×3 Latin square design. We assume these three gearboxes represent the same difficulty level to the participants because of the similar complexity, i.e. almost same numbers of assembly pieces and steps. Please find more discussion regarding the complexity of gearbox assembly in **Section 5.1**.

### 3.1 Apparatus

#### 3.1.1 3D-Printed Gearboxes

We 3D printed the target gearboxes (see **Figure 2A**) using the HP Jet Fusion 4200 3D printer with the material PA-12GB, which is Nylon 12 with glass bead reinforcement. Each gearbox piece is stiff and reusable. None of the pieces broke during the whole user study.





**FIGURE 3** | Exploded view of the three gearboxes.

**TABLE 2** | The numbers of components and assembly steps.

	Pieces	Steps: V	Steps: M	Steps: P
Gearbox 1	8	19	19 <sup>a</sup>	16 <sup>a</sup>
Gearbox 2	7	19	19	19
Gearbox 3	8	19	20 <sup>a</sup>	19 <sup>a</sup>

<sup>a</sup>indicates disparities between two VR conditions. V for Video, M for Mobile VR, and P for PC VR.

### 3.1.2 VR Headsets With Different Levels of Immersion

**Figure 2B** shows the two VR head-mounted displays used in the experiment, Google Cardboard and HTC VIVE Pro. As of November 2019, Google Cardboard costs around USD 15 while HTC VIVE Pro is priced at USD 799. Both headsets require an external device to drive the VR content and support stereoscopic rendering.

The two VR systems have many differences in graphics characteristics and input methods. HTC VIVE Pro has a better display with higher resolution, refresh rate, and field of view. The virtual environment rendering was driven by a powerful PC. HTC VIVE Pro also allows room-scale VR with direct 3D inputs and free walking in the space. In contrast, the phone-based VR system provided a limited field of view, had limited graphics processing power, and exploited the IMU on the smartphone to achieve simple inputs (see **Table 1**).

## 3.2 Gearbox Assembly

We chose functional gearboxes as our training targets. **Figure 3** shows the exploded views of these gearboxes and the corresponding 3D models, downloaded from Thingiverse<sup>1</sup>. These gearboxes were designed for training purposes and had several classical mechanisms. The gearbox 1 (left) had a double reduction gear mechanism that comprised of two pairs of gears. The gearbox 2 (center) mechanism had two bevel gears and two shafts that were

90° apart. The gearbox 3 (right) was a standard worm gearbox with both worm and worm gear at a gear ratio of 1:30.

All three gearboxes had similar numbers of components and assembly steps (see **Table 2**). During the virtual training, any kind of translation or rotation of the components was counted as one assembly step. The disparities in assembly steps between *Mobile VR* and *PC VR* were due to the different degree-of-freedom (DoF) in the input methods. The HTC Vive controller provided a 6 DoF control and the participant could control both the translation and rotation of the virtual piece in one step. While the *Mobile VR* control was limited to 3 DoF and thus required more steps to complete the gearbox assembly.

## 3.3 Virtual Training

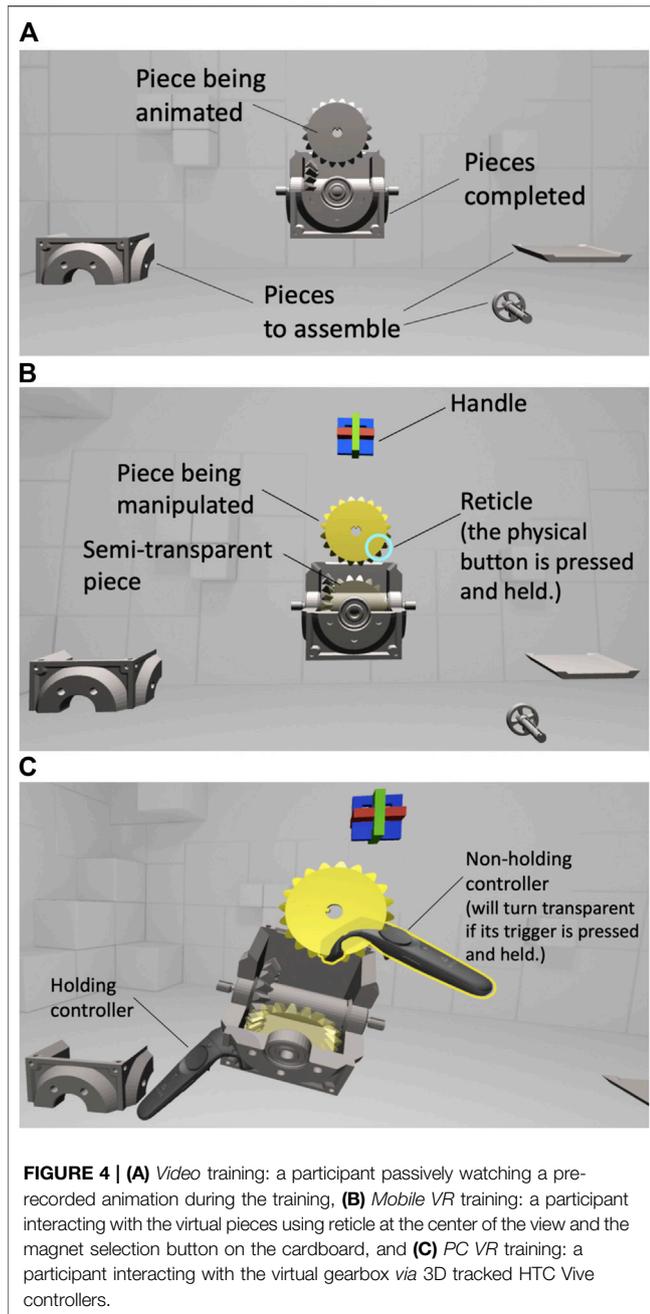
### 3.3.1 Video - Low Level of Immersion

This condition presented a video of animated assembly instructions for each gearbox (see **Figure 4A**). When the video begins, it shows all pieces of a gearbox positioned in a straight line in the lower part of the view. Then, according to a fixed sequence, they are moved to the center of the view and are assembled together step by step. Each step shows how pieces are translated to a proper position or rotated to a proper orientation. The participant passively watched the videos and memorized all the assembly moves. When the video is over, the participant can restart it by tapping its thumbnail in VLC. The training session ended when the participant announced that he/she was confident to start assembling the physical gearbox. Since this was the non-interactive baseline condition, the participant could not have any interaction (including video playback controls) while watching the assembly animation.

### 3.3.2 Mobile VR - Mid Level of Immersion

This condition used the same models and animated instructions in the VR application we developed (see **Figure 4B**). When the application was started, it showed all pieces of a gearbox with the same initial arrangement as the *Video* condition. Following the same sequence, it played the animated instruction of the assembly step with a semi-transparent version of the piece. The semi-transparent piece also emitted a blinking yellow light to

<sup>1</sup><https://www.thingiverse.com/thing:2995199>



attract the participant's attention. Once the animation was over, the application would wait for user's inputs. Participants needed to find the corresponding piece and manipulate it according to the animation. The application would keep checking the piece. If it had the same position and orientation as the semi-transparent one, the current step was completed and next step's animation would be played. After all the steps were completed, the gearbox would rotate by itself to indicate the success of assembly.

Interactions were achieved by using one physical button and head rotation. In the application, there was a small white dot (the reticle) at the center of the view. When users rotated their heads with the Google Cardboard headset, the reticle always remained

at the center. If it pointed to a gearbox piece, the white dot would become a bigger white circle and the color of the piece would change to yellow, indicating that this piece was interactive. Then users had to press and hold the physical button on the upper right corner of the headset to attach the piece to the reticle. If the attachment was successful, the color of the reticle would change to cyan and users could manipulate the piece with head rotation. Due to the limitation of the input method, in each step, a piece could either translate or rotate, but not both. The application decided the operation of current step and indicated it *via* the animation. In translation steps, pieces being manipulated would move with the reticle. In rotation steps, head rotation would be mapped to piece rotation accordingly. After the button was released, the head motion stopped affecting the piece and the color of the reticle changed back to white.

Besides gearbox pieces, there were several other objects that users could interact with. Two virtual buttons were implemented in the scene. One was the start button in front of the pieces and the other was the restart button below the pieces. Users could point to the virtual buttons and click the physical button to start or restart the animated instructions. Above the pieces, there was a small assembled burr puzzle, working as a handle. Rotating the handle would rotate all gearbox pieces together, so users could view instructions and assemble pieces from different angles if they wanted.

### 3.3.3 PC VR - High Level of Immersion

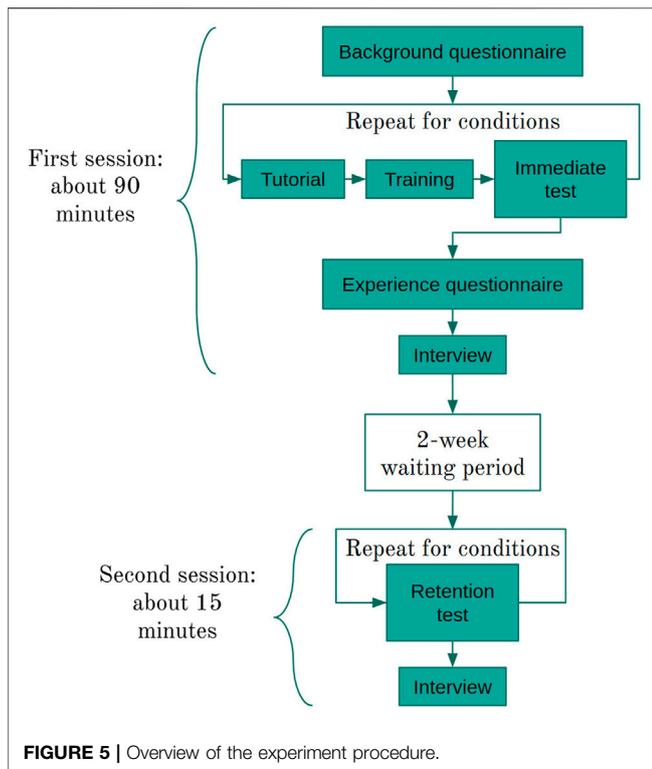
This condition had almost the same training contents as *Mobile VR* condition, expect for a few changes of interactions (see **Figure 4C**).

Since HTC VIVE Pro had motion tracking controllers, the reticle was removed and two models of the controllers were added in the scene. The virtual controllers had the same movements as the physical controllers. When a virtual controller collided with a gearbox piece, the piece would be highlighted by yellow color (just as *Mobile VR* condition). When the trigger of physical controller was pulled and held, the piece would be attached to the virtual controller. If the attachment was successful, the virtual controller would become transparent. Users then could use the physical controller to manipulate gearbox pieces with hand movements. If the trigger was released, the piece was detached and the virtual controller turned back to normal.

We introduced a compulsory bimanual mode in this condition. When the first step was completed, the completed piece would be attached to the controller. Subjects had to use the other controller to grab new pieces for the following steps and use both controllers to assemble them together. Subjects could change the controller for holding. They just need to use the non-holding controller to collide with the completed pieces and pull its trigger. Then the roles of the controllers were switched.

### 3.3.4 Virtual Training Limitation

Note that we did not enable the physical collision simulation in the virtual training. The size of virtual gearboxes and the interlocking nature of gears make the collision detection in Unity 3D unreliable. Instead, the application had a snap-to-fit mechanism (Carlson et al., 2015; Murcia-Lopez and Steed, 2018),



which would snap a piece to its semi-transparent guide if they were close enough. More specifically, the application would snap the piece to its semi-transparent guide if the piece was entirely inside the bounding box of the guide, which is 5% larger in size, and the difference between their orientations was within  $10^\circ$ .

All three conditions only allowed a linear progress of training. The participants could not disassemble the already assembled parts or skipped certain assembly steps. We intentionally set this constraint to ensure a similar training experience among participants.

## 4 METHODS

Written informed consent was obtained from the individuals for the publication of any potentially identifiable images or data included in this article. This study received approval from the University's Human Research Ethics Committee.

### 4.1 Participants

We recruited 23 adults, whose ages ranged from 19 to 38 (mean = 26.78, SD = 4.86), of which 9 were females. All subjects were reimbursed 30 dollars for their time. Among them, three subjects are excluded from the analysis, including one for the experience of assembly (having 5–10 h per week for assembly activities), one for misunderstanding the training protocol (not realizing that the training tasks could be completed multiple times) and one for the completion time in the immediate test (479.35 s, which is five times of the median value of 94.69 and more than two times of the second largest value of 217.38). About half of the participants (12

out of 20) had no or little experience using VR (less than 5 h) before the experiment, and five participants had studied mechanical engineering.

### 4.2 Experiment Procedure

The experiment consisted of two lab sessions with a 2-week waiting period. **Figure 5** shows a brief overview of the procedure. The first session trained the subjects in each experimental condition and tested their performance immediately after each training phase. In the second session, the subjects returned to the same room and were tested for their retention without undergoing any re-training. Before the experiment, the participants were asked to read an information sheet describing the assembly tasks and tests and to sign a paper copy of the consent form.

In the first session, the participant began by completing the online background questionnaire with the questions about their experiences of mechanical engineering, gearboxes, VR, 3D software, video games and assembly activities. These questionnaires evaluated their existing domain knowledge and their familiarity with the relevant technologies.

The participant then completed three virtual-training trials, one for each training condition. Three different gearboxes with similar assembly difficulties were used in the session, one for each trial. Each trial consists of three steps: *tutorial*, *training*, and *immediate-test* (see **Figure 5**). In the *tutorial* step, the subject learned the controls of the training conditions by completing a sample assembly task. In the *training* step, the subject completed the virtual training condition as described in **Section 3.3**. The subject could complete the training content as many times as they needed within 10 min and could end the training early. In the *immediate-test* step, the participant assembled the 3D-printed gearbox as fast as possible. There was no time limit for the test.

At the end of the first session, the participant completed an experience evaluation questionnaire to rate the gearbox assembly difficulty, instruction clarity, ease of use, usefulness and preference for each condition (see **Table 3**). The participant then participated in a semi-structured interview where they provided verbal responses about their training experiences and elaborated answers in the questionnaires.

In the second session, the participant first completed a retention test by reassembling the 3D-printed gearboxes using the same order in the first session. The participant was then interviewed about how they recall the assembly process and their thoughts about the effectiveness of previous virtual training.

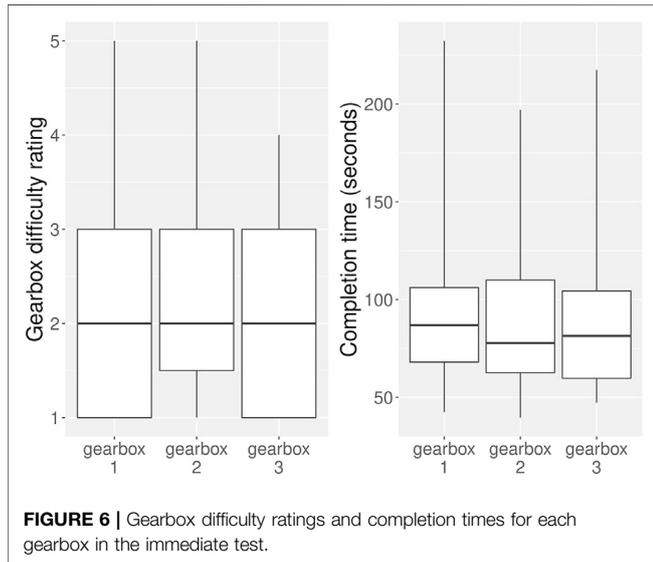
In both sessions, the 3D-printed pieces of the gearboxes were placed on a table and occluded from user's vision by a cover. The participants started the gearbox assembly by pressing a virtual button on a tablet and completed the process by pressing the same button again.

### 4.3 Data Collection

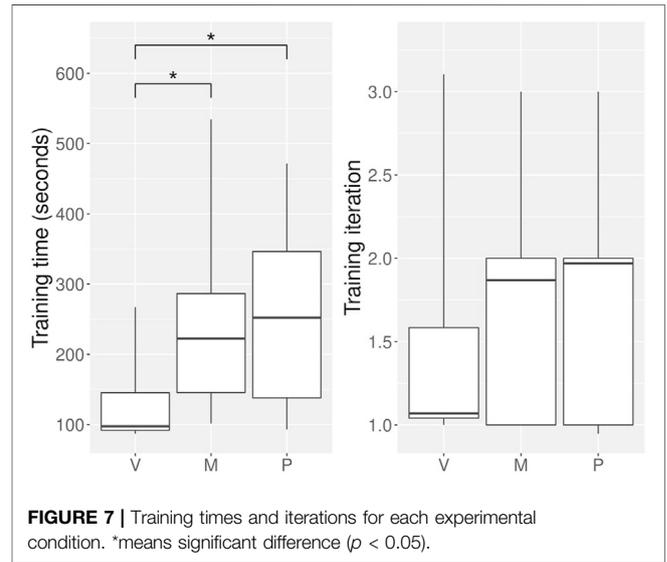
The VR application logged the training times of each assembly task completed in VR and the timestamps of each step. The training times spent on video were recorded by the researcher. These data were also used to calculate how many iterations of

**TABLE 3 |** Experience questionnaire.

Dimension	Question	Likert scale extremes
Difficulty	How difficult do you think are these gearboxes to assemble?	1 (very easy) - 5 (very hard)
Instruction clarity	Do you think the instructions are clear and easy to understand?	1 (very confusing) - 5 (very clear)
Ease of use	Do you think the applications are easy to use?	1 (very hard) - 5 (very easy)
Usefulness	Do you think the applications are helpful for memorizing the assembly steps?	1 (not helpful) - 5 (very helpful)
Preference	How likely do you think these three conditions will be your preferred learning method(s)?	1 (not likely) - 5 (very likely)



**FIGURE 6 |** Gearbox difficulty ratings and completion times for each gearbox in the immediate test.



**FIGURE 7 |** Training times and iterations for each experimental condition. \*means significant difference ( $p < 0.05$ ).

the training content the participant finished before the immediate test.

The completion times of physical assembly were recorded by the tablet. The assembled 3D-printed gearboxes were examined and disassembled by the researcher after the experiment and the misassembled pieces were recorded. The completion times and the numbers of assembly errors were used to evaluate the effectiveness of the virtual training.

The background questionnaire and the experience questionnaire were created and administered *via* Google Forms. All interviews were video recorded with the participant’s consent. All the collected data were saved in the CSV files (see **Supplementary Material**).

## 5 RESULTS

### 5.1 Gearbox Difficulty

We assumed all three gearboxes are of the same difficulty and the results appear to confirm our assumption with the participants rating the difficulty of each gearbox in the post-experiment questionnaire similarly. The completion time for each gearbox in the immediate test session was also similar (see **Figure 6**). Non-parametric analysis was performed and Friedman tests showed there was no statistically significant difference in ratings of difficulty or completion times between the different gearboxes.

### 5.2 Training Times and Iterations

**Figure 7** shows the training times and iterations of each training condition. Non-parametric analysis was performed.

A Friedman test showed there was an overall statistically significant difference in training times between the experimental conditions ( $\chi^2(2) = 24.4, p < 0.001$ ). A post hoc test using Wilcoxon signed-rank tests with Bonferroni correction showed the significant differences between *Video* and *Mobile VR* ( $Z = -3.920, p < 0.001, r = 0.877$ ) and between *Video* and *PC VR* ( $Z = -3.696, p < 0.001, r = 0.826$ ).

*Video* has the smallest median of 97.5 s (IQR = 91.75–145.25), followed by *Mobile VR* with a median of 222.6 s (IQR = 145.6–286.3) and *PC VR* has the largest median of 252.09 s (IQR = 138.01–346.25).

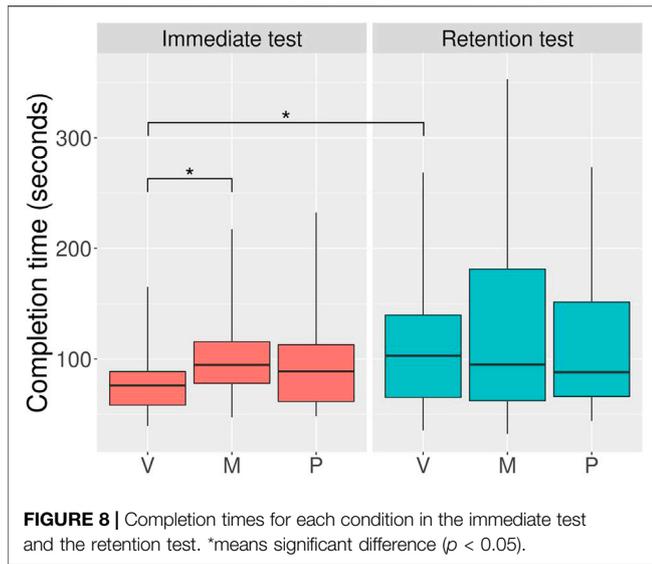
Although a Friedman test showed there was no overall statistically significant difference in training iterations, *Video* still has the smallest median of 1.07 iterations (IQR = 1.04–1.58), followed by *Mobile VR* and *PC VR* with 1.87 iterations (IQR = 1–2) and 1.97 iterations (IQR = 1–2) respectively.

### 5.3 Completion Times

**Figure 8** shows the completion times of the gearbox assembly in both the immediate test and the retention test. Non-parametric analysis was performed for completion times.

#### 5.3.1 Immediate Test

A Friedman test showed there was an overall statistically significant difference in completion times between the training



conditions in the immediate test ( $\chi^2(2) = 6.4, p = 0.041$ ). A post hoc test using Wilcoxon signed-rank tests with Bonferroni correction showed the significant difference between *Video* and *Mobile VR* ( $Z = -2.202, p = 0.027, r = 0.493$ ).

*Video* has the smallest median of 76.04 s (IQR = 58.36–88.69), followed by *PC VR* with a median of 88.9 s (IQR = 61.64–112.94) and *Mobile VR* has the largest median of 94.69 s (IQR = 78.08–115.6).

### 5.3.2 Retention Test

A Friedman test showed there was no overall statistically significant difference in completion times between the training conditions in the retention test.

*PC VR* has the smallest median of 88.04 s (IQR = 66.14–151.56), followed by *Mobile VR* with a median of 94.94 s (IQR = 62.3–181.39) and *Video* has the largest median of 103.05 s (IQR = 65.32–139.77).

### 5.3.3 Comparison Between Two Sessions

Wilcoxon signed-rank tests showed there was a significant difference in completion times of *Video* between the immediate test and the retention test ( $W = 42, Z = -2.352, p = 0.017, r = 0.526$ ) and no significant difference in completion times of the other two conditions between the sessions. The median value in the *Video* condition significantly increased by 35.53% in the retention test.

## 5.4 Assembly Errors

### 5.4.1 Qualification of Errors

All subjects were able to assemble the gearboxes in the immediate and retention tests. Some of the completed gearboxes had a few minor errors, which could be categorized into three types as follows:

- **Incorrect orientations of pieces:** the pieces were placed in the correct positions with the wrong orientations. This type of error does not affect how the gearboxes work. Each piece in this category is counted as one error. There is only one exception for

gearbox 1. It has three interconnected gears, which are treated as one for this category of error.

- **Misplaced pieces:** the pieces were placed in the incorrect positions. The pieces with this type of errors could still work together as gearboxes in ways which are a bit different from the ones of the original gearboxes. Each misplacement is counted as one error, even though it might involve multiple pieces. For instance, some participant swapped the positions of two gears, which is counted as one error instead of two errors.
- **Unconnected pieces:** the pieces were placed in the correct positions, but not connected to other pieces. The assembled gearboxes with this type of error could not work. However, it could be fixed easily. Each disconnection is counted as one error.

The maximum possible number of errors for a gearbox is the number of its total pieces (gearbox 1: 8, gearbox 2: 7, and gearbox 3: 8). **Figure 9** shows the numbers of assembly errors for each training condition in the both test sessions according to the qualification.

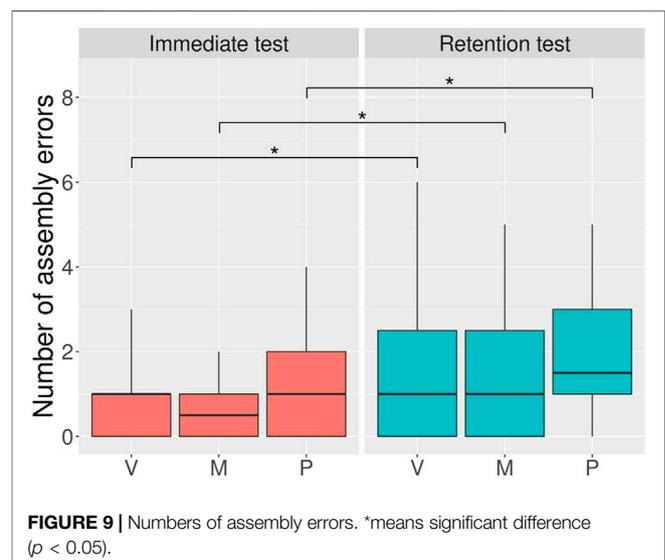
### 5.4.2 Comparison Within Each Session

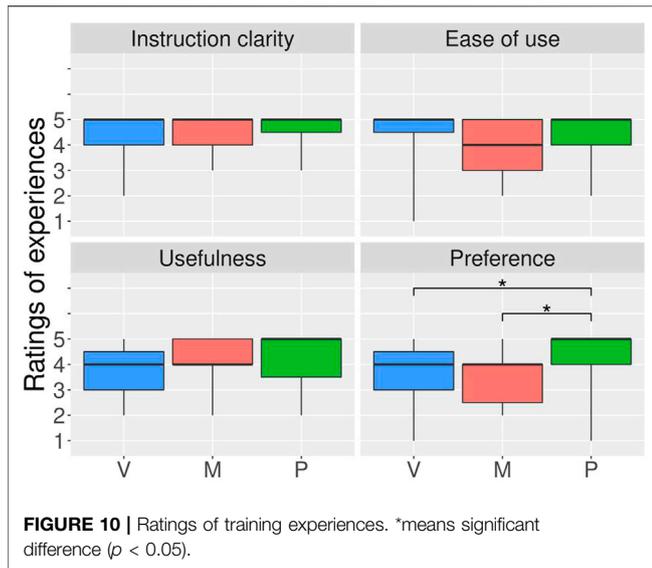
Non-parametric analysis was performed for numbers of errors. Within each session, a Friedman test showed there was no overall statistically significant difference in numbers of errors between the training conditions.

In the immediate test, *Mobile VR* has the smallest median of 0.5 (IQR = 0–1), followed by *Video* with a median of 1 (IQR = 0–1) and *PC VR* with 1 (IQR = 0–2). In the retention test, *Video* and *Mobile VR* have the same median value of 1 (IQR = 0–2.5) and *PC VR* has the largest median of 1.5 (IQR = 1–3).

### 5.4.3 Comparison Between Two Sessions

Wilcoxon signed-rank tests showed there was a significant difference in numbers of errors of each condition between the immediate test and the retention test (*Video*:  $W = 6, Z = -2.31, p = 0.021, r = 0.517$ ; *Mobile VR*:  $W = 25.5,$





$Z = -1.982, p = 0.049, r = 0.443$ ; and *PC VR*:  $W = 0, Z = -3.263, p < 0.001, r = 0.73$ ).

The numbers of errors in all experimental conditions increased significantly in the retention test. The median value increased by 0.5 in the *Mobile VR* and *PC VR* condition, while the one of the *Video* condition remained the same.

## 5.5 Questionnaire Ratings

Figure 10 shows the ratings of instruction clarity, ease of use, usefulness and preference from the experience questionnaire (see the rating of gearbox difficulty in Section 5.1). Non-parametric analysis was performed for ratings.

A Friedman test showed there was an overall statistically significant difference in rating of preference between the training conditions ( $\chi^2(2) = 11.03, p = 0.004$ ). A post hoc test using Wilcoxon signed-rank tests with Bonferroni correction showed the significant differences between *Video* and *PC VR* ( $Z = -2.607, p = 0.01, r = 0.582$ ) and between *Mobile VR* and *PC VR* ( $Z = -2.682, p = 0.005, r = 0.6$ ).

No overall statistically significant difference was found in the other three ratings between the training conditions.

## 5.6 Previous VR Experience

Based on their experiences of using VR, the participants were assigned to two groups: the group having 12 participants with less experience (0 or less than 5 h) and the group having 8 participants with more experience (more than 5 h). Within each group, the same statistical analysis was performed on the data of all measurements.

### 5.6.1 Training Times and Iterations (Grouped)

Within each group, a Friedman test showed there was an overall statistically significant difference in training times between the conditions (less experience:  $\chi^2(2) = 14, p < 0.001$ ; and more experience:  $\chi^2(2) = 13, p = 0.002$ ). A post hoc test using Wilcoxon signed-rank tests with Bonferroni correction showed

the significant differences between *Video* and *Mobile VR* (less experience:  $Z = -3.059, p < 0.001, r = 0.883$ ; and more experience:  $Z = -2.521, p = 0.008, r = 0.891$ ) and between *Video* and *PC VR* (less experience:  $Z = -2.667, p = 0.005, r = 0.770$ ; and more experience:  $Z = -2.521, p = 0.008, r = 0.891$ ).

Friedman tests showed there was no overall statistically significant difference in training iterations between the conditions for both groups.

### 5.6.2 Completion Times (Grouped)

Friedman tests showed there was no overall statistically significant difference in completion times between the conditions in the immediate test or the retention test for both groups.

Wilcoxon signed-rank tests showed a significant difference in completion times of *PC VR* between the immediate test and the retention test among the participants with less VR experience ( $W = 8, Z = -2.432, p = 0.012, r = 0.702$ , the median value increased by 28.07%) and a significant difference in completion times of *Video* between the two sessions among the participants with more VR experience ( $W = 0, Z = -2.521, p = 0.008, r = 0.891$ , the median value increased by 42.96%). There was no significant difference in completion times of the other conditions in the groups between the sessions.

### 5.6.3 Assembly Errors (Grouped)

Friedman tests showed there was no overall statistically significant difference in numbers of errors between the training conditions in the immediate test or the retention test for both groups.

Wilcoxon signed-rank tests showed there was a significant difference in numbers of errors of the *PC VR* condition between the immediate test and the retention test within the group having more VR experience ( $W = 0, Z = -2.369, p = 0.031, r = 0.838$ , the median value increased from 0 to 1). There was no significant difference in numbers of errors of the other conditions in the groups between the two sessions.

### 5.6.4 Questionnaire Ratings (Grouped)

Within the group having less VR experience, a Friedman test showed there was an overall statistically significant difference in rating of preference between the training conditions ( $\chi^2(2) = 8.048, p = 0.018$ ). A post hoc test showed the significant difference between *Mobile VR* and *PC VR* ( $Z = -1.970, p = 0.039, r = 0.569$ ).

Within the group having more VR experience, a Friedman test showed there was an overall statistically significant difference in rating of usefulness between the training conditions ( $\chi^2(2) = 7.583, p = 0.023$ ). A post hoc test showed the significant difference between *Video* and *PC VR* ( $Z = -2.369, p = 0.031, r = 0.838$ ).

No overall statistically significant difference was found in the other ratings between the conditions for both groups.

## 6 DISCUSSION

In summary, participants preferred virtual training with a higher level of immersion and believed they have learned more in those conditions (*H1* is true). In contrast, and

surprisingly, traditional virtual training with passive video playback had the best performance in terms of learning time and task completion time immediate after training ( $H2$  is false). However, video training was also the only virtual training condition that suffered a significant completion time increase in the retention session. We discuss these findings separately in the following subsections.

## 6.1 User Preference and Perception

Most participants appreciated the immersive experience provided by both HTC VIVE Pro and Google Cardboard. They agreed that training in an immersive environment was enjoyable during the interview. They also commented that the immersive training was memorable and easier to recall during the immediate test. Two participants even commented that they intentionally stayed longer in the virtual environment for the VR experience despite the fact that they had already learned the assembly procedure.

However, our hypothesis  $H1$ , that users would prefer training with a higher level of immersion, was only partially correct. Indeed, most participants (16 out of 20) favor virtual training with HTC VIVE Pro among all conditions. However, eight participants preferred virtual training with video to Google Cardboard. Among these eight participants, two commented that the gearbox assembly was simple [“I know gearboxes. So it (was) the knowledge (that) helped me. (I) can do this very well. I think it’s very easy.”] and watching video playbacks was sufficient [“When I was studying, what I learnt was to represent 3D objects in 2D. So it feels that (video training) was closer to my learning pattern and my way of understanding 3D. (What I liked the most was) video, for this reason.”] (they also rated virtual training with video higher than HTC VIVE Pro); while others commented that interacting with virtual objects using head movements in Google cardboard was “unnatural” and “uncomfortable”. It seems that the inconvenient interaction method of Google Cardboard interfered with the benefits afforded by the virtual immersive training environment. Besides, 5 of the 20 participants had experiences studying mechanical engineering. 4 of them preferred *Video* to *Mobile VR* and the other one rated *Video* as high as *Mobile VR* in terms of preference. Only one of them preferred *PC VR* to *Video*. On the other hand, most of the other participants preferred *PC VR* (12 out of 15) or *Mobile VR* (7 out of 15) to *Video*. It seems the participants with relevant education background tended to favor *Video* due to their domain knowledge and previous learning experiences while the others did not. Similarly, previous VR experiences might also affect user’s perception of virtual environments. In terms of rating, only the participants with more VR experience believed *PC VR* was significantly more helpful for training than *Video*.

## 6.2 How Immersive is Enough?

According to our data,  $H2$ , which hypothesizes that virtual training in VR would yield better performance (completion time), appears to be false. The *Video* condition had a significantly better performance than *Mobile VR*.  $H3$  is not

correct. The participants indeed preferred the training with bimanual 3D inputs, yet performance wise, there was no significant difference between *PC VR* condition with 3D inputs and *Video* condition with only passive video watching.

This resonates with the landmark paper from Bowman and McMahan (2007) which discussed several cases where a higher level of immersion did not result in better task performance. In particular, when tasks are simple or the visualization are less complex (Laha et al., 2012; Schuchardt and Bowman, 2007), less immersive systems might perform as well as the more immersive ones when considering the task of spatial understanding. Our result seems to exhibit the same phenomena for training of a different type of task. In addition, our study also fits the criterion well as 3D assembly demands spatial understanding and participants considered the assembly task relatively simple after the virtual training (see **Figure 6**). These seem to lead to a similar conclusion of previous works (Bowman and McMahan, 2007; Schuchardt and Bowman, 2007; Sowndararajan et al., 2008; Laha et al., 2012) in the context of learning transfer, i.e., for assembly tasks that are simple, less immersive virtual training systems might achieve similar level of learning transfer as the more immersive ones. This might help explain why *Video* led to better performance than *Mobile VR* in our experiment. The gearboxes were relatively simple, and the assembly steps were clear enough from the videos, especially for those with a mechanical engineering background. Most participants were probably more used to learning from videos, considering more than half of them had little experience using VR. The video training was sufficient and did not have the issue of inconvenient interaction methods as Google Cardboard, which eventually produced better results.

Previous works (Carlson et al., 2015; Murcia-Lopez and Steed, 2018) did not explicitly investigate virtual training with different levels of immersion. Instead, these works focused on the comparison between the effectiveness of virtual and physical training. In terms of training media, there are some similarities between our conditions (*Video* and *PC VR*) and those ( $PV_I$  using videos without assembly practice and  $V_{EA}$  using animations with virtual blocks for practice in PC-powered VR) in the work from Murcia-Lopez and Steed (2018), except the inclusion of paper instructions and the use of abstract 3D puzzles. Similarly, their work found no significant effect on the success rates and testing times among all puzzles except the most difficult one, where most participants who experienced only video-training failed to solve the puzzle in the given time. This result seems to imply the advantage of VR training over video one for the task with extremely high complexity e.g., 6-pieces 3D burr Puzzle. We believe it will be beneficial to examine if such finding still stands in other real-world training targets more complex than the gearboxes used in our experiment.

## 6.3 Immersive Virtual Training Might Help Recall

The *Video* condition is the only one having a significant difference in completion times between the immediate test

and the retention test. The descriptive statistics of task completion time in both tests are *median* = 76.04 and *median* = 103.05 respectively, an increase of 35.53% in the average completion time, while the median values in VR conditions remain almost the same (*Mobile VR*: from 94.69 to 94.94, and *PC VR*: from 88.9 to 88.04). This result shows that an immersive virtual training might be beneficial to the training retention. However, the participants' previous experiences of using VR might also affect the results. While the participants with more VR experience still had a significant increase of completion times in the *Video* condition between the two sessions (averagely 42.96%), the participants with less VR experience had a significant increase of completion times in the *PC VR* condition between the sessions (averagely 28.07%). It seems to suggest that lack of familiarity with training media could diminish the benefits of immersion for retention.

Compared to two previous works (Carlson et al., 2015; Murcia-Lopez and Steed, 2018) using a similar protocol, this result echos the finding of Carlson et al. (2015) where the performance of the VR-trained participants improved in a retention session 2 weeks after the training. The result from Murcia-Lopez and Steed (2018) is less conclusive because most participants failed to solve the puzzles in the retention session. Indeed, potential confounding factors, such as task complexity, learning times, and the amounts of time spent with the physical gearboxes during the immediate test, might have affected the participant's performance in the retention sessions. Further investigation is in dire need to understand how immersive virtual environments can be used as a tool to consolidate memory (Krokos et al., 2019).

## 6.4 Are Gearbox Assembly Tasks Difficult Enough?

Three functional gearboxes were used in our study to better represent a real-world assembly line task. Compared with the six-piece burr puzzle used in previous works (Carlson et al., 2015; Murcia-Lopez and Steed, 2018), these gearboxes were relatively simple and attracted the question of whether the training was effective or even necessary.

To answer this question, we recruited three untrained subjects for an informal study. For each gearbox, the participant was given two photos of the assembled physical gearbox from different angles. The participant could view the photos as long as they needed to memorize the internal structure and to guess how to assemble. When he/she was ready for the test, the photos were taken away and the 3D-printed pieces were given to start the assembly task.

The result (of 9 trials in total) showed that when compared with the trained subjects, these three untrained subjects took much longer to complete the assembly tasks (see **Table 4**) and committed more serious errors (see **Figure 11**). It indicated that the gearboxes were complex enough and the virtual training indeed increased the assembly performance of the participants.

## 6.5 Choice of Virtual Training

Despite the fact that our participants considered both VR conditions to be more engaging, the results of our study are in line with previous

**TABLE 4 |** The completion times of the untrained and trained subjects in the immediate test.

Conditions	Untrained	Video	Mobile VR	PC VR
Median	233.81	76.04	94.69	88.9

findings that a higher level of immersion does not always yield better performance (Adams et al., 2001; Hall and Horwitz, 2001; Sowndararajan et al., 2008; Gavish et al., 2015). The limited interaction capability provided by *Mobile VR* could undermine the benefits of the immersive experience and result in a poor training effectiveness. Whereas the traditional video-based training seems to be as effective as VR training for the assembly tasks with low-to-medium perceived difficulty, at least in the short-term.

To summarize, our study suggested that VR could provide more engaging learning experiences and prolonged training outcomes. While the traditional video training has lower costs and could be more time efficient during the training process **Figure 7**. The choice of training media depends on the difficulty of the task, the constraint on the budget, and the frequency of training. A hybrid training procedure that mixes the video-based and VR-based training could be worth investigating.

## 7 LIMITATION

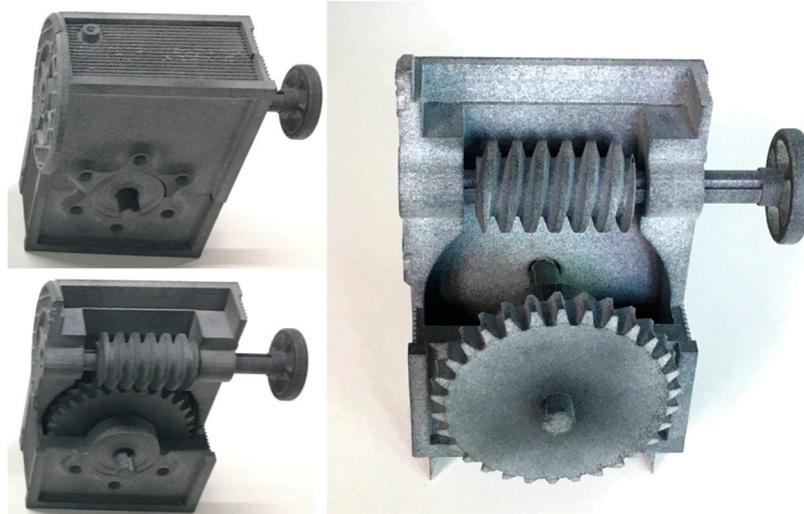
The within-subject design allows better control of each participant's background knowledge. However, the repeated exposures to gearbox assembly tasks inevitably incur learning effects. We believe the use of three different gearboxes consisting of different mechanisms should have mitigated the learning effect. Our result also concurs with previous works using between-subject designs (Carlson et al., 2015; Murcia-Lopez and Steed, 2018). Still, it is worth noting the limitation that it is difficult to disentangle the transfer of knowledge across conditions in the retention session.

VR conditions in our study lack physics constraints, just like the other assembly task studies (Carlson et al., 2015; Murcia-Lopez and Steed, 2018). The subjects might not get the correct perception of the motor demands for the physical assembly tasks in the testing phase.

The subjects took the immediate test right after the training phase. There was no distractor task (Carlson et al., 2015) or short break (Murcia-Lopez and Steed, 2018) between training and testing. The training performance might be positively affected by the recency effect.

The two tests only required the participants to complete the tasks once. From our observation, the subjects were sometimes stuck at some step due to motor demands even though they were aware of how to assemble it. The completion times might be influenced by these random events during the testing. To reduce the effect, subjects could be required to complete the tasks for multiple times and duplicate measurements need to be averaged.

However, participants might prefer VR conditions (*H1*) and consider it more helpful due to the novelty effect (Radu and Schneider, 2019). Thus, there is a definite need for a longitudinal study to better understand the effectiveness of VR or AR training.



**FIGURE 11** | The gearbox assembled by an untrained subject (**right**) and the correct example (**left**).

## 8 CONCLUSION AND FUTURE WORK

We have presented a study that compares the effectiveness of three different virtual training methods for bimanual gearbox assembly tasks. Each virtual training method represented a different level of immersion. The results showed that participants favored virtual training with a higher level of immersion and presumed that the immersive virtual experience helps the recall of assembly procedure. The performance of the *Video* condition was surprisingly good. The completion times of its participants in the immediate test were significantly better than the *Mobile VR* condition and had no significant difference against the *PC VR* condition, except that its performance during the retention test exhibited a significant drop and is worth further investigation. We believe that these results are important as it provides insights into how levels of immersion might affect training transfer of bimanual assembly tasks.

The directions of future studies could be evaluating how various factors affect virtual training with multiple levels of immersion, including physics constraints, domain knowledge, VR experiences and task complexity. Another potential future research could be exploring the design of a hybrid virtual environment that integrates multiple training media to utilize the benefits of different levels of immersion. Since tasks or trainees might require or prefer various immersion levels, another interesting direction of future work could be creating an authoring tool to effectively and easily generate training contents with multiple formats that could be used in different virtual environments.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and

accession number(s) can be found in the article/**Supplementary Material**.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of Technology Sydney Human Research Ethics Committee. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individuals for the publication of any potentially identifiable images or data included in this article.

## AUTHOR CONTRIBUTIONS

All authors contributed to this paper. SS, H-TC, and TL designed the study. SS performed the experiment and analyzed the data. SS and H-TC wrote the draft of the manuscript. WR and TL reviewed and edited the manuscript.

## FUNDING

SS was a PhD candidate at University of Technology Sydney. This research was supported by an Australian Government Research Training Program Scholarship and UTS Doctoral Scholarships.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/frvir.2021.597487/full#supplementary-material>

## REFERENCES

- Adams, R. J., Klowden, D., and Hannaford, B. (2001). Virtual Training for a Manual Assembly Task. *Haptics-e: Electron. J. Haptics Res.* 2, 1–7.
- Bailenson, J., Patel, K., Nielsen, A., Bajcsy, R., Jung, S.-H., and Kurillo, G. (2008). The Effect of Interactivity on Learning Physical Actions in Virtual Reality. *Media Psychol.* 11, 354–376. doi:10.1080/15213260802285214
- Berg, L. P., and Vance, J. M. (2017). Industry Use of Virtual Reality in Product Design and Manufacturing: a Survey. *Virtual Reality* 21, 1–17. doi:10.1007/s10055-016-0293-9
- Bhagat, K. K., Liou, W.-K., and Chang, C.-Y. (2016). A Cost-Effective Interactive 3d Virtual Reality System Applied to Military Live Firing Training. *Virtual Reality* 20, 127–140. doi:10.1007/s10055-016-0284-x
- Boud, A. C., Haniff, D. J., Baber, C., and Steiner, S. J. (1999). “Virtual Reality and Augmented Reality as a Training Tool for Assembly Tasks,” in 1999 *IEEE International Conference on Information Visualization (Cat. No. PR00210)*, 32–36. doi:10.1109/IV.1999.781532
- Bowman, D. A., and McMahan, R. P. (2007). Virtual Reality: How Much Immersion Is Enough?. *Computer* 40, 36–43. doi:10.1109/MC.2007.257
- Carlson, P., Peters, A., Gilbert, S. B., Vance, J. M., and Luse, A. (2015). Virtual Training: Learning Transfer of Assembly Tasks. *IEEE Trans. Vis. Comput. Graphics* 21, 770–782. doi:10.1109/TVCG.2015.2393871
- Dalgarno, B., and Lee, M. J. W. (2010). What Are the Learning Affordances of 3-D Virtual Environments?. *Br. J. Educ. Tech.* 41, 10–32. doi:10.1111/j.1467-8535.2009.01038.x
- de Moura, D. Y., and Sadagic, A. (2019). “The Effects of Stereopsis and Immersion on Bimanual Assembly Tasks in a Virtual Reality System,” in 2019 *IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (IEEE), 286–294. doi:10.1109/VR.2019.8798112
- Funk, M., Bächler, A., Bächler, L., Kosch, T., Heidenreich, T., and Schmidt, A. (2017). “Working with Augmented Reality?,” in *Proceedings of the 10th International Conference on Pervasive Technologies Related to Assistive Environments* (New York, NY, USA: ACM, PETRA), 17, 222–229. doi:10.1145/3056540.3056548
- Gavish, N., Gutiérrez, T., Weibel, S., Rodríguez, J., Peveri, M., Bockholt, U., et al. (2015). Evaluating Virtual Reality and Augmented Reality Training for Industrial Maintenance and Assembly Tasks. *Interactive Learn. Environments* 23, 778–798. doi:10.1080/10494820.2013.815221
- Gerbaud, S., Mollet, N., Ganier, F., Arnaldi, B., and Tisseau, J. (2008). “GVT: a Platform to Create Virtual Environments for Procedural Training,” in 2008 *IEEE Virtual Reality Conference* (IEEE), 225–232. doi:10.1109/VR.2008.4480778
- Gonzalez-Franco, M., Pizarro, R., Cermeron, J., Li, K., Thorn, J., Hutabarat, W., et al. (2017). Immersive Mixed Reality for Manufacturing Training. *Front. Robot. AI* 4, 1–8. doi:10.3389/frobot.2017.00003
- Gorecky, D., Khamis, M., and Mura, K. (2017). Introduction and Establishment of Virtual Training in the Factory of the Future. *Int. J. Comput. Integrated Manufacturing* 30, 1–9. doi:10.1080/0951192X.2015.1067918
- Hall, C. R., and Horwitz, C. D. (2001). Virtual Reality for Training: Evaluating Retention of Procedural Knowledge. *Jivr* 5, 61–70. doi:10.20870/IJVR.2001.5.1.2669
- Hoedt, S., Claeys, A., Van Landeghem, H., and Cottyn, J. (2017). The Evaluation of an Elementary Virtual Training System for Manual Assembly. *Int. J. Prod. Res.* 55, 7496–7508. doi:10.1080/00207543.2017.1374572
- Jiang, M. (2011). “Virtual Reality Boosting Automotive Development,” in *Virtual Reality & Augmented Reality in Industry*. Editors D. Ma, X. Fan, J. Gausemeier, and M. Grafe (Berlin, Heidelberg: Springer Berlin Heidelberg), 171–180. doi:10.1007/978-3-642-17376-9\_11
- Krokos, E., Plaisant, C., and Varshney, A. (2019). Virtual Memory Palaces: Immersion Aids Recall. *Virtual Reality* 23, 1–15. doi:10.1007/s10055-018-0346-3
- Laha, B., Sensharma, K., Schiffbauer, J. D., and Bowman, D. A. (2012). Effects of Immersion on Visual Analysis of Volume Data. *IEEE Trans. Vis. Comput. Graphics* 18, 597–606. doi:10.1109/TVCG.2012.42
- Leu, M. C., ElMaraghy, H. A., Nee, A. Y. C., Ong, S. K., Lanzetta, M., Putz, M., et al. (2013). CAD Model Based Virtual Assembly Simulation, Planning and Training. *CIRP Ann.* 62, 799–822. doi:10.1016/j.cirp.2013.05.005
- Mollet, N., and Arnaldi, B. (2006). “Storytelling in Virtual Reality for Training,” in *Technologies for E-Learning and Digital Entertainment*. Editors Z. Pan, R. Aylett, H. Diener, X. Jin, S. Göbel, and L. Li (Berlin, Heidelberg: Springer Berlin Heidelberg), 334–347. doi:10.1007/11736639\_45
- Mollet, N., Gerbaud, S., and Arnaldi, B. (2007). “STORM: a Generic Interaction and Behavioral Model for 3D Objects and Humanoids in a Virtual Environment,” in *Eurographics Symposium on Virtual Environments, Short Papers and Posters*. The Eurographics Association. Editors B. Froehlich, R. Blach, and R. van Liere doi:10.2312/PE/VE2007Short/095-100
- Murcia-Lopez, M., and Steed, A. (2018). A Comparison of Virtual and Physical Training Transfer of Bimanual Assembly Tasks. *IEEE Trans. Vis. Comput. Graphics* 24, 1574–1583. doi:10.1109/TVCG.2018.2793638
- Oren, M., Carlson, P., Gilbert, S., and Vance, J. M. (2012). “Puzzle Assembly Training: Real World vs. Virtual Environment,” in 2012 *IEEE Virtual Reality (VR)* (IEEE), 27–30. doi:10.1109/VR.2012.6180873
- Papachristos, N. M., Vrellis, I., and Mikropoulos, T. A. (2017). “A Comparison between Oculus Rift and a Low-Cost Smartphone VR Headset: Immersive User Experience and Learning,” in 2017 *IEEE 17th International Conference on Advanced Learning Technologies (ICALT)* (IEEE), 477–481. doi:10.1109/ICALT.2017.145
- Radu, L., and Schneider, B. (2019). “What Can We Learn from Augmented Reality (Ar)?,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA: ACM, CHI '19), 544, 1–544. doi:10.1145/3290605.3300774.12
- Ritter, F., Strothotte, T., Deussen, O., and Preim, B. (2001). Virtual 3D Puzzles: a New Method for Exploring Geometric Models in VR. *IEEE Comput. Graphics Appl.* 21, 11–13. doi:10.1109/38.946625
- Rodríguez, J., Gutiérrez, T., Sánchez, E. J., Casado, S., and Aguinaga, I. (2012). “Training of Procedural Tasks through the Use of Virtual Reality and Direct Aids,” in *Virtual Reality and Environments*. Editor C. S. Lanyi (Rijeka: IntechOpen), chap. 3. doi:10.5772/36650
- Ruthenbeck, G. S., and Reynolds, K. J. (2015). Virtual Reality for Medical Training: the State-Of-The-Art. *J. Simulation* 9, 16–26. doi:10.1057/jos.2014.14
- Schuchardt, P., and Bowman, D. A. (2007). “The Benefits of Immersion for Spatial Understanding of Complex Underground Cave Systems,” in *Proceedings of the 2007 ACM Symposium on Virtual Reality Software and Technology - VRST '07* (New York, New York, USA: ACM Press), 1, 121. doi:10.1145/1315184.1315205
- Shuralyov, D., and Stuerzlinger, W. (2011). “A 3D Desktop Puzzle Assembly System,” in 2011 *IEEE Symposium on 3D User Interfaces (3DUI)* (IEEE), 139–140. doi:10.1109/3DUI.2011.5759244
- Sowndararajan, A., Wang, R., and Bowman, D. A. (2008). “Quantifying the Benefits of Immersion for Procedural Training,” in *Proceedings of the 2008 Workshop on Immersive Projection technologies/Emerging Display Technologies - IPT/EDT '08* (New York, New York, USA: ACM Press), 1. doi:10.1145/1394669.1394672
- Stanney, K. M., Mourant, R. R., and Kennedy, R. S. (1998). Human Factors Issues in Virtual Environments: A Review of the Literature. *Presence* 7, 327–351. doi:10.1162/105474698565767
- Wang, X., Ong, S. K., and Nee, A. Y. C. (2016). A Comprehensive Survey of Augmented Reality Assembly Research. *Adv. Manuf.* 4, 1–22. doi:10.1007/s40436-015-0131-4
- Young, M. K., Gaylor, G. B., Andrus, S. M., and Bodenheimer, B. (2014). “A Comparison of Two Cost-Differentiated Virtual Reality Systems for Perception and Action Tasks,” in *Proceedings of the ACM Symposium on Applied Perception - SAP '14* (New York, New York, USA: ACM Press), 83–90. doi:10.1145/2628257.2628261
- Yuviler-Gavish, N., Yechiam, E., and Kallai, A. (2011). Learning in Multimodal Training: Visual Guidance Can Be Both Appealing and Disadvantageous in Spatial Tasks. *Int. J. Human-Computer Stud.* 69, 113–122. doi:10.1016/j.ijhcs.2010.11.005

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Shen, Chen, Raffaele and Leong. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.