



**HAL**  
open science

## Investigating Whether the Mass of a Tool Replica Influences Virtual Training Learning Outcomes

Julien Cauquis, Etienne Peillard, Lionel Dominjon, Thierry Duval, Guillaume Moreau

► **To cite this version:**

Julien Cauquis, Etienne Peillard, Lionel Dominjon, Thierry Duval, Guillaume Moreau. Investigating Whether the Mass of a Tool Replica Influences Virtual Training Learning Outcomes. Transactions on Visualization & Computer Graphics (TVCG), inPress. hal-04427994

**HAL Id: hal-04427994**





**<https://hal.science/hal-04427994>**

Submitted on 31 Jan 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Investigating Whether the Mass of a Tool Replica Influences Virtual Training Learning Outcomes

Julien Cauquis , Etienne Peillard , Lionel Dominjon, Thierry Duval , and Guillaume Moreau 

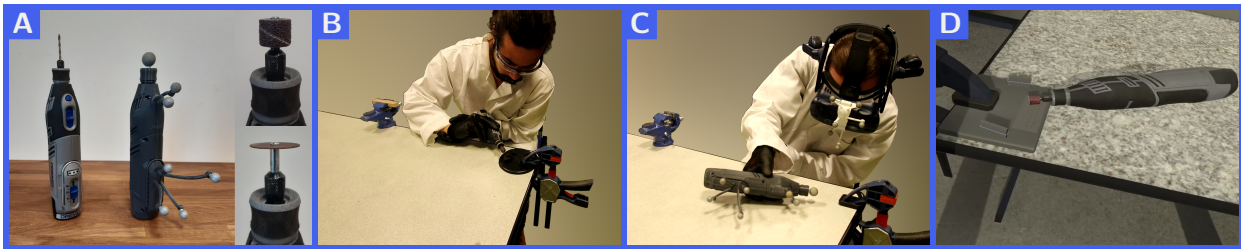


Fig. 1: **A:** The rotary tool (with the three tool bits used within the learning experience) and its replica with optical markers. **B:** A participant sanding in real conditions (pre and post-training). **C:** A participant during the immersive training phase. **D:** first-person view of the sanding task in virtual reality.

**Abstract**—Virtual Reality (VR) has emerged as a promising solution to address the pressing concern of transferring know-how in the manufacturing industry. Making an immersive training experience often involves designing an instrumented replica of a tool whose use is to be learned through virtual training. The process of making a replica can alter its mass, making it different from that of the original tool. As far as we know, the influence of this difference on learning outcomes has never been evaluated. To investigate this subject, an immersive training experience was designed with pre and post-training phases under real conditions, dedicated to learning the use of a rotary tool. 80 participants took part in this study, split into three groups: a control group performing the virtual training using a replica with the same mass as the original tool ( $m = 100\%$ ), a second group that used a replica with a lighter mass than the original tool ( $m = 50\%$ ), and a third group using a replica heavier than the original tool ( $m = 150\%$ ). Despite variations in the mass of the replica used for training, this study revealed that the learning outcomes remained comparable across all groups, while also demonstrating significant enhancements in certain performance measures, including task completion time. Overall, these findings provide useful insights regarding the design of tool replicas for immersive training.

**Index Terms**—Virtual Reality, Virtual Training, Prop Design, Weight Perception, User Study

## 1 INTRODUCTION

Virtual Reality (VR) offers endless possibilities to researchers and creators regarding the transfer of knowledge and experiences. Among the use cases this medium offers, immersive training appears to be one of interest. VR allows people to train in safe conditions and fully controlled environments, with infinite trials and repetitions. Furthermore, it can reduce resource consumption and can allow the creation of metrics that are not achievable in real conditions, such as the measure of the paint thickness in a vehicle painting simulator [38]. When learning the use of a tool thanks to virtual training, it is common to rely on an instrumented replica instead of the VR headset default controllers [22, 50]. Such a practice seeks to reinforce the realism of virtual training by augmenting sensory perceptions, which is said to lead to better knowledge retention regarding immersive learning [35]. Using a replica is an adequate solution when the original tool can not be instrumented, due to its fragility, cost of mobilization, or difficulty of acquisition. With the growth of 3D printing technologies, it is now quite easy to make such a replica that is cheaper and easier to instrument than the original tool. Guidelines regarding the design of such replicas are still in an early research stage and need to be explored considering the expansion of immersive training in VR. Instru-

menting a replica generally implies the addition of a tracking solution and electrical components, to allow this replica to be used instead of a controller within a VR application. Some sensors can also be added, to extract data concerning the replica or the trainee. This, coupled with differences between the materials used in the replica manufacturing process and those composing the original tool, can result in a different weighted replica. To the best of our knowledge, the influence of such discrepancies on the learning outcomes in a virtual training context is still unexplored.

Thus, a user study ( $n = 80$ ) is proposed to investigate how the mass of a replica used within virtual training may influence final learning outcomes. An immersive training application was designed, where learners had to learn how to perform three distinct tasks thanks to the use of a rotary tool. This training is composed of three phases: a pre-training phase in real conditions, a virtual training one, and a third post-training phase again in real conditions. Performances were assessed with both objective and subjective metrics. To deepen the analysis, the usability of the system, the perceived learning outcomes, and the cognitive load induced by the mass of a replica were also evaluated. Results from this study delivered guidelines that can be used to facilitate the design of tool replicas intended for use within immersive training applications. The study showed that the learning outcomes remained consistent regardless of which mass was used for the tool replicas. This suggests that tool replica mass did not have a noticeable impact on learning effectiveness, reinforcing the potential for flexibility and customization in immersive training programs.

The remainder of this paper is structured as follows: Sec. 2 describes related work. Sec. 3 presents the user study, while Sec. 4 presents the results. The discussion, limitations and future work, and conclusion are found in Sec. 5, Sec. 6, and Sec. 7 respectively.

- Julien Cauquis and Lionel Dominjon are with CLARTE.  
E-mail: [firstname.lastname@clarite-lab.fr](mailto:firstname.lastname@clarite-lab.fr)
- Julien Cauquis, Etienne Peillard, Thierry Duval, and Guillaume Moreau are with IMT Atlantique, Lab-STICC, UMR CNRS 6285.  
E-mail: [firstname.lastname@imt-atlantique.fr](mailto:firstname.lastname@imt-atlantique.fr)

Manuscript received xx xxx. 201x; accepted  
xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx.  
201x. For information on obtaining reprints of this article, please send e-mail  
to: [reprints@ieee.org](mailto:reprints@ieee.org). Digital Object Identifier: xx.xxx/TVCG.201x.xxxxxxx

## 2 RELATED WORK

### 2.1 Training in Virtual Reality

Immersive learning which “conceptualizes education as a set of active phenomenological experiences that are based on presence” is said to revolutionize the way one learns and experiences things [40]. It is now easier to learn procedural knowledge thanks to immersive training in VR and as a result, virtual training is a field in expansion. Training in VR yields several advantages. Thanks to this medium, it is possible to recreate high-fidelity environments [33] and train without putting the learner at risk. VR offers an effective way to minimize resource consumption, facilitate the repetitive practice of specific tasks or scenarios, and simulate situations or metrics that are not achievable in real conditions [39]. Moreover, VR can also offer a plurality of points of view, allowing one or more people to see things from different spatial and temporal perspectives. If learning theories support the fact that engagement and embodiment are vectors facilitating the retention of procedural knowledge and are emphasized by virtual applications [35], those do not always produce better learning performances [57] nor better knowledge retention than the one obtained through traditional media [25]. Anyway, VR usually does not degrade learning outcomes or performances compared to traditional media [45]. Indeed, some user studies comparing the benefits of VR to other media in a learning context did not reveal differences in learning outcomes. However, VR could be preferred by participants in terms of user experience [55]. Thus VR training advantages are numerous enough to justify its use among professionals to facilitate learning. Hence, it has a wide range of applications in a plethora of sectors [64] such as healthcare [45], education [28], military training [32], industry [46], social sciences [62], and others [12]. The application of VR could extend to acquiring knowledge about the operation of a specific tool.

### 2.2 Tool Learning thanks to Virtual Reality

When a tool is represented in an immersive training application designed for learning purposes, it is typically accompanied by a tangible counterpart that allows learners to manipulate it virtually. Design considerations regarding this tangible artifact depend on the specific needs of the learning application, such as cost or expected realism for instance. A dedicated haptic interface can be used, such as a force feedback robotic arm [51]. A simpler and thus more common approach consists of using directly the controllers of the VR headset and mapping a virtual tool to them, instead of using a replica. Tools like screwdrivers, fire hoses, enclosure tape, laser distance meters, drills, or cutter knives can thus be represented by their 3D representations [8], or even a 2D picture of them [53]. These visual-only solutions, supported by generic controllers, are easy to implement and inexpensive. However, they consider neither the affordances nor the properties of the tools. To reduce these discrepancies, a proposed solution is to augment the handheld controllers, adding mechanical parts and accessories to them, to modify their geometry or even their interaction modes. One can replace a button with a trigger [39], or add a few manufactured parts to one or two controllers to make them similar in shape to a paint sprayer [38].

Using a prop requires a tracking solution. A controller can be employed to track a physical prop within a virtual environment instead of a dedicated tracking system. In this case, the controller is affixed to the prop and solely utilized for its tracking functionality. Its interaction features (i.e., buttons, pads, and joysticks) are disregarded, and only the tracking capabilities are utilized [24]. Adding a controller to a prop often results in a difference in geometry between the original tool and its replica. Other discrepancies may be present regarding the original tool properties, such as its mass. This is the case when a tool smaller and lighter than a controller is equipped with one, such as a tooth drill stylus [60]. When employing a dedicated tracking system, optical and inertial trackers are the most widely used ones. Using such devices, it is possible to use a copy of the original tool as a replica, only adding the markers necessary for its tracking [2, 9, 58]. In this case, there is no need for the controllers. If a version of the original tool is unavailable, then an alternative replica can be made with a more or less advanced degree of realism [50]. However, when creating such replicas, the original properties of the tool and especially its mass are once again often not fully replicated.

The design of such tools dedicated to learning through virtual training is still emerging. There is an increasing number of studies proposing guidelines for haptic replicas, that can serve as a basis regarding tool design for learning thanks to virtual training.

### 2.3 Designing Props

Tool design is of interest to ergonomists, who suggest indirect design guidelines regarding the mass of tools, such as grasping affordances, grip strength, muscular fatigue, or the torque exerted by the tool on the wrist. Recommendations about mass among scholars are scarce and mainly subjective: “Tool weight should be determined according to the circumstances of the job” [61]. An equivalent literature specifically dedicated to tool design for immersive virtual training applications does not exist yet, only a few articles deal with this very topic. Li et al. introduced a design space for such tools, before proposing a framework to design haptic proxies with equivalent kinematics and movement patterns as the targeted tool [30]. They left aside the question of the mass of the replica, only minimizing it for the props they built. Sauter et al. investigated whether the tool representation within an immersive VR application should be 2D or 3D [53]. They used mere controllers as interfaces. They raised the question of the tool complexity which might induce differences in design for a VR training application.

Guidelines for designing tangible artifacts or interfaces for VR have recently emerged in the literature. Those props should fulfill a few criteria: complete co-location, compelling contact forces, and sufficient similarity [43]. Sufficient similarity gathers equivalences regarding both material and geometric properties, including mass. It is described in the literature under the term realism or fidelity. Realism can be either subjective or objective, as it refers to how the user perceives a simulation as real, and how close to reality this simulation is. Fidelity is objective and defined as how a system can replicate a real-world experience [19]. It has been shown that high-fidelity passive haptic feedback could improve task performance during virtual training [18]. However, in some cases, it is not possible to have a prop that weighs the same as the object it replaces. Thus, to increase realism, some studies proposed solutions to simulate masses for tangible interfaces. Indeed, several software-based and device-based simulation methods make it possible to render weight and thus object masses, as reviewed by Ye [65]. On one hand, software-based solutions mostly rely on sensory substitution and visual feedback such as pseudo-haptic [34]. On the other hand, device-based solutions are mostly divided into two types: those that simulate the static weight of the object and those that simulate its inertia [65]. For instance, Drag:on is a device said to render efficiently the weight of different original tools [66]. Another approach is to simulate the original tool properties by assembling several components of a given toolkit [17]. However, in this case, the mass of the replica is neglected to favor other properties of the tool such as its affordances.

When taken into consideration, the mass of the replica is often designed to be close to the mass of the original tool or chosen based on expert recommendations [50]. We questioned this realism-based approach, as some research revealed that the addition of sensory modalities on one hand decreased fidelity but on the other hand did not decrease knowledge transfer [10]. Regarding user experience, it has also been shown that a decrease in fidelity could be preferred over a realistic prop. In a user study comparing several props used as replicas of an object, including props with same and lighter masses compared to the original, participants preferred the lighter one [56].

### 2.4 Research Questions and Hypotheses

In the reviewed studies, tangible interfaces were not evaluated in a learning context, nor with different masses representing the same tool. Based on these observations, our research question is: does the mass of a replica influence the learning outcomes of virtual training? As it is commonly assessed and recommended that a lighter tool gives more precision and control on one’s movement [20], we made the following hypothesis:

- H1: The higher the mass of the replica, the lower the learning outcomes.

We also made some hypotheses regarding how this mass variation may affect user experience. Cognitive load refers to “any demands on

Table 1: Gender, Age, and Body Mass Index of Participants.

Tool	N	Gender	Age	Body Mass Index
Light	26	10 female	19-66, M=30, SD=12.5	M=25.1, SD=7.2
Normal	26	10 female	19-61, M=30, SD=13.2	M=24.1, SD=4.8
Heavy	28	10 female	20-52, M=29, SD=9.6	M=24.2, SD=3.3

Table 2: Participants' Familiarity with Virtual Reality, Video Games, and Do It Yourself. These means are obtained from five points Likert scales, "1: I have never used/played/practiced it" and "5: I use/play/practice it for a living".

Tool	N	VR Familiarity	Video Games Familiarity	DIY Familiarity
Light	26	M=2.1, SD=1.1	M=3.2, SD=0.9	M=2.7, SD=1.3
Normal	26	M=2.5, SD=1.2	M=3.2, SD=0.7	M=2.7, SD=0.8
Heavy	28	M=1.9, SD=1.2	M=3.1, SD=0.9	M=2.6, SD=1.0

working memory storage and processing of information" [54]. It is admitted that a high extraneous cognitive load has a negative impact on learning [59]. The physical learning environment, which refers to the physical properties of the learning environment (including tools used), is said to be a causal factor to cognitive load, potentially increasing extraneous load [7]. Therefore, we believe the mass of a tool could have an impact on cognitive load. In addition, this load could be calculated taking into account *physical demand* and *physical effort* factors [42], increasing with them. We thus formulate the hypothesis:

- H2: A higher mass of the replica implies a higher trainee cognitive load.

Moreover, we believe using a VR training application to learn can harm learning outcomes if the training system is judged as unusable, as it is the case for online learning environments [36]. We thus made an assumption regarding the usability of the developed training system:

- H3: A higher mass of the replica implies lower system usability.

Lastly, as we were in a learning context, we decided to investigate the perceived learning outcomes of the participants. In line with our other hypotheses, we proposed this assumption:

- H4: The higher the mass of the replica, the lower the trainee's perceived learning.

To test these hypotheses, an immersive training application dedicated to learning the usage of a tool was designed with several replicas of this tool with different masses.

### 3 USER STUDY

A user study was conducted to investigate the influence of the mass of a replica on learning outcomes in an immersive training context. To do so, a virtual training experience dedicated to learning the usage of a rotary tool was built with pre and post-training phases occurring under real conditions.

#### 3.1 Participants

An a priori sample size estimation performed with G\*power [16] recommended at least 22 participants per group. This number was exceeded, to ensure this minimal sample size after data cleaning. 80 participants voluntarily participated in the study. They were naive to the purpose of the experiment and had a normal or corrected-to-normal vision. The study was approved by the Institutional Advisory Board for the Protection of Human Subjects in Research at IMT Atlantique. All participants signed a consent form before participating in the study. They were pseudo-randomly assigned to obtain homogeneous groups in terms of gender. Tab. 1 and Tab. 2 present demographic data among the three groups.

#### 3.2 Experience Design

**Tasks.** To widen the application of the research, three tasks feasible with the same cordless rotary tool were selected, inspired by industrial applications: cutting, drilling, and sanding. The tool used was a Dremel 8220 (see Fig. 1 A), which meets several criteria: it can be handled with one hand, its mass ( $m=624g$ ) does not vary during its utilization, and

is greater than the one recommended by ergonomic design guidelines for tools. Indeed, East suggests a mass lighter than 500g for precision tools operated with one hand [15], while it should be lighter than 400g for the Canadian Center for Occupational Health and Safety [20].

**Design.** A between-subjects design was used, with three groups for three different masses of replica: a light one (Light L,  $m=312g$ , 50%), one with a normal mass (Normal N,  $m=624g$ , 100%), and a third one with a heavier mass (Heavy H,  $m=936g$ , 150%). Those masses were chosen in accordance with just noticeable difference (JND) regarding mass perception in VR as reviewed in [31]. A neutral condition, consisting solely of physical reality training, was not included since previous research has already delved into the efficiency of virtual training, as explained in Sec. 2. Each participant followed a three-phase procedure, with a pre-training phase using the tool in real conditions, then a virtual training phase using a replica, and a post-training phase again with the tool in real conditions. This design is inspired by [10]. To help reduce the length of the study, the number of trial repetitions during the training phase was limited to six.

#### 3.3 Dependent Variables

Subjective and objective variables were used to investigate the hypotheses. As the study was held in Brest, France, questionnaires were presented to the participants in French and English for non-French speakers. They were translated when no translation was available and pronoun adaptations were made to make them more inclusive when necessary. All questionnaires used within this study are provided as supplemental material.

##### 3.3.1 Performance Metrics

H1 was evaluated through objective and subjective metrics. Objective pre and post-training metrics included task completion time (TCT), and task-related metrics, derived from ISO Geometrical Product Specifications and Verification. Cutting was evaluated by comparing a cut profile photograph with a reference one. Photographs were converted to a binary format before being masked with a profile reference. This enabled a comparison of pictures on a pixel-count basis. All photographs were taken on the same day by the same person, with the camera and mechanical parts being in fixed positions. Drilling was evaluated with the resulting hole angle, the vertical axis of the tracking system used as a reference. A tracked stick with a diameter equal to the drill bits was used to make measurements. An objective assessment of the sanding task was impossible due to material properties and surface aspects. Surfaces as sanded by the participants were too coarse to make measurements with a roughness meter.

After all participant attempts, subjective evaluations were carried out to assess whether there was an improvement (i.e. something learned) between the pre-training and post-training phases in real conditions. Each pair of parts machined by participants was randomly and anonymously evaluated, on the same day by the same person. As it is a common practice in industrial environments with limited human resources and few dedicated trainers, only one rater was used. These evaluations resulted in binary scores (0: no improvement, 1: improvement). As holes drilled by participants were not visible in their wholeness, it was not possible to subjectively assess drilling improvement. Only the cutting and sanding tasks were subjectively evaluated.

Immersive training tasks were evaluated with objective metrics only: TCT and tasks-related ones, again derived from metrology standards. Cutting was evaluated thanks to the cutting profile mean height. Drilling was evaluated with three metrics: the angle difference between the drilled hole axis and the vertical axis of the virtual scene, the difference between the tool radius and the hole radius, and the hole center distance from the ideal hole location as indicated to participants by visual cues. Sanding was evaluated with the surface average roughness ( $S_a$ ) and the surface mean square roughness ( $S_q$ ), as defined by ISO 25178-2:2021 standards.

##### 3.3.2 Cognitive Load

Regarding H2, cognitive load was measured subjectively and objectively. The subjective assessment was made thanks to the NASA-RTLX questionnaire [6, 42].

Cognitive load was objectively assessed thanks to pupil diameter which increases under cognitive load [21, 41]. The Index of Pupillary



Activity (IPA) [14] and the Low/High Index of Pupillary Activity (LHIPA) [13] were used as metrics, as the authors made available their calculation processes. Pupil diameter was only measured during the immersive training phase, as the luminosity of the virtual environment was fully controlled. A wearable eye-tracking system was not used during the pre and post-training phases due to the need for security glasses. This decision was made on the basis that it would be inconvenient for participants who were already wearing prescription eyeglasses.

### 3.3.3 System Usability Scale

To investigate H3, the System Usability Scale (SUS) questionnaire [5] was utilized, as it is a widely used and validated questionnaire [29].

### 3.3.4 Cognitive Affective Psychomotor Learning

Following Bloom’s taxonomy [1], learning can be characterized in terms of Cognitive, Affective, and behavioral or Psychomotor (CAP) learning. These three domains are often evaluated individually [48], and in most cases in an education-related context. To our knowledge, there is no perceived learning questionnaire dedicated to learning thanks to immersive training. Thus we adapted the CAP Perceived Learning questionnaire from [49]. This short 9-item questionnaire evaluating the three said scales in terms of perceived learning was used to investigate H4. This adapted questionnaire is provided as supplemental material.

## 3.4 Procedure

According to health standards, all equipment was disinfected before every participant’s participation. The experience room was ventilated in consequence and was kept at a constant luminosity. Participants read an information leaflet before signing a consent form. Then, they were asked to wear safety gloves and glasses. As non-mandatory equipment, safety shoes were at disposal, as well as masks and protective jackets. An instruction video was then presented to the participants.

The experience began with a pre-training phase in real conditions (see Fig. 1 B). All participants used the same tool and performed once the three tasks. They were asked to hold the tool with their dominant hand like a pen. This grasping position prevented participants from holding the tip of the tool and accessing the speed and power buttons. At the end of this pre-training, they had to complete a NASA-RTLX questionnaire.

To begin the training phase, participants were immersed inside the virtual environment with an introductory period of around 3mins based on [37] recommendation. During this time, they had to familiarize themselves with the replica. As in pre-training, they were asked to hold it like a pen with their dominant hand. While the introduction was taking place, the eye-tracking system was calibrated utilizing the VR headset built-in calibration software. Participants were then provided with oral instructions describing the tasks execution order.

Then, a first training session happened, during which the three tasks were performed three times (see Fig. 1 C and D). Participants were then asked if they would like to take a maximum five-minute break (which was highly recommended for users not familiar with VR). After this optional break, participants had to perform a second training session, again performing the three tasks three times. The only difference between these two sessions was the virtual mechanical part to machine, which did not have the same shape, but similar features to perform the tasks. Participants were asked to keep safety gloves during this immersive training phase, to reduce the tactile perception discrepancies that might happen due to the roughness difference between the replica and the original tool. At the end of this two-session training phase, participants had to complete a NASA-RTLX and a SUS questionnaire.

Next, participants completed the three tasks one last time under real conditions, all using the same original tool again. After this post-training phase, participants were asked to complete several questionnaires: a third NASA-RTLX, a perceived learning questionnaire, and a demographic form. The experience ended with a debrief to answer participants’ questions. The procedure is summed up in Fig. 2. This user study lasted around 1h. Some participants completed it in less than 35mins.

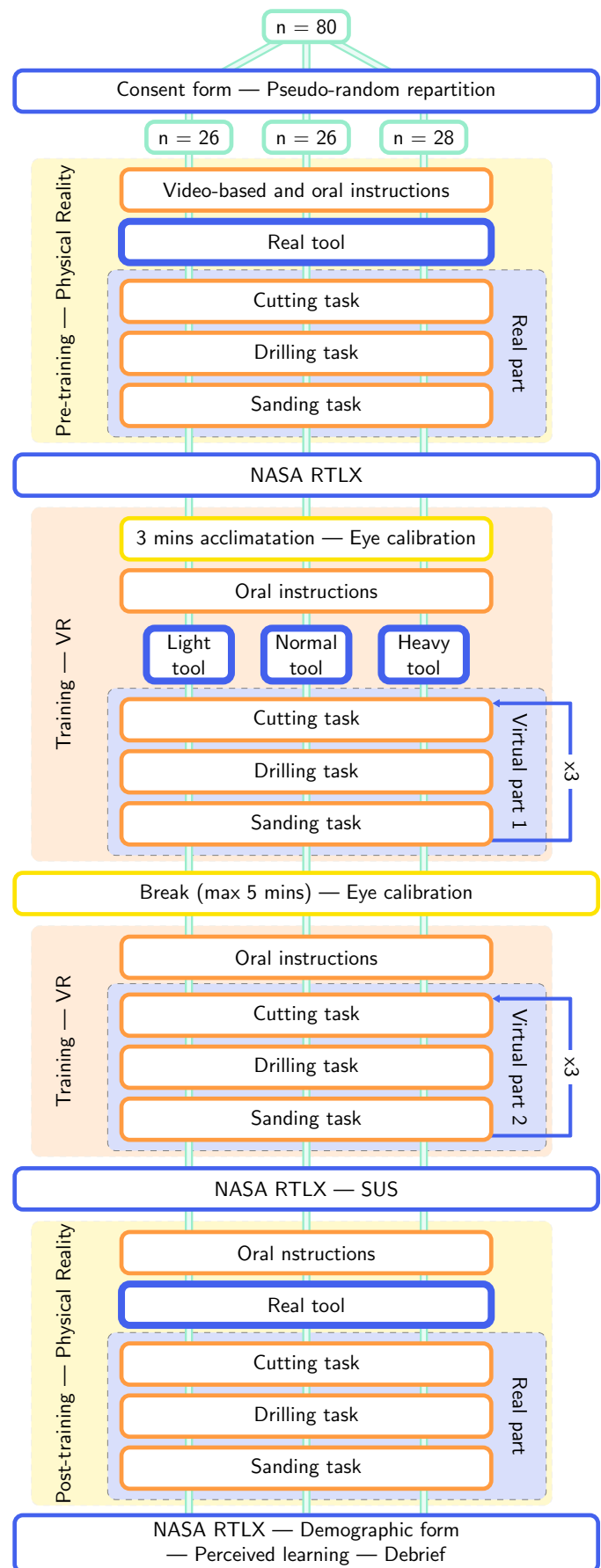


Fig. 2: Study Flow Diagram.

### 3.5 Apparatus

The application was deployed on an HTC VIVE Pro Eye headset with an integrated Tobii eye tracking system. The optical tracking was substituted with an Optitrack tracking system (24 Flex 13 cameras) to enable tracking of the replicas with insignificant mass markers. Regarding this setup, a usual solution would have been to use VIVE trackers, however, their 75g mass was too heavy for the purpose of this study. A first computer (Windows 7 Pro, 24GB RAM, Intel Xeon 2.80GHz) ran the Optitrack system, and a second one was used for the participants to answer the online questionnaires, and to display the video tutorial on a large screen (Sony 4K 75-inch). Another computer (Windows 10 Pro, 64Go RAM, Intel Xeon 3.00GHz, NVIDIA TITAN Xp) was used to run a VR application developed with Unity (2020.3.26f).

To prevent time loss due to tool bit changes, three Dremel 8220 equipped with three different tool bits (one for each task) were used. Several spare cutting blades, drilling bits, and sanding bands were stocked to prevent breakage and wear. When breakage occurred, the experience was paused to allow the examiner to change the tool bit. Moreover, the speed of the rotary tool was maintained constant.

Passive replicas of the tool were 3D printed with an Ultimaker 2 printer. Lead sheets were used as ballast weights. Particular attention has been paid to their balance. In accordance with JND regarding mass perception in VR as reviewed in [31], replicas were weighted with three different masses: 312g ( $m=50\%$ ), 624g ( $m=100\%$ , equals to the original tool mass), 936g ( $m=150\%$ ). Each replica was tracked thanks to seven passive markers, positioned to not disturb the pen-like grasping pose.

The working area included a desk, a clamp, and a vice. For safety concerns, moving parts of the vice were removed. Mechanical parts used to perform the tasks included paving support pads and pieces of wood. The pads were chosen for their geometry. They presented several protrusions used for the cutting task and some rough areas which allowed participants to perform the sanding task. Wood pieces shaped as parallelepipeds were used for the drilling task.

### 3.6 Virtual Environment

Participants were immersed inside a virtual room, with virtual counterparts for the tangible elements composing the working area. They were seated in front of a desk at a central position. A vice and a clamp holding the different virtual mechanical parts to machine were attached to this desk. Some purely virtual elements (a door, a window, a fire extinguisher, a heater) were added to the virtual environment to increase realism [19], reduce claustrophobia [3] and favor the feeling of presence [11].

To allow both right and left-handed participants to perform the tasks, the application was designed symmetrically. Visual cues directed the participants to task locations. Within the virtual application, participants' hands were not represented. As shown in [47], hand visualization does not always have an impact on performance when learning the usage of a tool in virtual reality. Participants did not have an avatar either, only the tool was tracked within the virtual environment.

When machining a part, visual stimuli were provided with the material being visually cut or machined. The areas to machine were rendered thanks to a GPU-based marching cubes algorithm. A connected component labeling technique was used to detect a cut piece. As it has been shown that multi-modal feedback in VR training could lead to an error rate decrease when the task is then performed in real conditions [9], pseudo-haptic [27] and auditory feedback were added to indicate whether the virtual tool was machining or not. The feed rate of the tool was virtually slowed down when machining, to produce a feeling of mechanical resistance from the materials. The original tool sounds at idle speed and when machining were recorded and integrated into the application, to indicate respectively when the tool was powered on and when a machining task was done. The application ran at 90FPS.

## 4 RESULTS

Statistical analyses were performed with R and  $\alpha = 0.05$ . One-way and mixed ANOVA were used when the normality assumption was met. For ANOVAs, Greenhouse-Geisser corrections were applied to DoFs when sphericity was violated. When the normality assumption was not met, an alternative was made to normalize the data through the use of a Box-Cox

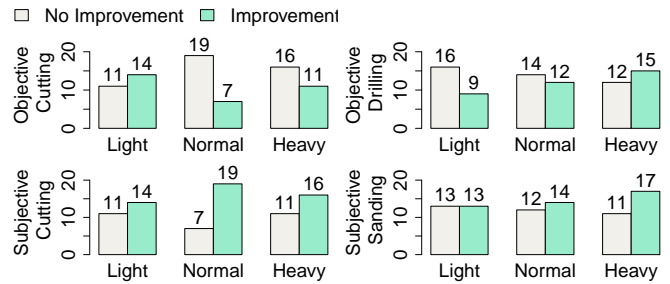


Fig. 3: **Top:** Objective Improvement Scores (**Left:** Cutting, **Right:** Drilling). **Bottom:** Subjective Evaluation (**Left:** Cutting, **Right:** Sanding).

transformation [4]. If it failed, data were processed using a Kruskal-Wallis test, or an Aligned Rank Transform (ART) [63]. Another approach was also used for mixed design models when the normality assumption was not met, analyzing the data with a Brunner-Langer test [44]. Post-hoc pairwise t-tests with Bonferroni corrections were used for after-effect analysis when normality was met. Otherwise, Mann-Whitney U test was used for independent variables and Wilcoxon signed-rank test for dependent ones. For some subsets of data, the outliers were also removed: values above  $Q3 + 1.5 \times (Q3 - Q1)$  or below  $Q1 - 1.5 \times (Q3 - Q1)$  (we considered them as errors regarding our experience design). Binary data were analyzed with binary logistic regression models and McFadden's pseudo R-squared value. A Principal Component Analysis (PCA) [23] was performed on the whole data set to detect eventual clusters of variables or participants. Finally, Pearson's and Spearman's correlation coefficients were used in an attempt to find correlations between variables.

A comparison was made between the pre-training and post-training conditions before the analysis of the immersive training was conducted.

## 4.1 Pre-training Versus Post-training

### 4.1.1 Performances

**Cutting:** neither an ART nor a Brunner-Langer test revealed any differences ( $p = 0.131$  and  $p = 0.124$ ). An improvement score calculated from these data is presented in Fig. 3. A binary logistic regression model for this improvement score was not well-fitted (McFadden's pseudo  $R^2 = 0.043$ ).

**Drilling:** no interaction was found between TOOL and ERROR ( $p = 0.620$ ) after a Box-Cox transformation. After removing four outliers, a mixed-design ANOVA did not reveal any significant effect ( $p = 0.470$ ), neither did a Brunner-Langer test performed on the whole data set ( $p = 0.383$ ). An improvement score calculated from these data is presented in Fig. 3. A binary logistic regression model for this improvement score was not well-fitted (McFadden's pseudo  $R^2 = 0.019$ ).

**Sanding:** as a reminder, an objective assessment of the sanding task was impossible due to surface textures being too coarse to allow a correct measure with a roughness meter.

### 4.1.2 Subjective Assessment

**Cutting & Sanding:** data are presented in Fig. 3. A binary logistic regression model for these data was not well-fitted (Cutting: McFadden's pseudo  $R^2 = 0.018$ ; Sanding: McFadden's pseudo  $R^2 = 0.006$ ).

**Drilling:** as a reminder, as holes drilled by participants were not visible in their wholeness, it was impossible to subjectively assess drilling improvement.

### 4.1.3 Learning Outcomes Score

To evaluate learning outcomes in physical reality, a global score was calculated based on objective results and subjective assessments from the pre and post-training phases. This score varies from 1 to 5 and is the addition of the four previously calculated improvement scores: objective and subjective cutting improvement scores, objective drilling improvement scores, and subjective sanding improvement scores. A 1 score means the participant did not improve at all, while a 5 score means

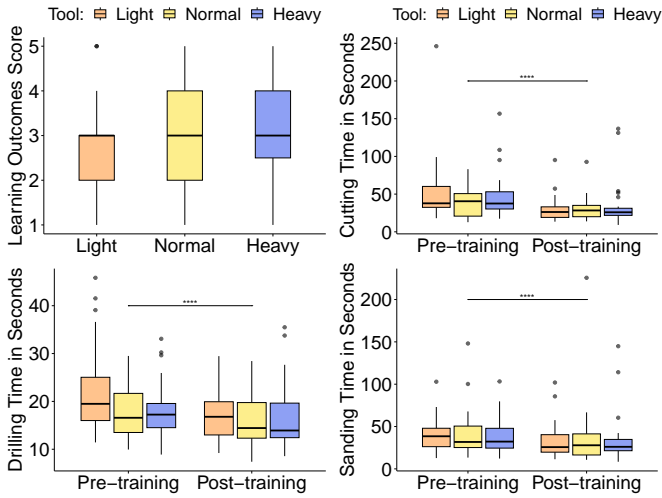


Fig. 4: **Top left:** Learning Outcomes Score among groups. **Top right:** TCT results for the Cutting task. **Bottom:** TCT results for the Drilling (**Left**) and the Sanding (**Right**) tasks. \*\*\*\*\* indicates significant difference with  $p < 0.001$ .

Table 3: Learning Outcomes Score Summary Statistics.

Tool	N	Mean	Sd	Median	Iqr
Light	25	2.96	1.02	3	1
Normal	26	3	1.02	3	2
Heavy	27	3.18	1.04	3	1.5

the participant made improvements regarding the four metrics. data are presented in Fig. 4 and in Tab. 3. A Kruskal-Wallis test on this global score did not reveal any difference between the three groups.

#### 4.1.4 Task Completion Time

**Cutting & Drilling:** a mixed-design ANOVA performed after a Box-Cox transformation revealed an effect on TIME (Cutting:  $F_{1,150} = 17.22$ ,  $p = 5.56 \times 10^{-5} < 0.001$ ,  $\eta_G^2 = 0.10$ , Drilling:  $F_{1,152} = 8.01$ ,  $p = 5.28 \times 10^{-3} < 0.05$ ,  $\eta_G^2 = 0.05$ ), and post-hoc pairwise t-tests indicated a significant difference between the pre and post-training (Cutting:  $p = 6.36 \times 10^{-9} < 0.001$  & Drilling:  $p = 7.78 \times 10^{-6} < 0.001$ ). A Brunner-Langer analysis also revealed a significant difference on TIME (Cutting:  $p = 8.58 \times 10^{-12} < 0.001$  & Drilling:  $p = 2.04 \times 10^{-6} < 0.001$ ), and a post-hoc Wilcoxon signed rank test confirmed this significant difference between pre and post-training (Cutting:  $p = 3.7 \times 10^{-8} < 0.001$  & Drilling:  $p = 5.9 \times 10^{-5} < 0.001$ ). Data are presented in Fig. 4.

**Sanding:** a mixed-design ANOVA performed after a Box-Cox transformation revealed an effect on TIME ( $F_{1,154} = 7.23$ ,  $p = 7.95 \times 10^{-3} < 0.05$ ,  $\eta_G^2 = 0.04$ ), and post-hoc pairwise t-tests indicated a significant difference between pre and post-training ( $p = 1.98 \times 10^{-2} < 0.001$ ). The same significant difference was observed when removing 10 outliers, with a mixed-design ANOVA performed after a Box-Cox transformation ( $F_{1,140} = 10.44$ ,  $p = 0.002 < 0.05$ ,  $\eta_G^2 = 0.07$ ), and with post-hoc pairwise t-test ( $p = 1.25 \times 10^{-5} < 0.001$ ). A Brunner-Langer analysis of the whole data set also revealed a significant difference on TIME ( $p = 8.27 \times 10^{-6} < 0.001$ ), and a Wilcoxon signed rank test confirmed the difference between the two phases ( $p = 8.5 \times 10^{-4} < 0.001$ ). Data are presented in Fig. 4.

## 4.2 Immersive Training

For the sake of brevity, only the first and last trials of the immersive training are depicted here. For the whole plots and results including the six trials for each metric, please refer to the supplemental material.

### 4.2.1 Performances

Performance results for training tasks are presented in Fig. 5 and in Tab. 5.

Table 4: Cutting, Drilling and Sanding TCT Summary Statistics.

Tool	Task	N	Mean	Sd	Median	Iqr
Light	Cutting	25	52.7	45.5	37.8	27.7
Normal	Cutting	26	37.6	18.2	40.5	29.8
Heavy	Cutting	27	47.7	30.7	37.5	22.8
Light	Cutting	25	30.4	17.1	26.3	14.0
Normal	Cutting	26	30.8	16.1	28.3	15.0
Heavy	Cutting	27	35.1	30.2	25.9	9.4
Light	Drilling	26	22.3	9.36	19.5	9.04
Normal	Drilling	26	17.5	5.20	16.6	8.17
Heavy	Drilling	27	18.4	5.72	17.2	5.06
Light	Drilling	26	17.2	5.31	16.8	6.94
Normal	Drilling	26	16.2	5.57	14.4	7.44
Heavy	Drilling	27	16.9	7.12	13.9	7.22
Light	Sanding	26	40.3	19.9	38.5	21.8
Normal	Sanding	26	41.7	29.2	31.8	25.4
Heavy	Sanding	28	40.4	22.4	32.2	23.4
Light	Sanding	26	33.4	21.7	25.7	20.7
Normal	Sanding	26	38.1	42.0	27.9	24.9
Heavy	Sanding	28	34.5	29.2	26.0	13.3

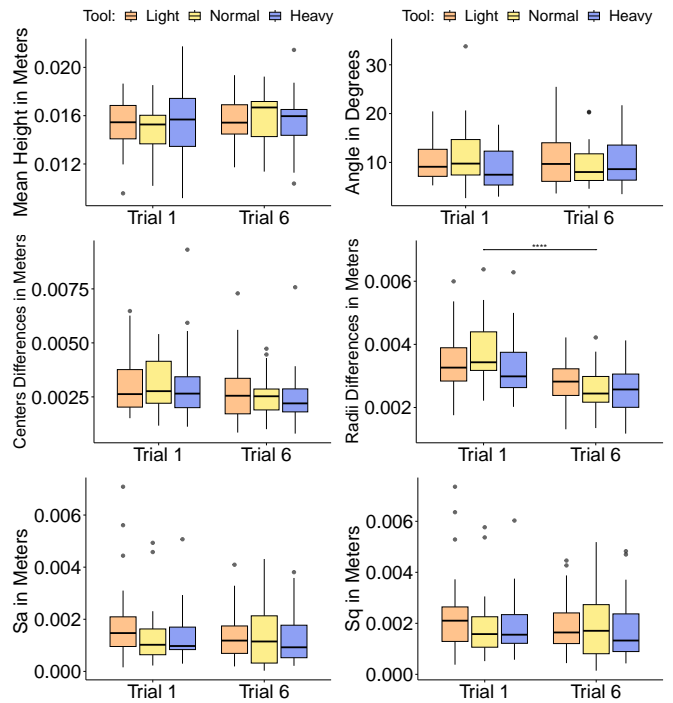


Fig. 5: Immersive Training Tasks Performance Results. **From Top to Bottom:** Cutting Mean Height, Drilling Angle, Drilling Centers, Drilling Radii, Sanding Sa, Sanding Sq. \*\*\*\*\* indicates significant difference with  $p < 0.001$ .

**Cutting Mean Height:** a mixed-design ANOVA performed after a Box-Cox transformation did not reveal any interaction ( $p = 0.948$ ). Neither did a Brunner-Langer analysis ( $p = 0.511$ ).

**Drilling Angle & Centers:** a mixed-design ANOVA performed after a Box-Cox transformation did not reveal any interaction (Angle:  $p = 0.927$ ; Centers:  $p = 0.784$ ). Neither did a Brunner-Langer analysis (Angle:  $p = 0.503$ , Centers:  $p = 0.307$ ).

**Drilling Radii:** a mixed-design ANOVA performed after a Box-Cox transformation revealed a significant difference on RADII ( $F_{5,450} = 9.97$ ,  $p = 4.62 \times 10^{-9} < 0.001$ ,  $\eta_G^2 = 0.10$ ). This was confirmed by a Brunner-Langer analysis ( $p = 1.35 \times 10^{-15} < 0.001$ ). Especially, there is a significant difference between the first and last trials ( $p = 2.12 \times 10^{-11} < 0.001$ ), as shown in Fig. 5. To see the whole plot with all trials and differences, please refer to the supplemental material.

Table 5: Immersive Training Summary Statistics.

Tool	Variable	Trial	N	Mean	Sd	Median	Iqr
Light	Mean Height	Trial 1	26	0.015	0.002	0.015	0.003
Normal	Mean Height	Trial 1	26	0.015	0.002	0.015	0.002
Heavy	Mean Height	Trial 1	26	0.015	0.003	0.016	0.004
Light	Mean Height	Trial 6	26	0.016	0.002	0.015	0.002
Normal	Mean Height	Trial 6	26	0.016	0.002	0.017	0.003
Heavy	Mean Height	Trial 6	26	0.015	0.002	0.016	0.002
Light	Drilling Angle	Trial 1	26	10.0	3.94	9.10	5.54
Normal	Drilling Angle	Trial 1	26	11.4	6.57	9.75	7.26
Heavy	Drilling Angle	Trial 1	26	8.74	4.08	7.47	6.95
Light	Drilling Angle	Trial 6	26	10.3	5.36	9.69	7.91
Normal	Drilling Angle	Trial 6	26	9.34	4.24	8.02	5.46
Heavy	Drilling Angle	Trial 6	26	10.5	5.53	8.62	7.19
Light	Drilling Center	Trial 1	26	0.003	0.001	0.003	0.002
Normal	Drilling Center	Trial 1	26	0.003	0.001	0.003	0.002
Heavy	Drilling Center	Trial 1	26	0.003	0.002	0.003	0.001
Light	Drilling Center	Trial 6	26	0.003	0.002	0.003	0.002
Normal	Drilling Center	Trial 6	26	0.003	0.001	0.003	0.001
Heavy	Drilling Center	Trial 6	26	0.002	0.001	0.002	0.001
Light	Drilling Radii	Trial 1	26	0.003	0.001	0.003	0.001
Normal	Drilling Radii	Trial 1	26	0.004	0.001	0.003	0.001
Heavy	Drilling Radii	Trial 1	26	0.003	0.001	0.003	0.001
Light	Drilling Radii	Trial 6	26	0.003	0.001	0.003	0.001
Normal	Drilling Radii	Trial 6	26	0.003	0.001	0.002	0.001
Heavy	Drilling Radii	Trial 6	26	0.003	0.001	0.003	0.001
Light	Sanding Sa	Trial 1	26	0.002	0.002	0.001	0.001
Normal	Sanding Sa	Trial 1	23	0.001	0.001	0.001	0.001
Heavy	Sanding Sa	Trial 1	26	0.001	0.001	0.001	0.001
Light	Sanding Sa	Trial 6	26	0.001	0.001	0.001	0.001
Normal	Sanding Sa	Trial 6	23	0.001	0.001	0.001	0.002
Heavy	Sanding Sa	Trial 6	26	0.001	0.001	0.001	0.001
Light	Sanding Sq	Trial 1	26	0.002	0.002	0.002	0.001
Normal	Sanding Sq	Trial 1	23	0.002	0.001	0.002	0.001
Heavy	Sanding Sq	Trial 1	26	0.002	0.001	0.002	0.001
Light	Sanding Sq	Trial 6	26	0.002	0.001	0.002	0.001
Normal	Sanding Sq	Trial 6	23	0.002	0.001	0.002	0.002
Heavy	Sanding Sq	Trial 6	26	0.002	0.001	0.001	0.001

**Sanding Sa & Sq:** a mixed-design ANOVA performed after a Box-Cox transformation revealed a significant effect on TOOL (Sa:  $F_{2,432} = 4.09$ ,  $p = 0.017 < 0.05$ ,  $\eta^2_G = 0.02$ ; Sq:  $F_{2,432} = 3.27$ ,  $p = 0.039 < 0.05$ ,  $\eta^2_G = 0.01$ ). Post-hoc comparisons showed a difference between the lighter one and the replica of normal mass (Sa:  $p = 0.0136 < 0.05$ ; Sq:  $p = 0.0391 < 0.05$ ). However, these results were not verified with a Brunner-Langer analysis (Sa:  $p = 0.496$ , Sq:  $p = 0.536$ ).

#### 4.2.2 Task Completion Time

For the three tasks, an ART revealed a significant effect on TIME (Cutting:  $F_{5,375} = 51.01$ ,  $p = 2 \times 10^{-16} < 0.001$ ; Drilling:  $F_{5,375} = 84.02$ ,  $p = 2 \times 10^{-16} < 0.001$ ; Sanding:  $F_{5,360} = 28.76$ ,  $p = 2 \times 10^{-16} < 0.001$ ), confirmed by a Brunner-Langer analysis (Cutting:  $p = 1.20 \times 10^{-46} < 0.001$ ; Drilling:  $p = 1.61 \times 10^{-66} < 0.001$ ; Sanding:  $p = 4.62 \times 10^{-23} < 0.001$ ). Results from the ART concerning the first and last trials (Cutting:  $p = 3.78 \times 10^{-33} < 0.001$ ; Drilling:  $p = 1.52 \times 10^{-48} < 0.001$ ; Sanding:  $p = 1.88 \times 10^{-21} < 0.001$ ) are shown in Fig. 6 and in Tab. 6. For the whole plots and differences, please refer to the supplemental material.

### 4.3 Cognitive Load

#### 4.3.1 Pupil Diameter

IPA data were normalized with a Box-Cox transformation and then processed with a one-way ANOVA. LHIPA data were analyzed with a Kruskal-Wallis test. In both cases, no significant effects were found (IPA:  $p = 0.906$ , LHIPA:  $p = 0.081$ ). Removing twelve outliers from the LHIPA data set did not lead to any significant effect, after performing a one-way ANOVA ( $p = 0.735$ ). Results are shown in Fig. 7 and in Tab. 7.

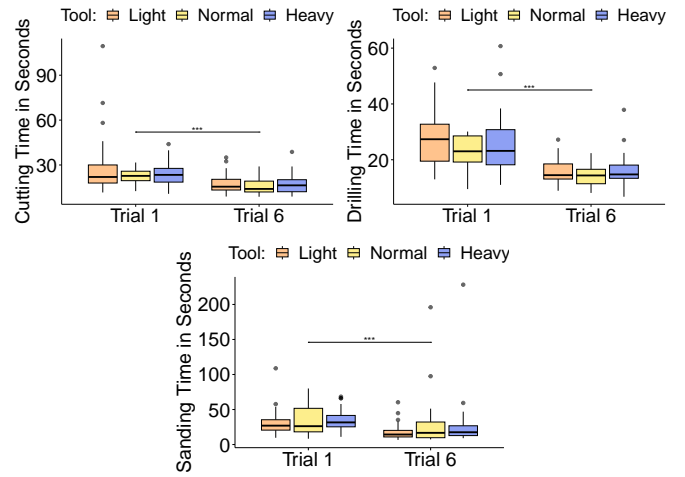


Fig. 6: TCT for the Immersive Training Tasks. \*\*\*\* indicates significant difference with  $p < 0.001$ .

Table 6: Immersive Training TCT Summary Statistics.

Tool	Task	Trial	N	Mean	Sd	Median	Iqr
Light	Cutting	Trial 1	26	29.8	21.2	22.0	12.1
Normal	Cutting	Trial 1	26	22.2	4.96	22.7	6.32
Heavy	Cutting	Trial 1	26	24.4	8.26	23.4	9.18
Light	Cutting	Trial 6	26	17.3	6.78	15.5	7.17
Normal	Cutting	Trial 6	26	15.3	5.19	14.0	7.25
Heavy	Cutting	Trial 6	26	17.2	6.98	16.4	8.03
Light	Drilling	Trial 1	26	27.5	10.2	27.4	13.2
Normal	Drilling	Trial 1	26	22.8	6.16	23.0	9.36
Heavy	Drilling	Trial 1	26	25.7	11.5	23.2	12.6
Light	Drilling	Trial 6	26	15.9	4.81	14.5	5.42
Normal	Drilling	Trial 6	26	14.6	4.00	14.4	5.18
Heavy	Drilling	Trial 6	26	16.0	6.38	14.7	4.74
Light	Sanding	Trial 1	26	31.7	19.7	27.0	14.8
Normal	Sanding	Trial 1	23	35.4	21.7	26.2	33.6
Heavy	Sanding	Trial 1	26	35.5	16.1	31.6	16.3
Light	Sanding	Trial 6	26	19.0	12.8	14.3	9.38
Normal	Sanding	Trial 6	23	31.8	41.6	16.7	22.4
Heavy	Sanding	Trial 6	26	29.0	42.4	17.5	14.0

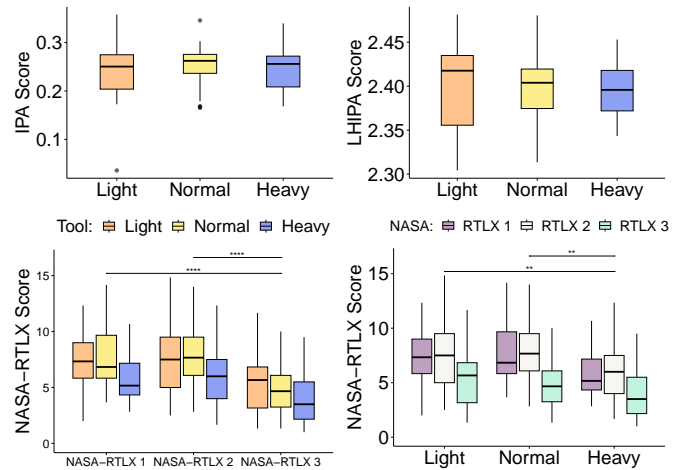


Fig. 7: **Top left:** a low IPA score indicates a low cognitive load. **Top right:** LHIPA scores after removing outliers, a high LHIPA score indicates a low cognitive load. **Bottom:** NASA-RTLX Pairwise Comparisons Differences after removing five outliers. NASA-RTLX scores range from 0 to 20, 20 meaning higher perceived cognitive load. \*\*\*\* indicates significant difference with  $p < 0.001$ , and \*\*\*\* indicates significant difference with  $p < 0.05$ .



Table 7: IPA and LHIPA Summary Statistics, without Outliers for LHIPA.

Tool	Variable	N	Mean	Sd	Median	Iqr
Light	IPA	24	0.241	0.064	0.25	0.071
Normal	IPA	23	0.25	0.047	0.262	0.039
Heavy	IPA	26	0.247	0.048	0.256	0.064
Light	LHIPA	17	2.41	0.05	2.42	0.079
Normal	LHIPA	21	2.40	0.041	2.40	0.045
Heavy	LHIPA	23	2.40	0.032	2.40	0.046

Table 8: NASA-RTLX without Outliers Summary Statistics.

Tool	Variable	N	Mean	Sd	Median	Iqr
Light	NASA-RTLX 1	25	7.21	2.70	7.33	3.17
Normal	NASA-RTLX 1	23	7.57	2.92	6.83	3.83
Heavy	NASA-RTLX 1	25	5.69	2.12	5.17	2.83
Light	NASA-RTLX 2	25	7.74	3.25	7.5	4.5
Normal	NASA-RTLX 2	23	7.81	2.87	7.67	3.42
Heavy	NASA-RTLX 2	25	6.07	2.61	6	3.5
Light	NASA-RTLX 3	25	5.33	2.59	5.67	3.67
Normal	NASA-RTLX 3	23	4.78	2.41	4.67	2.83
Heavy	NASA-RTLX 3	25	3.97	2.21	3.5	3.33

Table 9: SUS and CAP Summary Statistics.

Tool	Variable	N	Mean	Sd	Median	Iqr
Light	SUS	26	77.4	13.7	80	20
Normal	SUS	26	76.9	12.0	80	7.5
Heavy	SUS	28	80.4	11.9	82.5	18.8
Light	CAP	26	34.5	8.46	34.5	10.2
Normal	CAP	26	35.1	8.72	35.5	12.5
Heavy	CAP	28	37.2	6.62	37.5	7.75

### 4.3.2 NASA-RTLX

Data were normalized with a Box-Cox transformation. A mixed-design ANOVA shown a significant effect on NASA-RTLX answers ( $F_{2,225} = 20.43$ ,  $p = 7.01 \times 10^{-9} < 0.001$ ,  $\eta_c^2 = 0.15$ ), and a significant effect on TOOL ( $F_{2,225} = 4.63$ ,  $p = 1.07 \times 10^{-2} < 0.05$ ,  $\eta_c^2 = 0.04$ ). Pairwise comparisons revealed significant differences ( $p = 1.56 \times 10^{-13} < 0.001$ ,  $p = 7.71 \times 10^{-11} < 0.001$ ) between the third NASA-RTLX questionnaire and the two others, as well as a significant difference ( $p = 4.36 \times 10^{-2} < 0.05$ ) between the heavy replica and the one with normal mass. After removing five outliers within this data set, the significant effects previously found were accentuated. A mixed-design ANOVA performed after a Box-Cox transformation revealed significant effects on NASA-RTLX ( $F_{2,210} = 20.92$ ,  $p = 5.18 \times 10^{-9} < 0.001$ ,  $\eta_c^2 = 0.17$ ), and TOOL ( $F_{2,210} = 7.57$ ,  $p = 6.67 \times 10^{-4} < 0.001$ ,  $\eta_c^2 = 0.07$ ). As before, pairwise comparisons revealed significant differences between the third NASA-RTLX questionnaire and the two others (NASA-RTLX 1 & 3:  $p = 1.08 \times 10^{-13} < 0.001$ , NASA-RTLX 2 & 3:  $p = 7.95 \times 10^{-11} < 0.001$ ). A variation arose as significant differences between the heavy replica and the two others were found (LIGHT & HEAVY:  $p = 5.28 \times 10^{-3} < 0.05$ , NORMAL & HEAVY:  $p = 8.03 \times 10^{-3} < 0.05$ ). A Brunner-Langer analysis revealed a significant effect on NASA-RTLX ( $p = 1.24 \times 10^{-18} < 0.001$ ). Wilcoxon signed rank tests revealed significant differences between the third NASA-RTLX and the two others (NASA-RTLX 1 & 3:  $p = 2.5 \times 10^{-11} < 0.001$ , NASA-RTLX 2 & 3:  $p = 6.1 \times 10^{-10} < 0.001$ ) which is consistent with previous findings. With this non-parametric analysis, no effect was observed on TOOL ( $p = 0.106$ ). Results are shown in Fig. 7 and in Tab. 8.

### 4.4 System Usability Scale

Data met the normality assumption after a Box-Cox transformation and were analyzed with a one-way ANOVA. No significant effect was found ( $p = 0.550$ ). After removing six outliers from the data set, a one-way ANOVA did not reveal a significant effect ( $p = 0.588$ ). Results are shown in Fig. 8 and in Tab. 9.

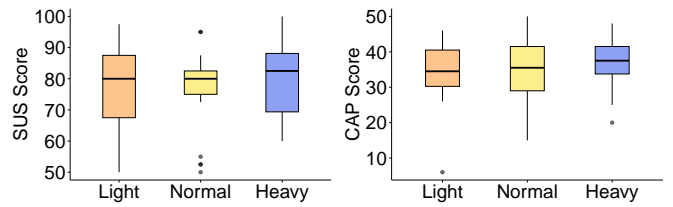


Fig. 8: **Left:** SUS scores range from 0 to 100, 100 meaning higher usability. **Right:** CAP scores range from 0 to 54, 54 meaning higher perceived learning.

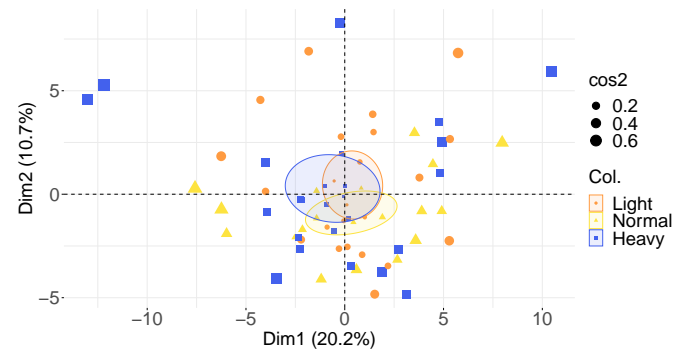


Fig. 9: PCA results. Dot size varies according to the participant's  $\cos^2$  value, the higher this value, the bigger the dot, and the better the participant is represented within these two dimensions.

### 4.5 Cognitive Affective Psychomotor Learning

CAP data were normalized with a Box-Cox transformation. A one-way ANOVA did not reveal any significant result ( $p = 0.471$ ). After removing two outliers, a one-way ANOVA performed on this new data set did not reveal any significant effect ( $p = 0.329$ ). Results are shown in Fig. 8 and in Tab. 9.

### 4.6 Correlations and Principal Component Analysis

Pearson's and Spearman's methods were applied to identify correlations among the data. Despite yielding significant results, no strong correlations were discovered.

To visualize and identify potential clusters of participants, a PCA was performed on a clean whole data set, removing all participants whose data subsets were incomplete, mainly due to issues with the eye tracking system. Some participants were not able to fulfill some tasks, thus they were also removed. When representing the three groups of participants (the three tools) within the two principal dimensions, no defined cluster appeared, as seen in Fig. 9. This notably indicates the groups were homogeneous enough in terms of gender, age, Body Mass Index, and former experiences with VR, video games, and Do It Yourself.

## 5 DISCUSSION

In this study, the influence of the weight of an object replica used during an immersive training was investigated. Task completion time measures demonstrated that the realization of the three tasks had improved, and it is noteworthy that there were no significant differences observed between the groups that used replicas of varying weights.

Regarding performance metrics, some significant results suggest a global improvement in the learning outcomes. Indeed, there were significant task completion time differences observed between the pre-training and post-training phases, as well as during the immersive training. During the training phase, this indicates that participants took longer to complete the task on their first attempt. This was expected, as they first must familiarize themselves with the system and the machining virtual processes. In real conditions, participants were also slower during the pre-training compared to the post-training. However, we

could not establish if this task completion time improvement was due to the immersive training, or only because of the two real phases repetitions. To investigate further the training influence, a control group with no VR training would have been necessary. Indeed, the time difference observed between the two trials during pre and post-training may potentially be explained by various factors, including the original tool discovery, its vibrations, force feedback from materials, and auditory feedback. Some other significant differences were found in the radius size of the virtual drilled hole. For this specific metric, participants reduced the size of the hole when repeatedly performing the task. Here, a training improvement was observed within the VR application. However, no notable differences were detected among the three groups in regard to this metric or any other. Thus, H1 is invalidated, as these results revealed the learning outcomes were not influenced by the different masses of the replica. This indicates that the learning outcomes are not adversely affected by a replica whose mass is different from that of the original tool.

Cognitive load was evaluated with both objective and subjective metrics. Objective metrics relied on the size of the pupil, which should be increased when one is under high mental demand. No significant differences were observed between the three groups with these metrics. However, significant differences were found when evaluating the perceived cognitive load with NASA-RTLX questionnaires. There were significant differences between the last questionnaire and the other two, suggesting that the perceived cognitive load is lower at the end of the post-training phase. Again, this reduction may be due to the immersive training or to the repetition of tasks in real-life conditions. The fact that there were no significant differences observed between the first and second questionnaires tends notably to show that performing the tasks within the immersive training system did not require more cognitive load than executing the tasks in real conditions. This suggests that the training application does not impede learning. Significant differences were found between the group that used the heavier replica and the others. These results state that the use of a heavier replica tends to reduce the perceived cognitive load. This is unexpected and as a consequence, H2 is invalidated. We assumed such a reduction could be due to the mechanical impedance of the hand and prop system, which when increased tends to absorb oscillations and other hand tremors. This leads us to make the following recommendation for conditions similar to the ones evaluated in this study: if the priority within an immersive training application is to decrease the perceived cognitive load, then using a replica with a heavier mass than the original tool is suggested.

It appeared that no matter the group, the usability scores were all around 80 suggesting good usability for the custom training application according to the Sauro-Lewis curved grading scale [52]. This suggests that the usability of the application was not a hindrance to knowledge transfer. Similarly, the perceived learning scores among the three groups were also high: all means greater than 34 out of 54, which suggests a good perception of total learning [49]. This indicates that the application had the potential to impart participants with a certain level of procedural knowledge and was relevant for training assessment. Regarding these two metrics, no significant results were found between groups. This suggests that the mass of the replica did not impact the usability of the training system, nor the perceived learning, and as a result both H3 and H4 are invalidated. If usability and perceived learning are the top priority for one who would like to design a VR training application, it could be stated that within a 300g - 900g range, for a one-handed rotary tool, and for tasks similar to the ones evaluated in this study, the mass of the replica is not important.

Taken together, these results suggest that the mass of the virtual replica is not a major factor for learning thanks to an immersive training application. This applies in the case of this study, with the specific tasks evaluated. On the one hand, the tool used is of relatively moderate weight. Particularly heavy tools were not used, which could have a greater impact on muscle fatigue. Furthermore, participants were able to lean on the table and perform the task in a relatively straightforward manner. There were no cluttering constraints or major obstructions. In addition, the tasks selected did not require extremely fine placement or control of the tool. When solving the task in real life, the participant could adjust his gesture to correct it. Thus, in the case of a task and an

object that do not involve high stakes in terms of weight or precision, the study tends to show that the weight of the virtual replica can vary significantly from the original tool while maintaining good learning performance, limited cognitive load, and reasonable usability.

## 6 LIMITATIONS AND FUTURE WORK

A one-handed rotary tool was used, with a non-powered replica. No actuators were included inside the replica as it is usual within VR training applications to build them in the simplest way possible. As vibrations could influence the perception of weight within a virtual environment [26], the use of a motored replica might lead to different results. Furthermore, the mass of the replica was in the 300g to 900g range. Expanding the findings to other ranges of masses of replicas could be interesting. If an influence of the mass had to be found for lighter or heavier replicas, then the question of the sufficient and necessary mass limit would be also of interest to facilitate the design of instrumented replicas.

One could also point out the difficulty of the three selected tasks. Indeed, with this experience design, it was not possible to assess the impact of the virtual training on the learning outcomes compared to a more traditional one. The study was limited to tasks that did not heavily rely on the user's proprioceptive feedback. More specific tasks, involving greater precision of gesture, a different grasp style for the tool, or greater muscle fatigue could be tested in order to extend these results to other types of learning.

When dealing with mass, balance often comes to one's mind. With the original tool as a reference, the replicas were meticulously balanced by closely aligning their center of gravity to ensure consistency with the original. To further investigate the influence of inertial properties, future research could examine the use of replicas that are intentionally unbalanced. If an effect were to be observed, the question of the optimal mass distribution for a replica would be worth of interest.

Moreover, most of the participants reported the lack of haptic feedback, arguing that it decreased the realism, and made the task harder to perform in VR because the resistance of the materials was not present and could not help them to guide the tool bit. Interestingly, among the 80 participants, only one wrote a comment regarding the mass of the tool: P46, using the heavy replica, stated that "the difference in weight between the real tool and the tool used in VR has a significant impact on the muscular effort required and on the grip of the tool, which in turn has an impact on the quality of the work rendered". This remark might indicate that the mass could actually affect the learning experience in non-quantitative ways. Further investigations are required to shed light on this possibility.

## 7 CONCLUSION

The study conducted in this paper to investigate the impact of a tool replica mass on learning outcomes gave interesting results: whereas different masses were used during the immersive training phase for replicas of the original tool, learning outcomes, cognitive load, usability, and participants' perceived learning remained comparable between the groups. We believe these results provide useful insights for designers and manufacturers who would like to make tool replicas for immersive training applications, as it could now be stated that, within a 300g - 900g range, for a one-handed rotary tool, and for tasks similar to the ones evaluated (cutting, drilling, and sanding) in this study, the mass of the replica could vary significantly from the original without influencing learning outcomes. To broaden the scope of these findings, additional factors should be explored, including proprioceptive cues or an extended range of masses, shapes, inertia, or grip styles for replicas.

## ACKNOWLEDGMENTS

The authors would like to thank all the participants for the time they spent in the completion of the user study, and the company NVIDIA which offered the TITAN Xp graphic card used in this project. This work was supported by a French government funding managed by the National Research Agency under the Investments for the Future program (PIA) grant ANR-21-ESRE-0030 (CONTINUUM).

## REFERENCES

- [1] B. S. Bloom. *Taxonomy of Educational Objectives: The Classification of Educational Goals*. New York, Longmans, Green, 1956. 4
- [2] E. Z. Borba, M. Cabral, A. Montes, O. Belloc, and M. Zuffo. Immersive and interactive procedure training simulator for high risk power line maintenance. In *ACM SIGGRAPH 2016 VR Village*, SIGGRAPH '16, p. 1. Association for Computing Machinery, New York, NY, USA, July 2016. doi: 10.1145/2929490.2929497 2
- [3] C. Botella, R. M. Baños, H. Villa, C. Perpiñá, and A. García-Palacios. Virtual reality in the treatment of claustrophobic fear: A controlled, multiple-baseline design. *Behavior Therapy*, 31(3):583–595, June 2000. doi: 10.1016/S0005-7894(00)80032-5 5
- [4] G. E. P. Box and D. R. Cox. An Analysis of Transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2):211–243, 1964. doi: 10.1111/j.2517-6161.1964.tb00553.x 5
- [5] J. Brooke. SUS: A 'Quick and Dirty' Usability Scale. In *Usability Evaluation In Industry*. CRC Press, 1996. 4
- [6] J. C. Byers, A. Bittner, and S. Hill. Traditional and Raw Task Load Index (TLX) Correlations: Are paired comparisons necessary? *Advances in Industrial Ergonomics and Safety I: Taylor and Francis.*, 1989. 3
- [7] H.-H. Choi, J. J. G. van Merriënboer, and F. Paas. Effects of the Physical Environment on Cognitive Load and Learning: Towards a New Model of Cognitive Load. *Educational Psychology Review*, 26(2):225–244, June 2014. doi: 10.1007/s10648-014-9262-6 3
- [8] A. Congès, A. Evain, F. Benaben, O. Chabiron, and S. Rebière. Crisis Management Exercises in Virtual Reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 87–92, Mar. 2020. doi: 10.1109/VRW50115.2020.00022 2
- [9] N. Cooper, F. Milella, I. Cant, C. Pinto, M. White, and G. Meyer. Augmented Cues Facilitate Learning Transfer from Virtual to Real Environments. In *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*, pp. 194–198, Sept. 2016. doi: 10.1109/ISMAR-Adjunct.2016.0075 2, 5
- [10] N. Cooper, F. Milella, I. Cant, M. D. White, and G. Meyer. Transfer of training—Virtual reality training with augmented multisensory cues improves user experience during training and task performance in the real world. *PLOS ONE*, 16(3):e0248225, Mar. 2021. doi: 10.1371/journal.pone.0248225 2, 3
- [11] B. Dalgarno and M. J. W. Lee. What are the learning affordances of 3-D virtual environments? *British Journal of Educational Technology*, 41(1):10–32, 2010. doi: 10.1111/j.1467-8535.2009.01038.x 5
- [12] S. Doolani, C. Wessels, V. Kanal, C. Sevastopoulos, A. Jaiswal, H. Nambiappan, and F. Makedon. A Review of Extended Reality (XR) Technologies for Manufacturing Training. *Technologies*, 8(4):77, Dec. 2020. doi: 10.3390/technologies8040077 2
- [13] A. T. Duchowski, K. Krejtz, N. A. Gehrler, T. Bafna, and P. Bækgaard. The Low/High Index of Pupillary Activity. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, pp. 1–12. Association for Computing Machinery, New York, NY, USA, Apr. 2020. doi: 10.1145/3313831.3376394 4
- [14] A. T. Duchowski, K. Krejtz, I. Krejtz, C. Biele, A. Niedzielska, P. Kiefer, M. Raubal, and I. Giannopoulos. The Index of Pupillary Activity: Measuring Cognitive Load vis-à-vis Task Difficulty with Pupil Oscillation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pp. 1–13. Association for Computing Machinery, New York, NY, USA, Apr. 2018. doi: 10.1145/3173574.3173856 4
- [15] J. East. Ergonomic Guidelines for Selecting Hand and Power Tools. <https://www.ehstoday.com/health/article/21908634/ergonomic-guidelines-for-selecting-hand-and-power-tools>, Dec. 2005. Last accessed: 25 May 2023. 3
- [16] F. Faul, E. Erdfelder, A. Buchner, and A.-G. Lang. Statistical power analyses using G\*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4):1149–1160, Nov. 2009. doi: 10.3758/BRM.41.4.1149 3
- [17] M. Feick, S. Bateman, A. Tang, A. Miede, and N. Marquardt. Tangible Proxies For Embodied Object Exploration And Manipulation In Virtual Reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 195–206, Nov. 2020. doi: 10.1109/ISMAR50242.2020.00042 2
- [18] A. Franzluebbers and K. Johnsen. Performance Benefits of High-Fidelity Passive Haptic Feedback in Virtual Reality Training. In *Proceedings of the 2018 ACM Symposium on Spatial User Interaction*, SUI '18, pp. 16–24. Association for Computing Machinery, New York, NY, USA, Oct. 2018. doi: 10.1145/3267782.3267790 2
- [19] G. Gonçalves, H. Coelho, P. Monteiro, M. Melo, and M. Bessa. Systematic review of comparative studies of the impact of realism in immersive virtual experiences. *ACM Computing Surveys*, Apr. 2022. doi: 10.1145/3533377 2, 5
- [20] Government of Canada Canadian Centre for Occupational Health and Safety. CCOHS: Hand Tool Ergonomics - Tool Design. <https://www.ccohs.ca/oshanswers/ergonomics/handtools/tooldesign.html>, Apr. 2023. Last accessed: 25 May 2023. 2, 3
- [21] E. Haapalainen, S. Kim, J. F. Forlizzi, and A. K. Dey. Psycho-physiological measures for assessing cognitive load. In *Proceedings of the 12th ACM International Conference on Ubiquitous Computing*, UbiComp '10, pp. 301–310. Association for Computing Machinery, New York, NY, USA, Sept. 2010. doi: 10.1145/1864349.1864395 3
- [22] K. Hinckley, R. Pausch, J. C. Goble, and N. F. Kassell. Passive real-world interface props for neurosurgical visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '94, pp. 452–458. Association for Computing Machinery, New York, NY, USA, Apr. 1994. doi: 10.1145/191666.191821 1
- [23] F. Husson, S. Le, and J. Pagès. *Exploratory Multivariate Analysis by Example Using R*. Chapman and Hall/CRC, New York, second ed., May 2017. doi: 10.1201/b21874 5
- [24] A. Ipsita, L. Erickson, Y. Dong, J. Huang, A. K. Bushinski, S. Saradhi, A. M. Villanueva, K. A. Peppler, T. S. Redick, and K. Ramani. Towards Modeling of Virtual Reality Welding Simulators to Promote Accessible and Scalable Training. In *CHI Conference on Human Factors in Computing Systems*, CHI '22, pp. 1–21. Association for Computing Machinery, New York, NY, USA, Apr. 2022. doi: 10.1145/3491102.3517696 2
- [25] A. D. Kaplan, J. Cruik, M. Endsley, S. M. Beers, B. D. Sawyer, and P. A. Hancock. The Effects of Virtual Reality, Augmented Reality, and Mixed Reality as Training Enhancement Methods: A Meta-Analysis. *Human Factors*, 63(4):706–726, June 2021. doi: 10.1177/0018720820904229 2
- [26] G. Kim, S. Okamoto, Y. Akiyama, and Y. Yamada. Weight illusion by presenting vibration to the fingertip. *Frontiers in Virtual Reality*, 3, 2022. 9
- [27] A. Lecuyer, S. Coquillart, A. Kheddar, P. Richard, and P. Coiffet. Pseudo-haptic feedback: Can isometric input devices simulate force feedback? In *Proceedings IEEE Virtual Reality 2000 (Cat. No.00CB37048)*, pp. 83–90. IEEE Comput. Soc, New Brunswick, NJ, USA, 2000. doi: 10.1109/VR.2000.840369 5
- [28] E. A.-L. Lee and K. W. Wong. A Review of Using Virtual Reality for Learning. In Z. Pan, A. D. Cheok, W. Müller, and A. El Rhalibi, eds., *Transactions on Entertainment I*, Lecture Notes in Computer Science, pp. 231–241. Springer, Berlin, Heidelberg, 2008. doi: 10.1007/978-3-540-69744-2\_18 2
- [29] J. R. Lewis. The System Usability Scale: Past, Present, and Future. *International Journal of Human-Computer Interaction*, 34(7):577–590, July 2018. doi: 10.1080/10447318.2018.1455307 4
- [30] N. Li, H.-J. Kim, L. Shen, F. Tian, T. Han, X.-D. Yang, and T.-J. Nam. HapLinkage: Prototyping Haptic Proxies for Virtual Hand Tools Using Linkage Mechanism. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, UIST '20, pp. 1261–1274. Association for Computing Machinery, New York, NY, USA, Oct. 2020. doi: 10.1145/3379337.3415812 2
- [31] W. N. Lim, K. M. Yap, Y. Lee, C. Wee, and C. C. Yen. A Systematic Review of Weight Perception in Virtual Reality: Techniques, Challenges, and Road Ahead. *IEEE Access*, 9:163253–163283, 2021. doi: 10.1109/ACCESS.2021.3131525 3, 5
- [32] J. Lu, A. Leng, Y. Zhou, W. Zhou, J. Luo, X. Chen, and X. Qi. An innovative virtual reality training tool for the pre-hospital treatment of cranialmaxillofacial trauma. *Computer Assisted Surgery*, 28(1):2189047, Dec. 2023. doi: 10.1080/24699322.2023.2189047 2
- [33] K. Lyu, A. Brambilla, A. Globa, and R. de Dear. An immersive multi-sensory virtual reality approach to the study of human-built environment interactions. *Automation in Construction*, 150:104836, June 2023. doi: 10.1016/j.autcon.2023.104836 2
- [34] A. Maehigashi, A. Sasada, M. Matsumuro, F. Shibata, A. Kimura, and S. Niida. Virtual Weight Illusion: Weight Perception of Virtual Objects Using Weight Illusions. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI EA '21, pp. 1–6. Association for Computing Machinery, New York, NY, USA, May 2021. doi: 10.1145/3411763.3451842 2
- [35] G. Makransky and G. B. Petersen. The Cognitive Affective Model of Immersive Learning (CAMIL): A Theoretical Research-Based Model of Learning in Immersive Virtual Reality. *Educational Psychology Review*,



- 33(3):937–958, Sept. 2021. doi: [10.1007/s10648-020-09586-2](https://doi.org/10.1007/s10648-020-09586-2) 1, 2
- [36] G. Meiselwitz and W. Sadera. Investigating the Connection between Usability and Learning Outcomes in Online Learning Environments. *MERLOT Journal of Online Learning and Teaching*, 4, July 2008. 3
- [37] I. Miguel-Alonso, B. Rodriguez-Garcia, D. Checa, and L. T. De Paolis. Developing a Tutorial for Improving Usability and User Skills in an Immersive Virtual Reality Experience. In L. T. De Paolis, P. Arpaia, and M. Sacco, eds., *Extended Reality*, Lecture Notes in Computer Science, pp. 63–78. Springer Nature Switzerland, Cham, 2022. doi: [10.1007/978-3-031-15553-6\\_5](https://doi.org/10.1007/978-3-031-15553-6_5) 4
- [38] M. Mulders. Vocational Training in Virtual Reality: A Case Study Using the 4C/ID Model. *Multimodal Technologies and Interaction*, 6(7):49, July 2022. doi: [10.3390/mti6070049](https://doi.org/10.3390/mti6070049) 1, 2
- [39] M. Mulders, J. Buchner, and M. Kerres. Virtual Reality in Vocational Training: A Study Demonstrating the Potential of a VR-based Vehicle Painting Simulator for Skills Acquisition in Apprenticeship Training. *Technology, Knowledge and Learning*, Nov. 2022. doi: [10.1007/s10758-022-09630-w](https://doi.org/10.1007/s10758-022-09630-w) 2
- [40] S. Mystakidis and V. Lympouridis. Immersive Learning. *Encyclopedia*, 3(2):396–405, June 2023. doi: [10.3390/encyclopedia3020026](https://doi.org/10.3390/encyclopedia3020026) 2
- [41] B. Naskrent, W. Grzywiński, K. Polowy, A. Tomczak, and T. Jelonek. Eye-Tracking in Assessment of the Mental Workload of Harvester Operators. *International Journal of Environmental Research and Public Health*, 19(9):5241, Jan. 2022. doi: [10.3390/ijerph19095241](https://doi.org/10.3390/ijerph19095241) 3
- [42] National Aeronautics and Space Administration. TLX @ NASA Ames - Home. <https://humansystems.arc.nasa.gov/groups/TLX/>. Last accessed: 25 May 2023. 3
- [43] N. C. Nilsson, A. Zenner, A. L. Simeone, D. Degraen, and F. Daiber. Haptic Proxies for Virtual Reality: Success Criteria and Taxonomy. In *CHI 2021 Workshop on Everyday Proxy Objects for Virtual Reality*, 2021. 2
- [44] K. Noguchi, Y. R. Gel, E. Brunner, and F. Konietzschke. nparLD: An R Software Package for the Nonparametric Analysis of Longitudinal Data in Factorial Experiments. *Journal of Statistical Software*, 50:1–23, Sept. 2012. doi: [10.18637/jss.v050.i12](https://doi.org/10.18637/jss.v050.i12) 5
- [45] K. Omori, N. Shigemoto, H. Kitagawa, T. Nomura, Y. Kaiki, K. Miyaji, T. Akita, T. Kobayashi, M. Hattori, N. Hasunuma, J. Tanaka, and H. Ohge. Virtual reality as a Learning Tool for Improving Infection Control Procedures. *American Journal of Infection Control*, June 2022. doi: [10.1016/j.ajic.2022.05.023](https://doi.org/10.1016/j.ajic.2022.05.023) 2
- [46] D. S. Patle, D. Manca, S. Nazir, and S. Sharma. Operator training simulators in virtual reality environment for process operators: A review. *Virtual Reality*, 23(3):293–311, Sept. 2019. doi: [10.1007/s10055-018-0354-3](https://doi.org/10.1007/s10055-018-0354-3) 2
- [47] A. Ricca, A. Chellali, and S. Otrmane. The influence of hand visualization in tool-based motor-skills training, a longitudinal study. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 103–112. IEEE, Lisboa, Portugal, Mar. 2021. doi: [10.1109/VR50410.2021.00031](https://doi.org/10.1109/VR50410.2021.00031) 5
- [48] V. P. Richmond, J. S. Gorham, and J. C. McCroskey. The Relationship Between Selected Immediacy Behaviors and Cognitive Learning. *Annals of the International Communication Association*, 10(1):574–590, Jan. 1987. doi: [10.1080/23808985.1987.11678663](https://doi.org/10.1080/23808985.1987.11678663) 4
- [49] A. P. Rovai, M. J. Wighting, J. D. Baker, and L. D. Grooms. Development of an instrument to measure perceived cognitive, affective, and psychomotor learning in traditional and virtual classroom higher education settings. *The Internet and Higher Education*, 12(1):7–13, Jan. 2009. doi: [10.1016/j.iheduc.2008.10.002](https://doi.org/10.1016/j.iheduc.2008.10.002) 4, 9
- [50] M. Saghafian, K. Laumann, R. S. Akhtar, and M. R. Skogstad. The Evaluation of Virtual Reality Fire Extinguisher Training. *Frontiers in Psychology*, 11, 2020. 1, 2
- [51] J. Salisbury and M. Srinivasan. Phantom-based haptic interaction with virtual objects. *IEEE Computer Graphics and Applications*, 17(5):6–10, Sept. 1997. doi: [10.1109/MCG.1997.1626171](https://doi.org/10.1109/MCG.1997.1626171) 2
- [52] J. Sauro and J. R. Lewis. *Quantifying the User Experience: Practical Statistics for User Research*. Elsevier, Morgan Kaufmann, Amsterdam Boston Heidelberg, 2nd edition ed., July 2016. 9
- [53] L. Sauter, A. Weigel, and T. Ludwig. It’s Tool Time: Exploring Tool Design Alternatives for Virtual Reality Trainings. *ECIS 2023 Research Papers*, May 2023. 2
- [54] W. Schnotz and C. Kürschner. A Reconsideration of Cognitive Load Theory. *Educational Psychology Review*, 19(4):469–508, Dec. 2007. doi: [10.1007/s10648-007-9053-4](https://doi.org/10.1007/s10648-007-9053-4) 3
- [55] S. Shen, H.-T. Chen, W. Raffe, and T. W. Leong. Effects of Level of Immersion on Virtual Training Transfer of Bimanual Assembly Tasks. *Frontiers in Virtual Reality*, 2, 2021. 2
- [56] A. L. Simeone, E. Velloso, and H. Gellersen. Substitutional Reality: Using the Physical Environment to Design Virtual Reality Experiences. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI ’15, pp. 3307–3316. Association for Computing Machinery, New York, NY, USA, Apr. 2015. doi: [10.1145/2702123.2702389](https://doi.org/10.1145/2702123.2702389) 2
- [57] A. Skulmowski and G. D. Rey. Embodied learning: Introducing a taxonomy based on bodily engagement and task integration. *Cognitive Research: Principles and Implications*, 3(1):6, Mar. 2018. doi: [10.1186/s41235-018-0092-9](https://doi.org/10.1186/s41235-018-0092-9) 2
- [58] P. L. Strandholt, O. A. Dogaru, N. C. Nilsson, R. Nordahl, and S. Serafin. Knock on Wood: Combining Redirected Touching and Physical Props for Tool-Based Interaction in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–13. Association for Computing Machinery, New York, NY, USA, Apr. 2020. 2
- [59] J. Sweller. CHAPTER TWO - Cognitive Load Theory. In J. P. Mestre and B. H. Ross, eds., *Psychology of Learning and Motivation*, vol. 55, pp. 37–76. Academic Press, Jan. 2011. doi: [10.1016/B978-0-12-387691-1.00002-8](https://doi.org/10.1016/B978-0-12-387691-1.00002-8) 3
- [60] T. Takano, K. Takashima, K. Fujita, H. Guang, K. Ikematsu, and Y. Kitamura. A Compact and Low-cost VR Tooth Drill Training System using Mobile HMD and Stylus Smartphone. In *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–3. ACM, Osaka Japan, Dec. 2021. doi: [10.1145/3489849.3489933](https://doi.org/10.1145/3489849.3489933) 2
- [61] E. TICHAUER and HOWARD. GAGE. Ergonomic principles basic to hand tool design. *American Industrial Hygiene Association Journal*, 38(11):622–634, Nov. 1977. doi: [10.1080/00028897708984406](https://doi.org/10.1080/00028897708984406) 2
- [62] M. Vasarainen, S. Paaavola, and L. Vetoshkina. A Systematic Literature Review on Extended Reality: Virtual, Augmented and Mixed Reality in Working Life. *International Journal of Virtual Reality*, 21(2):1–28, Oct. 2021. doi: [10.20870/IJVR.2021.21.2.4620](https://doi.org/10.20870/IJVR.2021.21.2.4620) 2
- [63] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’11, pp. 143–146. Association for Computing Machinery, New York, NY, USA, May 2011. doi: [10.1145/1978942.1978963](https://doi.org/10.1145/1978942.1978963) 5
- [64] B. Xie, H. Liu, R. Alghofaili, Y. Zhang, Y. Jiang, F. D. Lobo, C. Li, W. Li, H. Huang, M. Akdere, C. Mousas, and L.-F. Yu. A Review on Virtual Reality Skill Training Applications. *Frontiers in Virtual Reality*, 2, 2021. doi: [10.3389/frvir.2021.645153](https://doi.org/10.3389/frvir.2021.645153) 2
- [65] X. Ye. A Survey on Simulation for Weight Perception in Virtual Reality. *Journal of Computer and Communications*, 9(9):1–24, Sept. 2021. doi: [10.4236/jcc.2021.99001](https://doi.org/10.4236/jcc.2021.99001) 2
- [66] A. Zenner and A. Krüger. Drag: on: A Virtual Reality Controller Providing Haptic Feedback Based on Drag and Weight Shift. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI ’19, pp. 1–12. Association for Computing Machinery, New York, NY, USA, May 2019. doi: [10.1145/3290605.3300441](https://doi.org/10.1145/3290605.3300441) 2