

Article

Mapping or no Mapping: The Influence of Controller Interaction Design in an Immersive Virtual Reality Tutorial in Two Different Age Groups

Andreas Urech ^{1,2} , Pascal Valentin Meier ^{1,2}, Stephan Gut ³, Pascal Duchene ³ and Oliver Christ ^{1,2,*}

¹ Institute for Research and Development of Collaborative Processes, School of Applied Psychology, University of Applied Sciences and Arts Northwestern, 4600 Olten, Switzerland; andreas.urech@fhnw.ch (A.U.); pascal.meier@fhnw.ch (P.V.M.)

² Institute Humans in Complex Systems, School of Applied Psychology, University of Applied Sciences and Arts Northwestern, 4600 Olten, Switzerland

³ SBB AG, 3000 Bern, Switzerland; stephan.gut2@sbb.ch (S.G.); pascal.duchene@sbb.ch (P.D.)

* Correspondence: oliver.christ@fhnw.ch

Abstract: Navigation and interaction in the virtual world will occur via controllers. Previous literature suggests that not all buttons on these controllers are suitable for all functions and that unclear or uncomfortable operations can lead to frustration. This study aimed to determine if an operation with freely selectable buttons differs from one with mapped buttons, in which human-factor differences can be identified, and if there are differences between primary and advanced training. A field experiment with professionals employed by Swiss Federal Railways and apprentices and future apprentices ($N = 60$) was conducted in a VR tutorial that was previously developed using design cues from existing literature. Controller operation was varied in the groups. The results show significant differences and interaction effects, which indicate that a different operation is more appropriate for apprentices than for professionals in further education.

Keywords: immersive virtual reality; controller interaction; tutorial; age; user experience



Citation: Urech, A.; Meier, P.V.; Gut, S.; Duchene, P.; Christ, O. Mapping or no Mapping: The Influence of Controller Interaction Design in an Immersive Virtual Reality Tutorial in Two Different Age Groups. *Multimodal Technol. Interact.* **2024**, *8*, 59. <https://doi.org/10.3390/mti8070059>

Academic Editor: Mark Billinghurst

Received: 9 June 2024

Revised: 4 July 2024

Accepted: 5 July 2024

Published: 9 July 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The literature on immersive learning (immersive virtual reality/iVR) in curricular learning settings (higher education, continuing education courses, vocational training, etc.) identifies both advantages and disadvantages of the technology [1–9]. For example, in their meta-analysis, Wu et al. [9] analyzed 35 experimental or quasi-experimental studies that addressed the effects of iVR on learning. Despite a positive, albeit small, effect size of $ES = 0.24$ of iVR on learning success, 34% of the studies showed an adverse or no effect. The advantage of iVR using a head-mounted display (HMD, “VR glasses”) should, therefore, not be overestimated based on the results, and a look should be taken at possible influencing factors. In the review made in 2022 by Zhang and Song [10], the effects of sensory cues on immersive experiences to promote technology-supported sustainable behavior was investigated. The authors mention the following technical design factors: dimension, style (e.g., realism), perspective, color, feel, and the sound of cues, music, or dialogue. Regarding important factors for persuasion (scene and story), the authors see authority, sympathy, commitment, reciprocity, and brevity as essential for realization. Currently, factors like interaction with controllers and presumptions in humans like age are overlooked. To close the research gap, we conducted a study in which we empirically tested different interaction designs for iVR controllers in the working environment using quantitative and qualitative methods. The working environment in our research was local public transport in the rail sector. We designed and evaluated an introduction (controller learning and interaction in virtual space) to iVR safety training for the Swiss Federal Railways (SBB). iVR is a

relatively new way to experience potentially dangerous situations in safety training [11]. In iVR, an immersive three-dimensional world is simulated in which one can interact with the environment from a first-person perspective in real-time [12]. The medium is a head-mounted display (HMD, see Figure 1) that allows stereoscopic vision, a larger field of view than a PC monitor, and registers head movements and the viewing height of the person wearing it, which are considered factors that increase immersion [13].



Figure 1. Head-mounted display Meta Quest 2 with additional glass spacer and elite strap and corresponding controllers. Image taken by the authors.

A suitable environment can be simulated and adapted to different conditions through iVR. In addition, elements such as motor interaction (e.g., operation of tools), adaptive and time-controlled processes (e.g., weather or noise), or the integration of 360-degree videos can make training in iVR more realistic than in training on models or supported by 2D representation on the computer [14]. In practice, iVR training environments are used for different scenarios and target groups [11]. This poses particular challenges. Firstly, due to the widespread age and experience of the target group, there are different levels of previous knowledge and prerequisites in dealing with iVR, which must be covered. Secondly, the handling of a technical system must be learned quickly—the existing safety training is limited and should not be extended in duration. Since several people participate in the training simultaneously, close supervision by trainers is impossible, and therefore, the handling should be as intuitive as possible. A better understanding of immersive virtual reality must first be created to determine possible influencing variables on training in iVR and the dimensions that they affect. A virtual world often includes a level of interaction: different virtual tools or objects can be used [14], the environment can be customized [14], and a teleport function helps to cover distances in iVR. This is despite subjects not moving or moving minimally in the real world [15]. If the control of tasks in iVR is unclear or uncomfortable, this can lead to frustration [16]. Thus, controls may play a vital role in the experience of iVR. The most commonly used controls in iVR are hand gestures or the buttons on the iVR controllers. Hand gestures are considered more flexible because one does not need to hold a controller and is not dependent on the number of buttons. However, haptic feedback is lacking, precision is poorer [5], and operation is more complex and less reliable [17]. Muscle fatigue may occur depending on the type of gestures [18]. At the same time, some authors investigated which gestures are more suitable [19]. Y.-J. Huang et al. [5] investigated hybrid approaches by trying a hybrid input, i.e., a combination of controllers and gestures. Their results showed that movements requiring precision and stability should be performed with controllers. It can be concluded that teleport functions, which require aiming at a target, and functions, which require holding objects, work better

with a controller. Furthermore, users need additional cues to determine whether a gesture or the controller controls something. Kangas et al. [20] investigated whether certain input types (hand gestures only, controller and trigger button, or controller and grab button) behaved differently. There were significantly more error interactions in the task between the two controller input types, and they also differed in subjective ratings. For example, the controller and grab buttons felt more natural and gave users more freedom, while the controller and trigger buttons were rated more clear and accurate. This suggests that the position of a button on the controller and subjective preferences influence task performance. The question now arises of whether the assignment of a button on the controller to a function (mapping) differs from a condition without mapping, in which any button can be pressed, and the system recognizes the function. For example, it would be conceivable that pressing any button near or targeting objects that can be interacted with would trigger an interaction with the object. Otherwise, a navigation function such as teleport would be activated. Action theory [21] makes assumptions about these two mapping conditions, which can be investigated with usability measures.

To use iVR controllers, users need to learn the control system, which can be undertaken in different ways: instructions, explanations, trial and error, and tutorials are some of these methods. Since time will be a limited factor in later training, trial and error was eliminated. A guided tutorial is the most suitable, as it can include pictures, and the learning performance is better than words in a manual or explanation [22]. In the tutorial, users should mainly learn how to navigate and interact with the iVR controller in the virtual space. Therefore, it is essential to determine what can influence this learning negatively or positively. Regarding motor learning, it is helpful to learn by demonstrating movements [23]. An animated virtual character (subsequently called the virtual agent) can perform this task in a virtual space [24] and additionally deliver information verbally to users [25]. At the processing level, cognitive load theory explains how information processing load can be reduced in learning tasks [26]. Motivation, with its intrinsic and extrinsic components, also impacts the learning process [27] and should be considered.

Research Questions

The introduction clearly outlines the study's concerns. In advance, hints for designing an intuitive iVR tutorial were to be extracted from existing literature and implemented. Operationalizable theoretical constructs, which formed the basis of the design hints, were to be used as a measurable level in humans.

In a field experiment, differences in the experience of an iVR tutorial were determined by varying two influencing variables. One of these variables was the control of the controllers. In one mapping condition, a function was assigned to each button (With Mapping), while in the other, any button could be pressed (Without Mapping). Differences in age and professional status were the second influencing variable, differentiating between apprentices and professionals.

The following research questions were derived, whereby the procedure for answering Q1 was explorative.

Q1: At what levels are the influence of age and professional status and mapping condition best illustrated in humans?

Q2: Which mapping condition on the VR controllers is more intuitive and effective for users to experience?

Q3: Do the target groups of professionals and apprentices differ regarding the experience described in Q2?

This study is structured so that, to answer Q1, theoretical findings found in a literature search were first reported and considered relevant. Taking these findings into account, an iVR tutorial was created in an iterative process together with subject matter experts from SBB and project support from the University of Applied Sciences Northwestern Switzerland (FHNW), in which an experiment was conducted to answer Q2 and which represents the empirical part of the study. This document does not describe the iterative process in

detail; the final result can be found in the Methods section. The theoretical findings also support the derivation of hypotheses for Q2 and Q3 (see Section 2.7). The following Methods section describes the field experiment's structure and procedure, which were used to test the hypotheses for Q2 and Q3. Finally, the results of the experiment are reported and discussed.

2. Background

This part presents theories and theoretical assumptions in the literature that influence the experience of an iVR tutorial.

2.1. Cognitive Load Theory

When people learn new things, the information is processed in working memory and stored in long-term memory. Cognitive load theory (CLT; [28,29]) suggests that working memory has limited resources for processing and storing information. When the capacity of working memory is exceeded, cognitive overload occurs. This can lead to poorer learning [30] and physical stress responses such as muscular tension [31]. CLT is intended to support the learning process by providing guidelines for designing or presenting information.

In CLT, three different loads on working memory are distinguished [28,29,32]: intrinsic cognitive load (ICL) is described as the load due to the complexity of the learning task. The learner's prior knowledge and the number of objects to be processed simultaneously influence the ICL. Extraneous cognitive load (ECL) refers to claimed resources used in processing learning instructions. ECL stresses the ICL through shared working memory resources. Learning-related cognitive load (germane cognitive load; GCL) does not contribute to the overall load but shifts working memory resources from external to learning-related activities. This is performed by processing information that corresponds to the learning task.

The literature describes different methods for influencing the different cognitive loads. Klepsch et al. [33] describe complexity reduction as a way of reducing ICL, which is performed by building prior knowledge. Sweller et al. [26] note that changing ICL by changing the learning task and, thus, the extraneous cognitive load is possible.

In their systematic review, Noetel et al. [30] list several methods to reduce ECL and cite the reduction in ECL as the most important goal to reduce CL. For example, media combinations, addressing different modalities, omitting distracting or decorative information, or integrating information can impact ECL. When information cannot be reduced, cues or signals help focus on essential details. However, studies with VR show that too many instructions lead users too much, and the effectiveness of the training suffers. This is explained by an overly simple sequence of steps, which is associated with low cognitive effort [34].

To vary GCL, approaches can activate cognitive processes in the learner and promote the formation of mental models or schemas [35]. Concrete measures to promote GCL would be to edit learning material to encourage learners to explain the learning material themselves [33], repeat it, or organize it [35], which can also be prompted. Choices of learning difficulty [35] and teacher or instructional factors such as timing, autonomy, or familiarity [36] can also impact CLC.

The question remains whether iVR influences cognitive load in addition to the learning task. The VR/iVR technology and its possibilities examined in the literature differ, leading to divergent findings (see Table 1). This indicates that both the learning objective and the design of the learning process or the work task influence cognitive load, not necessarily VR/iVR as technology itself.

Table 1. Overview of different results on cognitive load in the literature.

Method	Results	Reference
2D vs. 3D (both on Screen)	Higher CL in 2D	[36]
iVR vs. AR vs. Screen	No difference in CL	[37]
iVR vs. Textbook vs. Video	Higher Engagement (GCL) in VR	[38]
3D static vs. 3D immersive (both on Screen)	Lower CL in 3D static	[39]
iVR vs. 2D (Screen)	Higher CL in iVR	[40]

Measuring Cognitive Load

Subjective methods such as the NASA-TLX [41] or the Paas Scale [42] are widely used because they are easy to use, do not require additional measuring devices, and are easy to apply in practice. The disadvantage is that they are often used after the learning process and usually do not differentiate between the three types of cognitive load. However, more recent scales, such as those by Leppink et al. [43] and Klepsch et al. [33], distinguish between these load types. Andersen and Makransky [44] further developed the Leppink et al. [43] scale. They optimized it for virtual environments by further subdividing ECL (ECL instructions, ECL interactions, and ECL environment), which is helpful for this study.

2.2. Action Theory

Action theory is a cognitive theory that deals with the processes between environmental influences and behavior. It seeks to understand the relationship between the performance of a task and the underlying thoughts. The theory identifies six action steps: goal development, orientation, planning, decision, execution control, and feedback. Each step contains parameters that influence its outcome [21]. According to Hacker (cited in [45]), there are three levels of action regulation on which these parameters can be arranged: the cognitive level, the level of flexible action patterns, and the sensorimotor level.

The action structure is organized hierarchically and sequentially, with action goals divided into subgoals and functional units. These are processed in a hierarchy from the intellectual level (determining the general action goal) to the level of flexible patterns (planning execution strategies) to the sensorimotor level (executing actions). Repeated practice leads to routinization of the action, which means that less effort, decisions, and feedback are required [21].

Action Errors and Problems in Action Regulation

Various errors can occur in the action process, which can be categorized into one of the six action steps and three levels of regulation (see Figure 2). The theory suggests that higher-level errors affect newer users, require more time to solve, and often need external help. In comparison, lower-level errors affect more experienced users, require less time to solve, and can be corrected by the users themselves [21].

	Goal development	Information integration	Prognosis	Plan development & decision	Monitoring	Feedback
Level of intellectual regulation	Goal setting errors - e.g. unclear goals	Mapping errors - e.g. missing information	Prognosis errors - e.g. difficulties system predict behavior	Thought errors - e.g. wrong decisions are made	Memory errors - e.g. parts of a plan are forgotten	Judgement errors - e.g. feedback cannot be understood
Level of flexible action patterns	Habit errors - e.g. correct action is performed in a wrong setting				Omission errors - e.g. interruption in task	Recognition errors - e.g. hints are confused with another one
Sensorimotor level of regulation	Movement errors - e.g. pressing wrong buttons, stumbling					

Error handling time ↑

Figure 2. Taxonomy of errors according to levels of regulation, adapted from Frese and Zapf (1994), p. 290.

In addition to errors in action, stressful situations can affect the action steps and regulation levels, so-called regulation problems. They can be divided into regulatory obstacles (e.g., interruptions and additional workload), regulatory uncertainty (e.g., delayed feedback, quantitative overload, and role conflict), and regulatory overload (e.g., time pressure and the concentration required) [21].

In this study, learning to control the iVR environment was designed using the theory of action so that the user needed to use as little cognitive effort as possible and there were as few handling errors as possible. The system should also help make assumptions about

the advantages and disadvantages of different mapping conditions in VR. Usability results should provide indications of handling errors or control problems.

2.3. Usability

Usability is defined as the extent “to which specified users can use a system, product or service to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context” [46]. According to this definition, each technical system must be considered in its context of use and with its group of users in terms of effectiveness, efficiency, and satisfaction. This is important because technical systems, their users, and even the context of use differ, and no system-wide or generalized design features can or should be applied [47]. To identify discrepancies between the requirements of the user tasks and the technical system at an early stage and thus to save effort by making changes in the advanced system, it is essential to evaluate the technical system—or prototypes of it—with the corresponding user group already in the development phase [48].

Measuring Usability

Usability evaluation of iVR technologies is challenging because traditional methods are not optimized. Bowman et al. [49] identify problems such as limited visibility through HMDs, evaluator interference, and difficulties with logging. Newer iVR technologies, such as Meta Quest 2, address these issues through cameras, screen transmission, multiplayer capabilities, and automatic logging [50]. According to Kamińska et al. [18], subjective survey methods have the disadvantage of being biased. Accordingly, users report their impressions after the VR interaction, not necessarily their experiences in the system, as the virtual information may no longer be present. This study considered objective measures such as time, goal achievement, and number of errors more useful.

2.4. Intrinsic Motivation

Motivation is generally thought of as the drive to act. Ryan and Deci [51] distinguish between extrinsic motivation (driven by external incentives) and intrinsic motivation (driven by internal satisfactions such as interest and enjoyment). Deci et al. [27] state that extrinsic incentives undermine intrinsic motivation, except verbal rewards, which can enhance it. However, the positive effect did not occur when the tasks were boring or when the type of verbal reward was perceived as controlling. In addition, the effect was less likely to occur in children [27]. According to Ryan and Deci [52], motivation plays an essential role in learning and can lead to high-quality education. Studies such as Ai-Lim Lee et al. [52] show positive effects on learning performance, effectiveness, and goal attainment. However, there is also evidence of adverse motivational effects, especially under the pressure of test situations [53].

As mentioned in Makransky et al. [40], several studies have found higher motivation scores in a desktop or iVR setup than in traditional methods but no consistent results on learning success in iVR. However, recent meta-studies find positive learning success when learning with iVR [9,54]. Wolfartsberger et al. [34] found that, depending on the design, the entertainment value of training in virtual environments can be higher than in the real world. However, the authors recommend a compromise between entertainment and learning because learning suffers.

It should be noted that the novelty of a new technical system such as iVR can lead to a temporary increase in engagement and interest and, thus, to an increase in motivation. However, this does not necessarily lead to internalizing the learned material or achieving goals. Such so-called “novelty effects” were found by Jeno et al. [55] in the context of learning with mobile devices. Nevertheless, the novelty effect in virtual reality does not appear to fade very quickly in virtual reality: W. Huang et al. [56] could not demonstrate any significant diminishing of the effect in three interventions within 8–16 days.

Measuring Intrinsic Motivation

Intrinsic motivation is a latent construct and cannot be measured directly. Questionnaires survey the subjective experience of respondents during an activity. There are

different scales used for general or specific activities, for example, the Sport Motivation Scale (SMS-II, [57]), the Motivated Strategies for Learning Questionnaire (MSLQ, [58]) or the Intrinsic Motivation Inventory (IMI, [59]). A short version of the IMI is the German “Kurzskala intrinsischer Motivation” (KIM, [60]), which uses a reduced number of 12 items in four constructs: “interest/enjoyment”, “perceived competence”, “perceived choice”, and “pressure/tension”. The questionnaire was developed to measure students’ intrinsic motivation when visiting an out-of-school learning site. In the iVR tutorial, such an extracurricular learning place was simulated in the context of the safety training at the SBB, which is why high reliability can be assumed. The questionnaire did not consider virtual agents, which can bring motivational effects.

2.5. Virtual Agents

Virtual agents often promote user motivation [61] or satisfaction with the learning process [62]. This occurs through a specific behavior (e.g., positive emotional expressions) or the agent’s appearance [61,63]. Baylor [61] emphasizes the importance of designing the virtual agent as an ideal social model, i.e., its visual presence and behavior adapted to the context of use. Evidence from other studies shows that virtual agents, especially those with human-like appearance, positively affect the learning experience (e.g., perceived task difficulty, entertainment, and maintenance of concentration; for a more detailed review, see [64]).

In addition to motivational effects, virtual agents offer more. In training situations, users often need external guidance. This can be a demonstration of a movement, which is particularly effective for motor learning [23], or verbal commands and information that address different modalities [30]. While in the real world, this is undertaken by a trainer, in computer-based learning or iVR settings, an avatar or virtual agent can take over these instructional tasks [24,25]. Virtual agents with high interactivity (e.g., gestures, animations, and speech) can create cognitive load (ECL), especially for people with little prior knowledge of the topic or system to be learned [62]. Fewer or non-interactive virtual agents seem more appropriate for people with little prior knowledge, but people with considerable previous knowledge benefit from existing interaction [62].

Measuring the Impact of the Virtual Agent

Often, the effects of a virtual agent on a construct, such as motivation or cognitive load, are measured using questionnaires and learning outcomes with follow-up. Interestingly, while some studies report positive effects of a virtual agent on motivation, few report positive learning outcomes [65]. To explain this, Baylor and Ryu [65] developed the Agent Persona Instrument (API), a questionnaire that measures affective interaction and information utility. The basic idea is that the virtual agent alone will not produce a positive learning effect if the affective interaction is not necessary or the information provided is invalid.

2.6. Age Differences in Learning with iVR

Verneau et al. [66] point to an age difference (22 vs. 58 years) in the type of motor learning (explicit vs. implicit). Explicit motor learning refers to the conscious learning of repetitive movement sequences, while implicit motor learning does not require the active acquisition of movement sequences. Their results showed that explicit learning declines with age—younger people benefited more from explicit learning instructions, which enabled them to cope better with an additional secondary task. For older people, explicit instruction had a positive effect when there was more time to practice the movement sequences [66].

In studies of learning or with learning-like tasks using computers or iVR, younger individuals appear to have an advantage: in navigation tasks through a 3D space on a computer, younger people are faster in comparison (18–46 years vs. 46+ years), and they also find it subjectively easier [67]. Visual cues significantly improved the times of older people but did not eliminate the age difference. When comparing an application on a computer and the same application in iVR, more errors were detected in older persons than younger ones [68].

In different tasks in iVR, Domin et al. [69] reported that all individuals had problems learning interactions in iVR. The reason identified was that—mainly by older individuals (over 30 years old)—virtual instruction texts were not understood. Faster processing times for the younger persons were attributed to prior knowledge: younger persons would already be familiar with similar controllers and especially knew the functions of the buttons.

Based on the described effect on humans through the design of iVR elements, research question Q1 is now to be answered, and hypotheses on Q2 and Q3 are to be formed.

2.7. Hypothesis

The design of iVR elements affects people on different levels, which were theoretically described in the previous section. The aim of question Q1 was to identify some critical levels and their measures where usability, cognitive load, and intrinsic motivation were perceived as necessary. With this theoretical knowledge and evidence gained, hypotheses will now be formulated for questions Q2 and Q3. Further, this will be used to develop the iVR tutorial (see Section 3.3).

H1: More invalid interactions in the With Mapping condition.

In the With Mapping condition, associated buttons must be learned for each function. This requires using potentially non-preferred fingers [20], leading to different action goals and a higher likelihood of errors. These errors are thought to occur at the sensorimotor level (movement errors) and the level of flexible action patterns (habit errors), as depicted in Figure 2. Results from Shibata et al. [70] show that, for example, already internalized button assignments cannot be optimally re-learned due to previous experience (e.g., with VR or game controllers). So-called overlearning (learning after maximizing performance improvement) of one task interferes with learning a new task. It is, therefore, hypothesized that when individuals work with a controller whose associated functions are mapped to individual buttons, more invalid interactions occur than when individuals work with a controller whose related functions are not mapped to separate buttons.

H2: Higher processing time in the Without Mapping condition.

The elimination of the different functions in the Without Mapping condition eliminates problems with errors at low levels of regulation. However, it is conceivable that omission and prognosis errors may occur when multiple buttons are pressed simultaneously or when all buttons are released. According to Frese and Zapf [21], these errors are at a higher level of regulation and therefore require more time to correct. This leads to the following hypothesis: when people work with a controller whose functions are not mapped to individual buttons, the processing time will be longer than when people work with a controller whose functions are mapped to separate buttons.

H3: Fewer buttons pressed in subjects with professional educational status.

The literature suggests that individuals of all ages face challenges when learning iVR interactions [69]. Given that professionals are familiar with the practical context of the tutorial, it is hypothesized that older professionals may require less trial and error, resulting in fewer button presses when learning iVR features compared to younger apprentices. Therefore, the following hypothesis is made: when older subjects who are already professionals learn functions for operation in iVR, they press fewer buttons than younger subjects who are apprentices.

H4: Higher cognitive load and lower subjective usability score in the With Mapping condition.

It is hypothesized that errors at higher levels of regulation require more cognitive resources, leading to more cognitive effort (measured by MCLSVE [44]) and thus to a lower subjective usability score (measured by SUS, Brooke, 1986). Therefore, the following hypothesis is formulated: when individuals work with a controller whose associated functions are mapped to individual buttons, they have a higher cognitive load and a lower subjective usability score than when they work with a controller whose related functions are not mapped to separate buttons.

H5: More positive evaluation of the virtual agent in subjects with apprentice educational status.

Apprentices had little or no prior knowledge of the practical context where the safety training takes place. Based on the findings of Schöbel et al. [62], it is hypothesized that a virtual agent guiding through this unfamiliar context will be perceived positively by apprentices. Therefore, the following hypothesis is proposed: when younger people who are still apprentices interact with a virtual agent in iVR, they will evaluate it more positively (measured API [65]) than when older people who are already professionals interact with a virtual agent in iVR.

H6: A higher value of intrinsic motivation in subjects with professional educational status.

Professionals are assumed to be more tense due to the safety training but have a higher intrinsic motivation due to the practical context. Therefore, the following hypothesis is made: when professionals experience the iVR tutorial in safety training, they will show higher motivation scores (measured with KIM [60]) than apprentices who experience the iVR tutorial in an extracurricular visit to a job fair.

3. Materials and Methods

The following describes the design, participants, material, procedure of the study, and data processing and analysis.

3.1. Design

A 2×2 between-subjects design was used with two different button mappings on the VR controller (With Mapping vs. Without Mapping) and other characteristics in age and professional status (apprentice vs. professional).

Processing time, faulty interactions, number of buttons pressed, and SUS, GUESS, API, KIM, and MCLSVE scores calculated from the follow-up questionnaire questions (see Section 3.3) were used as dependent variables. Prior experience with VR served as a control variable.

3.2. Participants

The required number of participants was calculated a priori by the software G*Power [71]. For the chosen statistical procedure (see Section 3.5), a sample size of $N = 60$ was calculated with an estimated moderate effect size of $f^2(V) = 0.15$, a test power of 0.80, four groups, and eight dependent variables. This sample size was achieved with a balanced group size ($n = 15$ each).

Of the 60 participants, 30 were professionals who had previously completed SBB regular safety training. The other 30 participants were apprentices and future apprentices recruited at a career fair at the stand of one of SBB's education partners (login) because they were interested in an apprenticeship in the railway industry. The average age of the professionals was 33.97 years, with a range of 22 to 56 years, while the average age of the apprentices was 14.57 years, with a range of 13 to 20 years. Three participants were excluded from the sample because they completed the tutorial early due to technical problems. All participants took part in the study voluntarily and received no financial compensation. Informed consent was obtained from all participants before participation.

3.3. Material

The following is a description of the material used, which includes the iVR tutorial, hardware, and a questionnaire.

3.3.1. iVR Tutorial

According to the theory of action [21], an action always needs a goal. To reach this goal step by step, actions are divided into smaller, hierarchically and sequentially structured sub-actions. In most cases, cognitive resources are addressed first, and an action is planned before the actions are muscularly executed. After repeated execution of an action, it can become automated, requiring fewer cognitive resources. The tutorial learning process uses this principle to learn actions in iVR.

To reduce processes on the intellectual level, users were given the goal of learning four functions: teleporting, grasping tools, operating a tool with both hands and answering a radio call with a radio. Six sub-goals, representing step-by-step instructions, were defined for these four functions.

Care was taken to ensure that the tasks of the subgoals were as simple as possible at the beginning (e.g., find button, press button) and did not involve any combinations. This way, errors at higher intellectual levels should be excluded. According to the CLT, this task division should also bring advantages: low ICL should be achieved by low complexity [33], and at the same time, prior knowledge could be built up this way. Subsequently, task difficulty was gradually increased (e.g., find the target and aim, then aim and press button), and then the task was practiced three times until a new function was presented as the next goal. They repeated the task thrice to internalize or routinize the procedures. Figure 3 shows the described hierarchical–sequential structure of the goals using two goals as examples.

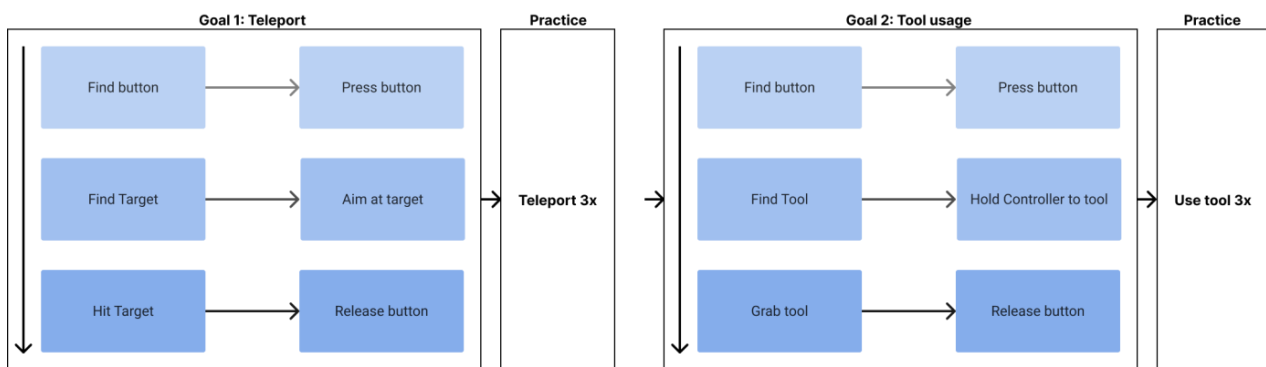


Figure 3. Hierarchical–sequential subdivision of the goals.

Further inputs from the literature were added: the combination of media and modalities [30] was implemented by text modules and a virtual agent that reproduced the text verbally [25]. Furthermore, it showed the users all tasks in parallel, including pressing buttons, aiming at imitation [24]. The agent was presented as an entity wearing a high-visibility vest and helmet, with the two VR controllers hovering in front of her (see Figure 4). In a user-centric design approach, we deliberately omitted a body in the virtual environment to prevent users from assuming a 3rd-person view. A female voice, introduced as Coach Emma, was used for the virtual agent, generated by a text-to-speech interface by default. This design choice enhanced the user’s immersion and engagement with the tutorial.

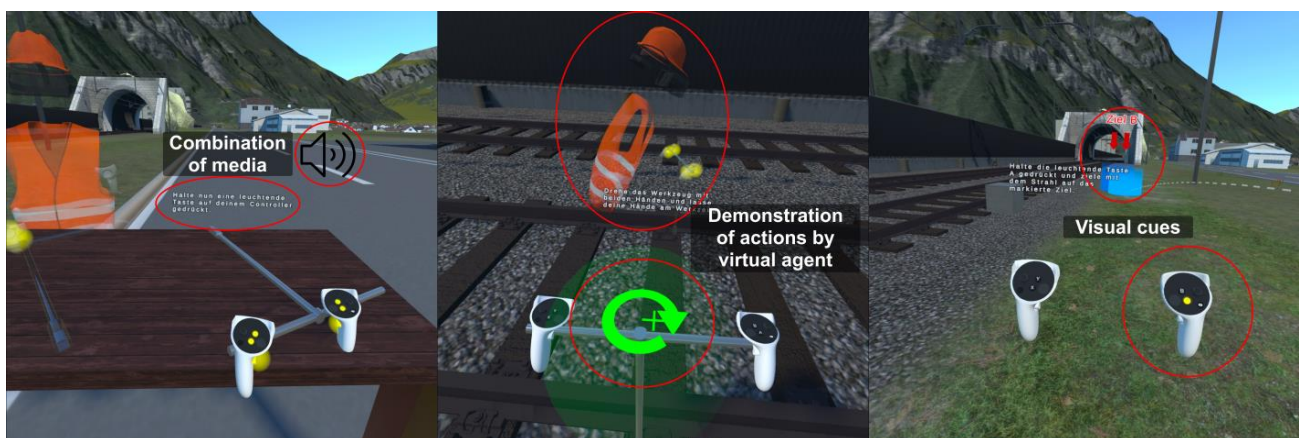


Figure 4. Virtual agent, visual cues, text blocks, highlighted buttons, and controllers.

Contrary to the recommendations of Noetel et al. [30], we retained decorative elements in the environment to recreate a real-life situation near the railroad tracks. This practical

design decision made the tutorial more relatable and engaging for the users. As Sayers suggested, visual cues, such as navigation markers and highlighted interactive objects, were implemented [68]. A map was not used as a navigation aid due to the manageable environment [67]. Two links can be found in the Supplementary Material, which provide an insight into the tutorial used.

3.3.2. Hardware

The comfort and support of the users were prioritized in the hardware selection. A Meta Quest 2 system in standalone mode (without connection to a PC) was used as the HMD. Glass spacers and elite straps were added to increase comfort and support for glasses wearers. Participants perceived audio through the built-in surround sound speakers, ensuring a seamless and comfortable learning experience.

The controllers that users held in their hands were visible in the virtual environment (see Figure 4). As suggested by Domin et al. [69], we refrained from mapping virtual hands. The buttons for each mapping condition can be seen in Figure 5 and were highlighted in a bright color in the virtual world (Figure 4). This improved interaction technique and highlighted instructions, both to lower ECL.

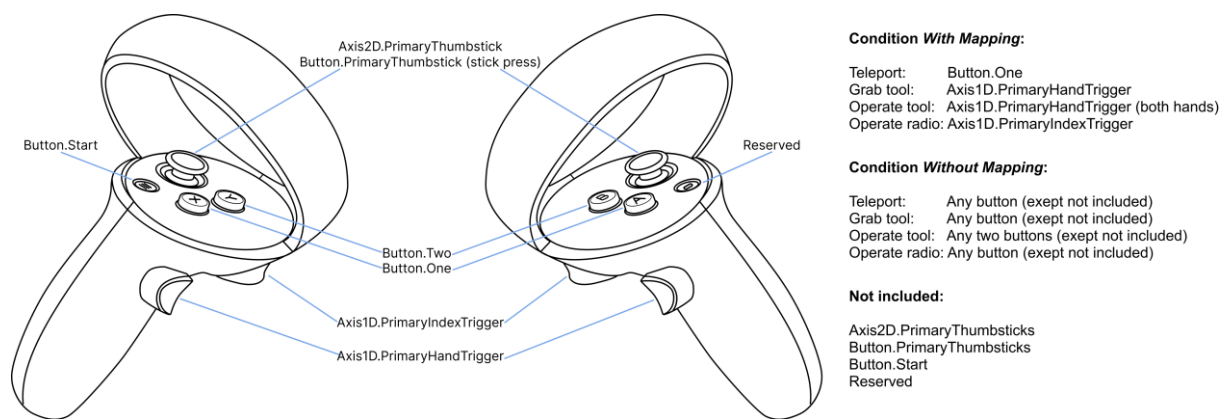


Figure 5. Controller and mapping. Use of the image authorized by the creator.

The hardware allowed usability metrics such as time, buttons pressed, and incorrect interactions to be recorded and exported directly. Invalid interactions were logged when prompts had to be repeated by the virtual agent or the system. This happened, for example, when the user aimed at a marked point for teleportation but did not release the button pressed or did not release the tool after being prompted to do so in the tool-grasping task. What were not recorded as incorrect interactions were buttons that were pressed incorrectly. As this was only possible in one mapping condition, this would have distorted or made the data unusable.

3.3.3. Questionnaire

An online survey tool [72] was used to create a questionnaire that users completed after completing tasks in VR. It contained 46 items on usability, CL, and motivation, 2 open-ended questions on likes and dislikes, and 5 on prior experience and demographic data. All items were translated into German and adapted to the context. A complete overview can be found in Appendix A. Some items and constructs were excluded (shaded grey in Appendix A). The following is an overview of the literature from which the questions were taken and an explanation of why some items were excluded.

The 10-item SUS was used to measure usability, which has a reliability of Cronbach's alpha = 0.91 and is widely used in research [73].

Eleven items of the MCLSVE questionnaire were used to measure cognitive load, with a reliability of 0.85 [44]. Only items measuring the different ECL (ECL instructions, interactions, and environment) were used because they were considered relevant to the study.

To evaluate the virtual agent, the “Facilitating Learning” and “Credible” subconstructs from the Agent Persona Instrument questionnaire [65] were used. These subconstructs have reliabilities of Cronbach’s alpha = 0.85 and 0.65, respectively, and assess the usefulness of the information conveyed [65]. The Human-Like and Engaging items were excluded because the virtual agent was not targeted for effective interaction.

For measuring motivation, “Interest/Pleasure”, “Perceived Competence”, and “Pressure/Tension” from the KIM [60] were used, which have reliability of Cronbach’s alpha = 0.79, except for the construct “Pressure/Tension”, which only has a reliability of Cronbach’s alpha = 0.54. Nevertheless, this construct was included because the safety training, as part of which the tutorial was to take place, was conducted in a short period and concluded with an examination—thus, a tense situation is quite conceivable. On the other hand, the construct “Perceived choice” was omitted because the tutorial was guided and the users did not influence the process flow, and thus no variability in the responses was expected.

3.4. Procedure

The study took place over seven days in 2022 at three locations—test sites A, B, and C (see Figure 6). At sites A and B, participants were working people who had attended an SBB safety training course on the same day. The rooms could accommodate nine (Site A) or ten (Site B) participants simultaneously, but this was never the case. Site C was located at a career fair attended by students and apprentices. This site differed from A and B because it had more limited space (max. two simultaneous participants) and a higher noise level. Technical limitations such as poor Internet connection, bright lights, or reflective materials that interfered with the HMD sensors were absent at all of the three locations.



Figure 6. Test sites (A–C).

The explanation of iVR included the spatial limit and the consequences of exceeding it, instructions on how to put on and take off the HMD, and the possibility of experiencing nausea. After drawing one of the two test conditions, With Mapping or Without Mapping, participants could freely choose one of the placed HMDs. In the With Mapping condition, all functions were assigned to a fixed button (refer to Figure 5). In contrast, it did not matter which button was pressed in the Without Mapping condition. By default, the teleport function was active, and the system detected whether the user was near an interactable object and aiming at it. If so, the function was automatically switched to grasp or operate, depending on the object.

When starting the application, the study management supported the participants; subsequently, the users worked through the tasks in the iVR tutorial independently and without a time limit. The participants had to move around using the teleport function in the first task. To do this, they had to press the primary or any other button, aim at a target, and then release the button. This was followed by the second task, where a tool lying on a table had to be grasped and released. One of the controllers had to be held up to the tool, and the grip or any other button had to be pressed (or released). In the third task, the previously learned tool had to be used. It had to be gripped the same way as before, and then held against a screw and turned while both hands grabbed the tool. Finally, in the last task, a radio call had to be returned via radio. The radio was hung around the virtual body;

it became visible when looking down. To operate it, one of the two controllers had to be held up to the radio, and the trigger button or any button had to be pressed. The system automatically detected when participants had completed a task and triggered the next task. The task instruction was repeated if an incorrect interaction was made, or no interaction occurred. After the last task, the system asked the participants to remove the HMD. Finally, the users completed the questionnaire on their smartphones or a provided laptop. During data collection, the test supervisors documented any unusual observations in writing and attributed them to the corresponding participants.

3.5. Data Preparation and Analysis

The raw data from the HMD export were processed so that processing time (s), number of buttons pressed, and errored interactions could be read as variables and assigned to a session ID. The HMD data and the data from the questionnaire were then merged and cleaned, and scores for the constructs of the dependent variables from the questionnaire were calculated.

A factorial MANOVA (multivariate analysis of variance) was chosen for the data analysis, which looks for linear trends in several dependent variables. Its advantage over multiple separate ANOVAs is that it does not have to account for the accumulation of alpha errors and shows the combined effect of the dependent variables as a result of the independent variables [74]. Field [75] indicates that only relevant dependent variables should be included in a MANOVA. The dependent variables included in this study all impact the experience of iVR, which is why this procedure is considered appropriate.

Using the MANOVA procedure, group differences can be detected that might not be apparent when looking at individual dependent variables—yet significant effects can still be examined individually [74] because the MANOVA will be calculated with two independent variables, primary and interaction effects can be detected. In addition, multiple independent variables in a MANOVA offer the advantage of at least attempting to reduce error variation [74]. Data analysis was conducted using SPSS 27 statistical software [76] and R [77–79].

4. Results

The results of the study are presented below. First, descriptive results are reported. Then, the examination of the prerequisites of the data analysis is described, followed by the results of the MANOVA.

4.1. Descriptive Results

In addition to providing an overview of the data, this part of the study also explores the data from the individual tasks in the tutorial and certain buttons. These data cannot be analyzed using statistical methods because preconditions are not given. In the With Mapping condition, for example, it is assumed that not just any buttons are pressed but those mentioned in the tutorial. Furthermore, the individual tasks are similar in structure but not identical in a way that would make comparisons meaningful. Nevertheless, these data may be included to provide possible explanations or recommendations for action.

4.1.1. Means and Standard Deviations of Dependent Variables

Table 2 shows the means and standard deviations of the dependent variables by mapping condition and professional status. There is no fixed range of values for Processing Time, Erroneous Interactions, and Number of Buttons Pressed, i.e., values from 0 to infinity are theoretically possible. The group of professionals with the Without Mapping condition stands out with more significant standard deviations in Processing Time and Erroneous Interactions. Also interesting are the different mean values in processing time(s) between the groups: while the apprentices in the With Mapping condition achieved higher values, the group of professionals in the Without Mapping condition did so too.

Table 2. Group sizes, means, and standard deviations of the dependent variables, buttons pressed simultaneously, and prior experience with iVR, broken down by professional status and mapping condition.

Variable	Professional Status	Mapping Condition					
		With Mapping			Without Mapping		
		n	M	SD	n	M	SD
Processing time (s)	Apprentice	15	401.27	88.90	15	371.27	80.60
	Professional	15	387.80	82.40	15	439.60	143.10
Invalid interactions	Apprentice	15	31.60	8.50	15	27.8	9.89
	Professional	15	23.30	6.18	15	34.1	16.18
Number of buttons pressed	Apprentice	15	164.0	61.12	15	133.8	44.95
	Professional	15	84.33	33.37	15	94.50	38.94
Number of buttons pressed simultaneously	Apprentice	15	66.30	40.06	15	61.50	30.73
	Professional	15	56.6	32.90	15	53.60	38.60
Score SUS	Apprentice	15	73.67	14.63	15	78.33	12.42
	Professional	15	84.50	12.86	15	78.00	7.75
	Apprentice	15	6.91	2.95	15	5.32	2.14
	Professional	15	4.53	1.36	15	4.85	1.15
Score KIM	Apprentice	15	11.07	1.33	15	11.07	1.56
	Professional	15	12.31	1.22	15	11.73	0.88
Score API	Apprentice	15	7.86	1.26	15	7.65	1.06
	Professional	15	7.43	0.90	15	7.24	0.89
Previous experience with iVR (dichotomous)		n	Yes (n)	No (n)	n	Yes (n)	No (n)
	Apprentice	15	10	5	15	9	6
	Professional	15	9	6	15	7	8

Annotation. N = 60. Time was measured in seconds: M = mean, SD = standard deviation, n = number of participants.

Limited ranges of values are possible for the remaining variables. In the SUS score, values between 0 and 100 can be achieved. Since the SUS is widely used in research, studies exist that help interpret the score. A score of approximately 70 is considered “Good”, while 85 is “Excellent”, and 100 is “Best Imaginable” [80], (p. 121). While values between 3 and 15 are possible in KIM and MCLSVE scores, they are 2 to 10 in API score. In these variables, it is noticeable that values in SUS score of over 73 are to be assessed as relatively high. The values of API score, with 7.24 in the group with the lowest value, are also classified as relatively high. In addition, previous experience with iVR, which was included as a control variable, should be considered. It showed similarly pronounced frequency values across all four groups.

4.1.2. Differences in Parts of the Tutorial and Individual Buttons

A breakdown was made according to the individual tasks to find out which buttons were pressed and how often, on average, in which part of the tutorial. Despite the high standard deviations in the mean, an attempt has been made in Figure 7 to give an overview of any trends and also to show the median.

In both mapping conditions, there is an apparent decrease in button presses on average from learning to trying. In addition, students press many buttons when learning to use the radio. Interestingly, students in the With Mapping condition pressed the Grip button more often ($M = 23.53$) than in the Without Mapping condition ($M = 19.86$), even though this button was not functional there. Furthermore, in the same group and mapping condition, there is a maximum mean of $M = 34.26$ for the activation of the trigger button, which is

more than three and a half times higher than for the professionals in the same condition and task ($M = 9.20$). It is also noticeable that the trigger and the grip buttons were pressed more often during the same task.

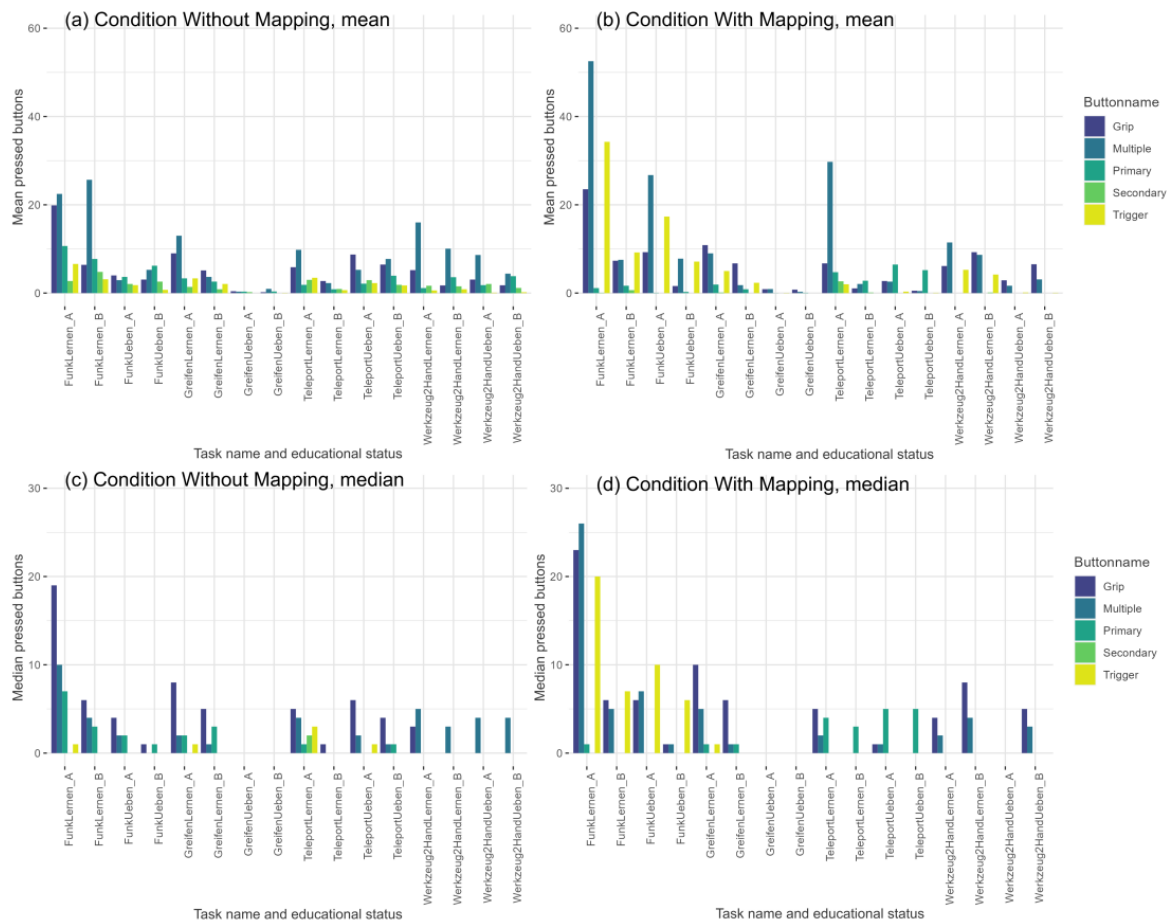


Figure 7. Mean and median number of buttons pressed, divided by task and button name. Multiple = several buttons pressed simultaneously. A = apprentices, B = professionals.

It is also noticeable that in the No Mapping condition, i.e., when the button was self-selectable, apprentices used the grip button most often in all tasks. Only when learning and practicing teleportation and grasping were more grip buttons pressed—when practicing grasping, the primary button was pressed again more often. On average, the secondary button was not used most often in any task in the No Mapping condition, and it was rarely used in the other mapping condition.

4.2. Checking the Requirements of MANOVA

The violation of the MANOVA requirements was checked. The results showed that the assumption of normal distribution was violated ($p > 0.05$) in some groups for the dependent variable processing time, incorrect interactions, number of buttons pressed, SUS score and MCLSVE score. A Johnson transformation [81] was used to bring the data closer to a normal distribution; the transformation was performed using online software [82]. After the transformation, no variable violated the assumption of normal distribution ($p < 0.001$).

4.3. Results of the MANOVA, Post Hoc Results, and Answering the Hypotheses

The MANOVA results showed a significant main effect of educational status $F(8, 49, p < 0.001, \text{partial } \eta^2 = 0.481, \text{ Pillai trace} = 0.481)$ and an interaction effect between educational status and mapping condition $F(8, 49, p < 0.034, \text{partial } \eta^2 = 0.275, \text{ Pillai trace} = 0.275)$. Since

there was no main effect in the mapping condition $F(8, 49, p = 0.506, \text{partial } \eta^2 = 0.131, \text{ Pillai trace} = 0.131)$, hypotheses H1 (more erroneous interactions in the With Mapping condition), H2 (higher processing time in the Without Mapping condition), and H4 (higher cognitive load and lower subjective usability score in the With Mapping condition) can be rejected at this point (see Table 3).

Table 3. Overview and accepting the hypotheses.

Hypothesis	<i>p</i> -Value	(<i>p</i>) η^2	Decision
H1: More invalid interactions in the With Mapping condition.	0.361	0.015	Rejected
H2: Higher processing time in the Without Mapping condition.	0.833	0.001	Rejected
H3: Fewer buttons pressed by subjects with professional educational status.	<0.001	0.301	Accepted
H4a: Higher cognitive load in the With Mapping condition.	0.319	0.018	Rejected
H4b: Lower subjective usability score in the With Mapping condition.	0.629	0.004	Rejected
H5: More positive evaluation of the virtual agent by subjects with apprentice educational status.	0.120	0.043	Rejected
H6: A higher value of intrinsic motivation in subjects with Professional educational status.	0.005	0.131	Accepted

Annotation. (*p*) η^2 = Partial Eta Squared.

A one-way ANOVA was calculated post hoc for all dependent variables. The results with means and standard deviations per group are shown in Table 4. In professional status, there were significant differences in the dependent variables, number of buttons pressed JS, KIM score, and MCLSVE JS score. In the interaction professional status* condition, there were significant differences in three dependent variables: error interaction JS, SUS JS score, and MCLSVE JS score. Since no significant effects were found in the dependent variables Editing Time JS, Guess score, and API score, hypotheses H1 (more incorrect interactions in the With Mapping condition) and H5 (more positive evaluation of the virtual agent by subjects with Apprentice professional status) can be rejected.

Table 4. MANOVA post hoc results (one-way ANOVA).

Variable	With Mapping			Without Mapping			ANOVA			
	n	M	SD	n	M	SD	Effect	<i>p</i> -Value	<i>F</i> (1, 56)	(<i>p</i>) η^2
Processing time JS										
Apprentice	15	−0.05	1.15	15	−0.39	1.06	OS	0.213	1.589	0.028
Professional	15	−0.11	0.87	15	0.34	1.04	CN	0.833	0.045	0.001
							OS × CN	0.147	2.162	0.037
Number of buttons pressed JS										
Apprentice	15	0.75	1.01	15	0.30	0.69	PS	<0.001	24.153	0.301
Professional	15	−0.65	0.75	15	−0.51	0.98	CN	0.476	0.514	0.009
							PS × CN	0.196	1.709	0.030
Invalid Interactions JS										
Apprentice	15	0.47	0.76	15	0.02	0.98	PS	0.357	0.862	0.015
Professional	15	−0.46	0.83	15	0.47	1.31	CN	0.361	0.850	0.015
							PS × CN	0.009	7.224	0.114
Score SUS JS										
Apprentice	15	−0.45	1.16	15	−0.05	1.01	PS	0.093	2.295	0.050
Professional	15	0.53	1.21	15	−0.13	0.59	CN	0.629	0.237	0.004
							PS × CN	0.050	4.013	0.067
Score GUESS										
Apprentice	15	20.27	2.24	15	18.58	2.27	PS	0.477	0.513	0.009
Professional	15	19.60	2.68	15	20.03	1.19	CN	0.262	1.285	0.022
							PS × CN	0.061	3.645	0.061
Score KIM										
Apprentice	15	11.07	1.34	15	11.07	1.56	PS	0.005	8.465	0.131
Professional	15	12.31	1.22	15	11.73	0.88	CN	0.383	0.772	0.014
							PS × CN	0.383	0.772	0.014

Table 4. Cont.

Variable	With Mapping			Without Mapping			ANOVA			
	n	M	SD	n	M	SD	Effect	p-Value	F(1, 56)	(p) η^2
Score MCLSVE JS										
Apprentice	15	0.73	1.03	15	−0.09	1.24	PS	0.020	5.945	0.093
Professional	15	−0.46	0.96	15	−0.16	0.79	CN	0.319	1.011	0.018
							PS × CN	0.038	4.503	0.074
Score API										
Apprentice	15	7.86	1.26	15	7.65	1.06	PS	0.120	2.494	0.043
Professional	15	7.43	0.90	15	7.24	0.89	CN	0.466	2.038	0.035
							PS × CN	0.970	0.001	0.000

Annotation. N = 60. ANOVA = analysis of variance; PS = professional status; CN = condition; × = interaction effect; bold = significant result; JS = Johnson transformation; (p) η^2 = Partial ETA Squared; M = mean, SD = standard deviation; n = number of participants.

Figure 8 shows the significant individual effects in graphical form. Figure 8A shows the main impact of professional status in the dependent variable number of buttons pressed BC. The apprentice group pressed significantly more buttons than the professional group (1.10, $p_{tukey} < 0.001$). As this effect is significant, hypothesis 3 (fewer buttons pressed by professionals) could be confirmed.

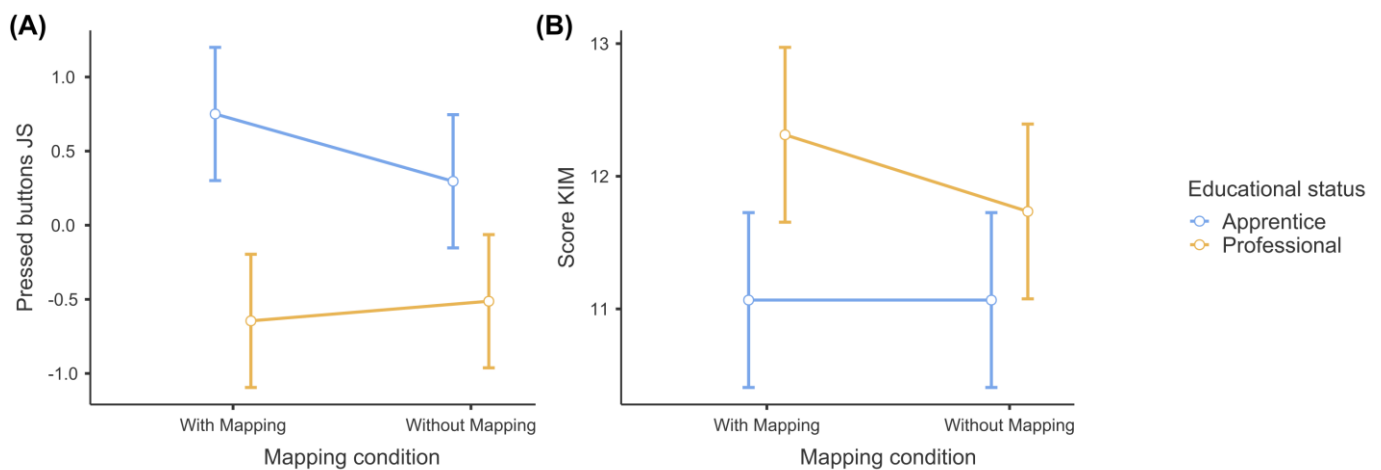


Figure 8. Results of post hoc ANOVA. Main effects in (A) pressed buttons JS and (B) KIM score. N = 60. Error bars correspond to a 95% confidence interval of the mean.

In Figure 8B, the main effect of professional status is also on the dependent variable KIM score. Here, apprentices scored significantly lower than professionals (-0.957 , $p_{tukey} = 0.005$). Therefore, the hypothesis H6 (higher value of intrinsic motivation in subjects with Professional educational status), which assumed this effect, can be confirmed.

Figure 9 shows the interaction effects found. Figure 9A shows a disordered interaction in the dependent variable, incorrect interactions JS. Apprentices in the With Mapping condition differed from professionals in the With Mapping condition by making more incorrect interactions, but the significance is lost due to Tukey correction (0.927 , $p_{tukey} = 0.062$). The same is true for both groups of professionals, with those in the With Mapping condition making fewer incorrect interactions than those in the Without Mapping condition (-0.93 , $p_{tukey} = 0.063$).

Figure 9B shows a disordinal interaction of the dependent variable Score SUS JS. Although the apprentices in the With Mapping condition have a lower score than the professionals in the With Mapping condition, Tukey correction also leads to a loss of significance (-0.98 , $p_{tukey} = 0.053$).

Finally, Figure 9C shows a hybrid interaction in the dependent variable MCLSVE JS score. This means that the main effect (in professional status) is larger than the interaction effect. For the main effect, this means that apprentices have a significantly higher score on the MCLSVE than professionals (0.63, $p_{tukey} = 0.020$). If we look at the interaction effect, we find a significant difference: apprentices in the With Mapping condition score higher than professionals in the Without Mapping condition (1.19, $p_{tukey} = 0.012$). Tukey correction eliminates the significance of other differences: apprentices in the With Mapping condition score higher than apprentices in the other condition (0.82, $p_{tukey} = 0.133$) and also than professionals in the Without Mapping condition (0.89, $p_{tukey} = 0.087$).

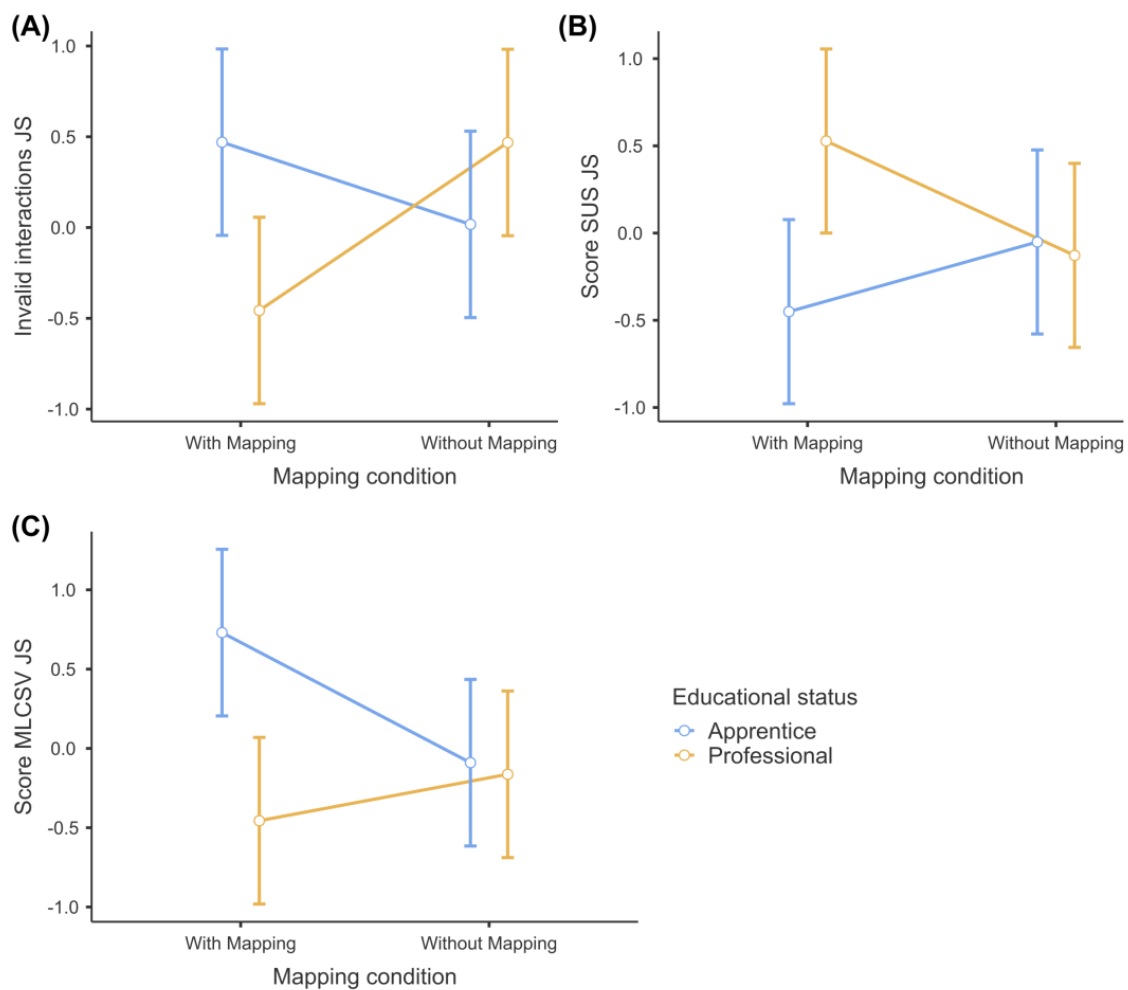


Figure 9. Results of post hoc ANOVA. Interaction effects in (A) faulty interactions JS, (B) SUS JS score, and main and interaction effect in (C) MCLSVE JS score. N = 60. Error bars correspond to the 95% confidence interval of the mean.

4.4. Qualitative Responses

The questionnaire contained two open questions about positive and negative features of the tutorial. A total of 73 responses were collected. The responses were analyzed using the software MAXQDA Analytics Pro 2022 [83]. In a first step, the responses were divided into groups according to condition and professional status. The responses were then explored and coded (marking sections in text documents), with the codings assigned to the two deductive categories of positive and negative. The codings were then further subdivided by deductive categorization. Table 5 gives an overview of the categories, the different groups and the corresponding number of codes.

Table 5. Evaluation of the qualitative responses according to categories, divided into groups.

Category	Coding per Groups			
	With Mapping Professional	With Mapping Apprentice	Without Mapping Professional	Without Mapping Apprentice
General/Negative	3	2	7	3
General/Positive	5	4	3	5
Controls/Teleport/Negative	2	1	1	0
Controls/Teleport/Positive	0	2	0	0
Controls/Tool operation/Negative	2	1	0	0
Controls/Tool operation/Positive	2	2	3	0
Controls/Radio/Negative	2	0	0	0
Virtual Agent and explanations/Negative	1	0	1	1
Virtual Agent and explanations/Positive	3	3	1	1

Annotation. Total number of responses = 73. Total number of codes = 61.

4.4.1. General Feedback

In general, the tutorial received some negative feedback. The statements were mainly related to the content or scope of the tutorial: people, especially professionals (see Table 5), would have liked to learn more activities or features, as the following quotes (translated from German) illustrate:

“a bit more tools”

“more different situations”

“Repeat the same task less often and do more different things”.

Two people had not put the HMD on correctly and had blurred vision because of this. Putting on the HMD was instructed, but according to one statement, there was still room for improvement:

“My vision was partly blurred and/or double. Only at the end did I realise that it was due to the adjustment of the glasses [HMD] to my head/eyes. Better to instruct at the beginning!”

The instruction took place in the same room as the experiment so that subjects had an HMD in front of them. As the devices were new to many people, it is possible that they paid less attention to the instructions. This should be taken into account in future studies.

There was a conspicuously large amount of negative feedback in the general/negative category from professionals in the no mapping condition.

Most positive feedback is also found in general feedback on the tutorial. Many found everything good (“I liked everything”); more concrete statements showed that the interactivity was well received (“That you could do something yourself”; “Great design in relation to the tasks”), that the technology worked perfectly (“Technology worked very well—Good graphics—Understandable tutorial”) or also that people enjoyed getting to know the iVR technology (“I liked looking into it, it’s something new for me”).

4.4.2. Controls

A separate category contained statements about the control system. In addition to two individual statements about the inaccuracy and non-functioning of a button, two people in the With Mapping condition reported that the buttons displayed in the virtual world were sometimes not in the same position as on the real controller:

“With the controller, it is not immediately clear to inexperienced users which buttons to use, as the virtual version is not exactly the same. Otherwise top.”

There were two positive responses to the teleport function, both from apprentices (“I thought the teleport was pretty cool”).

There were seven positive and three negative responses to the tool operation (gripping tools and using them with both hands). The handling or interaction with the tool was considered positive:

“Gripping something, holding the tool to the screw or turning it worked very easily”.

“Operating tool with two hands was very well synchronised.”

One negative feedback from the With Mapping condition was that they would like to use a different button:

“That you could also use the “A” [Button.One on the right controller] while tightening the screws.”

Other individual statements were made about the task of using tools with both hands to tighten screws. One person turned himself instead of the tool (“Turning screws, you shouldn’t have to turn [yourself]”) and another lost the tool (“I dropped the tool and then it was up to me to pick it up”).

Two professionals from the With Mapping condition commented on the radio operation. The controls were not clear:

“The control of the radio was not clear. Did you just have to press the button or touch the yellow box with the tool and then press?”

“Radio use so that it is assigned to a hand.”

4.4.3. Virtual Agent and Explanations

A separate category was also created for the virtual agent and explanations in the tutorial. There were eight positive and three negative responses. The negative answers were single statements: the appearance of the agent could be improved (“As a coach, a person, not just a waistcoat”), she restricted the procedure (“I was only allowed to take the tool in my hand when the virtual coach allowed me to. I didn’t really like that”) and the visual text block overlapped. The positive responses mainly referred to the good comprehensibility and simplicity of the instructions (“The instructions were clear and easy to follow”), but also to the virtual agent directly (“The virtual coach helped me very well with the explanation.”; “That I was well guided by the coach.”). Interestingly, the specific responses to the virtual agent came from apprentices (both conditions). One apprentice also commented that the displayed text was helpful:

“I found it helpful that the spoken text was also written down. That way, if I did not understand something, I could read it.”

5. Discussion

The results of the evaluation showed that there was no main effect of condition (With Mapping vs. Without Mapping), but that apprentices and professionals (professional status) differed significantly. There was also a significant interaction effect between professional status and condition.

5.1. Virtual Agent

The virtual agent was used to guide participants through the tutorial by giving verbal explanations and demonstrating functions. The rating of the virtual agent was insignificantly higher for apprentices than for professionals. A significant difference was expected, but given the high scores of both professional status groups, the virtual agent was also helpful for the professionals. Interestingly, despite the lack of statistical difference, most of the positive open-ended responses can be attributed to the co-mapping condition (e.g., “The virtual coach helped me very well with the explanation”). A negative influence of the virtual agent on the cognitive load of people with little prior knowledge or an interactive avatar cannot be ruled out [62]. The voice output of the avatar can be considered as interactivity. This was used to reduce the cognitive load only through instructions inserted via text modules, which worked at least in individual cases (“The virtual coach helped me very well with the explanation”). According to one statement, the relief also worked in the opposite direction, underlining the importance of addressing multiple modalities [30] (“I found it helpful that the spoken text was also written down. That way, if I didn’t understand something, I could read it”). The dissimilarity of the agent to a human was noticed and commented negatively by one person. Laine et al. [64] report benefits of human-like design. In this study, such a design was deliberately omitted, but it would be worth considering for future studies, for example to increase motivation.

5.2. Intrinsic Motivation

The lower motivation scores of the apprentices can be explained speculatively. The opposite effect, lower scores among professionals, was expected because it was assumed that they would be in a tense situation during the safety training [53]. However, the following issues arose during the field experiment:

Participation in the iVR tutorial at SBB was voluntary after the training, and the staff seemed relaxed because of no time limits or exam situation on this day. The apprentices, on the other hand, were at the career fair as part of an extracurricular stay and were supervised by a teacher. During the fair, it became clear that they had to collect information within a time limit and complete tasks (e.g., interviews) in order to complete work for school. This was not anticipated and may have been a reason for the rejected hypothesis.

In addition, the noise level at the career fair was higher than at the SBB premises—this was also negatively mentioned in the questionnaire. This disruptive factor could have led to distraction, which could have had an impact on perceived competence.

Another possible reason is that the apprentices found the activities in the tutorial less interesting or entertaining because the work context in which the tutorial took place was unknown to them. Novelty effects [55] probably occurred (“I liked looking into it, it’s something new for me”), but group differences that could be attributed to a novelty effect or previous experience with iVR can be ruled out, as the number of people with previous iVR experience was roughly equally distributed between apprentices and professionals.

5.3. Cognitive Load

Cognitive load refers to the limited capacity of working memory used. Measures of ECL specifically subdivided for iVR were measured [44], which refers to the resources used when processing learning instructions [26]. We restricted ourselves to the EL items of the MCLSVE [44], as this is directly influenced by the varied control assignment and represents the primary variable of interest. ICL refers to the inherent difficulty of the tasks themselves, whereas ECL is influenced by the type of presentation and comprehensibility of the control. We assumed that the tasks to be performed would be trivial due to the sequential hierarchical subdivision, meaning that the complexity of the tasks would be minimal and ICL would not have a significant impact on the results. Furthermore, reducing ECL has been identified in the literature as a key goal [30].

As we were particularly interested in the parts of the ECL, we measured CL using a survey with subdimensions, as recommended by Skulmowski [84]. We presented the questions after completing the tasks in the virtual world, as we did not want to interrupt the immersive experience during the tutorial. The use of a mobile measuring device (e.g., mobile EEG) is also mentioned by Skulmowski [84] for uninterrupted recording of CL, but we did not use it for two reasons: first, the measurement is not necessarily more accurate than in surveys [84], and second, we did not have the opportunity to use a mobile measuring device in this study.

Apprentices achieved a significantly higher cognitive load score, i.e., they used more of the limited resources. One possible explanation for the higher cognitive load is a lack of prior knowledge about the work context. Various methods to reduce cognitive load are based on prior knowledge [30,33]. Schöbel et al. [62] mentioned that virtual agents can generate high cognitive load without prior knowledge, and Noetel et al. [30] also pointed out decorative or distracting elements that can increase ECL. For the iVR tutorial, a track environment was created that was as realistic as possible. Apprentices are likely to be unfamiliar with the track environment and the tools and objects used in the tutorial, whereas SBB professionals are more familiar with them. Due to the lack of prior knowledge, the environment and the tools or objects themselves (not their operation) could have been perceived as distracting elements that had to be familiarized with through cognitive processing, thus generating a higher ECL. This assumption is supported by the fact that the professionals increasingly reported that the tutorial could be even more comprehensive in terms of the functions or tasks shown, which does not indicate a maximum cognitive load. The negative effect of the

environmental elements could not be sufficiently reduced by cues, especially by marking and highlighting important objects or buttons, as suggested by Noetel et al. [30].

The interaction effect between condition and professional status also shows that the group of apprentices in the With Mapping condition had significantly higher ECL scores. This suggests that the assigned button assignment increased the already high cognitive load of the apprentices. Statements from the open-ended questions in the questionnaire assigned to this group support this assumption (“With the control system, it is not immediately clear to inexperienced users which buttons are to be operated, since the virtual version is not exactly the same. Otherwise top [good]”). This also suggests that unclear controls led to more incorrect interactions. How cognitive load and incorrect interactions might be related is explained below.

5.4. Invalid Interactions and Processing Time

Both errors and time can be used as objective measures of usability and provide information about effectiveness and efficiency. The significant interaction effect in the dependent variable Erroneous Interactions showed that professionals in the With Mapping condition made fewer errors than the other groups. This result was surprising, as the results of Plechatá et al. [68] showed that younger people made fewer errors. If we now look at the interaction effect, i.e., cognitive load, the group of apprentices in the With Mapping condition stands out as having a higher load. One possible consideration would be to say that because of the higher cognitive load, more incorrect interactions were made.

However, for a number of reasons, it seems more likely that the opposite was true, and that higher cognitive load was caused by more incorrect interactions:

Firstly, increased cognitive load can lead to muscle tension [31], so it is assumed that there are associated errors at the sensorimotor level. These errors could manifest themselves in a variance in the number of buttons pressed—however, there is no significant effect that could be associated with the increased cognitive load group.

Secondly, the literature suggests that errors—especially at higher levels of regulation—cause cognitive load [21]. The nature of the errors could not be recorded by the system (see Section 3.3), but according to Frese and Zapf [21], errors at higher levels of regulation would manifest themselves in increased error correction time. No significant effects are found in the processing time, but if one would assume a generally higher processing time for professionals, as mentioned in the literature [68], a tendency towards more time needed by professionals in the No Mapping condition would be noticeable. This increased processing time could be an indication of errors at higher levels of regulation, which generate cognitive load.

Thirdly, statements from the questionnaire indicate that, due to lack of instruction, errors occurred not only at lower but also at higher levels of regulation—one person apparently had difficulty finding a dropped tool and another moved in circles to tighten a screw instead of just turning the tool. Both of these can be attributed to a higher level of regulation than the sensorimotor level [21].

5.5. Number of Buttons Pressed and Button Preferences

The significantly higher number of buttons pressed by the apprentices was surprising, as the literature had suggested the opposite. In order to come closer to an explanation, the descriptive results, broken down by tutorial task and by individual button, are considered here.

The majority of the buttons pressed by the apprentices can be assigned to the tasks ‘getting to know the radio’ or ‘practising the radio’. It is noticeable that the value of simultaneously pressed buttons was also highest in this task. Taking into account the comments from the questionnaire, it is clear that this task was not sufficiently explained. A possible explanation for the difference between the younger apprentices and the older professionals relates to age differences in implicit and explicit motor learning [66]: apprentices benefit from explicit instruction. In the radio task, insufficient instruction could have led to an attempt to apply what had been learned in the previous task (operating the tool with the grip button). However, the required step was to hold the iVR controller to the radio.

Professionals may have already gained experience with radios through their work and therefore have prior knowledge to draw on.

To identify trends in preferred buttons per task in the Without Mapping condition, descriptive results were presented in Section 4.1. It was found that there were high standard deviations, indicating different preferences. In addition to the mean, the median was also taken into account. The grip button was used most frequently across all tasks, but due to its ergonomic position on the controller and the fact that it was often accompanied by similarly high values for buttons pressed multiple times, it is not possible to draw a clear conclusion as to whether it was actually used or pressed by mistake. This should be taken into account when designing future tasks—especially for functions such as teleportation, where a button has to be released specifically, an accidental button press could block the function from being triggered.

In the radio and grab tasks, the primary button was often used in addition to the grab button (in the With Mapping condition, it was the trigger button) and in the teleport task, the trigger button (in the With Mapping condition, it was the primary button). The preference for the primary button in the grasping task was also noted in the questionnaire (“[Improvement wish:] That the “A” [Button.One on the right controller] could also be used when pulling the screws”).

5.6. Subjective Usability

In addition to objective measures of usability, usability can also be measured subjectively. Subjective usability was measured by the score of the SUS. The apprentices in the With Mapping condition reported a lower score than the professionals in the same condition. The results fit well with the picture that emerged from the effects of incorrect interactions and cognitive load. It is important to note that all four groups achieved scores between “good” and “excellent” [80]. This should not be taken for granted, given the comments made in Section 2.3 and the findings of Domin et al. [69] that all people had problems with iVR.

5.7. Answering the Research Questions

Question Q2 was as follows: which mapping on the iVR controllers is more intuitive and effective for users to experience?

Based on the results discussed in the focus, it can be assumed that iVR controllers that do not map associated functions to individual buttons are more intuitive for apprentices in a field scenario, as fewer cognitive resources are required, and more effective, as fewer incorrect interactions occur. Processing time is also shorter for apprentices in this condition, at just over six minutes, although not significantly so.

For the professionals, it is assumed that both mapping conditions are experienced intuitively in a field scenario, but that the variant of the iVR controllers on which associated functions are mapped to individual buttons is more effective because fewer incorrect interactions occur. The processing time for professionals in this condition is not quite six and a half minutes shorter, but the difference is not significant.

Question Q3 was as follows: do the target groups of professionals and apprentices differ with regard to the experiences described in Q2?

Parts of this question have already been answered with the answer to Q2: There are differences depending on the design of the iVR controller mapping and the target groups in terms of cognitive load and subjective as well as objective usability values. In addition, a higher intrinsic motivation was found among professionals, which is purely due to their professional status.

5.8. Generalization of the Findings

First of all, it should be noted that it was important to include the actual target groups of SBB. This implies a high degree of external validity, meaning the results are transferable to apprentices and professionals involved in track safety training. However, the findings are not necessarily transferable to other contexts or laboratory research with more controlled

conditions. Further studies in other industrial contexts and laboratory-controlled studies will show if our results are replicable.

5.9. Limitations

This study is not without its limitations. First of all, it should be noted that it was important to include the actual target groups of SBB. This implies a high degree of external validity, but it also means that the findings are not necessarily transferable to other contexts or laboratory research with more controlled conditions.

Furthermore, for the dependent variables, the question arises as to whether other measurement instruments would have been more accurate. As stated in the literature, objective measurements for CL are not necessarily more accurate than surveys [85], and we did not have any mobile measuring devices available. In contrast to a controlled experiment in the laboratory, it was considered more important to carry out a field experiment with the actual user groups, i.e., professionals from the SBB environment and apprentices. Nevertheless, predictable confounding variables such as noise could have been collected and taken into account.

Another limitation of this study is the intrinsic cognitive load. ICL was not taken into account in this study because the tasks to be learned in the tutorial were considered trivial, and ECL has been identified in the literature as the most important factor influencing CL. In follow-up studies dealing with actual safety training or other more complex tasks, both ICL and GCL should be considered. Firstly, this is because the GCL items describe the subjective increase in procedural, conceptual and content knowledge [44], which was not needed for our study, but can describe the success of the intervention in safety training. Secondly, according to both Leppink et al. [43] and Andersen and Makransky [44], the ICL items are designed for topic-specific assessment (e.g., “The topic/topics covered in the activity was/were very complex”). Again, this was not important for our study, but may be very relevant for studies with training scenarios, e.g., in mining [86] or construction [87]. CL has not yet been measured in the studies mentioned above. However, since, according to our data, CL can vary for apprentices and professionals, and the work in virtual space and the content to be learnt in these studies is considerably more complex than in our study, we recommend that ICL and GCL also be measured for safety training courses of this type. Furthermore, it is important for such studies to find suitable variables and methods for measuring CL. A selection of suitable variables/procedures for decision-making can be found in Skulmowski [84].

A further limitation of the study is that we have not yet understood the extent of the influence of the degree of reality in the representation of the tutorial. Makransky et al. [40] argue that although immersive virtual environments can enhance a topic and increase the sense of presence, this does not necessarily have to be related to the learning objective and therefore requires unnecessary ECL. However, it is also clear from the qualitative responses that auditory and visual effects (e.g., the train passing through) were perceived positively. Wolfartsberger et al. [34] suggest that a compromise should be found between entertainment and learning success. Skulmowski and Rey [85] found a higher load on the ECL with more realistic representations in VR, but also better retention of learning material, which is a contradiction. Where possible, they recommend using realistic imagery to focus the user’s attention on important content.

We assume that a high degree of realism is required for safety training in the railway sector in order to experience realistic scenarios. However, for a tutorial to learn the controls, the degree of realism of the virtual world could be adapted to the user’s prior knowledge.

Also, we found that apprentices had significantly higher CL values compared to professionals (see Table 4, MCLSVE score). This cannot be attributed to a lack of previous experience with VR glasses, as more apprentices (19 people) than professionals (16 people) reported previous experience with VR. It should therefore be investigated whether the virtual environment and its design could have an influence here. This is because either the novelty of the virtual elements (tracks, railroad, tools) or their detailed graphic representation could

have resulted in a possible overload. Experienced professionals may have experienced less cognitive overload as they were already familiar with the virtual elements. This is also shown by the qualitative data (Table 5). Here, the professionals reported back that the tutorial could have been even more comprehensive in terms of the functions or tasks shown, which does not indicate maximum cognitive load. Skulmowski and Rey [85] were able to show that it is possible to influence cognitive load due to different graphic details. Thus, we recommend that future studies could also vary the degree of realism in order to examine the interaction between condition and educational status in a more controlled way.

6. Conclusions

The aim of the study was to empirically test different interaction designs for iVR controllers during an iVR tutorial in a working environment using quantitative and qualitative methods. Factors like different interactions with controllers and presumptions in humans, such as age, are overlooked if we look at meta-analytic research or systematic reviews [1–9]. To close the research gap, we conducted a study in which we empirically determined which mapping conditions of the iVR controller were more intuitive and effective for people with different educational statuses (age groups). For this purpose, an iVR tutorial for safety iVR training was developed that introduced users to different functions by guiding them step by step and supporting them in different ways. With the help of action theory [21], the actions to be learnt were subdivided in such a way that they could be learnt step by step, practiced, and eventually become routine, thus keeping the cognitive effort low. In addition, cognitive load was minimized by using information design principles from CLT, such as the use of multiple modalities and the use of cue stimuli. Finally, the use of a virtual agent was designed to promote motor learning through demonstration and explanation.

Significant differences in cognitive load and objective as well as subjective usability showed that mapping control is unsuitable for apprentices, whereas it is recommended for professionals. In the experience of the iVR tutorial, differences in intrinsic motivation were found between the two target groups, independent of the control. Due to our study, the results by Zhang and Song [10] may be enhanced in the future, because not only dimensions like style (e.g., realism), perspective, color, feel, sound of cues, music or dialogue are important. Also, differences in interaction technology seem to be important factors when it comes to sensory cues. Further studies will show if our results will be verified or falsified. Also, we will see if the evaluation of tutorials in areas other than public transport safety training will show the same results or not.

Supplementary Materials: The following supporting information can be downloaded at: Video S1: <https://tube.switch.ch/videos/XQxDqKdsy8>; Video S2: <https://tube.switch.ch/videos/tNSrvkILbw>. All accessed on 8 June 2024.

Author Contributions: Conceptualization, A.U. and O.C.; methodology, A.U., P.V.M., S.G., P.D. and O.C.; software, A.U., P.V.M., S.G., P.D. and O.C.; validation, A.U., S.G., P.D. and O.C.; formal analysis, A.U. and O.C.; investigation, A.U., P.V.M. and O.C.; resources, S.G., P.D. and O.C.; writing—original draft preparation, A.U. and O.C.; writing—review and editing, O.C. and P.V.M.; visualization, A.U.; supervision, O.C.; project administration, O.C. and S.G.; funding acquisition, O.C. and S.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by SBB Research Fund on management in the transport sector represented by the Institute for Systemic Management and Public Governance (IMP-HSG) at the University of St. Gallen, Dufourstrasse 40a, 9000 St. Gallen.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Quantitative Data are available for researchers on request.

Conflicts of Interest: A.U., P.V.M. and O.C., declare no conflicts of interest. S.G. and P.D. are employed by SBB, AG Bern Wankdorf, Switzerland. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A

Table A1. Items used in the survey.

Category	Item Original	Item Adapted and Translated (DE)	Reference
Usability	I think that I would like to use this system frequently. I found the system unnecessarily complex. I thought the system was easy to use. I think that I would need the support of a technical person to be able to use this system.	Ich denke, dass ich die Simulation häufig nutzen möchte. Ich fand die Simulation unnötig kompliziert. Ich fand die Simulation einfach zu bedienen. Ich glaube, ich bräuchte die Unterstützung einer technischen Person, um die Simulation nutzen zu können.	[47]
	I found the various functions in this system were well integrated. I thought there was too much inconsistency in this system. I would imagine that most people would learn to use this system very quickly.	Ich fand, dass die verschiedenen Funktionen in der Simulation gut integriert waren. Ich fand, dass die Simulation zu inkonsequent war. Ich könnte mir vorstellen, dass die meisten Menschen sehr schnell lernen würden, mit der Simulation umzugehen.	
	I found the system very cumbersome to use. I felt very confident using the system. I needed to learn a lot of things before I could get going with this system.	Ich fand die Simulation sehr umständlich zu bedienen. Ich fühlte mich sehr sicher im Umgang mit der Simulation. Ich musste eine Menge Dinge lernen, bevor ich mit der Simulation loslegen konnte.	
Motivation/interest enjoyment	Die Tätigkeit in der Ausstellung hat mir Spaß gemacht Ich fand die Tätigkeit in der Ausstellung sehr interessant. Die Tätigkeit in der Ausstellung war unterhaltsam.	Die Tätigkeit in der Simulation hat mir Spass gemacht Ich fand die Tätigkeit in der Simulation sehr interessant. Die Tätigkeit in der Simulation war unterhaltsam.	[60]
Motivation/perceived competence	Mit meiner Leistung in der Ausstellung bin ich zufrieden. Bei der Tätigkeit in der Ausstellung stellte ich mich geschickt an. Ich glaube, ich war bei der Tätigkeit in der Ausstellung ziemlich gut.	Mit meiner Leistung in der Simulation bin ich zufrieden. Bei der Tätigkeit in der Simulation stellte ich mich geschickt an. Ich glaube, ich war bei der Tätigkeit in der Simulation ziemlich gut.	[60]
Motivation/perceived choice	Ich konnte die Tätigkeit in der Ausstellung selbst steuern Bei der Tätigkeit in der Ausstellung konnte ich wählen, wie ich es mache Bei der Tätigkeit in der Ausstellung konnte ich so vorgehen, wie ich es wollte	Ich konnte die Tätigkeit in der Simulation selbst steuern Bei der Tätigkeit in der Simulation konnte ich wählen, wie ich es mache Bei der Tätigkeit in der Simulation konnte ich so vorgehen, wie ich es wollte	[60]
Motivation/pressure tension	Bei der Tätigkeit in der Ausstellung fühlte ich mich unter Druck. Bei der Tätigkeit in der Ausstellung fühlte ich mich angespannt. Ich hatte Bedenken, ob ich die Tätigkeit in der Ausstellung gut hinbekomme.	Bei der Tätigkeit in der Simulation fühlte ich mich unter Druck. Bei der Tätigkeit in der Simulation fühlte ich mich angespannt. Ich hatte Bedenken, ob ich die Tätigkeit in der Simulation gut hinbekomme.	[60]
Cognitive Load/IL	The topic covered in the simulation was very complex. The simulation covered procedures that I perceived as very complex. The simulation covered concepts and definitions that I perceived as very complex.		[44]
Cognitive Load/EL ins	The instructions and/or explanations used in the simulation were very unclear. The instructions and/or explanations used in the simulation were, in terms of learning, very ineffective. The instructions and/or explanations used in the simulation were full of unclear content.	Die in der Simulation verwendeten Anweisungen und/oder Erklärungen waren sehr unklar. Die in der Simulation verwendeten Anweisungen und/oder Erklärungen waren im Hinblick auf den Lernerfolg sehr ineffektiv. Die in der Simulation verwendeten Anweisungen und/oder Erläuterungen waren inhaltlich unklar.	[44]

Table A1. Cont.

Category	Item Original	Item Adapted and Translated (DE)	Reference
Cognitive Load/EL int	The interaction technique used in the simulation was very unclear. The interaction technique used in the simulation was, in terms of learning, very ineffective. The interaction technique used in the simulation made it harder to learn. The interaction technique used in the simulation was difficult to master.	Die in der Simulation verwendete Interaktionstechnik war sehr unklar. Die in der Simulation verwendete Interaktionstechnik war im Hinblick auf das Lernen sehr ineffektiv. die in der Simulation verwendete Interaktionstechnik erschwerte das Lernen. Die in der Simulation verwendete Interaktionstechnik war schwer zu meistern.	[44]
Cognitive Load/EL env	The elements in the virtual environment made the learning very unclear. The virtual environment was, in terms of learning, very ineffective. The virtual environment was full of irrelevant content. It was difficult to find the relevant learning information in the virtual environment.	Die Elemente in der Simulation machten das Lernen sehr unklar Die Simulation war, was das Lernen angeht, sehr ineffektiv. Die Simulation war voll von unwichtigen Inhalten. Es war schwierig, die relevanten Lerninformationen in der Simulation zu finden.	[44]
Cognitive Load/GL	The simulation really enhanced my understanding of the topics covered The simulation really enhanced my knowledge and understanding of lab safety The simulation really enhanced my understanding of the procedures covered. The simulation really enhanced my understanding of concepts and definitions		[44]
Facilitating Learning	The agent led me to think more deeply about the presentation. The agent made the instruction interesting. The agent encouraged me to reflect what I was learning. The agent kept my attention. The agent presented the material effectively. The agent helped me to concentrate on the presentation. The agent focused me on the relevant information. The agent improved my knowledge of the content. The agent was interesting. The agent was enjoyable.	Der virtuelle Coach hat mich dazu gebracht, tiefer über das Tutorial nachzudenken. Der virtuelle Coach machte das Tutorial interessant. Der virtuelle Coach ermutigte mich, das im Tutorial Gelernte zu reflektieren. Der virtuelle Coach hat meine Aufmerksamkeit aufrechterhalten. Der virtuelle Coach hat das zu Lernende effektiv präsentiert. Der virtuelle Coach half mir, mich auf das Tutorial zu konzentrieren. Der virtuelle Coach hat mich auf die relevanten Informationen konzentriert. Der virtuelle Coach hat mein Wissen über das Tutorial verbessert. Der virtuelle Coach war interessant. Der virtuelle Coach war unterhaltsam.	[65]
Credible	The agent was knowledgeable. The agent was intelligent. The agent was useful. The agent was useful. The agent was instructor-like.	Der virtuelle Coach war kompetent. Der virtuelle Coach war intelligent. Der virtuelle Coach war nützlich. Der virtuelle Coach war hilfreich. Der virtuelle Coach war wie eine Lehrerin.	[65]
Human-like	The agent has a personality. The agent's emotion was natural. The agent was human-like. The agent's movement was natural. The agent showed emotion.		[65]
Engaging	The agent was expressive.		[65]

Note: Fields shaded in grey indicate excluded items.

References

1. Checa, D.; Bustillo, A. A Review of Immersive Virtual Reality Serious Games to Enhance Learning and Training. *Multimed. Tools Appl.* **2020**, *79*, 5501–5527. [\[CrossRef\]](#)
2. Coban, M.; Bolat, Y.I.; Goksu, I. The Potential of Immersive Virtual Reality to Enhance Learning: A Meta-Analysis. *Educ. Res. Rev.* **2022**, *36*, 100452. [\[CrossRef\]](#)
3. Di Natale, A.F.; Repetto, C.; Riva, G.; Villani, D. Immersive Virtual Reality in K-12 and Higher Education: A 10-year Systematic Review of Empirical Research. *Br. J. Educ. Technol.* **2020**, *51*, 2006–2033. [\[CrossRef\]](#)
4. Hamilton, D.; McKechnie, J.; Edgerton, E.; Wilson, C. Immersive Virtual Reality as a Pedagogical Tool in Education: A Systematic Literature Review of Quantitative Learning Outcomes and Experimental Design. *J. Comput. Educ.* **2021**, *8*, 1–32. [\[CrossRef\]](#)
5. Huang, Y.-J.; Liu, K.-Y.; Lee, S.-S.; Yeh, I.-C. Evaluation of a Hybrid of Hand Gesture and Controller Inputs in Virtual Reality. *Int. J. Hum.-Comput. Interact.* **2021**, *37*, 169–180. [\[CrossRef\]](#)
6. Pellas, N.; Dengel, A.; Christopoulos, A. A Scoping Review of Immersive Virtual Reality in STEM Education. *IEEE Trans. Learn. Technol.* **2020**, *13*, 748–761. [\[CrossRef\]](#)
7. Radianti, J.; Majchrzak, T.A.; Fromm, J.; Wohlgenannt, I. A Systematic Review of Immersive Virtual Reality Applications for Higher Education: Design Elements, Lessons Learned, and Research Agenda. *Comput. Educ.* **2020**, *147*, 103778. [\[CrossRef\]](#)
8. Rojas-Sánchez, M.A.; Palos-Sánchez, P.R.; Folgado-Fernández, J.A. Systematic Literature Review and Bibliometric Analysis on Virtual Reality and Education. *Educ. Inf. Technol.* **2023**, *28*, 155–192. [\[CrossRef\]](#)
9. Wu, B.; Yu, X.; Gu, X. Effectiveness of Immersive Virtual Reality Using Head-mounted Displays on Learning Performance: A Meta-analysis. *Br. J. Educ. Technol.* **2020**, *51*, 1991–2005. [\[CrossRef\]](#)
10. Zhang, Y.; Song, Y. The Effects of Sensory Cues on Immersive Experiences for Fostering Technology-Assisted Sustainable Behavior: A Systematic Review. *Behav. Sci.* **2022**, *12*, 361. [\[CrossRef\]](#)
11. Scorgie, D.; Feng, Z.; Paes, D.; Parisi, F.; Yiu, T.W.; Lovreglio, R. Virtual Reality for Safety Training: A Systematic Literature Review and Meta-Analysis. *Saf. Sci.* **2024**, *171*, 106372. [\[CrossRef\]](#)
12. Mikropoulos, T.A.; Natsis, A. Educational Virtual Environments: A Ten-Year Review of Empirical Research (1999–2009). *Comput. Educ.* **2011**, *56*, 769–780. [\[CrossRef\]](#)
13. Cummings, J.J.; Bailenson, J.N. How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence. *Media Psychol.* **2016**, *19*, 272–309. [\[CrossRef\]](#)
14. Ziegler, C.; Papageorgiou, A.; Hirschi, M.; Genovese, R.; Christ, O. Training in Immersive Virtual Reality: A Short Review of Presumptions and the Contextual Interference Effect. In *Human Interaction, Emerging Technologies and Future Applications II*; Ahram, T., Taiar, R., Gremaux-Bader, V., Aminian, K., Eds.; Advances in Intelligent Systems and Computing; Springer International Publishing: Cham, Switzerland, 2020; Volume 1152, pp. 328–333. ISBN 978-3-030-44266-8.
15. Buttussi, F.; Chittaro, L. Locomotion in Place in Virtual Reality: A Comparative Evaluation of Joystick, Teleport, and Leaning. *IEEE Trans. Vis. Comput. Graph.* **2021**, *27*, 125–136. [\[CrossRef\]](#)
16. Faric, N.; Potts, H.W.W.; Hon, A.; Smith, L.; Newby, K.; Steptoe, A.; Fisher, A. What Players of Virtual Reality Exercise Games Want: Thematic Analysis of Web-Based Reviews. *J. Med. Internet Res.* **2019**, *21*, e13833. [\[CrossRef\]](#)
17. Nyssönen, T.; Helle, S.; Lehtonen, T.; Smed, J. A Comparison of Gesture and Controller-Based User Interfaces for 3D Design Reviews in Virtual Reality. In Proceedings of the 55th Hawaii International Conference on System Sciences, Maui, HI, USA, 4–7 January 2022.
18. Kamińska, D.; Zwoliński, G.; Laska-Leśniewicz, A. Usability Testing of Virtual Reality Applications—The Pilot Study. *Sensors* **2022**, *22*, 1342. [\[CrossRef\]](#)
19. Rempel, D.; Camilleri, M.J.; Lee, D.L. The Design of Hand Gestures for Human–Computer Interaction: Lessons from Sign Language Interpreters. *Int. J. Hum. Comput. Stud.* **2014**, *72*, 728–735. [\[CrossRef\]](#)
20. Kangas, J.; Kumar, S.K.; Mehtonen, H.; Järnstedt, J.; Raisamo, R. Trade-Off between Task Accuracy, Task Completion Time and Naturalness for Direct Object Manipulation in Virtual Reality. *Multimodal Technol. Interact.* **2022**, *6*, 6. [\[CrossRef\]](#)
21. Frese, M.; Zapf, D. Action as the Core of Work Psychology: A German Approach. In *Handbook of Industrial and Organizational Psychology*; Triandis, H.C., Dunette, M.D., Hough, L.M., Eds.; Consulting Psychologists Press: Palo Alto, CA, USA, 1994; pp. 272–340.
22. Schumacher, V.; Martin, M. Lernen und Gedächtnis im Alter. In *Gedächtnisstörungen*; Bartsch, T., Falkai, P., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 31–39. ISBN 978-3-642-36992-6.
23. Witte, K. *Grundlagen der Sportmotorik im Bachelorstudium (Band 1)*; Springer: Berlin/Heidelberg, Germany, 2018; ISBN 978-3-662-57867-4.
24. Lampen, E.; Liersch, M.; Lehwald, J. Towards Motor Learning in Augmented Reality: Imitating an Avatar. In *HCI International 2020—Late Breaking Posters*; Stephanidis, C., Antona, M., Ntoa, S., Eds.; Communications in Computer and Information Science; Springer International Publishing: Cham, Switzerland, 2020; Volume 1294, pp. 181–188. ISBN 978-3-030-60702-9.
25. Graesser, A.C.; Moreno, K.; Marineau, J.; Adcock, A.; Olney, A. AutoTutor Improves Deep Learning of Computer Literacy: Is It the Dialog or the Talking Head? In Proceedings of the 11th International Conference on Artificial Intelligence in Education, Sydney, Australia, 20–24 July 2003; Volume 4754, pp. 47–54.
26. Sweller, J.; van Merriënboer, J.J.G.; Paas, F. Cognitive Architecture and Instructional Design: 20 Years Later. *Educ. Psychol. Rev.* **2019**, *31*, 261–292. [\[CrossRef\]](#)
27. Deci, E.L.; Koestner, R.; Ryan, R.M. Extrinsic Rewards and Intrinsic Motivation in Education: Reconsidered Once Again. *Rev. Educ. Res.* **2001**, *71*, 1–27. [\[CrossRef\]](#)

28. Sweller, J.; van Merriënboer, J.J.G.; Paas, F.G.W.C. Cognitive Architecture and Instructional Design. *Educ. Psychol. Rev.* **1998**, *10*, 251–296. [[CrossRef](#)]
29. Sweller, J. Cognitive Load Theory: Recent Theoretical Advances. In *Cognitive Load Theory*; Plass, J.L., Moreno, R., Brünken, R., Eds.; Cambridge University Press: Cambridge, UK, 2010; pp. 29–47. ISBN 978-0-521-67758-5.
30. Noetel, M.; Griffith, S.; Delaney, O.; Harris, N.R.; Sanders, T.; Parker, P. Multimedia Design for Learning: An Overview of Reviews with Meta-Analysis. *Rev. Educ. Res.* **2022**, *92*, 413–454. [[CrossRef](#)]
31. Leyman, E.L.; Mirka, G.A.; Kaber, D.B.; Sommerich, C.M. Cervicobrachial Muscle Response to Cognitive Load in a Dual-Task Scenario. *Ergonomics* **2004**, *47*, 625–645. [[CrossRef](#)]
32. Paas, F.; van Gog, T.; Sweller, J. Cognitive Load Theory: New Conceptualizations, Specifications, and Integrated Research Perspectives. *Educ. Psychol. Rev.* **2010**, *22*, 115–121. [[CrossRef](#)]
33. Klepsch, M.; Schmitz, F.; Seufert, T. Development and Validation of Two Instruments Measuring Intrinsic, Extraneous, and Germane Cognitive Load. *Front. Psychol.* **2017**, *8*, 1997. [[CrossRef](#)]
34. Wolfartsberger, J.; Zimmermann, R.; Obermeier, G.; Niedermayr, D. Analyzing the Potential of Virtual Reality-Supported Training for Industrial Assembly Tasks. *Comput. Ind.* **2023**, *147*, 103838. [[CrossRef](#)]
35. Klepsch, M.; Seufert, T. Understanding Instructional Design Effects by Differentiated Measurement of Intrinsic, Extraneous, and Germane Cognitive Load. *Instr. Sci.* **2020**, *48*, 45–77. [[CrossRef](#)]
36. Dan, A.; Reiner, M. EEG-Based Cognitive Load of Processing Events in 3D Virtual Worlds Is Lower than Processing Events in 2D Displays. *Int. J. Psychophysiol.* **2017**, *122*, 75–84. [[CrossRef](#)]
37. Wenk, N.; Penalver-Andres, J.; Buetler, K.A.; Nef, T.; Müri, R.M.; Marchal-Crespo, L. Effect of Immersive Visualization Technologies on Cognitive Load, Motivation, Usability, and Embodiment. *Virtual Real.* **2023**, *27*, 307–331. [[CrossRef](#)]
38. Allcoat, D.; von Mühlent, A. Learning in Virtual Reality: Effects on Performance, Emotion and Engagement. *Res. Learn. Technol.* **2018**, *26*, 1–13. [[CrossRef](#)]
39. van der Land, S.; Schouten, A.P.; Feldberg, F.; van den Hooff, B.; Huysman, M. Lost in Space? Cognitive Fit and Cognitive Load in 3D Virtual Environments. *Comput. Hum. Behav.* **2013**, *29*, 1054–1064. [[CrossRef](#)]
40. Makransky, G.; Terkildsen, T.S.; Mayer, R.E. Adding Immersive Virtual Reality to a Science Lab Simulation Causes More Presence but Less Learning. *Learn. Instr.* **2019**, *60*, 225–236. [[CrossRef](#)]
41. Hart, S.G.; Staveland, L.E. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology*; Elsevier: Amsterdam, The Netherlands, 1988; Volume 52, pp. 139–183. ISBN 978-0-444-70388-0.
42. Paas, F. Training Strategies for Attaining Transfer of Problem-Solving Skill in Statistics: A Cognitive-Load Approach. *J. Educ. Psychol.* **1992**, *84*, 429–434. [[CrossRef](#)]
43. Leppink, J.; Paas, F.; Van der Vleuten, C.P.M.; Van Gog, T.; Van Merriënboer, J.J.G. Development of an Instrument for Measuring Different Types of Cognitive Load. *Behav. Res. Methods* **2013**, *45*, 1058–1072. [[CrossRef](#)]
44. Andersen, M.S.; Makransky, G. The Validation and Further Development of a Multidimensional Cognitive Load Scale for Virtual Environments. *J. Comput. Assist. Learn.* **2021**, *37*, 183–196. [[CrossRef](#)]
45. Oesterreich, R.; Leitner, K.; Resch, M. *Analyse psychischer Anforderungen und Belastungen in der Produktionsarbeit*; Hogrefe: Göttingen, Germany, 2000.
46. *ISO Stand. No 9241-112018*; International Organization for Standardization Ergonomics of Human-System Interaction—Part 11: Usability: Definitions and Concepts. ISO: Geneva, Switzerland, 2018.
47. Brooke, J. SUS—A Quick and Dirty Usability Scale. In *Usability Evaluation in Industry*; CRC Press: Boca Raton, FL, USA, 1986.
48. Mutschler, B.; Reichert, M. Usability-Metriken als Nachweis der Wirtschaftlichkeit von Verbesserungen der Mensch-Maschine-Schnittstelle. In Proceedings of the IWSM/MetriKon Workshop on Software Metrics (IWSM/MetriKon’04), Königs Wusterhausen, Germany, 2–5 November 2004; pp. 407–418.
49. Bowman, D.A.; Gabbard, J.L.; Hix, D. A Survey of Usability Evaluation in Virtual Environments: Classification and Comparison of Methods. *Presence Teleoperators Virtual Environ.* **2002**, *11*, 404–424. [[CrossRef](#)]
50. Oculus Oculus Developer Blog. Available online: <https://developer.oculus.com/blog/> (accessed on 19 August 2022).
51. Ryan, R.M.; Deci, E.L. Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemp. Educ. Psychol.* **2000**, *25*, 54–67. [[CrossRef](#)]
52. Ai-Lim Lee, E.; Wong, K.W.; Fung, C.C. How Does Desktop Virtual Reality Enhance Learning Outcomes? A Structural Equation Modeling Approach. *Comput. Educ.* **2010**, *55*, 1424–1442. [[CrossRef](#)]
53. Grolnick, W.S.; Ryan, R.M. Autonomy in Children’s Learning: An Experimental and Individual Difference Investigation. *J. Pers. Soc. Psychol.* **1987**, *52*, 890–898. [[CrossRef](#)]
54. Villena-Taranilla, R.; Tirado-Olivares, S.; Cózar-Gutiérrez, R.; González-Calero, J.A. Effects of Virtual Reality on Learning Outcomes in K-6 Education: A Meta-Analysis. *Educ. Res. Rev.* **2022**, *35*, 100434. [[CrossRef](#)]
55. Jenö, L.M.; Vandvik, V.; Eliassen, S.; Grytnes, J.-A. Testing the Novelty Effect of an M-Learning Tool on Internalization and Achievement: A Self-Determination Theory Approach. *Comput. Educ.* **2019**, *128*, 398–413. [[CrossRef](#)]
56. Huang, W.; Roscoe, R.D.; Johnson-Glenberg, M.C.; Craig, S.D. Motivation, Engagement, and Performance across Multiple Virtual Reality Sessions and Levels of Immersion. *J. Comput. Assist. Learn.* **2021**, *37*, 745–758. [[CrossRef](#)]
57. Pelletier, L.G.; Rocchi, M.A.; Vallerand, R.J.; Deci, E.L.; Ryan, R.M. Validation of the Revised Sport Motivation Scale (SMS-II). *Psychol. Sport Exerc.* **2013**, *14*, 329–341. [[CrossRef](#)]

58. Pintrich, P.R.; Smith, D.A.F.; Garcia, T.; Mckeachie, W.J. Reliability and Predictive Validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educ. Psychol. Meas.* **1993**, *53*, 801–813. [[CrossRef](#)]
59. Deci, E.L.; Eghrari, H.; Patrick, B.C.; Leone, D.R. Facilitating Internalization: The Self-Determination Theory Perspective. *J. Pers.* **1994**, *62*, 119–142. [[CrossRef](#)]
60. Wilde, M.; Bätz, K.; Kovaleva, A.; Urhahne, D. Überprüfung einer Kurzsкала intrinsischer Motivation (KIM). *Z. Didakt. Naturwissenschaften* **2009**, *15*, 31–45.
61. Baylor, A.L. Promoting Motivation with Virtual Agents and Avatars: Role of Visual Presence and Appearance. *Philos. Trans. R. Soc. B Biol. Sci.* **2009**, *364*, 3559–3565. [[CrossRef](#)]
62. Schöbel, S.; Janson, A.; Mishra, A. A Configurational View on Avatar Design—The Role of Emotional Attachment, Satisfaction, and Cognitive Load in Digital Learning. In Proceedings of the Fortieth International Conference on Information Systems, Munich, Germany, 15–18 December 2019; pp. 1–17. [[CrossRef](#)]
63. Martha, A.S.D.; Santoso, H. The Design and Impact of the Pedagogical Agent: A Systematic Literature Review. *J. Educ. Online* **2019**, *16*, n1. [[CrossRef](#)]
64. Laine, J.; Lindqvist, T.; Korhonen, T.; Hakkarainen, K. Systematic Review of Intelligent Tutoring Systems for Hard Skills Training in Virtual Reality Environments. *Int. J. Technol. Educ. Sci.* **2022**, *6*, 178–203. [[CrossRef](#)]
65. Baylor, A.L.; Ryu, J. The Psychometric Structure of Pedagogical Agent Persona. *Technol. Instr. Cogn. Learn. (TICL)* **2005**, *2*, 291–315.
66. Verneau, M.; van der Kamp, J.; Savelsbergh, G.J.P.; de Looze, M.P. Age and Time Effects on Implicit and Explicit Learning. *Exp. Aging Res.* **2014**, *40*, 477–511. [[CrossRef](#)]
67. Sayers, H. Desktop Virtual Environments: A Study of Navigation and Age. *Interact. Comput.* **2004**, *16*, 939–956. [[CrossRef](#)]
68. Plechatá, A.; Sahula, V.; Fayette, D.; Fajnerová, I. Age-Related Differences with Immersive and Non-Immersive Virtual Reality in Memory Assessment. *Front. Psychol.* **2019**, *10*, 1330. [[CrossRef](#)]
69. Shibata, K.; Sasaki, Y.; Bang, J.W.; Walsh, E.G.; Machizawa, M.G.; Tamaki, M.; Chang, L.-H.; Watanabe, T. Overlearning Hyperstabilizes a Skill by Rapidly Making Neurochemical Processing Inhibitory-Dominant. *Nat. Neurosci.* **2017**, *20*, 470–475. [[CrossRef](#)]
70. Domin, M.; Janneck, M.; Grimm, S. Altersbezogene Unterschiede bei der Interaktion mit einem Virtual-Reality-System. In Proceedings of the Communities in New Media: Researching the Digital Transformation in Science, Business, Education and Public Administration—Proceedings of 22nd Conference GeNeMe, Dresden, Germany, 10–11 October 2019; Koehler, T., Schoop, E., Kahnwald, N., Eds.; TUDpress: Dresden, Germany, 2019; pp. 24–34.
71. Faul, F.; Erdfelder, E.; Lang, A.G.; Buchner, A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* **2007**, *39*, 175–191. [[CrossRef](#)]
72. Tivian XI GmbH Enterprise Feedback Suite 2022. Available online: <https://www.tivian.com/de/> (accessed on 8 June 2024).
73. Lewis, J.R. The System Usability Scale: Past, Present, and Future. *Int. J. Hum.-Comput. Interact.* **2018**, *34*, 577–590. [[CrossRef](#)]
74. Hahs-Vaughn, D.L. *Applied Multivariate Statistical Concepts*; Routledge: New York, NY, USA, 2016; ISBN 978-0-415-84235-8.
75. Field, A. *Discovering Statistics Using IBM SPSS Statistics*, 5th ed.; SAGE Publications: Thousand Oaks, CA, USA, 2017; ISBN 978-1-5264-1952-1.
76. IBM Corp. *IBM SPSS Statistics for Windows 2020*; IBM Corp.: Armonk, NY, USA, 2020.
77. R Core Team. *R: A Language and Environment for Statistical Computing 2021*; R Core Team: Vienna, Austria, 2021.
78. Wickham, H.; Averick, M.; Bryan, J.; Chang, W.; McGowan, L.; François, R.; Grolemund, G.; Hayes, A.; Henry, L.; Hester, J.; et al. Welcome to the Tidyverse. *J. Open Source Softw.* **2019**, *4*, 1686. [[CrossRef](#)]
79. RStudio Team. *RStudio: Integrated Development for R. RStudio*; PBC: Boston, MA, USA, 2020; Available online: <http://www.rstudio.com/> (accessed on 8 June 2024).
80. Bangor, A. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. *J. Usability Stud.* **2009**, *4*, 114–123.
81. Johnson, N.L. Systems of Frequency Curves Generated by Methods of Translation. *Biometrika* **1949**, *36*, 149–176. [[CrossRef](#)]
82. Hemmerich, W. StatistikGuru: Johnson Transformation Berechnen. 2016. Available online: <https://statistikguru.de/rechner/johnson-transformation-berechnen.html> (accessed on 8 June 2024).
83. VERBI Software. MAXQDA Software for qualitative data analysis. Consult. Sozialforschung GmbH: Berlin, Germany. 2022. Available online: <https://www.maxqda.com/> (accessed on 8 June 2024).
84. Skulmowski, A. Guidelines for Choosing Cognitive Load Measures in Perceptually Rich Environments. *Mind Brain Educ.* **2023**, *17*, 20–28. [[CrossRef](#)]
85. Skulmowski, A.; Rey, G.D. The Realism Paradox: Realism Can Act as a Form of Signaling despite Being Associated with Cognitive Load. *Hum. Behav. Emerg. Technol.* **2020**, *2*, 251–258. [[CrossRef](#)]
86. Liang, Z.; Zhou, K.; Gao, K. Development of Virtual Reality Serious Game for Underground Rock-Related Hazards Safety Training. *IEEE Access* **2019**, *7*, 118639–118649. [[CrossRef](#)]
87. Nykänen, M.; Puro, V.; Tiikkaja, M.; Kannisto, H.; Lantto, E.; Simpura, F.; Uusitalo, J.; Lukander, K.; Räsänen, T.; Heikkilä, T.; et al. Implementing and Evaluating Novel Safety Training Methods for Construction Sector Workers: Results of a Randomized Controlled Trial. *J. Saf. Res.* **2020**, *75*, 205–221. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.