# Sense of Presence, Realism, and Simulation Sickness in Operational Tasks: A Comparative Analysis of Virtual and Mixed Reality

Giorgio Ballestin[a] and Heike Diepeveen[b]
*Department of Maritime Operations - Human Factors, MARIN (Maritime Research Institute Netherlands), Haagsteeg, Wageningen, Netherlands*

Keywords: Human Computer Interaction, Empirical Studies in HCI, Mixed Reality, Sense of Presence, Simulation Sickness.

Abstract: Virtual Reality (VR) training is often used to replicate real-world situations, which brings many challenges. Interaction in VR is notoriously more difficult with respect to real world interaction, and VR wearables can sometimes decrease the quality of the interaction, leading to less effective simulations. While reliability and stability are important considerations, it is also crucial sometimes to have a high level of physical realism in order to develop proper muscle memory. This study aims to investigate whether Video-See-Through Augmented Reality (VST AR) can enhance the interaction during an operational task, and its impact on the immersivity and sense of presence, when compared to a VR based approach. Our results show that VR seems to perform better than AR in terms of sense of presence, with a much lower impact on sickness symptoms, suggesting that it is important to strive for a balance between reliability and realism in order to create an immersive and effective training environment.

## 1 INTRODUCTION

Virtual reality (VR) head-mounted displays (HMDs) have made significant progress in offering a natural vision experience. However, one of the current major challenges in the field is still to achieve a natural interaction with the virtual world. Proper hand interaction through hand-tracking is believed to be one of the requirements which enables natural interaction, albeit the lack of haptic feedback creates further difficulties. Hand-tracking technology is rapidly evolving, but still faces several problems, such as pose estimation during hand-to-hand self occlusion, or lack of tracking when the hands move outside the HMD tracking camera's Field of View (FoV), to name a few.

The trend with recent commercial VR HMDs is steering towards Mixed Reality (MR) designs which enable VST AR, in which real objects and tools can be used to interact with the virtual environment. In this study we focused on VST AR, but since the procedure can be generalized to the full MR continuum (as defined by (Milgram et al., 1995) (Skarbez et al., 2021)), for practical purposes we will use the terms VST AR and MR interchangeably. In some applications, such as simulators, the users' hands do not necessarily need to be tracked - but the general assumption is that the ability to see your own real hands (i.e. through depth-masking, chroma-key, or similar techniques) should make the interaction easier and more natural. However, VST AR is known to introduce latency and other optical mismatches with respect to natural vision.

Using tracking techniques to overlap real tools with their virtual 3D model, when applicable, is an effective approach to improve the interaction, as feeling the real object will provide the correct haptic feedback to the user. That being said, tracking the objects is sometimes not practical as it usually requires either the use of trackers (such as Vive Trackers, or motion capture systems), or vision-based pose-estimation systems, which needs to be fine-tuned specifically for every object that needs to be overlapped in VR.

Comparative studies between different MR and VR devices are frequent in the literature (Müller et al., 2019)(Rolland and Fuchs, 2000) (Tang et al., 2004)(Ballestin et al., 2021) . The approach is usually to collect various metrics, which are typically used to evaluate the effectiveness of different systems or technologies. These metrics can fall into sev-

[a] https://orcid.org/0000-0002-1435-3224
[b] https://orcid.org/0009-0008-3642-7037



Ballestin, G. and Diepeveen, H.
Sense of Presence, Realism, and Simulation Sickness in Operational Tasks: A Comparative Analysis of Virtual and Mixed Reality.
DOI: 10.5220/0012256300003660

In *Proceedings of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2024) - Volume 1: GRAPP, HUCAPP and IVAPP*, pages 341-349
ISBN: 978-989-758-679-8; ISSN: 2184-4321

eral categories, including usability (such as cognitive load, sense of presence, task completion time, and error rate), physiological factors (such as depth perception, simulator sickness symptoms, and optical aberrations), technological considerations (such as frame rate, latency, field of view, and resolution), or a combination of these factors. When simulating operational scenarios, some of the most important aspects are the sense of presence, the development of sickness symptoms, the cognitive load overhead required to operate the simulation and its realism.

Spatial presence, also known as sense of presence, is a psychological state that describes the feeling of authenticity and physicality experienced in virtual environments, as opposed to actual physical locations. It's a subjective experience that's closely linked to the user's perception and overall experience. In simpler terms, it has been defined as the subjective experience of being in one place or environment, even when one is physically situated in another place or environment (Witmer and Singer, 1998)(Heater, 1992). The presence of one's own body plays a significant role in enhancing both interaction and sense of presence, as these two concepts are closely intertwined. In fact, numerous studies have been conducted on the effects of Embodiment/Avatars, which are extensively documented in the literature. In VR environments, questionnaires are often used to measure presence (Youngblut, 2003)(Schuemie et al., 2001)(Riva et al., 2003): the IGroup Presence Questionnaire, the Presence Questionnaire, and the Slater, Usoh, and Steed (SUS) Questionnaire are well-established standardized surveys used to measure the sense of presence. The literature is less rich with respect to the rest of the MR spectrum. In (Wagner et al., 2009) Wagner et al. argue that the narrow psychological explanation of sense of presence in VR is not as meaningful in MR, when the experience steers away from laboratory settings. In (Toet et al., 2021)(Abbey et al., 2021), questionnaires to evaluate presence in MR are proposed by merging questions and definitions from past questionnaires, but their stability and sensitivity is not validated. Sometimes, questionnaires intended for VR (like the IPQ (Schubert et al., 2001)) are used in MR environments (Schaik et al., 2004)(Ballestin et al., 2021). Pushing the user to give its own interpretation to questions not particularly fit to a MR case is likely to invalidate the purpose of the questionnaire (Slater, 2004)(Usoh et al., 2000). The concept of "presence" is thus a complex and intangible psychological state that can be challenging to measure accurately. Despite ongoing efforts, there is currently no widely accepted consensus on the most effective methodology for measuring presence, especially in MR environments. Nevertheless, obtaining accurate measures of presence is critical for various applications in the simulation field.

Cybersickness is a common problem associated with the whole MR spectrum, and is often the cause of dropout in simulated training. The sensory conflict, poison, and postural instability theories attempt to explain why this occurs. The sensory conflict theory (Reason and Brand, 1975) suggests that discrepancies between visual and vestibular systems cause a perceptual conflict which is proportional to the experienced symptoms. The poison theory (Bouchard et al., 2011) suggests that these discrepancies trigger a physical response similar to the body's response to being poisoned, for an evolutionary factor. The postural instability theory (Riccio and Stoffregen, 1991) suggests that the sensory conflict caused by VR/MR environments hampers motion control leading to a situation of postural instability which is directly related with cybersickness. When comparing different HMDs, the level of visual conflict experienced can be more or less pronounced depending on the device's optical characteristics. Factors such as resolution, FoV, registration delays, and refresh rates can vary significantly across different HMDs. While VR devices used as VST MR devices will have equal resolution, FoV, and refresh rates, the MR case may still experience a co-locational mismatch due to differences in processing speeds between the video stream from the cameras and virtual imagery. Additionally, parallax-induced distortions caused by the offset between the camera sensors and the user's actual eyes may occur in the MR case. Although it is possible to predict which condition may perform better in a specific use case, sickness can be caused by several factors. Therefore, it is still essential to validate assumptions through operational testing.

Cognitive load is a psychological measure describing the mental effort an individual is experiencing during the execution of a certain task. A high cognitive task load can negatively impact metrics such as performance and transfer of training in a VR environment. Therefore, cognitive load is sometimes used as a metric in comparative studies, to infer the usability of the system (Armougum et al., 2019)(Frederiksen et al., 2020)(Collins et al., 2019). The construct can be assessed in three distinct ways, namely through performance measures, physiological measures and subjective measures. Performance measures are task oriented and include variables such as speed, accuracy and approach of task completion (Casner and Gore, 2010). Examples of physiological measures are heart rate variability and event related potentials (Delliaux et al., 2019)(Ghani et al., 2021). Subjective measures are the most common method and include questionnaires/scales, such as NASA Task Load Index (NASA TLX) (Hart and Staveland, 1988)(Hart, 2006), Subjective mental effort questionnaire (SMEQ) (Zijlstra and Van Doorn, 1985), the Modified Cooper-Harper Scale (Wierwille and Casali, 1983) and Instantaneous Self-assessment of Workload (ISA) (Tattersall and Foord, 1996). The NASA TLX consists of six sub scales, designed with the intent to conceptualize different definitions of workload. The ISA, Modified Cooper-Harper Scale and SMEQ on the other hand are one-dimensional and therefore cover less facets of workload. The measures are however validated and frequently used in simulator experiments due to their fast and easy administration (Luong et al., 2020)(Leggatt, 2005).

The main focus of this research is to assess whether VST MR can enhance the effectiveness of interaction in complex scenarios without disrupting the sense of presence in the simulation, or increasing the development of simulator sickness symptoms. To evaluate this, we performed a comparative experiment between pure VR and MR cases, in which the user had to perform many head movements, to mimic the behavior of operatives that need to use many monitors and devices at once. We recorded objective metrics such as the time to task completion and user questionnaires, such as the Simulator Sickness Questionnaire (SSQ) (Kennedy et al., 1993) and the aforementioned IPQ, to gather subjective feedback on the experience. Since the IPQ questionnaire was considered not completely suited for this scenario, an adapted version has been proposed.

The paper is organized as follows. Section 2 describes the experiment. More specifically, in Section 2.1, the equipment used during the experiment is listed. In Section 2.2 we describe the experimental setup, while in Sections 2.3 and 2.4 the participants and research methodology. The quantitative and qualitative results of our study are shown in Section 2.5. Finally, in Section 3, we discuss the obtained results and provide our conclusions.

# 2 THE EXPERIMENT

## 2.1 Equipment & Apparatus

The experimental setup was developed in Unreal Engine 5. The system used was a PC with a 12th Gen Intel(R) Core(TM) i9-12900K processor (3200 Mhz, 16 Cores), 64 GB of RAM, and a NVIDIA GeForce RTX 3080 GPU. The OS used was Microsoft Windows 10 Pro.

The device used for displaying Virtual Reality is the Varjo XR-3 [1]. The headset has a refresh rate of 90 Hz and a 115° field of view. The device uses two OLED panels, one per eye, each having a display resolution of 1920 x 1920 (at 70 pixels per degree) inside the (27°x 27°) Focus area and 2880 x 2720 in the peripheral area (which has 30 pixels per degree). The HMD features 200Hz eye tracking with sub-degree accuracy and 1-dot calibration for foveated rendering and two 12-Megapixel 90Hz video pass-through cameras for MR use. Hand tracking is provided by the embedded Ultraleap gemini (v5), but for the interaction task a controller from the HTC vive has been used. The HMD is tracked by means of the SteamVR 2.0 tracking system (Lighthouse).

The Varjo offers three ways to achieve a MR experience: depth based segmentation, model based, and chroma-key masking. The depth based segmentation works by showing the real world only up to a specified range, but the depth sensor has been considered too noisy to achieve a smooth blend. The model based segmentation allows to specify mask-objects within the graphical engine on which the video feed will be displayed instead. However, this kind of segmentation would not work properly with the user's hand, because although the built-in Ultraleap hand tracking provides skeletal joint locations - it does not provide the hands' accurate silhouette detection. This would lead to a mismatched overlap of the video feed to the actual user hands. The most stable segmentation, in controlled experimental settings, was deemed to be the chroma-key based one. For the sake of this experiment, a small portable green-screen has been used, with two diffuse lights that ensured uniform lighting of the scene.

## 2.2 Experimental Setup

The experimental setup consists of two conditions: the first one (VR condition) immerse the user in a completely virtual environment. The second one (MR condition) is showing the same environment of the VR condition, however only a portion of it is virtual. Since the HMD used for both conditions was the same, the main perceptual difference in the MR condition was given by the latency and noise introduced by the pass-through camera. One of our research question was to assess whether MR and VR were causing the development of simulator sickness symptoms in a similar way. To assess this, we designed the experiment to force the user to move the head around, similarly as in an operational scenario (e.g. when inside a cockpit). To reduce the duration

[1] https://varjo.com/products/xr-3

of each user session, we forced a higher frequency of movements with respect to a practical use case, to increase the speed at which adverse symptoms arise.

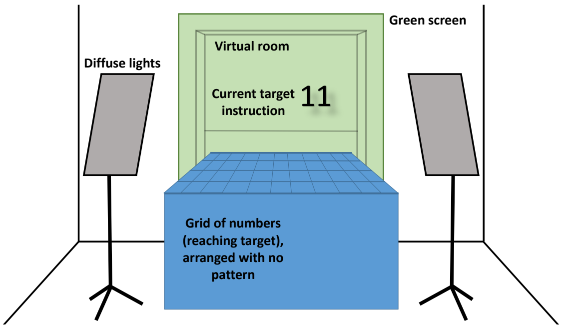


Figure 1: The MR experimental setup. A green-screen is used to display another room behind the (real) wall.

The scene overview of the experiment can be seen in Figure 1. We arranged a table in the middle of a room, with a grid (4 rows, 11 columns) of 44 numbers (between 1 and 44) displaced without any noticeable pattern. Each cell was a square of side 12cm. In front of the table, there was a wall with a door-less doorframe, with another empty closed room on the other side. In the MR case, the whole room was displayed through the pass-through camera, while the door and the second (virtual) room was displayed through the green screen. In the VR case, a virtual table was overlapped to the real one. The table thus acts as a physical anchor (landmark) to the real environment, to avoid the complete loss of the real physical spatial presence. The view from the two conditions can be seen in Figure 2.

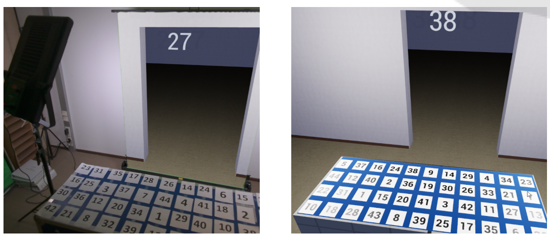

Figure 2: On the left: The MR Condition. On the right: The VR Condition.

The test subjects were asked to perform a seek and a reaching task in both experimental conditions, i.e., to find and touch a displayed number on the table. This forced the user to continuously move the head around. The indication of the current target is displayed through a 3D text which appears in random positions inside the virtual room. This also caused the numbers to sometimes appear partially covered by the walls, forcing the user to move and peek inside the room, to further increase the amount of head movements.

The reaching task was performed by means of an HTC Vive Controller. The test subject could see the controller mesh in the VR case. On the top of the controller, an invisible collider was placed to register the end effector position. On top of the current grid cell target, a 10 $cm^3$ invisible cubic collider triggered when the controller reached the correct position. Upon reaching the correct number, the next target was instantly displayed, and the time passed between each reaching was recorded.

All the 44 numbers were displayed once, in a randomized sequence. When the task was complete, all the collected data was serialized and exported into an external JSON file.

For both cases, the test subject was asked to fill out the Simulator Sickness Questionnaire (SSQ). One was administered before the start of the task, as a precondition, and one after the task had finished, as a post-condition. Since we expected the workload to be similar for both conditions, as the interaction task was simple, we decided to assess the mental workload through the Instantaneous Self-assessment of Workload (ISA) instead of using more elaborate scales such as the NASA-TLX (Hart and Staveland, 1988), to reduce the experimental complexity. This was only administered after the task, for each case.

Finally, a modified version of the Igroup Presence Questionnaire (IPQ) was used to measure spatial presence, immersion and realism. This questionnaire was also administered after the task, for both of the cases. We modified the questionnaire since the questions used in the original IPQ are focused on full virtual reality. For this reason, the decision was made to rephrase the questions in such a way, they could apply to VR, MR and simulator studies (see Table 1). When interpreting the results, it is therefore important to keep in mind that the renewed version of the questionnaire has yet to be validated.

The reasoning behind the rephrasing will be briefly discussed. In questions 1, 10, 12 and 13, ”*computer generated world/virtual world*” have been replaced with ”*simulation*”, to better represent MR situations where both real and virtual worlds might be present. Question 2 (*”Somehow I felt that the virtual world surrounded me”*) was considered confusing in MR environments, as the virtual world might appear only in front of the user, as it was happening in our case. It has thus been rephrased to assess whether the user spatial presence was also applied to virtual scenery. Question 3 (*”I felt like I was just perceiving pictures.”*) has been changed as it seemed outdated for current hardware. Commercially available VR/MR HMDs usually reach rendering rates of at least 90FPS (unless the application is badly optimized), leading most people to always answer ”fully disagree”. We thus replaced the question with another one, which we believe contributes to spatial presence, i.e., the perceived spatial registration error/lag between the real and virtual worlds (which is still a technological limitation with current devices). Question 5 (”*I had a sense of acting in the virtual space, rather than operating something from outside*”) has been slightly changed to better display the MR situation where virtual and real worlds are not registered correctly. Questions 7 and 8 in its original versions made little sense in a MR scenario (where you can see the real world surroundings), thus we rephrased them to encapsulate the broader concept of losing track of your real context situation while immersed in the simulated experience. Question 9 (”*I still paid attention to the real environment.*”) was also changed, and since it contributed to the involvement measurement, we tried to assess whether the user ”snaps back to reality” after the simulation ended. We feel like the new phrasing is easier to interpret in MR simulations, i.e. vehicle simulations with projectors and motion platforms. Question 11 (*”How real did the virtual world seem to you?”*) was slightly changed to be more clear in case only a couple elements of the scene are virtual. Finally, question 14 (*”The virtual world seemed more realistic than the real world.”*) was slightly changed as in MR participants often struggle to comprehend how the virtual world could be considered ”more realistic” compared to the tangible reality they can physically perceive.

To see a comprehensive overview of the variations to the original IPQ questionnaire, refer to Tab. 1.

## 2.3 Participants

13 volunteers (9 males, 4 females) participated in the experiment. The experiment has been performed within subject. To avoid any learning bias, half of the subjects tested the VR condition first, while the others tested the MR condition first.

They were aged between 28-55 years (mean 35 $\pm$ SD 10 years) and between 160-203 cm tall (mean 180.0 $\pm$ 12.4cm). The IPD of subjects was calibrated through the Varjo-XR3 automatic calibration procedure (which uses the HMD built in eye tracker). 6 users claimed they never used VR HMDs nor MR HMDs before. 4 had only tried it a couple of times, and with older HMDs. Finally, 3 were expert users (from gaming, or VR developers).

The experiment has been conducted at our research facility. No biometric data was collected, and every participant gave informed consent complying with the statutes of the Declaration of Helsinki (2008).

## 2.4 Methodology

The experiments were conducted as follow. First, the test subject was asked to fill the SSQ PRE questionnaire, to measure the baseline conditions. Then, the user was informed about the task. A brief explanation was given on how to properly wear and tighten the various adjustment wheels of the HMD, to ensure a snug fit was achieved. The automatic IPD calibration was then performed. The user was given a few moments to look around and get acquainted to the new environment, and a quick recap of the task was given again. When the user was ready, the experiment started from external input from the researcher. The test subject had then to perform 44 seek-and-reach interactions. When the test was complete, the user was helped to remove the HMD, and was then asked to fill the SSQ POST, the modified IPQ and ISA questionnaires. After filling out the questionnaires, the subject was asked to elaborate on their experiences through a short interview. The second condition, composed again by another set of 44 seek-and-reach interactions, was usually tested in a separate day. 2 participants performed it on the same day, but since task performance is not the main aim of this study, and the numbers were arranged without a pattern and in different positions in the two conditions, we deem no bias was introduced.

## 2.5 Results

We performed a Wilcoxon Signed Rank Test between the PRE and POST conditions of the SSQ questionnaire (see Fig. 3). The results show significant difference between every PRE and POST condition in the MR case (i.e. the paired, two-sided test rejects the null hypothesis of zero median in the difference between paired samples at the 5% significance level). The VR condition, by contrast, did not display a significant increase on any scale.

We performed a two sample t-test on the perceived workload score, which revealed no significant difference between the MR and VR case (see Fig. 4).

In Fig. 5 we plotted the result of the modified IPQ analysis. We retained the original weights from the initial questionnaire for each question, as the overall essence of the questions remained unchanged. Only minor adjustments were made to enhance readability and comprehension.

Responses to the interview questions showed a strong preference for VR. All participants reported a

Table 1: The rephrased IPQ questionnaire used in this study. In the original version, many questions are hard to interpret in a MR scenario, where the real world can also be seen (e.g. question 14). Questions with ”-” have not been changed.

| Question | IPQ questionnaire | Modified IPQ questionnaire |
|---|---|---|
| 1 | In the computer generated world I had a sense of ”being there” | In the simulation I had a sense of ”being there” |
| 2 | Somehow I felt that the virtual world surrounded me. | Somehow I felt like I was in the context/environment of the simulation. |
| 3 | I felt like I was just perceiving pictures. | I felt like virtual objects were displayed in the wrong position. |
| 4 | I did not feel present in the virtual space. | - |
| 5 | I had a sense of acting in the virtual space, rather than operating something from outside. | I had a sense of acting in a single space, rather than in two different systems. |
| 6 | I felt present in the virtual space. | - |
| 7 | How aware were you of the real world surrounding while navigating in the virtual world? (i.e. sounds, room temperature, other people, etc.)? | How much were you aware of your real context while immersed in the new simulated context? |
| 8 | I was not aware of my real environment. | I was not aware of my real context. |
| 9 | I still paid attention to the real environment. | When the simulation finished, I felt disoriented at first, before remembering where I was. |
| 10 | I was completely captivated by the virtual world. | I was entirely concentrated on the simulation. |
| 11 | How real did the virtual world seem to you? | How realistic did the virtual world/objects seem to you? |
| 12 | How much did your experience in the virtual environment seem consistent with your real world experience? | How much did your experience in the simulation seem consistent with your real world experience |
| 13 | How real did the virtual world seem to you? | How real did the simulation seem to you? |
| 14 | The virtual world seemed more realistic than the real world. | The virtual world seemed as realistic as the real world. |

substantial discomfort difference upon the first moment of experiencing the other condition, indicating no bias due to order effects. The most important argument for this, reported by 11 out of 13 participants, was that the camera view in MR that represented the real world was lagging behind, which led to a blurry vision and discomfort. VR was generally considered more immersive and easier to perform the task in (as reported by 10 of the participants). However, there was also critique from 5 participants on the

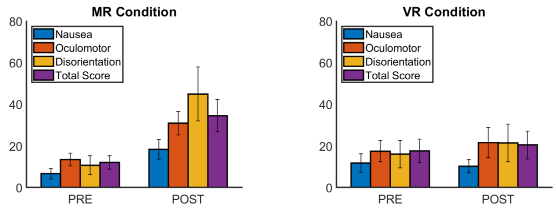


Figure 3: The results of the SSQ questionnaire for the MR and the VR conditions (left and right, respectively). Each plot shows the mean values, averaged on all the participants, and the associated standard deviation, for 3 subscales and the total score. The SSQ questionnaire has been filled before (PRE) and after the experiment (POST).


Figure 4: The ISA workload scale.

way the peripheral vision tended to become blurry. The average TTC was similar in the two conditions (mean 5.19 ± 4.4 seconds in the MR case, and mean 5.19 ± 4.0 seconds in the MR case)

## 3 CONCLUSIONS

Our result shows that in situations where many frequent fast head rotations and movements are performed, VR leads to a lower impact on the developed sickness symptoms, and a better sense of immersion. It must be noted that, since the frequency of the head movements was higher with respect to most practical scenarios, it is possible that this difference would not be perceivable in normal settings. That being said, from the interviews participants expressed how the discomfort arose even before starting the task, thus considering the relative difference between the two techniques it should still be advisable to choose VR when applicable. Moreover, in our experiment the MR condition was composed by mostly real environment, while the virtual augmentation composed a smaller fraction of the perceived image. Our assumption is that by minimizing the amount of real elements in the scene, the results would improve, as the mismatch between the expected and perceived visuals would be reduced due to the lower latency required

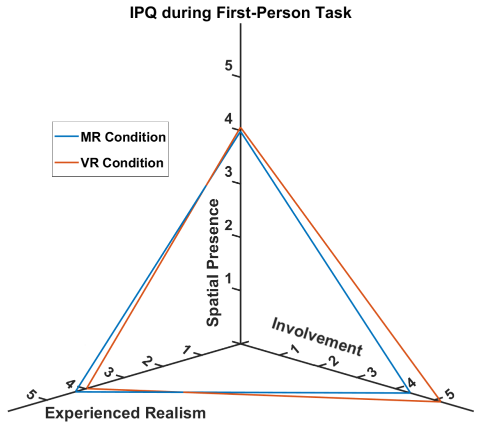


Figure 5: The modified IPQ results. The VR condition scores higher involvement, comparable spatial presence and slightly lower experienced realism. VR scored an average of 5.85 ± 0.83 in General Presence, while MR scored an average of 4.38 ± 1.55.

to render virtual imagery w.r.t. images captured by the cameras. In these specific settings, it is therefore questionable whether there is any added value of MR using VST technology, which leads to additional problems such as scene illumination or legibility problems, over a VR approach, where real items are overlapped with their 3D virtual model (some would call it Augmented Virtuality, or AV). The authors of this study thus suggest using MR only for displaying the user's real hands inside an AV setting, minimizing the negative effects of the VST delays, and allowing proper hand interactions that would otherwise be difficult with the current instability of hand tracking softwares. Achieving a good quality hand segmentation is however still challenging by itself. We obtained good results using the chroma-key technique, which has the limitation of requiring all the environment to be of the same key color (green). In a controlled simulation environment however this can be an acceptable compromise.

It must be noted how the experienced difference between VR and MR conditions should not be generalized across technologies as our experiment was carried out with a specific HMD brand. Different HMDs might have different optical characteristics which might yield different results. That being said, the experienced issues (e.g., camera latency) are common across most currently available HMDs, which suggest caution whenever choosing between VST-AR and VR modes.

A modified version of the IPQ has been proposed to better fit a VR-MR comparison. We are aware that changing the questions partially invalidates the way that the IPQ scores are computed, as the new questions potentially do not overlap perfectly the original questions intents. However, the original questionnaire questions would have been most likely misinterpreted by all the test subjects. Moreover, the interviews we collected confirm the results of the collected questionnaires. We recommend to further refine this questionnaire into an up-to-date version which more generically encapsulates the meaning of immersivity in the complete mixed-reality continuum, rather than being restricted to VR. Some recent studies that investigated the subject include the work from (Brübach et al., 2022)(Westermeier et al., 2023). It would be required to merge and adapt several of them into a singular questionnaire, which could be first evaluated with a card sorting procedure and finally reduced to a minimum number of items before being further validated, e.g. through Bayesian Exploratory Factor Analysis (BEFA) and Bayesian Confirmatory Factor Analysis (BCFA). Finally, this study showed the results obtained by using a specific HMD brand (Varjo XR3), a possibly interesting comparison would be to show the differences between all the HMDs currently available on the market.

## REFERENCES


Abbey, A., Porssut, T., Herbelin, B., and Boulic, R. (2021). Assessing the impact of mixed reality immersion on presence and embodiment. In *Proceedings of the 14th ACM SIGGRAPH Conference on Motion, Interaction and Games*, pages 1–10.

Armougum, A., Orriols, E., Gaston-Bellegarde, A., Joie-La Marle, C., and Piolino, P. (2019). Virtual reality: A new method to investigate cognitive load during navigation. *Journal of Environmental Psychology*, 65:101338.

Ballestin, G., Chessa, M., and Solari, F. (2021). A registration framework for the comparison of video and optical see-through devices in interactive augmented reality. *IEEE Access*, 9:64828–64843.

Bouchard, S., Robillard, G., Renaud, P., and Bernier, F. (2011). Exploring new dimensions in the assessment of virtual reality induced side effects. *Journal of computer and information technology*, 1(3):20–32.

Brübach, L., Westermeier, F., Wienrich, C., and Latoschik, M. E. (2022). Breaking plausibility without breaking presence-evidence for the multi-layer nature of plausibility. *IEEE Transactions on Visualization and Computer Graphics*, 28(5):2267–2276.

Casner, S. M. and Gore, B. F. (2010). Measuring and evaluating workload: A primer. *NASA Technical Memorandum*, 216395:2010.

Collins, J., Regenbrecht, H., Langlotz, T., Can, Y. S., Ersoy, C., and Butson, R. (2019). Measuring cognitive load and insight: A methodology exemplified in a virtual reality learning context. In *2019 IEEE International symposium on mixed and augmented reality (ISMAR)*, pages 351–362. IEEE.

Delliaux, S., Delaforge, A., Deharo, J.-C., and Chaumet, G. (2019). Mental workload alters heart rate variability, lowering non-linear dynamics. *Frontiers in physiology*, 10:565.

Frederiksen, J. G., Sørensen, S. M. D., Konge, L., Svendsen, M. B. S., Nobel-Jørgensen, M., Bjerrum, F., and Andersen, S. A. W. (2020). Cognitive load and performance in immersive virtual reality versus conventional virtual reality simulation training of laparoscopic surgery: a randomized trial. *Surgical endoscopy*, 34:1244–1252.

Ghani, U., Signal, N., Niazi, I. K., and Taylor, D. (2021). Efficacy of a single-task erp measure to evaluate cognitive workload during a novel exergame. *Frontiers in Human Neuroscience*, 15:742384.

Hart, S. G. (2006). Nasa-task load index (nasa-tlx); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 50, pages 904–908. Sage publications Sage CA: Los Angeles, CA.

Hart, S. G. and Staveland, L. E. (1988). Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, volume 52, pages 139–183. Elsevier.

Heater, C. (1992). Being there: The subjective experience of presence. *Presence Teleoperators Virtual Environ.*, 1(2):262–271.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., and Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3):203–220.

Leggatt, A. (2005). Validation of the isa (instantaneous self assessment) subjective workload tool. In *Proceedings of the International Conference on Contemporary Ergonomics*, pages 74–78.

Luong, T., Argelaguet, F., Martin, N., and Lécuyer, A. (2020). Introducing mental workload assessment for the design of virtual reality training scenarios. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 662–671. IEEE.

Milgram, P., Takemura, H., Utsumi, A., and Kishino, F. (1995). Augmented reality: A class of displays on the reality-virtuality continuum. In *Telemanipulator and telepresence technologies*, volume 2351, pages 282–292. Spie.

Müller, J., Zagermann, J., Wieland, J., Pfeil, U., and Reiterer, H. (2019). A qualitative comparison between augmented and virtual reality collaboration with handheld devices. In *Proceedings of Mensch und Computer 2019*, pages 399–410.

Reason, J. T. and Brand, J. J. (1975). *Motion sickness.* Academic press.

Riccio, G. E. and Stoffregen, T. A. (1991). An ecological theory of motion sickness and postural instability. *Ecological psychology*, 3(3):195–240.

Riva, G., Davide, F., and IJsselsteijn, W. A. (2003). *Being there: Concepts, effects and measurements of user presence in synthetic environments.* Ios Press.

Rolland, J. P. and Fuchs, H. (2000). Optical versus video see-through head-mounted displays in medical visualization. *Presence*, 9(3):287–309.

Schaik, P. V., Turnbull, T., Wersch, A. V., and Drummond, S. (2004). Presence within a mixed reality environment. *CyberPsychology & Behavior*, 7(5):540–552.

Schubert, T., Friedmann, F., and Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence: Teleoperators & Virtual Environments*, 10(3):266–281.

Schuemie, M. J., Van Der Straaten, P., Krijn, M., and Van Der Mast, C. A. (2001). Research on presence in virtual reality: A survey. *Cyberpsychology & behavior*, 4(2):183–201.

Skarbez, R., Smith, M., and Whitton, M. C. (2021). Revisiting milgram and kishino's reality-virtuality continuum. *Frontiers in Virtual Reality*, 2:647997.

Slater, M. (2004). How colorful was your day? why questionnaires cannot assess presence in virtual environments. *Presence*, 13(4):484–493.

Tang, A., Biocca, F., Lim, L., et al. (2004). Comparing differences in presence during social interaction in augmented reality versus virtual reality environments: An exploratory study. *Proceedings of PRESENCE*, pages 204–208.

Tattersall, A. J. and Foord, P. S. (1996). An experimental evaluation of instantaneous self-assessment as a measure of workload. *Ergonomics*, 39(5):740–748.

Toet, A., Mioch, T., Gunkel, S. N., Niamut, O., and van Erp, J. B. (2021). Assessment of presence in augmented and mixed reality: Presence in augmented and mixed reality. In *Conference on Human Factors in Computing Systems, CHI 2021*. ACM SigCHI.

Usoh, M., Catena, E., Arman, S., and Slater, M. (2000). Using presence questionnaires in reality. *Presence*, 9(5):497–503.

Wagner, I., Broll, W., Jacucci, G., Kuutii, K., McCall, R., Morrison, A., Schmalstieg, D., and Terrin, J.-J. (2009). On the role of presence in mixed reality. *Presence*, 18(4):249–276.

Westermeier, F., Brübach, L., Latoschik, M. E., and Wienrich, C. (2023). Exploring plausibility and presence in mixed reality experiences. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):2680–2689.

Wierwille, W. W. and Casali, J. G. (1983). A validated rating scale for global mental workload measurement applications. In *Proceedings of the human factors society annual meeting*, volume 27, pages 129–133. Sage Publications Sage CA: Los Angeles, CA.

Witmer, B. G. and Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence*, 7(3):225–240.

Youngblut, C. (2003). Experience of presence in virtual environments. Technical report, INSTITUTE FOR DEFENSE ANALYSES ALEXANDRIA VA.

Zijlstra, F. and Van Doorn, L. (1985). *The construction of a scale to measure perceived effort*. University of Technology.