Using Presence Questionnaires in Virtual Reality

Valentin Schwind
University of Stuttgart
Stuttgart, Germany
valentin.schwind@acm.org

Pascal Knierim
LMU Munich
Munich, Germany
pascal.knierim@ifi.lmu.de

Nico Haas
University of Stuttgart
Stuttgart, Germany
nicohaasni@gmx.de

Niels Henze
University of Regensburg
Regensburg, Germany
niels.henze@ur.de

ABSTRACT

Virtual Reality (VR) is gaining increasing importance in science, education, and entertainment. A fundamental characteristic of VR is creating *presence*, the experience of 'being' or 'acting', when physically situated in another place. Measuring presence is vital for VR research and development. It is typically repeatedly assessed through questionnaires completed after leaving a VR scene. Requiring participants to leave and re-enter the VR costs time and can cause disorientation. In this paper, we investigate the effect of completing presence questionnaires directly in VR. Thirty-six participants experienced two immersion levels and filled three standardized presence questionnaires in the real world or VR. We found no effect on the questionnaires' mean scores; however, we found that the variance of those measures significantly depends on the realism of the virtual scene and if the subjects had left the VR. The results indicate that, besides reducing a study's duration and reducing disorientation, completing questionnaires in VR does not change the measured presence but can increase the consistency of the variance.

CCS CONCEPTS

• **Human-centered computing → HCI design and evaluation methods**; **Virtual reality**; *User studies*;

KEYWORDS

Virtual reality; presence; questionnaire; evaluation.

ACM Reference Format:
Valentin Schwind, Pascal Knierim, Nico Haas, and Niels Henze. 2019. Using Presence Questionnaires in Virtual Reality. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland UK.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3290605.3300590

ACM ISBN 978-1-4503-5970-2/19/05...$15.00
https://doi.org/10.1145/3290605.3300590

1 INTRODUCTION

The key characteristic of virtual reality (VR) is the ability to create a sense of *presence* [11, 26, 43], the feeling of *being* or *acting* in a place, even when one is physically situated in another location [1, 24]. To create immersive VR experiences and to study the interaction in VR, it is, therefore, crucial to reliably measure *presence*. Previous work developed increasingly sophisticated approaches to assess presence. While multiple physiological measures have been proposed [18], validated questionnaires are still the most common method for measuring this construct [10]. Using different items and subscales such questionnaires provide scores, which reflect the level of felt presence in the virtual world (*cf.* Table 1).

The use of validated presence questionnaires is omnipresent in research and industry. These standardized questionnaires are filled in using pen and paper. Typical studies repeatedly assess presence, especially when multiple VR experiences are compared. As current VR experiences are presented through head-mounted displays (HMDs), users must remove the headset and leave the VR before completing the questionnaire. In doing so, the person has to re-orientate in the real-world which causes a so-called "break-in-presence (BIP)" [11]. Slater and Steed describe it as a moment when "a report can be given that a break has occurred without this in itself disturbing the sense of presence, which of course has already been disturbed" [36]. This means that surveying subjects about their feeling of presence potentially causes the BIP and compromises the phenomenon that the questionnaire is supposed to be measuring [26, 34, 36].

Leaving VR can cause BIPs which distort the phenomenon that presence questionnaires measure [26, 34, 36]. Furthermore, leaving and re-entering the VR takes time not only because it requires removing and putting on the HMDs but also reorientation in the real-world and when entering the VR experience again.

Instead of requiring participants to leave the VR to fill in questionnaires, we propose to survey participants directly within the VR using existing questionnaires. Surveying participants during the VR experience have a number of potential advantages:

- Assessing their sense of presence becomes easier for participants when BIPs are reduced.
- Staying in VR reduces study duration and avoids the need to adjust after a BIP.

- Distracting or biasing cues from the real-world, such as the experimenter, are avoided.

Recent work has already started to assess presence by surveying people directly within the VR [29, 30]. However, existing presence questionnaires have been developed and validated over more than two decades by asking participants to fill them *after* leaving the VR. Therefore, the effect of filling questionnaires within the VR is unclear, and it is unknown if existing questionnaires provide meaningful results when filled without leaving the VR.

In this paper, we investigate if answering questionnaires in VR has an effect on three standardized presence questionnaires, originally developed to be answered in the real world. We conducted a mixed-design study with 36 participants who experienced two different immersion levels. We found no significant differences between the mean scores of the three questionnaires, which we measured inside or outside the VR. However, we found that the variance of those scores significantly depends on the realism of the scene when the participants answered the questionnaires after leaving the VR. Furthermore, the variance is consistent when presence scores are recorded in VR. This indicates that the feeling of presence depends on the contrast in the visual realism of the virtual environments (VEs) and *where* the BIP occurred. We argue to use presence questionnaires in VR, provide design implications, and provide directions for optimizing presence questionnaires for in-situ use.

2 RELATED WORK

In 1980, Minsky introduced the concept of telepresence describing human operators interacting through a remote video robot [19]. Convinced that this is technically feasible, researchers and engineers developed systems that creating an immersive illusion of being somewhere else and it became evident that the concept of presence also exists for VR systems [15, 25, 37]. Over more than two decades, previous work developed, refined and validated questionnaires to measure presence created by VR which we summarize in the following. Afterward, we discuss initial work that used questionnaires for measuring presence within the application causing the experience.

Measuring Presence

Barfield and Weghorst proposed one of the earliest approaches to measuring presence for VEs [3]. They investigated the effect of varying the update rate of a computer-generated simulation on the sense of presence within stereoscopic VEs [3]. They developed a 6-item questionnaire to measure presence and a 7-item questionnaire to measure the fidelity of the interaction [3]. Some items, however, contained experiment-specific questions and, thus, had to be further refined.

Based on previous work and a theoretical approach by Sheridan [31], who determined the underlying factors of presence (sensory information, sensor control, motor control), Wittmer and Singer [42, 43] developed a 32-item presence questionnaire. The authors identified three subscales which they labeled: *involvement/control*, *natural*, and *interface quality*. They also developed the immersive tendency questionnaire (ITQ) to determine characteristics of subjects, which potentially cause biases and affect subsequent judgments of presence [7]. However, the questionnaire by Witmer and Singer was criticized for the subjectively defined factors and the low number of items directly assessing presence [27, 33]. Nevertheless, the presence questionnaire by Witmer and Singer (in the following referred to as WS) is currently the most cited presence questionnaire on Google scholar (*cf.* Table 1).

Another approach to measuring presence was developed by Slater et al. [35, 39] and Usoh et al. [38] in multiple studies. The Slater-Usoh-Steed (in the following referred to as SUS) questionnaire is based on questions that are variations of one of the three themes: (1) *the sense of being in the VE*, (2) *the extent to which the VE becomes the dominant reality*, and (3) *the extent to which the VE is remembered as a "place"*. The current version of the questionnaire has six items and is the second most cited presence questionnaire applicable for VEs. A study by Usoh et al. revealed that WS and SUS are not able to discriminate between presence in a VE and physical reality [39]. Furthermore, Sanchez-Vives and Slater argued that using questionnaires causes a "methodological circularity" as surveying about presence in VR might one bring about the phenomenon of presence that the questionnaire is supposed to be measuring [26].

Slater stated that a scientific basis for presence could not be established on the basis of post-experience presence questionnaires [34]. He concluded that "presence researchers must move away from questionnaires in order to make any progress in this area" [34]. It is conceivable that behavioral or physiological measures are the more reliable measures of presence. While there has been little research into whether behavioral measures are reliable enough [10], there are promising results with physiological measures such as heart rate [18]. However, physiological measurements require a baseline comparison for each participant, which means a considerable effort in some study designs. Furthermore, it has been shown that additional equipment to measure physiological responses can be "the greatest cause of breaks in presence" [18].

Subjective questionnaires are currently the most common method for measuring presence and have been shown to be

Table 1: Overview and comparison of 15 published presence questionnaires

Authors	Year	Citations*	Items	Usage
Banos et al. [2]	1998	146	77	VE
Barfield & Hendrix [3]	1995	186	5+1	VE
Cho et al. [4]	2003	34	4	VE
Dinh et al. [6]	1999	365	13+1	VE
Gerhard et al. [9]	2001	57	19+4	SVE
Kim & Biocca [12]	1997	664	8	VE
Krauss et al. [13]	2001	8	42	VE
Lombard & Ditton [16]	2000	205	103	NA
Lombard & Weinstein (TPI) [17]	2009	120	4-8	CM
Lessiter et al. (ITC-SOPI)[14]	2001	861	44	CM
Nichols et al. [20]	2000	158	9	VE
Nowak & Frank [21]	2003	569	9	SVE
Schubert et al. [23, 27, 28] (IPQ)	**2001**	**758**	**14**	**VE**
Usoh/Slater et al. [36, 39] (SUS)	**1994/2000**	**853/466**	**3/6**	**VE**
Witmer & Singer [43] (WS)	**1998**	**3569**	**32**	**VE**

VE = Virtual Environment, CM = Cross-Media,
SVE = Shared Virtual Environment, NA = items not listed
* determined using Google Scholar, Sept 2018

sensitive enough to find differences in presence [10]. Schubert et al. investigated the cognitive processes leading to a model of presence and explored each component of the construct [27]. They identified spatial-construction, attention, and judgments of realness as distinct components of presence. Building on these results the authors developed and verified a 13-item presence scale consisting of three independent components called: *spatial presence*, *involvement*, and *experienced realism*. One item with a mild double loading on *involvement* was added in their final version of the iGroup presence questionnaire (IPQ)[1].

Further presence questionnaires have been developed focusing on specific applications. Examples include the ITC-Sense of presence inventory (ITC-SOPI) [14] and the temple presence inventory (TPI) [17] for non-interactive media as well as questionnaires focusing on social environments [9, 21]. Particularly the concepts of social or co-presence were investigated using a considerable number of measurement methods and questionnaires [5, 22, 41], but they are beyond the scope of our work.

Integrated Questionnaires

The idea to integrate questionnaires directly into the application causing the experience has already been facilitated in other domains. Shute, for example, recommended less obtrusive methods for computer-based video games to gain feedback during the runtime of the game [32]. This was further investigated by Frommel et al., who determined how presence in video games for PCs could be maintained for self-reports [8]. They compared presence measures of the Witmer and Singer presence questionnaire (WS) in games obtained using simple user interfaces with active game elements representing the questionnaire items (*e.g.*, a 7-lane road instead of Likert-items). The presence scores were measured with desktop computers and were higher for questionnaires represented by interactive gaming elements.

Previous work already used presence questionnaires in VR: To investigate the effects of gender and the avatar hand appearance on presence in VR, Schwind et al. used integrated questionnaires, which were presented on virtual walls [30]. Using the gesture recognition of the Leap Motion sensor, the subjects answered the questions using the virtual hands whose effect was measured. The underlying assumption was that having the manipulated aspect, the appearance of the hands, in view when measuring its effect would lead to more reliable results. Similarly, the same procedure was used in a study by Schwind et al. [29] to examine the effect of having fewer fingers in VR. Unfortunately, both studies did not compare their measures to the established approach for filling out presence questionnaires in the real-world [29, 30].

Summary

Previous research highlights the importance of presence as an outcome of immersive virtual environments [3, 31, 43] and multiple approaches to measuring presence have been suggested [27, 33, 35, 38, 39, 43]. Researchers have serious concerns that presence questionnaires are not able to discriminate between presence in VE and the real-world and there have been vigorous debates over measuring presence [33, 34, 36].

Presence questionnaires used today have been carefully designed and refined over more than two decades. While asking participants to fill our questionnaires directly in VR might have advantages, work on video games suggests that measuring presence within the application causing the experience might lead to different results [8, 32]. Recent work has already used questionnaires in VR [29, 30] without considering its potential effects. It is unclear if measuring presence in VR differs from measuring presence in the real-world.

3 METHOD

Study Design

To investigate the effect of presence questionnaires integrated into VR, we conducted a study with the two independent variables (IVs) Environment and Virtual Realism. Environment has the two levels *inside* and *outside the VR*. Virtual Realism has the two levels *abstract* and *realistic scene*. We measured presence through three different Questionnaires: Slater-Usoh-Steed questionnaire (SUS) [36, 39], igroup presence questionnaire (IPQ) [23, 27, 28], and the

[1] http://www.igroup.org/pq/ipq/

Figure 1: Virtual (l) and real (r) environment with questionnaire and input controller.

questionnaire by Witmer and Singer (WS) [43]. As the participants should not experience the *abstract* or *realistic* scene twice, the Environment in which the questionnaires were completed was a between-subject variable. To prevent any order effects we counterbalanced the questionnaires using a 3 × 3 Latin Square, doubled by the conditions *in-* or *outside the VR*, doubled by the two possible sequences of the *abstract* and *realistic* scene, which resulted in a total of 12 conditions, which were covered in triplicate by the number of our participants (N=36).

Virtual Scenes and Tasks

To study the effects of Environment on the three questionnaires, we developed a first-person drone shooter game with two levels of Virtual Realism (see Figure 2). In both conditions, the participants were located in an open space equipped with a weapon to strike drones that exploded and disappeared when hit. Drones served as moving targets and flew a predefined route. Enlarged hitboxes encasing the drones remained the same for both scenes to keep the difficulty of striking drones equal among the conditions. They appeared in three to four waves of three drones each.

To induce different feelings of presence, we varied the visual and auditory fidelity of the scenes (see Figure 1). The *abstract* environment included only simple untextured cubes of different sizes forming walls and a blue floor. Flat cubes represented the drones while simple octagon rods served as a weapon. Further, only a simple explosion effects without sound, smoke, and sparks was displayed when a drone was hit. In contrast, the *realistic* scene was based on the high fidelity *Adam Exterior Environment*[2] with detailed drones[3] as targets.

For the *inside the VR* condition, we designed a virtual representation of the laboratory in our institution. Using the virtual replica of a real world laboratory was already proposed for smooth transitions between the real world and immersive environments in order to reduce the state of confusion [40]. Real and virtual laboratory scene included an iMac for displaying and answering the presence questionnaires (see Figure 1). Similar to the game scenes, the Oculus Touch controller was used as an input device since it gave visual cues for natural interaction within the VR. This controller was also utilized in the *outside the VR* condition with the distinction that the participant was not immersed in VR and filled in the questionnaires sitting in front of a real iMac, which displayed the questions.

All VR scenes were designed to be explored in a seated situation and presented to the participants using an Oculus Rift CV1. To ensure fluent VR experiences and that the refresh rate remained constant for the HMD (90 FPS) in all scenes, we used state-of-the-art gaming hardware using an NVIDIA GeForce GTX 1060 - 6 GB video card, an Intel i7-4790 CPU running at 3.60 GHz, and 16 GB of memory.

Measures

Presence scores of the three questionnaires were obtained using a total of 52 items (6 SUS, 14 IPQ, 32 WS) on 7-point scales (see Figure 1). All items were presented as described in the original work. Response times were recorded by our software. As we hypothesized that the ability to recall the presence state after BIP decreases the consistency between our measures, we also used the absolute difference from the means as the measure of variability. To assess the perceived workload using VR and the real-world questionnaires, we used the NASA TLX with 6 items.

Procedure

After welcoming the participants, we asked them to sign the consent form and take a seat. We explained the course of the study to the participants and gave them a brief introduction into VR. We further explained how to interact within the designed scenes to strike drones and navigate through the questionnaires using the Oculus Touch controller. Then, we adjusted the HMD to the participant's head and calibrated it to the participant's inter-pupil distance for best visual results.

[2]https://assetstore.unity.com/packages/essentials/tutorial-projects/adam-exterior-environment-74969
[3]https://assetstore.unity.com/packages/3d/characters/robots/drone-controller-pc-joystick-mobile-61327

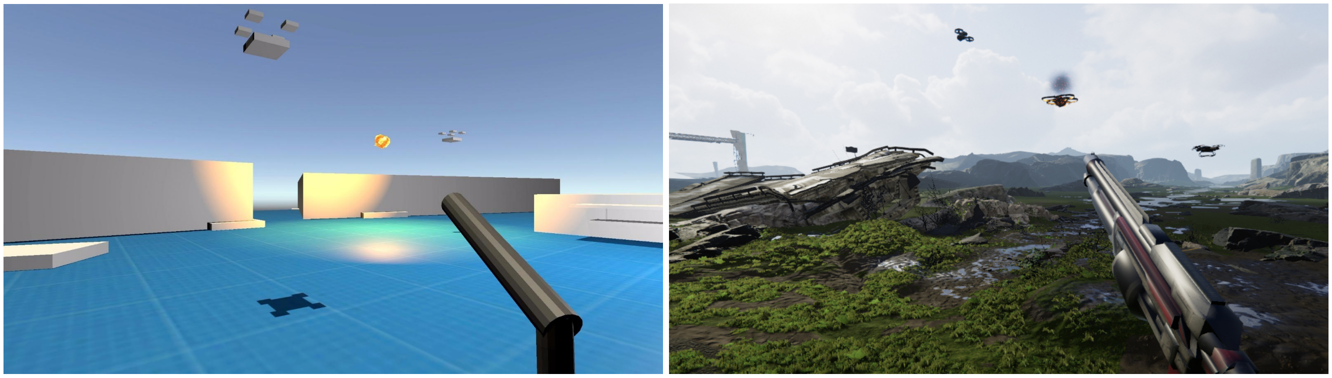

Figure 2: Abstract and realistic scene of a first person shooter game developed to induce different levels of presence.

Then participants entered the VR and explored either the *immersive* or *abstract* scene first. After spending approximately 2.5 minutes in VR striking drones, we asked them to answer the presence questionnaires. Depending on the condition, they filled the questionnaire *in the VR* scene representing our lab on a virtual iMac, or the participant was asked to exit the VE and answer the question *outside the VR*. All questionnaires were presented in a counterbalanced order to prevent any sequence effects. Progress between the items was shown in percent. After the first condition, the process was repeated with the other level of VIRTUAL REALISM. After experiencing both scenes and answering all questionnaires inside or outside the VR, we asked participants to answer the NASA TLX questionnaire concerning filling the previous questionnaires, not the drone game. The NASA TLX was always presented on paper. Finally, we collected demographic data and debriefed the participants. On average, participants completed the study in 24.2 minutes ($SD = 6.7$).

Participants

We recruited 36 participants (24 male, 12 female) via our universities' mailing lists with ages ranging from 19 to 32 years ($M = 23.81, SD = 2.81$). All of them had a technical background in computer science or engineering. Participants received either 5 EUR or course credits as compensation for their participation. A number of 13 participants wore glasses, 20 had previous VR experience. The study received ethics clearance according to the ethics and privacy regulations of our institution.

4 RESULTS

Questionnaire Scores

We obtained the presence scores for each of the three questionnaires by averaging their 7-point scores. As the questionnaires originate from ordinal scales, we performed non-parametric tests for statistical evaluation. We used linear mixed-model analysis using ARTools by Wobbrock et al. [44] to perform non-parametric tests with the two independent variables VIRTUAL REALISM and ENVIRONMENT. Participant was entered as a random factor.

We found a main effect of VIRTUAL REALISM, $F(1, 34) = 11.518, p = .002$; however, not of ENVIRONMENT, $F(1, 34) = .002, p = .958$, and no interaction effect of VIRTUAL REALISM × ENVIRONMENT, $F(1, 34) = .007, p = .933$ on the scores of the SUS. Similarly, there was a main effect of VIRTUAL REALISM, $F(1, 34) = 15.552, p < .001$; however, not of ENVIRONMENT, $F(1, 34) = .013, p = .909$, and no interaction effect of VIRTUAL REALISM × ENVIRONMENT, $F(1, 34) = .004, p = .946$ on the scores of the IPQ. Furthermore, there was a main effect of VIRTUAL REALISM, $F(1, 34) = 17.078, p < .001$; however, not of ENVIRONMENT, $F(1, 34) < .001, p = 1.000$, and no interaction effect of VIRTUAL REALISM × ENVIRONMENT, $F(1, 34) < .001, p = .759$ on the scores of the WS presence questionnaire.

Thus, the results did not reveal that leaving the VR significantly influenced the mean scores of the questionnaires. The mean scores between the questionnaires only showed significant differences when VIRTUAL REALISM changed. All means of the questionnaire scores are shown in Figure 3.

Score Variance

The absolute difference from the means was taken as a measure of variance. Means of the scores' variance are shown in Figure 3. Figure 4 shows the Gaussian distribution curve fits of those differences. We performed a multivariate analysis of variance (MANOVA) to assess the overall effect on the scores' variance and to determine if the variance measures are independent. As the statistical results are based on parametric data ARTool was not used.

We found no statistically significant differences for VIRTUAL REALISM, $F(1, 68) = 1.897, p = .138$, Wilk's $\lambda = .920$, $\eta_p^2 = .016$, or ENVIRONMENT, $F(1, 68) = .148, p = .930$,Wilk's

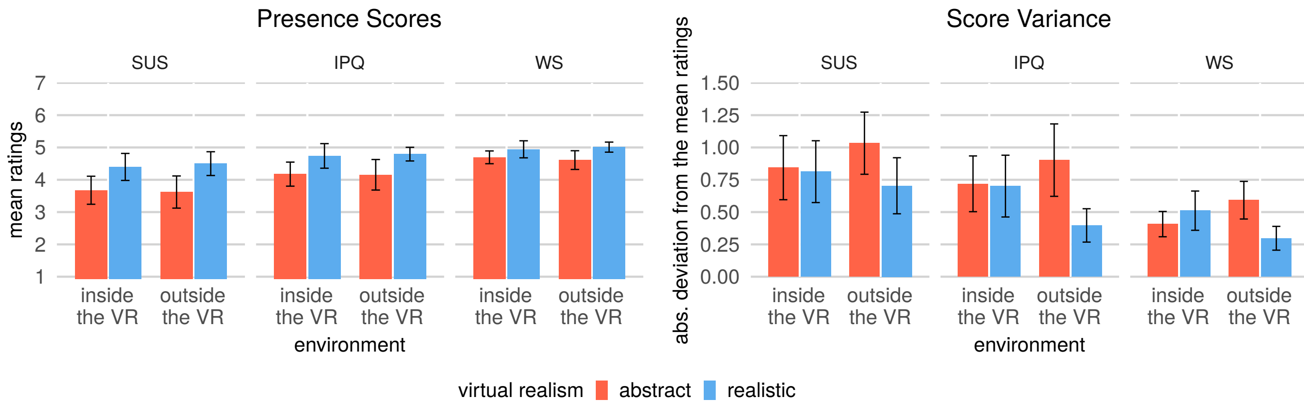

Figure 3: Means and score variances of the presence measures. While presence scores of the three questionnaires did not differ significantly between in- and outside the VR, variance measures show significant interaction effects between the questionnaire environment and the realism during the VR experience. Error bars show 95% confidence intervals (CI95)

$\lambda = .993$, $\eta_p^2 = .061$; however, there was a significant interaction effect of VIRTUAL REALISM × ENVIRONMENT, $F(1, 68) = 2.826, p = .045$, Wilk's $\lambda = .866, \eta_p^2 = .133$. Univariate analysis of variance (ANOVA) revealed no main or interaction effect on the variance of the SUS score, a significant interaction effect of VIRTUAL REALISM × ENVIRONMENT on the IPQ, $F(1, 34) = 6.792, p = .013$, and a significant interaction effect on the variance of the WS, $F(1, 34) = 10.981, p = .002$.

To determine if there were significant differences between the variance measures of the three questionnaires, we used QUESTIONNAIRE as additional IV and performed a three-way ANOVA.

We found significant main effects of QUESTIONNAIRE, $F(2, 204) = 11.860, p < .001$, VIRTUAL REALISM, $F(1, 204) = 7.199, p = .007$, and a significant interaction effect of VIRTUAL REALISM × ENVIRONMENT, $F(1, 204) = 8.762, p = .003$. There were no further significant main or interaction effects (all with $p > .645$). Bonferroni-corrected pairwise comparisons revealed a significant difference between the SUS ($M = .848, SD = .576$) and the WS ($M = .451, SD = .318$), ($p < .001$), as well as the IPQ ($M = .679, SD = .562$) and the WS ($p = .013$).

Overall, the results show that the variance of the three questionnaires increased when participants left the VR after an *abstract* scene and decreased when they left the VR after a *realistic* scene. This was not the case when participants filled in the questionnaires *in VR*. As there were no interaction effects of the other factors with QUESTIONNAIRES, we showed that the variance of the WS was generally lower than for the SUS or the IPQ questionnaire.

Subscales

The IPQ and WS questionnaires use subscales, which allow a more nuanced analysis and focusing on more components of presence. The means of those subscales are shown in Figure 5.

IPQ subscales. Considering general presence (GP) of the IPQ, there was a significant main effect of VIRTUAL REALISM, $F(1, 34) = 10.206, p = .003$; however, not of ENVIRONMENT, $F(1, 34) = .040, p = .842$, and no interaction effect of VIRTUAL REALISM × ENVIRONMENT, $F(1, 34) = .008, p = .927$. Spatial presence (SP) of the IPQ was neither affected by VIRTUAL REALISM, $F(1, 34) = 2.093, p = .157$ nor by ENVIRONMENT, $F(1, 34) = .062, p = .804$, and there was no interaction effect of VIRTUAL REALISM × ENVIRONMENT, $F(1, 34) = .087, p = .768$. Involvement (INV) was affected by VIRTUAL REALISM, $F(1, 34) = 6.851, p = .013$; however, not of ENVIRONMENT, $F(1, 34) = .279, p = .600$, and no interaction effect of VIRTUAL REALISM × ENVIRONMENT, $F(1, 34) = 3.030, p = .090$. And, for the realism (REAL) measure of the IPQ, there was a significant main effect of VIRTUAL REALISM, $F(1, 34) = 24.082, p < .001$, and on ENVIRONMENT, $F(1, 34) = 1.010, p < .001$, and a significant interaction effect of VIRTUAL REALISM × ENVIRONMENT, $F(1, 34) = 1.033, p < .001$.

WS subscales. Considering the subscales of the WS presence questionnaire, there was a significant main effect of VIRTUAL REALISM, $F(1, 34) = 6.963, p = .012$; however, not of ENVIRONMENT, $F(1, 34) = .055, p = .814$, and no interaction effect of VIRTUAL REALISM × ENVIRONMENT, $F(1, 34) = 1.051, p = .312$ on involvement (INV). On the natural (NAT) subscale of the WS questionnaire, there was also a main

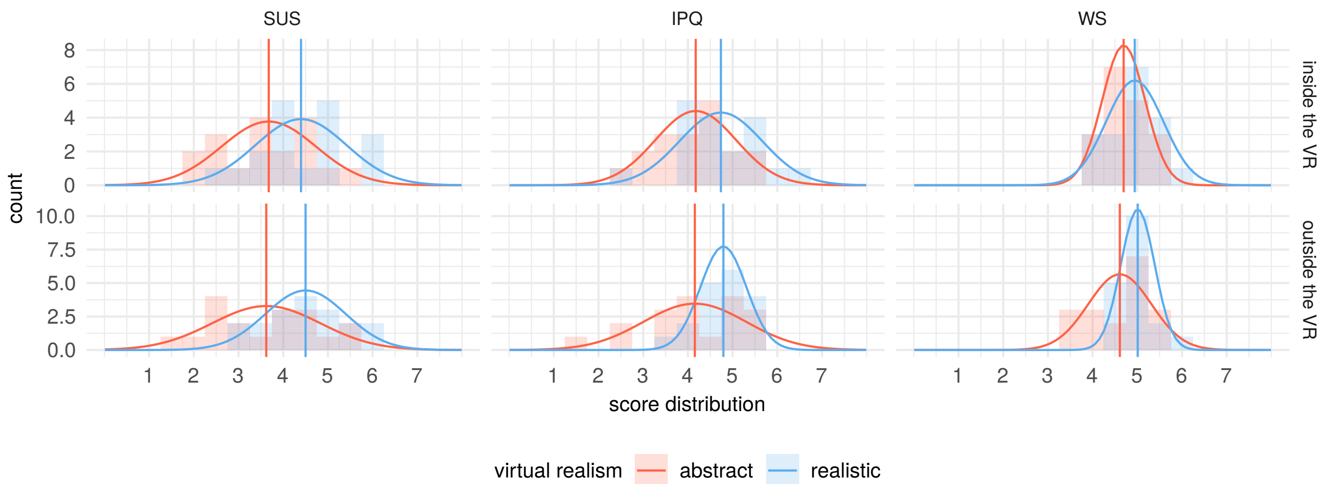

Figure 4: Histogram, means, and fitted Gaussian distribution using the average scores of all questionnaires. While the distributions remain nearly constant when the questionnaires are used in VR, there is a significant difference between the variances in the real-world responses after BIP using abstract and after BIP using a realistic virtual scene.

effect of Virtual Realism, $F(1,34) = 14.908, p < .001$; however, not of Environment, $F(1,34) = .209, p = .650$, and no interaction effect of Virtual Realism × Environment, $F(1,34) = .448, p = .507$. The auditory (AUD) scale of the WS was significantly affected by Virtual Realism, $F(1,34) = 20.104, p < .001$, and Environment, $F(1,34) = .637, p < .001$. We also found an interaction effect of Virtual Realism × Environment, $F(1,34) = .412, p < .001$. Haptics (HAPTC) was neither significantly affected by Virtual Realism, $F(1,34) = .307, p = .582$, nor Environment, $F(1,34) = .542, p = .466$. There was no interaction effect of Virtual Realism × Environment, $F(1,34) = 1.692, p = .202$. Resolution (RES) of the WS was neither affected by Virtual Realism, $F(1,34) = 0.255, p = .582$, nor Environment, $F(1,34) = .542, p = .466$, and no interaction effect of Virtual Realism × Environment, $F(1,34) = 1.692, p = .202$. Interface quality (IFQUAL) of the WS was neither affected by Virtual Realism, $F(1,34) = .671, p = .582$; however, not of Environment, $F(1,34) < .001, p = .993$, and no interaction effect of Virtual Realism × Environment, $F(1,34) = .012, p = .911$.

In sum, we observerd that most of the subscales within the IPQ and WS were not affected by the environment of the questionnaires. Scores of realism of the IPQ and auditory features of the WS showed main and interaction effects with virtual realism and environment, implying that those measures depend on the virtual scene presented beforehand and on the environment in which the sample was taken.

Reliability Analyses

Cronbach's alpha estimates of reliability were used to determine correlations between the questionnaires' items and to assess internal consistency of each questionnaire. Internal consistency was acceptable for SUS ($\alpha = .75$), good for IPQ ($\alpha = .87$), good for WS ($\alpha = .82$), and excellent when all items were used ($\alpha = .91$). The questions with the highest inter-correlations considering all questionnaire items were found in the IPQ questionnaire: IPQ6 with *"I feel present in the virtual space"* ($\rho = .786$), IPQ1 with *"In the computer generated world I had a sense of 'being there'"* ($\rho = .735$), and IPQ10 with *"I was completely captivated by the virtual world."* ($\rho = .719$).

Workload

To investigate if virtual questionnaires cause higher workload than questionnaires in the real-world, we performed a one-way ANOVA on the NASA-Task Load Index (TLX) scores, which were 33.16 ($SD = 20.96$) in VR and 37.77 ($SD = 19.26$) outside the VR. There were no significant effects, neither on the overall workload, $F(1,34) = .472, p = .497$, nor on its sub-scales (mental, physical, temporal, performance, effort, and frustration, all with $p > .111$). This indicates that completing the virtual questionnaire in VR had a negligible effect on the participants' perceived workload.

Questionnaire Completion Times

As the three questionnaires had different numbers of items (SUS=6, IPQ=14, WS=32), a three-way ANOVA showed significant main effects of Questionnaire on the completion

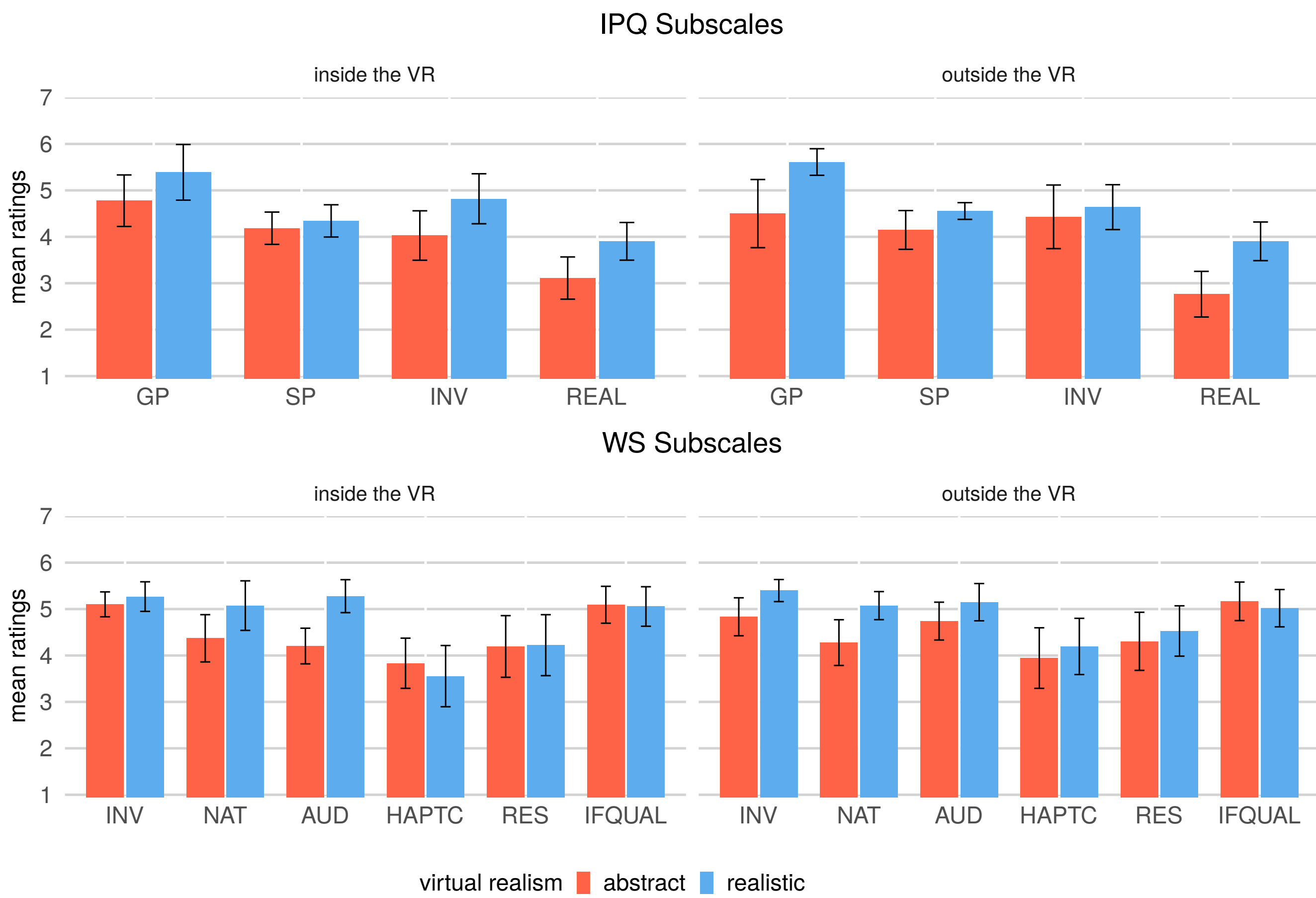

Figure 5: Subscales of the IPQ and WS presence questionnaires. IPQ: General Presence (GP), spatial presence (SP), involvement (INV), and realism (REAL); WS: Involvement (INV), natural (NAT), auditory (AUD), haptics (HAPTC), resolution (RES), and interface quality (IFQUAL).

times as shown in Figure 6, $F(2, 203) = 150.608$, $p < .001$. Bonferroni-corrected pairwise comparisons showed significant differences between the SUS ($M = 1.107mins$, $SD = .440$), the IPQ ($M = 2.449mins$, $SD = 1.054$), and the WS ($M = 5.982mins$, $SD = 2.733$), all with $p < .001$. However, there were no further main or interaction effects of Virtual Realism or Environment, all with $p > .275$. This means that the speed with which the participants completed the questionnaire did not significantly change through leaving the VR.

To determine potential effects of gender and previous VR experience, all analyses were repeated considering both factors as between-subject variables. None of the tests revealed statistically significant effects.

5 DISCUSSION

We found a systematic increase of the questionnaires' variance when the participants filled in the questionnaires in the real-world after experiencing an abstract scene and a decrease after they experienced a realistic scene. This variance remained constant when the participants filled in the questionnaires in VR. The mean variance of the questionnaires was higher after experiencing a realistic scene than after an abstract scene; however, the questionnaire variance in VR delivered similar results to the real-world and did not differ significantly. Thus, all presence questionnaires used (SUS, IPQ, WS) were still reliable tools to measure presence even when participants were surveyed in VR. However, as the interaction effects of the variance measures show, the increased variability after experiencing an abstract scene and after leaving the VR indicates that, potentially, more

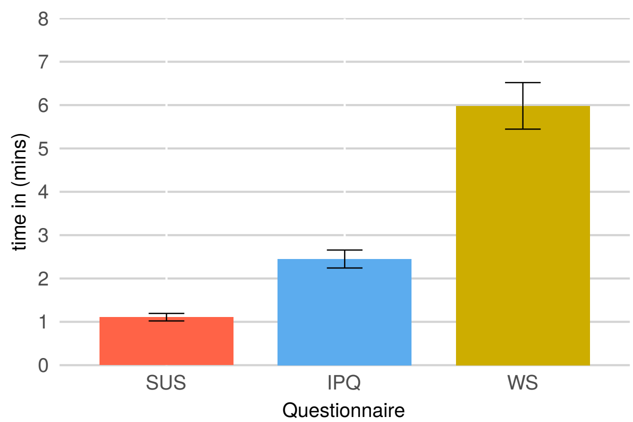

Figure 6: Completion times of the SUS, IPQ, and WS questionnaire. Error bars show CI95.

samples (and participants) are needed to reveal statistically significant results using questionnaires which are presented in the real-world. Hence, the three questionnaires can provide more reliable and consistent results when measuring presence in VR.

We assume that removing the HMD and/or reorienting in the real-world increases the probability of a BIP occurring and potentially causes a state of confusion or uncertainty. This increases the likelihood that ratings in the real-world are higher (or lower) than in VR. This finding was supported by the variances of presence scores in VR, which were all higher after experiencing an abstract scene and lower after experiencing a realistic scene. This was also partially evident by interaction effects of two subscale measures of the IPQ and WS questionnaire: differences of the visual realism subscale (IPQ) were lower in VR than outside the VR, while differences of the auditory subscale measure were higher in VR than outside the VR. Thus, the real-world responses increase the contrast of presence ratings. However, it should be noted that the virtual copy of our laboratory can be a confound in our presence measures. The environment was a real and neutral setup, however, the effect of the virtual questionnaire environment may have influenced those ratings.

There are previous discussions about which questionnaire is the most appropriate measurement tool to determine presence [33]. Objective signs of the reliability of a questionnaire are variance and item correlation. The lowest variance was found for the questionnaire by Witmer and Singer [43]. This is probably due to the large number of items. While the SUS and the IPQ questionnaire consist of 6 or 14 items, respectively, the WS questionnaire uses 32 items. A higher number of items potentially reduces the score variance, however, significantly increases the time a participant needs to answer the questions as shown in our results (Figure 6). It potentially causes fatigue effects using long questionnaires. The lower variance of the WS could also be caused by the conditions, which are not exercising some of the factors. Nevertheless, we found that the IPQ questionnaire includes three items with the highest inter-correlations among all items of the three questionnaires. These questions (1, 6, and 10) are directly related to the feeling of being present in the virtual world. Thus, we assume that the IPQ questionnaire best reflects the construct of presence.

Furthermore, we argue that our results justify a renewed discussion of the question if a single score is sufficient to describe presence. The increased variability of the responses and the contrast between a realistic and a less realistic scene potentially indicates that a BIP occurs (*cf.* [36]). The likelihood that a difference between different levels of presence being detected is higher after experiencing an abstract scene than after a realistic scene. Participants reflected that a BIP occurred using a higher contrast in their ratings leading to higher variance in their presence scores. Hence, it is conceivable that a BIP counter or a variance measure can be an additional measure while using questionnaires [36].

Some researchers argue that presence is a binary state – a moment of time when a participant is experiencing the sense of presence or not [36]. Consequently, the probability that a BIP is being perceived will be lower with scenarios similar to the real-world and can eventually no longer be determined using perfect technology. It is conceivable that the probability of a BIP occurring is potentially the more reliable measure of presence than a simple score. However, we explicitly highlight that a BIP may occur not only after leaving the VR and through removing the headset but also after leaving a scene within the VR. This means that the procedure of *removing the headset* is potentially not the only trigger for a BIP occurring. Nevertheless, we argue that the variance in responses from real-world questionnaires is potentially an additional measure for the sense of presence when using presence questionnaires.

Limitations and Future Work

Currently, post-test questionnaires are the most frequently used measures of presence in previous work. However, one disadvantage of such questionnaires is that they rely on the subjects' memories of the VR experience. Such memories can reflect an inconsistent and incomplete picture of the VR experience. Our questionnaires were presented in VR, however, the participants were asked for the VR experience after a scene change. Therefore, we recommend, for future work, to use a measure for immersion at the moment when it is perceived. As our results show, questionnaires in VR can reduce the score variance and avoid a state of confusion or uncertainty after leaving the VR and removing the HMD.

We used two different scenes at different levels of realism to manipulate the sense of presence. As presence can be affected by many factors such as engagement, involvement, and auditive cues, further work should investigate to which degree related factors contribute to the outcome. Furthermore, we found that the differences between the scores were relatively low despite the large audiovisual difference between the two virtual scenes. The least difference was shown in the WS questionnaire, followed by IPQ and SUS. Therefore, we suggest that future work should consider revised and optimized questionnaires for use in VR. We also note that the virtual copy of our laboratory could have been a confounding variable in our presence measure. Despite the fact that both environments were clean and simple settings, the effect of realism in the virtual questionnaire environment may have influenced the ratings. We highlight that measuring presence in VR potentially differs from measuring presence when experiencing the actual condition.

We assume that there is a relationship between the probability that a BIP occurs and the increased variance of presence scores. As a BIP may occur not only after leaving the VR but also after scene changes within the VR, more research is needed to determine the precise context of this relationship to develop a reliable measure of presence in VR. We found no effect on the perceived workload between questionnaires presented in the real-world and in VR, however, the workload between the questionnaires itself was not determined. Future studies could investigate if and how the number of items of a questionnaire affects the presence and how the interaction with questionnaires can be optimized to avoid a BIP through the questionnaire itself. Furthermore, to compare virtual and real questionnaires, we used the Oculus Touch controller for questionnaire input. However, in the real-world, presence questionnaires are usually presented on paper and answered using a pen. A comparison of input techniques was not performed in our study and could also be subject of future work.

6 CONCLUSION

Presence questionnaires in VR can help researchers in their experiments in multiple ways: Participants do not have to remove their HMD or put it on again to experience the next condition in VR. Presence questionnaires in VR can reduce the time for adjusting in the real and virtual world again. Distracting or biasing cues from the real-world, such as the experimenter or experimental setup, can be reduced when the participant remains in the VR during the experiment. The results of our study which used three standardized questionnaires indicate that answering questionnaires in VR does not affect the means of their measures. As questionnaires in VR decrease the chance of a BIP occurring, the variance of the measure from the VR questionnaire remains nearly constant for the conditions. This was not the case using questionnaires in the real-world. As the increased variance of real-world questionnaires potentially requires a larger sample to reveal statistically significant results between two or more conditions, we recommend using presence questionnaires in VR.

Finally, we recommend the IPQ questionnaire as the measure of presence as it provides the highest reliability within a reasonable timeframe. Furthermore, we conclude that the variance of the scores can potentially be utilized as an additional measure of presence as it can help to assess if a BIP occurred. More research is needed to adapt presence questionnaires for VR applications.

ACKNOWLEDGMENTS

This work was supported by the German Federal Ministry of Education and Research as part of the project Be-greifen (Grant No. 16SV7527) and the German Research Foundation (DFG) within project C04 of SFB/Transregio 161.

REFERENCES

[1] Rosa María Baños, Cristina Botella, Mariano Alcañiz, Víctor Liaño, Belén Guerrero, and Beatriz Rey. 2004. Immersion and Emotion: Their Impact on the Sense of Presence. *CyberPsychology & Behavior* 7, 6 (2004), 734–741. https://doi.org/10.1089/cpb.2004.7.734

[2] Rosa María Baños, Cristina Botella, Azucena Garcia-Palacios, Helena Villa, Concepción Perpiñá, and Mariano Alcaniz. 2000. Presence and Reality Judgment in Virtual Environments: A Unitary Construct? *CyberPsychology & Behavior* 3, 3 (2000), 327–335. https://doi.org/10.1089/10949310050078760

[3] Woodrow Barfield and Suzanne Weghorst. 1993. The Sense of Presence within Virtual Environments: A Conceptual Framework.. In *HCI (2)*, Gavriel Salvendy and Michael J. Smith (Eds.). Elsevier, 699–704. 0-444-89540-X.

[4] Dongsik Cho, Jihye Park, G. J. Kim, Sangwoo Hong, Sungho Han, and Seungyong Lee. 2003. The dichotomy of presence elements: the where and what. In *IEEE Virtual Reality, 2003. Proceedings.* 273–274. https://doi.org/10.1109/VR.2003.1191155

[5] Elizabeth Dean, Joe Murphy, and Sarah Cook. 2009. Social presence in virtual world surveys. In *Proceedings of The 12th Annual International Workshop on Presence.*

[6] Huong Q. Dinh, Neff Walker, Chang Song, Akira Kobayashi, and Larry F. Hodges. 1999. Evaluating the Importance of Multi-sensory Input on Memory and the Sense of Presence in Virtual Environments. In *Proceedings of the IEEE Virtual Reality (VR '99).* IEEE Computer Society, Washington, DC, USA.

[7] Jonathan Freeman, S. E. Avons, Don E. Pearson, and Wijnand A. IJsselsteijn. 1999. Effects of Sensory Information and Prior Experience on Direct Subjective Ratings of Presence. *Presence: Teleoperators and Virtual Environments* 8, 1 (1999), 1–13. https://doi.org/10.1162/105474699566017

[8] Julian Frommel, Katja Rogers, Julia Brich, Daniel Besserer, Leonard Bradatsch, Isabel Ortinau, Ramona Schabenberger, Valentin Riemer, Claudia Schrader, and Michael Weber. 2015. Integrated Questionnaires: Maintaining Presence in Game Environments for Self-Reported Data Acquisition. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '15).* ACM, New York, NY, USA, 359–368.

[9] Michael Gerhard, David J. Moore, and Dave J. Hobbs. 2001. Continuous Presence in Collaborative Virtual Environments: Towards a Hybrid Avatar-Agent Model for User Representation. In *Proceedings of the Third International Workshop on Intelligent Virtual Agents (IVA '01)*. Springer-Verlag, London, UK, UK, 137–155. http://dl.acm.org/citation.cfm?id=648034.744242

[10] Brent Insko. 2003. Measuring presence: Subjective, behavioral and physiological methods. (2003).

[11] Jason Jerald. 2016. *The VR Book: Human-Centered Design for Virtual Reality*. Association for Computing Machinery and Morgan & Claypool, New York, NY, USA.

[12] Taeyong Kim and Frank Biocca. 1997. Telepresence via television: Two dimensions of telepresence may have different connections to memory and persuasion. *Journal of computer-mediated communication* 3, 2 (1997). https://doi.org/10.1111/j.1083-6101.1997.tb00073.x

[13] Michael Krauss, Rainer Scheuchenpflug, Walter Piechulla, and Alf Zimmer. 2001. Measurement of presence in virtual environments. *Experimentelle Psychologie im Spannungsfeld von Grundlagenforschung und Anwendung Proceedings* (2001), 358–362.

[14] Jane Lessiter, Jonathan Freeman, Edmund Keogh, and Jules Davidoff. 2001. A Cross-Media Presence Questionnaire: The ITC-Sense of Presence Inventory. *Presence: Teleoperators and Virtual Environments* 10, 3 (2001), 282–297. https://doi.org/10.1162/105474601300343612

[15] Matthew Lombard and Theresa Ditton. 1997. At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication* 3, 2 (1997). https://doi.org/10.1111/j.1083-6101.1997.tb00072.x

[16] Matthew Lombard, Theresa B Ditton, Daliza Crane, Bill Davis, Gisela Gil-Egui, Karl Horvath, Jessica Rossman, and S Park. 2000. Measuring presence: A literature-based approach to the development of a standardized paper-and-pencil instrument. In *Third international workshop on presence, delft, the netherlands*, Vol. 240. 2–4.

[17] Matthew Lombard, Theresa B. Ditton, and Lisa Weinstein. 2009. Measuring presence: the temple presence inventory. In *Proceedings of the 12th Annual International Workshop on Presence*. 1–15.

[18] Michael Meehan, Brent Insko, Mary Whitton, and Frederick P. Brooks, Jr. 2002. Physiological Measures of Presence in Stressful Virtual Environments. In *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '02)*. ACM, New York, NY, USA, 645–652. https://doi.org/10.1145/566570.566630

[19] Marvin Minsky. 1980. Telepresence. *Omni* (1980), 45–51.

[20] Sarah Nichols, Clovissa Haldane, and John R. Wilson. 2000. Measurement of presence and its consequences in virtual environments. *International Journal of Human-Computer Studies* 52, 3 (2000), 471 – 491. https://doi.org/10.1006/ijhc.1999.0343

[21] Kristine L. Nowak and Frank Biocca. 2003. The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators & Virtual Environments* 12, 5 (2003), 481–494.

[22] Sandra Poeschl and Nicola Doering. 2015. Measuring Co-Presence and Social Presence in Virtual Environments - Psychometric Construction of a German Scale for a Fear of Public Speaking Scenario. *Studies in health technology and informatics* 219 (2015), 58–63. https://doi.org/10.3233/978-1-61499-595-1-58

[23] Holger Regenbrecht and Thomas Schubert. 2002. Real and Illusory Interactions Enhance Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments* 11, 4 (2002), 425–434. https://doi.org/10.1162/105474602760204318

[24] Holger Regenbrecht, Thomas Schubert, and Frank Friedmann. 1998. Measuring the Sense of Presence and its Relations to Fear of Heights in Virtual Environments. *International Journal of HumanâĂŞComputer Interaction* 10, 3 (1998), 233–249. https://doi.org/10.1207/s15327590ijhc1003_2

[25] Louis Barry Rosenberg. 1994. *&Ldquo;Virtual Fixtures&Rdquo;: Perceptual Overlays Enhance Operator Performance in Telepresence Tasks*. Ph.D. Dissertation. Stanford, CA, USA. UMI Order No. GAX95-08441.

[26] Maria V. Sanchez-Vives and Mel Slater. 2005. From presence to consciousness through virtual reality. *Nature reviews. Neuroscience* 6, 4 (2005), 332–339. https://doi.org/10.1038/nrn1651

[27] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. 2001. The Experience of Presence: Factor Analytic Insights. *Presence: Teleoperators and Virtual Environments* 10, 3 (2001), 266–281. https://doi.org/10.1162/105474601300343603

[28] Thomas W Schubert. 2003. The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realness. *Zeitschrift für Medienpsychologie* 15, 2 (2003), 69–71.

[29] Valentin Schwind, Pascal Knierim, Lewis Chuang, and Niels Henze. 2017. "Where's Pinky?": The Effects of a Reduced Number of Fingers in Virtual Reality. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '17)*. ACM, New York, NY, USA, 507–515. https://doi.org/10.1145/3116595.3116596

[30] Valentin Schwind, Pascal Knierim, Cagri Tasci, Patrick Franczak, Nico Haas, and Niels Henze. 2017. "These Are Not My Hands!": Effect of Gender on the Perception of Avatar Hands in Virtual Reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 1577–1582. https://doi.org/10.1145/3025453.3025602

[31] Thomas B. Sheridan. 1992. Musings on Telepresence and Virtual Presence. *Presence: Teleoper. Virtual Environ.* 1, 1 (1992), 120–126.

[32] Valerie J Shute. 2011. Stealth assessment in computer-based games to support learning. *Computer games and instruction* 55, 2 (2011), 503–524.

[33] Mel Slater. 1999. Measuring Presence: A Response to the Witmer and Singer Presence Questionnaire. *Presence: Teleoper. Virtual Environ.* 8, 5 (1999), 560–565. https://doi.org/10.1162/105474699566477

[34] Mel Slater. 2004. How Colorful Was Your Day? Why Questionnaires Cannot Assess Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments* 13, 4 (2004), 484–493. https://doi.org/10.1162/1054746041944849

[35] Mel Slater, John McCarthy, and Francesco Maringelli. 1998. The Influence of Body Movement on Subjective Presence in Virtual Environments. *Human Factors* 40, 3 (1998), 469–477. https://doi.org/10.1518/001872098779591368

[36] Mel Slater and Anthony Steed. 2000. A Virtual Presence Counter. *Presence: Teleoperators and Virtual Environments* 9, 5 (2000), 413–434. https://doi.org/10.1162/105474600566925

[37] Mel Slater and Martin Usoh. 1993. Presence in immersive virtual environments. In *Proceedings of IEEE Virtual Reality Annual International Symposium*. 90–96. https://doi.org/10.1109/VRAIS.1993.380793

[38] Martin Usoh, Kevin Arthur, Mary C. Whitton, Rui Bastos, Anthony Steed, Mel Slater, and Frederick P. Brooks, Jr. 1999. Walking > Walking-in-place > Flying, in Virtual Environments. In *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '99)*. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 359–364. https://doi.org/10.1145/311535.311589

[39] Martin Usoh, Ernest Catena, Sima Arman, and Mel Slater. 2000. Using Presence Questionnaires in Reality. *Presence: Teleoperators and Virtual Environments* 9, 5 (2000), 497–503. https://doi.org/10.1162/105474600566989

[40] Dimitar Valkov and Steffen Flagge. 2017. Smooth Immersion: The Benefits of Making the Transition to Virtual Environments a Continuous Process. In *Proceedings of the 5th Symposium on Spatial User Interaction (SUI '17)*. ACM, New York, NY, USA, 12–19. https://doi.org/10.1145/3131277.3132183

[41] Minjuan Wang, Sabine Lawless-Reljic, Marc Davies, and Victor Callaghan. 2011. Social Presence in Immersive 3D Virtual Learning Environments. In *Ambient Intelligence - Software and Applications*, Paulo Novais, Davy Preuveneers, and Juan M. Corchado (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 59–67.

[42] Bob G. Witmer, Christian J. Jerome, and Michael J. Singer. 2005. The Factor Structure of the Presence Questionnaire. *Presence: Teleoper. Virtual Environ.* 14, 3 (2005), 298–312. https://doi.org/10.1162/105474605323384654

[43] Bob G. Witmer and Michael J. Singer. 1998. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments* 7, 3 (1998), 225–240. https://doi.org/10.1162/105474698565686

[44] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only Anova Procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 143–146. https://doi.org/10.1145/1978942.1978963