Inconsistencies of Presence Questionnaires in Virtual Reality

Sarah Graf
Media Informatics Group
University of Regensburg, Germany
sarah.graf@stud.uni-regensburg.de

Valentin Schwind
Frankfurt University of Applied Sciences
Frankfurt, Germany
valentin.schwind@acm.org

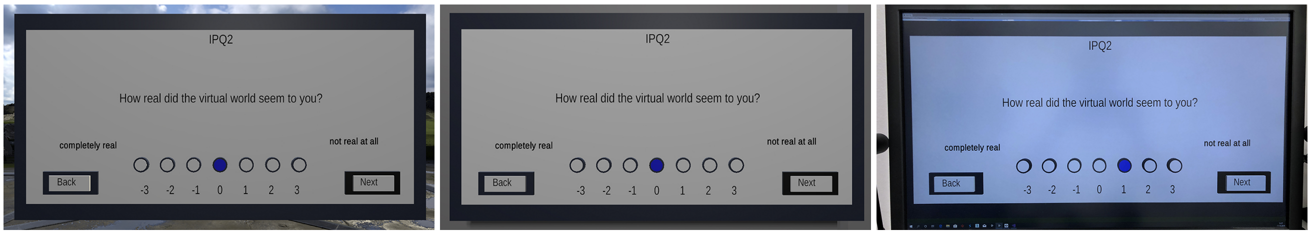

Figure 1: Questionnaire environments in the virtual scene, in the virtual lab, and in the real lab.

ABSTRACT

Presence in virtual reality (VR) is typically assessed through questionnaires in the real world and after leaving an immersive experience. Previous research suggests that questionnaires in VR reduce biases caused by the real-world setup. However, it remains unclear whether presence questionnaires still provide valid results when subjects are being surveyed while the construct is perceived. In a user study with 36 participants, two standardized presence questionnaires (IPQ, SUSa) were either completed in the real lab, in a virtual lab scene, or in the actual scene after a virtual gaming experience. Our results show inconsistencies between the measurements and that main scores, as well as subscales of the presence measures are significantly affected by the subjects' environment. As presence questionnaires have been designed to be answered after an immersive experience, we recommend revising those tools for measuring presence in VR.

CCS CONCEPTS

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

KEYWORDS

Virtual reality; presence; questionnaire; break in presence.

ACM Reference Format:
Sarah Graf and Valentin Schwind. 2020. Inconsistencies of Presence Questionnaires in Virtual Reality. In *26th ACM Symposium on Virtual Reality Software and Technology (VRST '20), November 1–4, 2020, Virtual Event, Canada.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3385956.3422105

VRST '20, November 1–4, 2020, Virtual Event, Canada

ACM ISBN 978-1-4503-7619-8/20/11.
https://doi.org/10.1145/3385956.3422105

1 INTRODUCTION AND BACKGROUND

Designers of virtual reality (VR) experiences seek for creating a consisting sense of presence – the feeling of being or acting within an environment even when one is physically situated in another place [6]. Thus, to create immersive experiences and to assess the quality of any VR setup, it is important to measure presence [10]. Researchers primarily use standardized and validated (pen-and-paper) post experiment questionnaires to measure presence [6, 12, 14, 15]. Those presence questionnaires are typically presented *after* the VR experience and in the real world. As leaving the VR after using an head-mounted display (HMD) requires removing the headset and causes a state of disorientation and confusion while "returning" to the real-world [1, 3, 8], the procedure increases the probability that confounds interfere with the subjective ratings on items of presence questionnaires [4]. Participants are not able to ignore the mediating technology, novelty effects, and excitement can potentially compromise the results in human subject studies [17]. Thus, it is conceivable that the typical procedure of measuring presence compromises the measure when subjects leave the VR in order to fill in those questionnaires and inevitably go through a so-called "break-in-presence (BIP)" [13].

In a between-group user study, two standardized questionnaires were either completed in the real lab, in a virtual lab scene, or in the scene, in which the experience – a virtual archery game – was being perceived. In addition to the conditions used by Schwind et al. [8], participants were surveyed *during* the immersion (without scene change) and BIPs were triggered to modulate presence using external cues. Our results show that the main scores, as well as subscales of two standardized questionnaires (IPQ, SUSa) are affected by the environment, in which the questions have been answered. The ratings significantly differ and even reveal contradictory scores among the conditions. We conclude that presence questionnaires show inconsistent results when being answered during the actual experience compared to results measured after a scene change or in the real world. Possible reasons for inconsistencies of integrated questionnaires can be found in the formulation of the items and that the sensation of presence should only be assessed after the experience has happened.

2 METHOD

We conducted a VR user study, to investigate the validity of current standardized presence measures. We used the three independent variables (IVs) Scene Realism (SR), BIP, and Questionnaire Environment (QE). The within-subject variables SR and induced BIPs were used to modulate presence scores of standardized questionnaires, while QE (Figure 1) has been operationalized to determine if the surrounding environment affects the presence scores. SR consisted of the two levels *realistic* and *abstract*, while BIPs using external auditory cues from the real world were either presented or not. In line with previous work [8], QE was used as between-subject variable that served to counteract potential order effects, so that all participants completed their questionnaires either outside the VR (*real lab*), inside the VR (*virtual lab*), or inside the *virtual scene*. SR and BIPs were counterbalanced and used as within-subject variables. We recorded if participants recognized a transition from the virtual to the real world using the verbal Transition-to-Reality Instructions [2, 11, 16] by Slater and Steed [13]. Questionnaire scores were obtained using the igroup presence questionnaire (IPQ) as well as the Slater-Usoh-Steed questionnaire (SUSa) and were completed in the corresponding environment. NASA-Task Load Index (TLX) and the System Usability Scale (SUSb) after each experiment were used to assess if the workload and the perceived usability score differ between the environments.

We developed an archery game to facilitate an immersive VR experience with Unity3D. The VR setup consisted of the Oculus Rift CV1 components with two position tracking sensors, two Oculus Touch controllers, and the HMD. Each participant held the left Oculus Touch controller to adjust the bow and the right one to spawn and shoot an arrow by pressing a button. After seven minutes of performing the archery game, the participants had either to take off their HMD or to stay in the VR to complete the questionnaires in the virtual environment. In all cases, they used the Oculus touch controller to navigate through the questionnaires and to select their answers. We recruited 36 participants (18 f, 18 m), from 20 to 54 ($M = 27.9$, $SD = 9.9$) via mailing lists of our institution as well as social networks.

3 RESULTS

A mixed-design three-way ANOVA showed significant main effects of SR ($p = .003$), BIP ($p < .001$), and a significant interaction effect of QE × BIPs ($p = .017$) on the IPQ scale. Further effects were not found (all with $p > .062$). Bonferroni-corrected pairwise t-test comparisons revealed that the IPQ scores obtained in the *virtual lab* ($p = .001$, $d = 1.104$) as well as *real lab* ($p = .019$,$d = .544$) were able to determine significant differences between the two BIP levels, however, not when participants were tested in the *virtual scene* ($p = 1$, $d = .240$). The tests only revealed significant differences between the two levels of SR in the *virtual scene* ($p = .044$, $d = −.729$).

IPQ subscales were also affected. An ART ANOVA [18] revealed a significant main effect of SR ($p = .01$) and BIP ($p = .031$) on the general presence (GP) subscale of the IPQ questionnaire. There was also a significant interaction effect of QE × BIP ($p = .002$). Further main or interaction effects were not found (all with $p > .134$). Bonferroni corrected pairwise Wilcoxon rank sum comparisons between both BIP levels were significant within the *virtual lab* ($p = .006$, $d = −.991$). Furthermore, main effects were found for SR ($p = .034$), QE ($p = .032$), and BIP ($p = .016$) and SR × QE ($p = .047$) on spatial presence (SP). The involvement (INV) subscale of the IPQ was significantly affected by BIP ($p < .001$). Further main or interaction effects on the INV subscale were not found (all with $p > .116$). As excepted SR, ($p < .001$), and BIP, ($p = .003$), significantly affected the realism (REAL) subscale of the IPQ. We also found an interaction effect of SR × BIP ($p = .038$). Further main or interaction effects were not found (all with $p > .083$).

There was a significant three-way interaction effect ($p = .032$) on the SUSa scale. Bonferroni corrected pairwise comparisons on factor interactions (using the phia package in R) were not able to determine between which combinations of conditions the three-way interaction effect occurred. Further, participants did not report on any BIP when no BIPs occurred. Thus, the statistical analysis were performed on conditions, in which participants experienced any BIPs. A two-way ANOVA revealed a significant main effect of SR ($p = .034$) on the number of BIPs. Further main or interaction effects were not found (all with $p > .555$). A one-way ANOVA revealed a significant effect of the TLX score ($p = .039$), however, Tukey's test was not able to reveal any significant differences between the three levels of QEs (all with $p > .062$). A one-way ANOVA revealed no significant effect of QE ($p = .086$) on SUSb.

4 DISCUSSION

The findings in our study revealed a "methodological circularity" [5]. Scales and subscales of two standardized questionnaires were significantly affected by the surrounding VR environment, in which the questionnaires have been completed. Thus, a virtual environment compromises the subjective ratings and leads to contradictory results as participants are not able to ignore it. The inconsistencies become particularly evident through the interaction effects with the environment. Significant differences were though found in the virtual and real lab environments, however, with higher effect sizes in the virtual lab conditions. Interestingly, the IPQ main scores obtained in the virtual scene were not able to differentiate between the two BIP levels. In contrast, significant changes in scene realism could only be revealed in the virtual scene but not after switching into a virtual lab or through leaving the VR. Considering the results within each environment they are plausible for its own, however, taking into account *where* they have been obtained indicates that surveying participants in the VR does not lead to the same results as in the real world.

Standardized questionnaires have been designed and employed before VR user interfaces allowed the integration of surveys during the experience. The results of our study support the finding that they do not provide valid results when they are being integrated into the VR. We conclude that they must be revised and optimized for surveying subjects during their perception of presence. Inconsistencies of presence questionnaires integrated into the VR raise fundamental questions about the actual construct. We assume that questionnaires do not measure the construct but the *contrast* of a subjectively perceived quality of an experience of the virtual world compared to previous experiences from the real one. For this reason, researchers measure, for example, reduced presence while experiencing less realistic avatars [7, 9], even when participants accept the illusion and should, therefore, feel present.

REFERENCES

[1] Dmitry Alexandrovsky, Susanne Putze, Michael Bonfert, Sebastian Höffner, Pitt Michelmann, Dirk Wenig, Rainer Malaka, and Jan David Smeddinck. 2020. Examining Design Choices of Questionnaires in VR User Studies. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–21. https://doi.org/10.1145/3313831.3376260

[2] Andrea Brogni, Mel Slater, and Anthony Steed. 2003. More breaks less presence. In *Presence 2003: The 6th Annual International Workshop on Presence*. 1–4.

[3] Susanne Putze, Dmitry Alexandrovsky, Felix Putze, Sebastian Höffner, Jan David Smeddinck, and Rainer Malaka. 2020. Breaking The Experience: Effects of Questionnaires in VR User Studies. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3313831.3376144

[4] Georg Regal, Jan-Niklas Voigt-Antons, Steven Schmidt, Johann Schrammel, Tanja Kojić, Manfred Tscheligi, and Sebastian Möller. 2019. Questionnaires embedded in virtual environments: reliability and positioning of rating scales in virtual environments. *Quality and User Experience* 4 (12 2019). https://doi.org/10.1007/s41233-019-0029-1

[5] 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

[6] 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 arXiv:https://doi.org/10.1162/105474601300343603

[7] 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

[8] 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)* (2019-05-04). ACM, Glasgow, Scotland Uk, 12. https://doi.org/10.1145/3290605.3300590

[9] 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

[10] Richard Skarbez, Frederick P. Brooks, Jr., and Mary C. Whitton. 2017. A Survey of Presence and Related Concepts. *ACM Comput. Surv.* 50, 6, Article 96 (Nov. 2017), 39 pages. https://doi.org/10.1145/3134301

[11] Mel Slater, Andrea Brogni, and Anthony Steed. 2003. Physiological responses to breaks in presence: A pilot study. 157 (10 2003).

[12] 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

[13] 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 arXiv:https://doi.org/10.1162/105474600566925

[14] 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

[15] 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

[16] Vinoba Vinayagamoorthy and Andrea Brogni. 2004. An Investigation of Presence Response across Variations in Visual Realism. *Proceedings of Presence 2004: The 7th Annual International Workshop on Presence*, 119–126.

[17] Alexandra Voit, Sven Mayer, Valentin Schwind, and Niels Henze. 2019. Online, VR, AR, Lab, and In-Situ: Comparison of Research Methods to Evaluate Smart Artifacts. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 12. https://doi.org/10.1145/3290605.3300737

[18] 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