# "Never The Same": Systematic Analysis of the Methodological Issues in the Presence Studies That Employ Questionnaires

Eugene Kukshinov
eugene.kukshinov@uwaterloo.ca
Stratford School of Interaction Design and Business, University of Waterloo
Canada

Joseph Tu*
joseph.tu@uwaterloo.ca
Stratford School of Interaction Design and Business, University of Waterloo
Canada

Kata Szita
kata.szita1@ucd.ie
Insight Centre, School of Computer Science, University College Dublin
Dublin, Ireland

Kaushall Senthil Nathan*
k3senthilnathan@uwaterloo.ca
Stratford School of Interaction Design and Business, University of Waterloo
Canada

Lennart E. Nacke
lennart.nacke@acm.org
Stratford School of Interaction Design and Business, University of Waterloo
Canada

## ABSTRACT

Presence is a psychological state that is usually measured via questionnaires. Many presence researchers assume self-report questionnaires are standardized. However, we do not know how reliable they are. This knowledge gap impacts the accuracy and validity of data collected through these questionnaires. Reliable and accurate data collection is crucial to trust findings in presence research. Inaccurate or unreliable data could lead to incorrect conclusions. This impacts theoretical understanding and practical applications. To address this, we conducted a comprehensive systematic review of 100 empirical quantitative presence studies. Our goal was to uncover the underlying issues with these self-report questionnaires. So, we explored the employment of these questionnaires and analyzed the specific reasons for measuring presence. We show patterns and inconsistencies in the current methodologies, as presence questionnaires are frequently utilized in a non-standardised manner. We will propose well-grounded improvements and constructive approaches based on our findings. These will improve the validity and effectiveness of these questionnaires and lead to more consistent and replicable results in presence research.

## CCS CONCEPTS

• **Human-centered computing → HCI theory, concepts and models**.

## KEYWORDS

Presence, Questionnaires, Methods, Reliability, Systematic Analysis

*Authors are also affiliated with the Department of Systems Design Engineering, University of Waterloo


*CHI EA '24, May 11–16, 2024, Honolulu, HI, USA*

ACM ISBN 979-8-4007-0331-7/24/05
https://doi.org/10.1145/3613905.3651074

**ACM Reference Format:**
Eugene Kukshinov, Joseph Tu, Kata Szita, Kaushall Senthil Nathan, and Lennart E. Nacke. 2024. "Never The Same": Systematic Analysis of the Methodological Issues in the Presence Studies That Employ Questionnaires. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3613905.3651074

## 1 INTRODUCTION

*Presence* is a psychological state that results from interacting with technology. It may represent an extent of not recognizing the technology's role in an individual's experience. As a psychological construct, it is an important part of the HCI field [28]. However, being a complicated phenomenon, it has many issues in research [8, 13, 15, 24]. One key issue is the lack of clear conceptualization; this affects the tools used to measure presence. Similar measures are often grounded in different conceptual frameworks; this casts doubt on fundamental presence research and its findings [8, 15].

Using presence questionnaires is the most dominant way to measure presence [13]. There are many, often incomparable, measurements of presence [13, 15]: they vary in the number of scale items from a single question to dozens to measure this complex psychological state [15, 21]. Also, there are concerns over the validity of presence questionnaires because they may not measure presence itself [15]. A study by Nannipieri [21] showed that among 38 analyzed questionnaires, 21 measured a single dimension and 17 used multidimensional scales. The questionnaire items are dominantly related to user characteristics, devices, and virtual environment characteristics; not to presence itself. Researchers acknowledge the issues of presence questionnaires [17]. However, they may not see an alternative to them and continue to use them despite these challenges.

Validity is not the only problem with presence questionnaires. Presence questionnaires are considered standardized measures [15]. However, we question this assumption. In our analysis, we consider some of the most used and impactful questionnaires. These include the Presence Questionnaire (PQ) [35], the ITC-Sense of Presence Inventory (ITC) [18], and the presence scale by Nowak and Biocca (or NB) [23].

In this study, we offer a critical reflection on established methods of presence studies. We focus on their limitations. Presence scholarship relies on self-report measures and questionnaires. The main advantage of these is standardization. Yet, they are not used in a standardized manner. We discuss major methodological issues in presence studies. We also describe future directions for advancing presence scholarship. Our identified issues include noticeable problems in the HCI field, such as questionnaire modifications and methodological transparency in research papers [1, 26]. We analyzed potential modifications in these scales and reviewed corresponding methods. This led us to propose areas for improvement that future studies can use to improve presence measures.

## 2 RELATED WORK

The concept of presence, initially termed "*telepresence*," has gained increased attention in academic discourse with the advancement of VR technologies [19, 28]. Minsky [20] initially introduced "*telepresence*" to define systems enabling remote control or tele-operation with a sense of being physically present. The scholarly community eventually shortened the term to "presence" and applied it to other (media) technologies, such as XR/VR, video games, computer-mediated communication (CMC), or robots [19]. Presence is often defined as a "sense of being there" or a "perceptual illusion of non-mediation"[27].

### 2.1 Presence Measurements

There are many ways to measure presence. In presence research, there are two broadly established types of measures: subjective and objective [16]. Objective measures usually include behavioural and physiological measures and task performance [4, 16]. As Freeman et al. [9] concluded, measures based on task performance may be useful to assess training outcomes, but they do not necessarily measure presence. Physiological measures are usually either costly or its data are hard to interpret for presence [13]. Physiological responses (at least by themselves) might not explain the reason for emotions or feelings that constitute presence [31]. Behavioral measures might be the most accurate tools, but they are hard to develop.

Subjective measures include, for example, questionnaires, continuous assessment techniques, psychophysical measures, and qualitative measures including interviews, and ethnographic methods. While researchers recommend using multiple measures to increase the validity of the findings [8], post-test self-report questionnaires are the dominant way to measure presence [13]. They are easy to design and administer [8], and they are intended to be standardized.

### 2.2 Presence Questionnaires

There are many presence questionnaires. For instance, Nannipieri [21] identified 38 major presence questionnaires. Among them, the Presence Questionnaire by Witmer and Singer [35] is the most popular (VR) presence questionnaire up to this day [13]. However, there are multiple issues with the validity of presence questionnaires. Their limited scope or inability to capture continuous presence is the gravest one [8]. This is why some authors question the fundamental assumption of whether they measure presence at all [21].

*2.2.1 Conceptualization Issues.* Any measures should rely on the theory-driven intrinsic value of the variable, but it is one of the biggest issues with presence questionnaires [16]. Conceptualization of presence is often too vague [8, 15], along with the conceptualization of the scale items [13]. The construct's dimensions lack explication. This does not provide enough guidance for successful operationalization and measurement [2]. As a result, self-report measurements often do not indicate presence itself. For instance, "communication qualities, such as composure, spontaneity, positivity, richness, and evaluation may represent not social presence itself, but rather markers of the quality of communication that transpires when social presence is realized or not" [2, p. 467]. Alternatively, presence questionnaires can concentrate on previous behaviour or factors of presence representing variables that correlate to social presence rather than social presence per se [2]. Nannipieri [21] found that among the 38 questionnaires, most of the scale items were not related to presence itself. This is true as well for the PQ [35] because some of its dimensions directly state that they measure factors, such as the perceived quality of a device's interface.

*2.2.2 Continuity Issues.* In contrast to the post-hoc questionnaires, which do not work during VR (or other media) experiences [12], continuous measures of presence are more valid tools. They consider the dynamic nature of presence and breaks that can appear during interaction with technology [11]. In other words, continuous tools allow assessment of "variations in the subjective experience of presence, which are likely to occur not only through changes in the stimulus but also through the participant (e.g., increasing fatigue during exposure) and to overcome limitations associated with post-rating measures" [34, p. 239].

*2.2.3 Issues with Sensation vs. Knowledge.* Presence is a complicated psychological state. It is not a self-evident sensorial experience, such as perceived temperature. Thus, it cannot easily be described [9]. People generally have a limited ability to observe their cognitive processing. They are also not aware of perceptual and memory processes [22], So, they cannot deliver higher-order thoughts (for instance, about perceptual illusions) verbally [7]. While people may be only partially conscious of their state of presence, self-reports cannot give a full picture without being complemented by objective measurements [16]. It is also true that subjective measures may be unstable and biased by previous experience and interest [9], and also post-test evaluations are generally confounded by the inaccurate recall and memory effect [10].

Another significant issue concerning the internal validity of self-reports relates to the conflict between sensation and knowledge. As highlighted in a study by Freeman et al. [9], observers themselves can inadvertently shape participants' feelings. In essence, "the problem is that we cannot rule out the possibility that presence in a VE [virtual environment] may seem to exist in our experimental subjects simply because questions are asked about it" [30, p. 486]. The phrasing of items in a questionnaire might suggest the expected effect. People who are not familiar with the concept could be more sensitive to demand characteristics in the experiment [14].

*2.2.4 Standardization Issues.* There is an overall assumption that questionnaires used for presence are standardized [32]. However, presence cannot be measured fully in a standardized manner. These

measures should be media-specific. Presence is transient and varies with medium and other factors [2]. Depending on the design of the medium/simulation (or at least the type of media), it can produce different experiences. Thus, the measurement should reflect on them [16]. This is true for any mediated experiences that derive from diverse designs and result in diverse social consequences [19]. The circumstances can be so different that it may not be a question of adaptation but rather of creating a different measure.

In light of the issues with presence questionnaires explained above, we argue for the need to analyze how various presence studies use questionnaires. We aim to provide an understanding of the consistency of self-report measures in presence studies through this. Thus, this paper answers the following research questions:

- **RQ1:** How are the most common presence questionnaires used in studies to measure presence?
- **RQ2:** How consistently are they used?

## 3 METHODS

In our study, we identified and analyzed empirical (quantitative) presence studies that used three specific presence scales [18, 23, 35]. We selected these particular scales because of their established relevance and effectiveness in measuring physical and social presence [6, 8, 13, 15, 29], particularly in the domains of VR/XR, video games, and mediated communication. Their widespread adoption within these specific contexts made them particularly suitable for our study (see their citation count in Table 1).

| Database | PQ [35] | ITC [18] | Nowak and Biocca [23] |
|---|---|---|---|
| Scopus | 3510 | 709 | 644 |
| Web of Science | 2663 | 554 | 470 |
| APA | 869 | 225 | 185 |
| ACM | 829 | 155 | 166 |
| IEEE | 438 | 54 | 52 |

**Table 1: Number of citations per scale in the key databases as of April 19, 2023.**

We started with 8253 records, and 580 screened papers are currently ready for analysis. For this late-breaking work paper, we randomly selected (to avoid bias) a subset of the screened sample (n=100) for a preliminary analysis. Our methodology for the review's search, screening, and data extraction stages was based on the PRISMA [25] approach in the Covidence software[1] (mainly focused on clear reporting of the search process and the flow diagram; see Figure 1).

### 3.1 Protocol and Search

We systematically collected an extensive compilation of academic papers referencing the specified scales. This formed the foundation of our paper corpus. We also identified the main databases where these scales are cited [13]. The major ones are Scopus, Web of Science, APA, ACM, and IEEE (see Table 1). Following this compilation, we conducted a systematic review within the constructed corpus. At this stage, the key inclusion criteria were the citation

[1]Covidence (Veritas Health Innovation Ltd.), https://www.covidence.org/, last accessed December 22, 2023

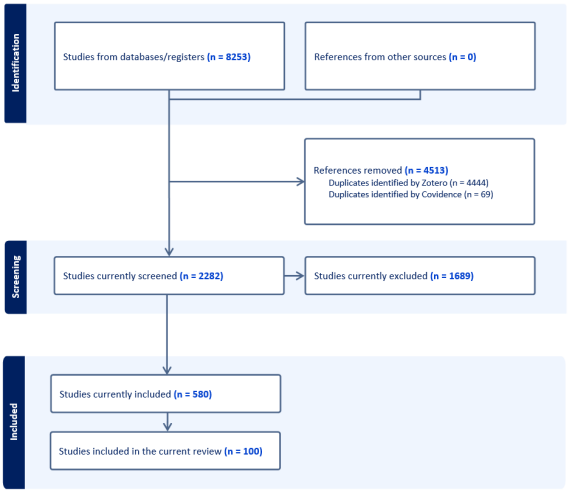


**Figure 1: This figure illustrates the PRISMA [25] flow diagram reviews which included searches of databases and registers only.**

of at least one of the papers [18, 23, 35]. For this late-breaking work study, we selected academic journal and conference papers in English published between 2003 and 2023. We chose this time frame to include the most recent and relevant advancements in the field. This allowed for an up-to-date analysis of the studies that used the scale by Nowak and Biocca [23]. The selection criteria excluded theses/dissertations, pilot studies, proposals, systematic reviews, theoretical papers, literature reviews, and the development of new scales (see Table 2). The search intentionally excluded certain subject areas within Scopus and Web of Science (WOS). It specifically excluded topics related to mathematics, business, and natural sciences. As a result, we arrived at the sample of $n = 8253$.

### 3.2 Duplicate Removal and Screening Phase

We removed duplicates from the automated duplication percent match from Zotero and Covidence (any abstracts and titles with 95% similarity). Additionally, we checked these manually for any errors from $n = 4513$ entries. After eliminating duplicates, we arrived at $n = 3740$ unique records that we identified as suitable for the next phase. In preparation for this paper's screening process, we pairwise evaluated the first 10 papers. The primary aim of this preliminary assessment was to establish a consensus level of no less than 80% agreement among the reviewers/coders—equivalent to a minimum of eight out of the 10 articles before proceeding further. During this early screening process, the first author resolved any emerging disagreements or conflicts. Three or more reviewers thoroughly evaluated every article from the corpus. This improved the rigour and reliability of our paper review process.

*3.2.1 Eligibility Scoping Criteria.* In alignment with the PRISMA protocol, the initial stage often involves screening abstracts and titles. However, in our study, we took a more rigorous approach by directly screening full papers to ensure they met the criteria. We

| **Inclusion Criteria** | |
|---|---|
| **Citation Requirement** | The item must contain the citation of at least one of the papers: Witmer and Singer [35], Nowak and Biocca [23], and/or Lessiter et al. [18]. |
| **Accessibility** | The item must be accessible in its full paper form. |

| **Exclusion Criteria** | |
|---|---|
| **Literature Review** | The item should not be a literature review (including systematic reviews, scoping reviews, etc). |
| **New Scale Development** | The item should not involve the development of new scales. |
| **Dissertation or Thesis** | The item should not be a dissertation or thesis. |
| **Book or Full Proceedings** | The item should not be a book or part of full proceedings. |
| **Extended Abstract** | The item should not be part of extended abstracts, including works-in-progress, posters, case studies, pilots, etc. |
| **Subject area** | The item should not be part of mathematics, natural sciences, and economic sciences. |
| **Language** | The item should be written in English; items in other languages are excluded. |
| **Year** | The paper must be published between 2003 and 2023. |

**Table 2: Table of Inclusion and Exclusion criteria**

acknowledge that this method is time-intensive. However, the articles selected from the early 2000s included titles and abstracts that frequently lacked the requisite information for accurate screening. Specifically, certain papers mentioned one or more of the original (i.e., not modified by someone else) scales but did not employ them, necessitating a more thorough examination through full-paper screening. For example, a paper might reference the scale in its content without actually applying the scale in any substantive manner. At the same time, some of the studies used the scales to measure something other than presence. These were excluded from the analysis. Finally, we included papers that involve research on extended reality (XR), computer-mediated communication (CMC), and video games, as sampled scales were majorly designed for these media.

### 3.3 Data Extraction and Synthesis

From the screened papers, we randomly selected $n = 100$ papers for the current analysis. Four coders participated in the process, with each paper undergoing independent double-coding by two separate coders. Following the coding process, six significant disagreements arose regarding the interpretation of modifications. These were primarily because of unclear reporting in the given papers. An example includes a paper stating the revision of the ITC scale without specifying the nature of the revision, coupled with a discrepancy in the reported number of questions used, indicating one more question than in the original scale. Additionally, eight papers were initially marked as uncertain by one coder, necessitating reconciliation among the coders. To address these disagreements, a third coder was consulted when needed. Table 2 lists the criteria for coding where each coder also justified their answers. Our objective is not to undermine any completed research; rather, to contribute to establishing a standardized protocol for presence research within the academic community.

## 4 FINDINGS

The analyzed sample had the following distribution of papers: 63 papers used the PQ, 29 papers used the ITC, 6 papers used the scales by Nowak and Biocca (NB), and, finally, 2 papers used PQ together with ITC or NB. Most of the studies (94 papers) were related to XR, two to video games, one to CMC, and three to other simulators.

Based on the present analysis, **67%** of the reviewed sample of papers (n=100) included modifications of the analyzed questionnaires to measure presence in one way or another. Most of the time, the number of questions or questionnaire units were reduced. Seventeen of these studies did not report the number of questions they used. Six papers did not include any details, only stating that the questionnaires were "adapted," "revised," "based on," "modified," "derived from," or "culled from" the original questionnaires. Twenty-two studies out of 67 reported that they paraphrased the questions. Only 12 studies out of 67 calculated and reported high internal consistency of modified scales, and two reported low reliability. Finally, only four studies out of 67 re-validated the modified scales.

Only 37 of these 67 papers provided any explanation for the scale modifications. Among those who specified the reasons, the most frequent explanation was that specific subsets of the questionnaires were excluded. This was because they were deemed irrelevant to the studied experiences or media. For example, questions related to sound or haptics were removed from the PQ [35]. ITC [18] was often used via one subscale. Many authors, for some reason, framed this as "spatial presence" instead of the "sense of the physical space" as if it were a separate measure. Some authors also argued that some subscales did not represent presence, contrary to how the authors of the ITC scale argue. In our subsequent analysis which is ongoing at the moment, we analyze the range of the modifications, reported reasons for them, and the use of other non-questionnaire presence measurements to provide methodological solutions for advancing the presence field.

| | |
|---|---|
| **Covidence ID** | List the CrossReference ID on Covidence: 1234 |
| **DOI** | List the DOI of the paper |
| **Title** | List the Title of Paper |
| **Year** | List the Year of when the paper is published |
| **Media type** | List is it "VR, XR, Video Games, Medicated Communication, etc" |
| **The scale** | List all the scales used in the paper |
| **Modification and How** | List if and how the modification is reported |
| **Reason for modification** | List the justification for the modification, if reported |
| **How many questions are used?** | List amount of questions used from the scale |
| **Paraphased** | List the which parts are paraphrased and why |
| **Internal consistency of the whole scale** | List how the scale is administered, and reported |
| **Re-validation, e.g., Factor analysis** | List if any re-validated is completed |
| **Results/details of revalidation** | List the results and details of revalidations. Eg; constructs |
| **Other presence scales used (citation)** | List any other presence scales that are used |
| **Other measures of presence used** | List other (non-questionnaire) measures of presence used |

**Table 3: The coding criteria**

## 5 DISCUSSION

Presence questionnaires are widely adopted [3], or, at least, a recognized method to measure presence, although their validity is questioned in research [17, p. 551]. As our study shows, there are concerns about their reliability as well. In the following subsections, we discuss the most important methodological issues in the analyzed presence studies and describe the future directions to advance presence scholarship.

### 5.1 Possibilities for and Consequences of Modifications

Questionnaires are frequently applied in dynamic contexts; be they social or technological. They often require adjustments for valid measurements. However, these changes should always be justified, and measures be re-validated—a step typically overlooked in presence research. Validity concerns emerge when presence scales—meticulously developed and validated for specific user experience aspects—are modified. This can compromise scale validity, potentially yielding inaccurate or unreliable results. These modifications also give rise to comparability issues, hindering the comparison of results across studies and affecting the consistency needed for accumulating knowledge. Alterations to the original scales—especially when unexplained—impact not only its reliability [5] and potentially reduce measurement consistency but also introduce uncertainty in results interpretations and challenge the replicability of presence studies.

### 5.2 The Range of Modifications

As our results showed, presence questionnaires are not consistently used in a standardized manner. Most of the time, they undergo reduction, revision, paraphrasing, or other modifications. While we observe some commonalities in these changes, they remain quite diverse. The reported reasons for these modifications are also varied (as we explore in more detail in our subsequent research). This raises the question if we can consider presence questionnaires de facto standardized. We argue that—in their current state—presence questionnaires are not valid and reliable measures of presence. Especially if we consider presence as a highly contextual and subjective psychological state dependent on specific content and perceived affordances of a given technology. As a result, the current approach to operationalizing presence is ineffective.

### 5.3 Reporting Problem

Our current results demonstrated that studies measuring presence often lacked clarity in reporting. This makes it challenging to evaluate the studies' methodological choices and perform a systematic analysis. Frequently, the necessity of the modifications was left unreported, or descriptions of scale modifications lacked details. In such circumstances, understanding what variable is measured and why it is measured in a certain way becomes difficult. To enhance clarity, we recommend authors consolidate all methodological information in the Methods section. This avoids the dispersion of methodological choices throughout the paper. This information should include the specific measure used. It should also explain why the authors chose that measure. It should describe any changes they made. And, it should include statistics testing the reliability and validity of the new measures. This will improve the validity of their changes and clarify their research objectives. Ideally, they should report these details in the Appendix for transparency and increased replicability.

### 5.4 The Impact Problem

It is notable that a significant part of the analyzed papers that used modified presence scales failed to validate them. Despite this, these studies have received many citations from other researchers. This has spread potentially inaccurate knowledge. It is crucial to highlight that some of these studies are not recent. The sample under the current analysis spans publication dates from 2007 to 2022. However, the field of presence scholarship has many more studies. Some of these may also not have measured presence accurately. This issue is compounded by other studies that rely on these inadequately validated works. They use them to make assumptions and

develop theories. This perpetuates a cycle of confusion and misunderstanding about the basic aspects of presence. In the end, these studies normalize the practice of making invalid modifications and encourage non-standardized use of questionnaires.

### 5.5 Limitations

In our study, we focused on three scales, chosen for their significant impact as evidenced by existing research and citation numbers. These scales can capture the landscape of presence scholarship. However, we acknowledge that this representation is still partial. While there still may be exceptions among presence questionnaires, major questionnaires often exhibit variations in their application. We did not explore the alterations in scaling, representing a potential methodological gap. Additionally, we did not scrutinize how well presence concepts in studies aligned with the chosen measurements. This aspect warrants exploration in future research. Also, in cases where authors did not report changes, we assumed the use of the original scale. This approach enhanced objectivity in our analysis but constrained our understanding of the full scope of methodological flaws in presence studies.

### 5.6 Future Directions

In the future, we plan to explore more deeper with the scale modifications and the reasons behind them. Our goal is to improve the accuracy of presence measurements. We plan to investigate patterns in modifications, seeking to identify common mistakes to avoid. Additionally, we will analyze the reasoning behind the modifications. We aim to discern contexts in which each questionnaire can or cannot be used, considering factors such as technological advancements.

Our subsequent research will also explore the use of alternative measurements, such as physiological measures or other self-report methods. We will investigate why some studies choose to combine them with questionnaires. Some researchers advocate for the triangulation of methods and using many presence measurements [8], but it remains unclear whether this is an optimal solution to the problems associated with presence questionnaires. If it is, we must work out how they should be used together. Interviews can offer a less standardized yet insightful exploration of different facets of presence experiences. This is particularly helpful in experiences characterized by variability in expressiveness and descriptiveness. We could gain a deeper understanding with these other measurements [33]. Thus, while behavioural measures may dominate in testing presence, qualitative research methods can uncover nuanced aspects of presence [16]. Ultimately, self-report questionnaires remain irreplaceable for measuring individual factors of presence and control variables. However, in their current state, they do not seem to be an appropriate measure of presence itself.

## 6 CONCLUSION

In this study, we have uncovered fundamental methodological challenges and limitations. These include widespread modifications, inconsistent reporting practices, and a lack of validation of altered scales. Together, these issues make current major presence measures unreliable and questionable. They compromise the credibility of individual studies. But they permeate the broader landscape of presence scholarship. The scale of modifications prompts fundamental questions about the standardization of presence questionnaires. We have emphasized an arising need for a more profound understanding of the contextual and subjective nature of presence. The reporting problem underscores the necessity for transparency in methodological choices. The impact problem shows how crucial it is to revalidate scales to prevent inaccurate knowledge from spreading. In essence, our study serves as a call to action for researchers to assess and improve the methodologies used to study presence. Overall, this study contributed to developing more robust and reliable presence measures. These will improve our understanding of presence experiences in the future.

## ACKNOWLEDGMENTS

This study was supported by the SSHRC INSIGHT Grant (grant number:435-2022-0476), NSERC Discovery Grant (grant number: RGPIN-2023-03705) and CFI John R. Evans Leaders Fund or CFI JELF (grant number: 41844). Special thanks to the members of the Games Institute at the University of Waterloo for all their support.

## REFERENCES

[1] Lena Fanya Aeschbach, Sebastian A.C. Perrig, Lorena Weder, Klaus Opwis, and Florian Brühlmann. 2021. Transparency in Measurement Reporting: A Systematic Literature Review of CHI PLAY. *Proceedings of the ACM on Human-Computer Interaction* 5, CHI PLAY (Oct. 2021), 1–21. https://doi.org/10.1145/3474660

[2] Frank Biocca, Chad Harms, and Judee K. Burgoon. 2003. Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria. *Presence: Teleoperators and Virtual Environments* 12, 5 (Oct. 2003), 456–480. https://doi.org/10.1162/105474603322761270

[3] Carmen Bisogni, Lucia Cascone, Aniello Castiglione, and Ignazio Passero. 2021. Deep learning for emotion driven user experiences. *Pattern Recognition Letters* 152 (Dec. 2021), 115–121. https://doi.org/10.1016/j.patrec.2021.09.004

[4] Yasra Chandio, Noman Bashir, Victoria Interrante, and Fatima M. Anwar. 2023. Investigating the Correlation Between Presence and Reaction Time in Mixed Reality. *IEEE Transactions on Visualization and Computer Graphics* Early access (2023), 1–13. https://doi.org/10.1109/TVCG.2023.3319563

[5] Lee Anna Clark and David Watson. 2019. Constructing validity: New developments in creating objective measuring instruments. *Psychological Assessment* 31, 12 (Dec. 2019), 1412–1427. https://doi.org/10.1037/pas0000626

[6] James J Cummings and Erin E Wertz. 2022. Capturing social presence: concept explication through an empirical analysis of social presence measures. *Journal of Computer-Mediated Communication* 28, 1 (Nov. 2022), zmac027. https://doi.org/10.1093/jcmc/zmac027

[7] Zoltan Dienes. 2004. Assumptions of Subjective Measures of Unconscious Mental States: Higher Order Thoughts and Bias. *Journal of Consciousness Studies* 11, 9 (Jan. 2004), 25–45.

[8] William M. Felton and Russell E. Jackson. 2022. Presence: A Review. *International Journal of Human–Computer Interaction* 38, 1 (Jan. 2022), 1–18. https://doi.org/10.1080/10447318.2021.1921368

[9] Jonathan Freeman, S. E. Avons, Ray Meddis, Don E. Pearson, and Wijnand IJsselsteijn. 2000. Using Behavioral Realism to Estimate Presence: A Study of the Utility of Postural Responses to Motion Stimuli. *Presence: Teleoperators and Virtual Environments* 9, 2 (April 2000), 149–164. https://doi.org/10.1162/105474600566691

[10] 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 (Feb. 1999), 1–13. https://doi.org/10.1162/105474699566017

[11] Maia Garau, Doron Friedman, Hila Ritter Widenfeld, Angus Antley, Andrea Brogni, and Mel Slater. 2008. Temporal and Spatial Variations in Presence: Qualitative Analysis of Interviews from an Experiment on Breaks in Presence. *Presence: Teleoperators and Virtual Environments* 17, 3 (June 2008), 293–309. https://doi.org/10.1162/pres.17.3.293

[12] Sarah Graf and Valentin Schwind. 2020. Inconsistencies of Presence Questionnaires in Virtual Reality. In *26th ACM Symposium on Virtual Reality Software and Technology*. ACM, Virtual Event Canada, 1–3. https://doi.org/10.1145/3385956.3422105

[13] Simone Grassini and Karin Laumann. 2020. Questionnaire Measures and Physiological Correlates of Presence: A Systematic Review. *Frontiers in Psychology* 11 (2020), 1–21. https://www.frontiersin.org/articles/10.3389/fpsyg.2020.00349

[14] Wijnand IJsselsteijn, Huib De Ridder, Jonathan Freeman, S. E. Avons, and Don Bouwhuis. 2001. Effects of Stereoscopic Presentation, Image Motion, and Screen Size on Subjective and Objective Corroborative Measures of Presence. *Presence: Teleoperators and Virtual Environments* 10, 3 (June 2001), 298–311. https://doi.org/10.1162/105474601300343621

[15] Karel Kreijns, Kate Xu, and Joshua Weidlich. 2022. Social Presence: Conceptualization and Measurement. *Educational Psychology Review* 34, 1 (March 2022), 139–170. https://doi.org/10.1007/s10648-021-09623-8

[16] Jari Laarni, Niklas Ravaja, Timo Saari, Saskia Böcking, Tilo Hartmann, and Holger Schramm. 2015. Ways to Measure Spatial Presence: Review and Future Directions. In *Immersed in Media: Telepresence Theory, Measurement & Technology*, Matthew Lombard, Frank Biocca, Jonathan Freeman, Wijnand IJsselsteijn, and Rachel J. Schaevitz (Eds.). Springer International Publishing, Cham, 139–185. https://doi.org/10.1007/978-3-319-10190-3_8

[17] Cha Lee, Gustavo A. Rincon, Greg Meyer, Tobias Hollerer, and Doug A. Bowman. 2013. The Effects of Visual Realism on Search Tasks in Mixed Reality Simulation. *IEEE Transactions on Visualization and Computer Graphics* 19, 4 (April 2013), 547–556. https://doi.org/10.1109/TVCG.2013.41

[18] 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 (June 2001), 282–297. https://doi.org/10.1162/105474601300343612

[19] Matthew Lombard, Frank Biocca, Jonathan Freeman, Wijnand IJsselsteijn, and Rachel J. Schaevitz (Eds.). 2015. *Immersed in Media*. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-10190-3

[20] Marvin Minsky. 1980. Telepresence. *OMNI magazine* 2, 9 (1980), 45–51. https://cir.nii.ac.jp/crid/1570854175773549312

[21] Olivier Nannipieri. 2022. Do Presence Questionnaires Actually Measure Presence? A Content Analysis of Presence Measurement Scales. In *Extended Reality*. Vol. 13445. Springer International Publishing, Cham, 273–295. https://doi.org/10.1007/978-3-031-15546-8_24

[22] Richard E. Nisbett and Timothy D. Wilson. 1977. The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology* 35, 4 (April 1977), 250–256. https://doi.org/10.1037/0022-3514.35.4.250

[23] 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 and Virtual Environments* 12, 5 (Oct. 2003), 481–494. https://doi.org/10.1162/105474603322761289

[24] Seok Hee Oh, Jung Woon Park, and Seong-Jin Cho. 2022. Effectiveness of the VR Cognitive Training for Symptom Relief in Patients with ADHD. *Journal of Web Engineering* 21, 3 (March 2022), 767–788. https://doi.org/10.13052/jwe1540-9589.21310

[25] Matthew J. Page, Joanne E. McKenzie, Patrick M. Bossuyt, Isabelle Boutron, Tammy C. Hoffmann, Cynthia D. Mulrow, Larissa Shamseer, Jennifer M. Tetzlaff, Elie A. Akl, Sue E. Brennan, Roger Chou, Julie Glanville, Jeremy M. Grimshaw, Asbjørn Hróbjartsson, Manoj M. Lalu, Tianjing Li, Elizabeth W. Loder, Evan Mayo-Wilson, Steve McDonald, Luke A. McGuinness, Lesley A. Stewart, James Thomas, Andrea C. Tricco, Vivian A. Welch, Penny Whiting, and David Moher. 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 372 (March 2021), n71. https://doi.org/10.1136/bmj.n71

[26] Sebastian A. C. Perrig, Nicolas Scharowski, and Florian Brühlmann. 2023. Trust Issues with Trust Scales: Examining the Psychometric Quality of Trust Measures in the Context of AI. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–7. https://doi.org/10.1145/3544549.3585808

[27] Giuseppe Riva, Fabrizia Mantovani, Claret Samantha Capideville, Alessandra Preziosa, Francesca Morganti, Daniela Villani, Andrea Gaggioli, Cristina Botella, and Mariano Alcañiz. 2007. Affective Interactions Using Virtual Reality: The Link between Presence and Emotions. *CyberPsychology & Behavior* 10, 1 (Feb. 2007), 45–56. https://doi.org/10.1089/cpb.2006.9993

[28] Giuseppe Riva, John Waterworth, and Dianne Murray. 2014. *Interacting with Presence: HCI and the Sense of Presence in Computer-mediated Environments*. DE GRUYTER OPEN, Poland. https://doi.org/10.2478/9783110409697

[29] Richard Skarbez, Frederick P. Brooks, Jr., and Mary C. Whitton. 2018. A Survey of Presence and Related Concepts. *Comput. Surveys* 50, 6 (Nov. 2018), 1–39. https://doi.org/10.1145/3134301

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

[31] Mel Slater, Christoph Guger, Guenter Edlinger, Robert Leeb, Gert Pfurtscheller, Angus Antley, Maia Garau, Andrea Brogni, and Doron Friedman. 2006. Analysis of Physiological Responses to a Social Situation in an Immersive Virtual Environment. *Presence: Teleoperators and Virtual Environments* 15, 5 (Oct. 2006), 553–569. https://doi.org/10.1162/pres.15.5.553

[32] Stella Sylaiou, Katerina Mania, Athanasis Karoulis, and Martin White. 2010. Exploring the relationship between presence and enjoyment in a virtual museum. *International Journal of Human-Computer Studies* 68, 5 (May 2010), 243–253. https://doi.org/10.1016/j.ijhcs.2009.11.002

[33] Phil Turner and Susan Turner. 2006. Place, Sense of Place, and Presence. *Presence: Teleoperators and Virtual Environments* 15, 2 (April 2006), 204–217. https://doi.org/10.1162/pres.2006.15.2.204

[34] Bartholomäus Wissmath, David Weibel, and D. Stricker. 2008. When and how to assess subjective overall judgments of presence?. In *Wissmath, Bartholomäus; Weibel, David; Stricker, D. (2008). When and how to assess subjective overall judgments of presence? In: Spagnolli, Anna; Gamberini, Luciano (eds.) Proceedings of the 11th Annual International Workshop on Presence, Padova, 16-18 October 2008 (pp. 238-243). Padova: CLEUP*, Anna Spagnolli and Luciano Gamberini (Eds.). CLEUP, Padova, 238–243. https://boris.unibe.ch/33931/

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