

# Widespread yet Unreliable: A Systematic Analysis of the Use of Presence Questionnaires

Eugene Kukshinov<sup>1,\*</sup>, Joseph Tu<sup>1</sup>, Kata Szita<sup>2</sup>, Kaushal Senthil Nathan<sup>1</sup> and Lennart E. Nacke<sup>1</sup>

<sup>1</sup>Stratford School of Interaction Design and Business, University of Waterloo, 200 University Ave W, M5H 3C6, Ontario, Canada

<sup>2</sup>School of Communications, Dublin City University, Collins Ave Ext, Whitehall, Dublin 9, Ireland

\*Corresponding author: [e.kukshinov@gmail.com](mailto:e.kukshinov@gmail.com)

Presence, as a psychological state, is typically assessed using questionnaires. While many researchers in this field assume that these self-report instruments are standardized, the reliability of such questionnaires remains uncertain. This knowledge gap challenges the accuracy and validity of data derived from studies assessing presence. Ensuring reliable and precise data collection and reporting is essential for the credibility of findings in presence research, because inaccuracies may cause errors in conclusions, which affects theoretical understandings, methodological approaches and practical applications. To address this issue, we conducted a systematic analysis of 397 empirical quantitative studies on presence. We investigated the use of presence scales, including applications, modifications, a variety of measures and reporting practices. We found that the majority of the presence studies modify questionnaires, do not re-validate them and improperly report their methods. Based on these findings, we propose solutions to enhance transparency and validation of the presence measurements.

**Keywords:** *presence; questionnaires; measurement; self-report; systematic review.*

## 1 INTRODUCTION

Presence can be described as a psychological state induced by interactions with technology, often indicating a level of ‘unawareness’ of mediation or simulation in one’s experience (Kukshinov, 2024). Within Human–Computer Interaction (HCI), particularly in virtual reality research, presence is crucial for assessing user engagement with computer-generated environments (Riva et al., 2014). However, its complexity presents numerous research challenges (Felton & Jackson, 2022, Grassini & Laumann, 2020, Kreijns et al., 2022, Oh et al., 2022). A major issue is the lack of clear conceptualization, which impacts the measurement tools used for presence. The measures often originate from diverse conceptual foundations, casting doubt on the core research and its outcomes (Felton & Jackson, 2022, Kreijns et al., 2022).

The predominant method for assessing presence is through questionnaires (Grassini & Laumann, 2020). However, these numerous questionnaires often produce incomparable measurements of presence (Grassini & Laumann, 2020, Kreijns et al., 2022). They vary widely in the number and content of scale items, ranging from single inquiries to extensive sets designed to capture this complex psychological state (Kreijns et al., 2022, Nannipieri, 2022). Additionally, concerns remain regarding the validity of these questionnaires, as they may not accurately measure presence itself (Kreijns et al., 2022). A study by Nannipieri (Nannipieri, 2022) found that out of 38 analyzed questionnaires, 21 measured a single dimension and 17 used multidimensional scales. Often, questionnaire items focus on user characteristics, devices and virtual environment characteristics rather than

presence itself. Many researchers recognize the issues with the presence questionnaires (Lee et al., 2013). However, due to an arguable lack of alternatives, they continue to use them despite these challenges.

Validity is not the only issue with presence questionnaires. Although they are considered standardized measures (Allman-Farinelli et al., 2019, Faas et al., 2014, Kreijns et al., 2022), we challenge this assumption. Our analysis focuses on some of the most widely used and influential questionnaires, including the Presence Questionnaire (PQ) (Witmer & Singer, 1998), the ITC-Sense of Presence Inventory (ITC) (Lessiter et al., 2001) and the presence scales by Nowak and Biocca (NB) (Nowak & Biocca, 2003). By systematically examining presence studies that utilize these questionnaires, particularly in terms of modifications and their combination with other measures, we provide a critical reflection on established methods in the field. Thus, the objective of the study is not to determine whether presence measurement scales accurately measure presence, but rather, on a methodological level, how rigorously these existing scales are applied.

Our analysis highlights the lack of standardization in these questionnaires, which is often suggested as their primary advantage over other measures. We discuss multiple issues related to the use of measurements, their modifications and the reasons for these changes. Additionally, we address the lack of re-validation for these modifications, the use of other measurement tools and the problems with reporting methodological details in presence studies. Finally, we propose future directions for advancing presence research based on our findings.

Received: May 29, 2024. Revised: October 8, 2024. Accepted: December 23, 2024

© The Author(s) 2025. Published by Oxford University Press on behalf of The British Computer Society.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License

(<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

## 2 RELATED WORKS

The notion of presence, originally labeled as ‘telepresence’, has gained high attention in scholarly discourse due to advancements in Virtual Reality (VR) technologies (Lombard et al., 2015, Riva et al., 2014). Coined by Minsky (Minsky, 1980), ‘telepresence’ initially referred to systems facilitating remote control or tele-operation with a feeling of presence within real but mediated physical environment. Over time, the academic community shortened the term to ‘presence’ and broadened its application to encompass various (media) technologies, such as XR/VR, video games, computer-mediated communication (CMC) or robots (Lombard et al., 2015). Presence is often characterized as a ‘sense of being there’ or a ‘perceptual illusion of non-mediation’ (Riva et al., 2007).

### 2.1 Presence Measurements

Various methods exist for measuring presence; they are broadly categorized into subjective and objective measures (Laarni et al., 2015). Objective measures typically encompass behavioural, physiological and task performance assessments (Chandio et al., 2023, Laarni et al., 2015). However, as Freeman et al. (Freeman et al., 2000) noted, while task performance-based measures may be valuable for evaluating training outcomes, they may not inherently capture presence. Physiological measures, although potentially insightful, are often expensive, and their data can be challenging to interpret within the context of presence (Grassini & Laumann, 2020). Moreover, physiological responses alone may not explain the underlying feelings or sensations constituting presence (Slater et al., 2006). Behavioural measures may offer more precise insights, yet their development is often more complicated and can depend on specific contexts.

Subjective measures include tools such as questionnaires, continuous assessment techniques, psychophysical measures and qualitative methods like interviews and ethnographic measures. While researchers advocate for employing multiple measures to enhance the validity of findings (Felton & Jackson, 2022), post-test self-report questionnaires remain the predominant method for measuring presence (Grassini & Laumann, 2020). They are convenient to create and administer (Felton & Jackson, 2022) and are intended to adhere to standardized protocols, i.e. they are not supposed to ‘require substantial (or any) modification to be used with any type of experience’ (Kreijns et al., 2022, p.96:28].

### 2.2 Presence Questionnaires

Numerous presence questionnaires are available: Nannipieri (Nannipieri, 2022) identified 38 prominent ones. Among these, the Presence Questionnaire by Witmer and Singer (Witmer & Singer, 1998) stands out as the most widely used (VR) presence questionnaire to date (Grassini & Laumann, 2020). However, concerns remain regarding the validity of any presence questionnaires, with their limited scope or incapacity to capture continuous presence being particularly significant (Felton & Jackson, 2022). Consequently, some authors question the fundamental assumption of whether these questionnaires truly measure presence at all (Nannipieri, 2022). Nevertheless, while recognizing these issues, researchers still use them because they are, ‘at least, recognized questionnaires’ (Lee et al., 2013) that are ‘popular’ in the field (Bouchard et al., 2008).

#### 2.2.1 Conceptualization Issues

Any measures should rely on the theory-driven intrinsic value of a variable, but it is one of the biggest shortcomings of presence

questionnaires (Laarni et al., 2015). The conceptualization of presence is frequently defined ambiguously (Felton & Jackson, 2022, Kreijns et al., 2022), along with a lack of clarity in the conceptualization of scale items (Grassini & Laumann, 2020). The dimensions of the construct often lack thorough explanation, thus providing insufficient guidance for effective subsequent operationalization and applications. This leads to the consequence that self-report measures often fail to capture presence accurately. For instance, as Biocca et al. (Biocca et al., 2003) noted, certain attributes, such as communication qualities (composure, spontaneity, positivity, richness and evaluation), might reflect indicators of communication quality rather than social presence (Biocca et al., 2003, p. 467]. Alternatively, presence questionnaires may focus on past behaviours or presence-related factors that correlate with social presence rather than social presence itself (Biocca et al., 2003). Nannipieri (Nannipieri, 2022) found that among the 38 questionnaires examined, the majority of scale items were not directly linked to presence itself. This observation extends to the Presence Questionnaire (PQ) as well (Witmer & Singer, 1998), wherein certain dimensions explicitly state that they measure factors such as the perceived quality of a device’s interface.

#### 2.2.2 Continuity Issue

In contrast to post-hoc questionnaires, which seem ineffective during VR (or other media) experiences (Graf & Schwind, 2020), continuous presence measures emerge as more valid tools. These measures account for the dynamic nature of presence and the breaks that arise during interaction with technology (Garau et al., 2008). To put this simply, continuous tools facilitate the evaluation 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),’ thus surmounting limitations associated with post-rating measures (Wissmath et al., 2008, p. 239].

#### 2.2.3 Issue with Sensation vs. Knowledge

Presence represents a complex psychological state, unlike self-evident sensorial experiences such as perceived temperature, making it challenging to articulate (Freeman et al., 2000). Individuals generally possess limited insight into their cognitive processing and are often unaware of perceptual and memory mechanisms (Nisbett & Wilson, 1977). This hinders their ability to articulate higher-order thoughts, particularly regarding perceptual illusions, verbally (Dienes, 2004). In other words, it is possible to feel presence but not to be aware of it. As individuals may only have partial consciousness of their state of presence, self-reports alone may not provide a comprehensive understanding without supplementation from objective measures (Laarni et al., 2015). Additionally, subjective measures can be susceptible to instability and bias stemming from prior experiences and interests (Freeman et al., 2000), while post-test evaluations are frequently affected by inaccurate recollection and memory effects (Freeman et al., 1999). Overall, there is a natural gap between thought, emotion and expression based on language.

A significant concern regarding the internal validity of self-reports is the potential conflict between sensation and knowledge. As highlighted in a study by Freeman et al. (Freeman et al., 2000), observers 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’ (Slater, 2004, p. 486]. The phrasing of questionnaire items may suggest expected effects, potentially influencing

participants' responses. Individuals unfamiliar with the concept may be particularly sensitive to demand characteristics in experimental settings (Ijsselstein et al., 2001, Schrader & Bastiaens, 2012). In particular, 'there is a risk that instead of experiencing different levels of virtual presence, learners may not understand the terminology used in the rating scale or may have only differed in their capability and willingness to report their emotional state' (Schrader & Bastiaens, 2012).

### 2.2.4 Standardization Issues

It is a prevalent assumption that questionnaires used for measuring presence are standardized (Sylaiou et al., 2010). However, it does not seem that presence can be comprehensively measured in a standardized manner, as it is inherently transient and varies depending on the medium, the participants and other contextual factors (Biocca et al., 2003). The design of the medium or simulation can promote diverse experiences, requiring measurement approaches tailored to these variations (Laarni et al., 2015). This is true for any mediated/simulated experiences stemming from diverse designs and resulting in various social consequences (Lombard et al., 2015). The circumstances surrounding these experiences may differ to such an extent that adaptation of measurement instruments may not suffice; rather, novel measures may need to be developed.

Given the aforementioned challenges with presence questionnaires, we advocate for an examination of how various presence studies employ these instruments. We aim to gain insights into the consistency of self-report measures across presence studies. Therefore, this paper addresses the following research questions:

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

### 2.3 Modifications of Measurements

While re-validations of modified measures in studies are not that common (Finn & Kayande, 2004), modifications to the measures can happen if they are justified. Any modified or adapted measurements should be re-validated (Clark & Watson, 2019), starting with reliability calculations. It is essential because changing the (wording of a) scale item can alter the construct that is being measured or its validity (Juniper, 2009). Also, modifications to scales can impact reliability by changing the internal consistency of the scale items: 'By narrowing the scale content, the scope and nature of the assessed construct is itself changed; in particular, it increases item redundancy, thereby reducing the total amount of construct-related information the test provides' (Clark & Watson, 2019, p. 1416). It is possible to use shorter questionnaires that better fit studies' designs without changing the existing measures (Lull & Bushman, 2016). Finally, modifications to scales can make it difficult to compare results across different studies or groups of respondents, affecting their replicability. To better understand key issues and possible solutions to the methodological gaps in the presence studies, this paper also addresses the following questions:

- **RQ3:** If presence questionnaires are modified, how and why does it happen?
- **RQ4:** What other presence measurements are used and why?

## 3 METHOD

In our research, we examined empirical studies that used at least one of the three presence scales presented above (Lessiter

et al., 2001, Nowak & Biocca, 2003, Witmer & Singer, 1998). PQ, initially, consists of 32 items, however, Witmer and Singer (Witmer & Singer, 1998) also offered another shorter version with 19 items. ITC consists of 43 items in total, however, it is based on four subscales that can be used separately to measure presence, as suggested by Lessiter et al. (Lessiter et al., 2001). Nowak and Biocca (Nowak & Biocca, 2003) offered four different scales (for different types or subtypes of presence) that, in total, consisted of 24 questions. We selected these scales because of their established effectiveness in measuring physical and social presence (Cummings & Wertz, 2022, Felton & Jackson, 2022, Grassini & Laumann, 2020, Kreijns et al., 2022, Skarbez et al., 2018), particularly in VR/XR, video games and computer-mediated communication platforms. Their widespread adoption within these contexts makes them well-suited for our study (see Table 1).

Our dataset includes 8253 records, 4708 papers as duplicates and 3545 papers ready for screening. Our methodology for the review's search, screening and data extraction stages was based on the PRISMA (Page, 2021) approach (see Figure 1). The screening was run in the Covidence software<sup>1</sup>. In the end, we analyzed  $N = 397$  papers.

### 3.1 Protocol and Search

We systematically gathered an extensive compilation of scholarly articles referencing at least one of the specified scales, which formed the basis of our paper corpus. Additionally, we identified the primary databases where these scales are cited (Grassini & Laumann, 2020). The key databases include Scopus, Web of Science, APA, ACM and IEEE (refer to section 1). Subsequently, we conducted a systematic review protocol within this compiled corpus. During this phase, the primary inclusion criterion was the citation of at least one of the papers (Lessiter et al., 2001, Nowak & Biocca, 2003, Witmer & Singer, 1998). We focused on academic journal and conference papers (as they often include original empirical studies) in English published between 2003 and 2023 to encompass the most recent and relevant developments in the field. This time frame enabled us to perform an up-to-date analysis of studies employing the scale by Nowak and Biocca (Nowak & Biocca, 2003). Our selection criteria excluded theses/dissertations and other formats that were not journal and conference papers (see Table 2). Moreover, the search deliberately omitted certain subject areas within Scopus and Web of Science (WOS), particularly those related to mathematics, business and natural sciences (for the full list of excluded areas, see the Appendix). Consequently, our sample size amounted to  $n = 8253$ .

### 3.2 Screening Phase

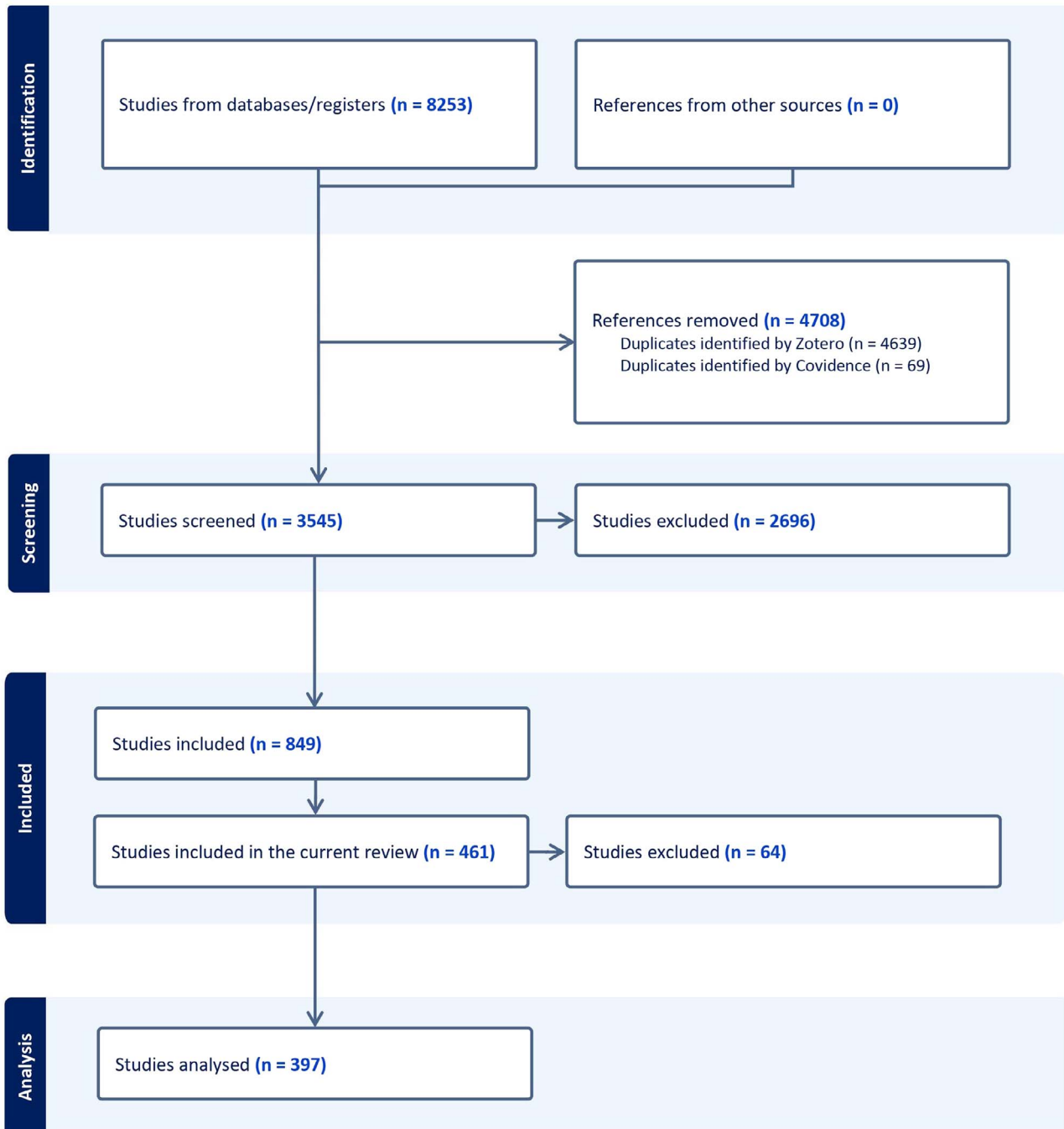
#### 3.2.1 Duplicate Removal

We eliminated duplicates identified through automated duplication checks in Zotero and Covidence, removing any entries with 95% similarity in abstracts and titles. Additionally, we checked these manually for any errors from  $n = 4708$  entries. Following this de-duplication process, we identified  $n = 3545$  unique records deemed suitable for the subsequent phase. To prepare for the screening process, we conducted pairwise evaluations of the first ten papers. The primary objective of this initial assessment was to achieve a consensus level of at least 80% agreement among the reviewers/coders, equivalent to a minimum of eight out of the ten articles, before proceeding. Any disagreements or conflicts arising

<sup>1</sup> Covidence (Veritas Health Innovation Ltd.), <https://www.covidence.org/>, last accessed March 15, 2024

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

Database	PQ (Witmer & Singer, 1998)	ITC (Lessiter et al., 2001)	NB (Nowak & Biocca, 2003)
Scopus	3510	709	644
Web of Science	2663	554	470
APA	869	225	185
ACM	829	155	166
IEEE	438	54	52

**FIGURE 1.** This figure illustrates the PRISMA (Page, 2021) flow diagram reviews, which included searches of databases and registers only.

during this early screening phase were resolved by the first author. Each article from the corpus underwent a thorough evaluation by two or more reviewers, enhancing the rigour and reliability

of our paper review process. While four coders were consistently involved in the analysis, two additional coders participated temporarily.

**TABLE 2.** Table of Inclusion and Exclusion criteria in search.

<b>Inclusion Criteria</b>	
<b>Citation Requirement</b>	The item must contain the citation of at least one of the papers: PQ (Witmer & Singer, 1998), NB (Nowak & Biocca, 2003) and/or ITC (Lessiter et al., 2001).
<b>Accessibility</b>	The item must be accessible in its full paper form.
<b>Exclusion Criteria</b>	
<b>Dissertation or Thesis</b>	The item should not be a dissertation or thesis.
<b>Book or Full Proceedings</b>	The item should not be a book or part of full proceedings.
<b>Extended Abstract</b>	The item should not be part of extended abstracts, including works-in-progress, posters, case studies, pilots, etc.
<b>Subject area</b>	The item should not be part of mathematics, natural sciences and economic sciences.
<b>Language</b>	The item should be written in English; items in other languages are excluded.
<b>Year</b>	The paper must be published between 2003 and 2023.

**TABLE 3.** Table of Inclusion and Exclusion Criteria in Screening Phase of the protocol.

<b>Inclusion Criteria</b>	
<b>Citation Requirement</b>	The study must use at least one of the original scales to measure presence quantitatively: PQ (Witmer & Singer, 1998), NB (Nowak & Biocca, 2003) and/or ITC (Lessiter et al., 2001).
<b>Accessibility</b>	The study must be accessible in its full paper form.
<b>Media type</b>	The study must be based on one of the media types: VR/XR, CMC, video games.
<b>Exclusion Criteria</b>	
<b>Different scale version</b>	The item should not use or cite an existing (adapted) version of one of the scales.
<b>Measured variable</b>	The item should not measure any other construct but presence, such as user experience (UX), with the scales.
<b>Literature Review</b>	The item should not be a literature review (including systematic reviews, scoping reviews, etc).
<b>New Scale Development</b>	The item should not involve the development of new scales.
<b>Pilot studies</b>	The item should not be a pilot study.

### 3.2.2 Eligibility Scoping Criteria

Following the PRISMA protocol, the initial stage involves screening abstracts and titles. However, in our study, we opted for a more rigorous approach by directly screening full papers to confirm they met the criteria. We recognize that this method is more time-consuming. However, many articles had titles and abstracts that often lacked sufficient information for accurate screening. In particular, some papers mentioned one or more of the original scales but did not actually use them, requiring a more comprehensive examination through full-paper screening. Also, studies were excluded if they did not use the original scale but its adaptation, including an existing validated translation, such as the one by UQO Cyberpsychology Lab (Robillard et al., 2002), or the updated version of PQ by the original authors (Witmer et al., 2005). See all screening criteria in Table 3. A study was included if it translated an original scale, counting this as modification. Any conflicts during screening were reconciled by a third coder (which mostly happened due to different access to the papers among the coders). The full screening took ten months, and then we compiled a dataset of  $N = 849$  papers for the subsequent analysis.

### 3.3 Data Extraction and Synthesis

For the final analysis, we randomly coded around half of the full sample,  $N = 461$ . We stopped coding the full-screened sample when we concluded that no new results emerged from the sample. Four coders participated in the process, with each paper undergoing independent double-coding by two coders. Table 4 lists the criteria for coding.

After coding the data, we encountered 24 significant disagreements about how to interpret the reported modifications in the studies. This was primarily due to unclear information within the papers themselves. For example, one paper mentioned revising a scale (ITC) but didn't explain the specific changes made. We also found a discrepancy in the reported number of questions used in

one study, with one more question than what was reported for the original scale. Additionally, one coder initially marked eight papers as uncertain. To address these disagreements, we brought in a third coder to review the cases and make a final decision (of the eight papers). For minor disagreements, such as how to handle paraphrased items, we discussed the issue among ourselves and mediated the differences.

Lessiter et al. (Lessiter et al., 2001) suggested that researchers can measure presence with separate sub-dimensions of their questionnaire, such as 'Sense of Physical Space', which was usually referred to as 'Spatial Presence' by some researchers. In the case of the PQ, Witmer and Singer (Witmer & Singer, 1998) presented two versions of their questionnaire. As a result, the use of sub-scales or a second version was not considered modification unless it was modified some other way.

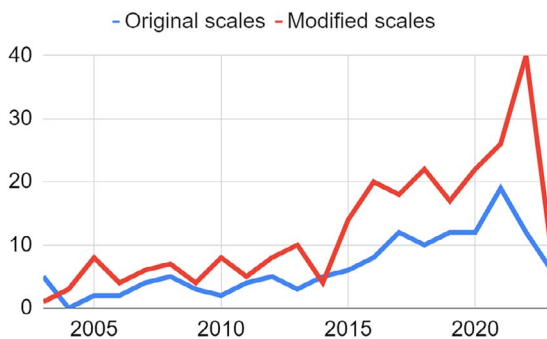
Finally, during the analysis, we excluded an additional 64 papers. These included papers that were out of the publication year scope or were pilot studies but remained in the sample by mistake. The main reasons for these exclusions are discussed in 5.4 Limitations section.

## 4 RESULTS

The analyzed sample ( $n=397$ ) had the following distribution of papers: 278 papers used the PQ (Witmer & Singer, 1998), 88 papers used the ITC (Lessiter et al., 2001), 28 papers used the scales by Nowak and Biocca (NB) (Nowak & Biocca, 2003) and 3 papers used PQ together with ITC or NB. Most of the studies ( $n=375$ ) were related to XR, 19 to video games, 9 to CMC and four to other simulators. Based on the current analysis, 65% ( $n=257$ ) of the coded sample included modifications of the analyzed questionnaires to measure presence in one way or another (there was also one uncertain case due to vague reporting). Modifications were found in 192 papers out of 278 with PQ, 39 out of 88 papers with ITC

**TABLE 4.** The coding criteria.

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

**FIGURE 2.** Distribution of papers across years.

and 25 out of 28 papers with NB. See Figure 2 for the distributions across years.

#### 4.1 Forms of And Reasons for Modifications

Most of the time, the number of questions or questionnaire items was modified ( $n=151$ ), which includes six papers where authors added extra questions. In 145 cases, authors reported a reduced number of questions (this excludes the cases where studies used separate unmodified sub-scales of the questionnaires). This includes papers where modifications were not directly addressed but were evidenced in tables or annexes. Notably, in 48 cases, the authors did not report any details; they only used 'shortened' or 'reduced' versions of the given questionnaire(s). In 63 papers, authors reported only the number of the items used from the original scales, but only in 34 cases were these items specified. For example, only the items related to certain concepts or phenomena, such as haptics or realism, were used. Any number of the scale items could have been used (sometimes studies used 1-2 items, which could have been combined with items from other questionnaires). On average, when reported, studies in our sample used 15 items out of the original 32 from the PQ and 22.7 items out of the original 43 from the ITC. BN included multiple scales, which complicated the calculation of the average used item number. Overall, 86 papers did not include any details of modifications besides that the questionnaires were 'adapted,' 'revised,' 'based on,' 'modified,' 'derived from,' 'inspired by' or 'culled from' the original questionnaires. Eight studies reported that they reworded (generally or specifically) questions; 11 studies reported translating original questionnaires.

Only 89 (35%) of the papers that reported modifications provided any reasons or justification for the modifications. Among those with the reasons specified, the most frequent explanation was that specific subsets of the questionnaires were excluded because they were deemed irrelevant to the studied context or media ( $n=57$ ). For example, in 20 cases, items related to sound or haptics were removed as they were unrelated to the experience being evaluated or because their Virtual Environments (VEs) did not provide these features and did not allow direct manipulation. One study, in addition to the auditory and haptic items, excluded items on the interface quality, because these items were less relevant to the visual experiences of viewers [the focus of their study]. In addition, four studies excluded items about interaction with VE, because their systems did not involve this feature. Overall, fifteen of the studies with modifications only generally stated that not all items were 'relevant,' 'applicable' or 'useful' to their studies' scope, context or focus. Sixteen papers specified the context, for instance, video games, virtual reality therapy or odour focus. Only twenty-six studies (mostly using the PQ) gave specific details on which factors/items were irrelevant.

The second major reason for modifications was the length of the questionnaires. Some researchers reduced the number of questions to be able to use multiple measures to avoid fatigue among participants, or simply because they wished to keep it short. Six papers only used questionnaire items that contributed to the reliability of the scale. In other cases, some items were cut from the original questionnaires because of precedent; that is, questions were excluded in previous studies or had overlaps with the other used questionnaires.

Arguably, the most justifiable reason for modifications was to cater to specific groups of participants to improve comprehension of the question items ( $n=11$ ). In particular, 3 studies paraphrased original questionnaires so that children or students can understand them. Eight studies translated questionnaires for their respective non-English speaker participant pool. However, although justified, modifications predominantly lacked re-validations.

#### 4.2 Statistics Reporting

Only 72 studies (28% of the total sample) calculated and reported the internal consistency of the modified scales: 63 of them reported high reliability, and 9 reported low-reliability scores, where 0.8 Cronbach's alpha was considered a high level of scale reliability. Only 14 papers of the total sample (5%) re-validated

their modified scales. Six of these did not report any statistics or details, only stating that they re-validated scales. Eight of them reported assessing validity via various statistical methods.

### 4.3 Use of Other Measures

Out of our sample, 68 studies used at least one extra presence scale: this includes papers without modified scales ( $n=21$ ) and with modifications of one of the analyzed scales ( $n=47$ ). As additional measures besides our analyzed questionnaires, the SUS questionnaire (Usoh et al., 2000) was the most frequently used one ( $n=12$ ). Twenty-eight papers used other presence measures than questionnaires as an extra measure (12 without modifications; 17 with modifications of the analyzed scales). These methods involved structured or semi-structured interviews ( $n=9$ ), behavioural measures ( $n=6$ ), tasks and physiological measures. In terms of physiological measures, three studies used heart rate (HR), three electroencephalogram (EEG) and two electrodermal activity (EDA). Also, two studies used a combination of behavioural and physiological measures and three used a combination of different physiological measures.

Only twenty-four papers motivated the use of other measurements. Behavioural and physiological measures were contrasted to questionnaires as more objective measures of presence. They also used these objective measures to triangulate, cross-examine or validate subjective measures, arguing that one questionnaire may not be enough to process presence. Other studies used objective measures to avoid subjective biases or inherent problems, such as anchoring effects, imprecise memories and the inability to describe the subjective variations of presence. Interviews were used for the in-depth explanation of the participants' experiences that questionnaires could not offer. Overall, 'the inclusion of multiple measures, such as including an objective corroborative measure (e.g. incorporating a behavioural measure) in conjunction with a post-immersion questionnaire, could provide additional confidence in questionnaire results' (Felton & Jackson, 2022).

The use of additional questionnaires was rarely motivated ( $n=6$ ). The authors of four of these papers reported using additional shorter questionnaires because they could have used them as in-session measures during the experiments. One study used multiple presence questionnaires to compare them, and another did it because, as the authors claimed, no one had done it before.

## 5 DISCUSSION

Questionnaires are widely adopted (Bisogni et al., 2021) and recognized methods to measure presence, although their validity is questioned in research (Lee et al., 2013, p. 551). As our findings show, 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 Key Findings

#### 5.1.1 The Range of Modifications

Our results revealed that presence questionnaires are not consistently used in a standardized manner. They frequently undergo reduction, revision, paraphrasing or other modifications. While there are some commonalities in these changes, they remain quite diverse, and the reported reasons for these modifications vary widely. This raises the question of whether we can consider presence questionnaires as *de facto* standardized. We argue that,

given the current state of the art, presence questionnaires are not valid and reliable measures of presence. This is especially true if we consider presence as a highly contextual and subjective psychological state that is dependent on specific content, perceived affordances of a given technology and personal factors. Consequently, the current approach to operationalizing and assessing presence is ineffective. This disagreement exemplifies a fundamental issue in the presence research: primarily the lack of a universally accepted definition and measurement tool for presence. Presence is a complex construct encompassing not just the feeling of being physically located within the virtual environment (VE) but also a sense of self-extension into the virtual world and the perception of social interaction with others (Lessiter et al., 2001, Nowak & Biocca, 2003, Witmer & Singer, 1998). The current state of presence measurement tools often struggle to capture these multifaceted aspects.

In addition, while Lessiter et al. (Lessiter et al., 2001) suggested that it is acceptable to use separate subscales of the ITC. However, many scholars have framed the subscale of the 'sense of physical space' as 'spatial presence.' In doing so, they argued that only this subscale measures presence, which contradicts the original authors' intentions for the ITC that was designed to assess different factors of presence, including but not limited to the 'sense of physical space.' This represents an uncertain state of presence measures and a major confusion regarding what scholars consider presence.

#### 5.1.2 The Reasoning Behind Modifications

While translating or adapting questionnaires for certain target demographics (e.g. children) are justifiable modifications, these were rare in our sample ( $n=11$ ). Rather, most reported modifications were related to the specific use cases and technologies examined in the studies. Although adapting questionnaires for specific technologies or content is common practice, there is no evidence that technological advancements would drive modifications. Contrary to what one might expect with technological developments, the major reason for reducing the number of questionnaire items was due to the limitations of a given technology or stimuli used in particular studies, such as haptics or audio. In these cases, researchers often treated certain items as 'optional.' However, some presence scholars argue that all these sensory cues are integral parts and strong indicators of the sensation of presence in certain media (e.g. VR) (Witmer & Singer, 1998), suggesting that these elements should also be reflected in the questionnaires. If these sensory cues and features are essential to the sense of presence, items representing them cannot be omitted. As a result, the absence of these elements indicates a weaker sense of presence in the evaluated technological context. Sometimes, researchers recognize that their modifications affect the validity of the measurements. However, they still modify them as the original measure is not relevant enough to their conditions (Price & Anderson, 2007).

There is evidence that presence questionnaires, such as the PQ, do not measure presence itself but rather its factors (Nannipieri, 2022), which may vary widely or be minimal during a media experience to induce presence. Consequently, it is reasonable to suggest that presence questionnaires cannot capture an overall presence sensation based on these factors. This includes the most basic features, such as 360-degree view. For instance, some items in the Ingroup Presence Questionnaire (Schubert et al., 2001) ask about the feeling of 'being surrounded,' which specifically targets the use of head-mounted displays (HMDs) and cannot be compared with other media conditions, such as flat-screen.

This is problematic because presence is not confined to HMD use. Therefore, any comparisons between immersive conditions made using such questionnaires lack validity. Given their variety and issues, we argue that presence questionnaires cannot be standardized, as presence heavily depends on specific affordances that each study design involves. In other words, it depends on specific aspects that are simulated and/or mediated.

### 5.1.3 The Use of Multiple Presence Measures

Many studies in our sample used more than one presence measure, whether it was another questionnaire or a different tool, such as physiological or behavioural measures. When using multiple measures, some studies reduced the number of items to avoid lengthy questionnaires and biases due to fatigue. However, given the diversity of different questionnaires, we argue it is more reasonable to use a shorter, validated measure or fewer measures overall instead of modifying existing ones. Another questionable practice is the blending of several questionnaires, which was common in studies that used multiple scales. This practice raises concerns as these questionnaires often rely on distinct conceptualizations. While a few studies explained the necessity of using several presence scales simultaneously, it appears that researchers were unsure about the capability of these measures to capture presence on their own. It is possible that certain questionnaires measure only one side of presence (if any at all). This uncertainty also led them to use additional measures, such as physiological tools or interviews, to validate or explain questionnaire results. At this stage of questionnaire development, employing different forms of measurement seems to be the most reasonable approach.

## 5.2 Issues of and Impact on Presence Research

### 5.2.1 The Outcomes of Modifications

Questionnaires are frequently used in real-life dynamic contexts as both technology and human that uses it evolve over time. Since no single questionnaire can fit all (emerging) situations, they often need modifications or complementary methods to measure specific mediated contexts, technologies and experiences. However, our results indicate that the adaptation of existing scales often lacks justification or detailed explanations (such as disclosing the exact items used) and re-validation. This is particularly concerning because modifying questionnaires—originally developed and validated for specific user-experience aspects—raises validity concerns and can yield inaccurate or unreliable results. These modifications also create comparability issues, making it difficult to compare results across studies and affecting the consistency required to build cumulative knowledge. Unexplained alterations to the original scales impact not only the scale's reliability and consistency but also introduce uncertainty in result interpretations and challenge the replicability of presence studies (Clark & Watson, 2019). While re-validations of the existing scales require large samples, it does not mean they should be neglected. If an existing validated measure cannot be used in its original form or re-validated when modified, it should not be used. Instead, researchers can explore other presence measures.

### 5.2.2 The Reporting Problem

Our results demonstrate that studies measuring presence often lack clarity in reporting, which may account for text length limitations, brevity or incorrect practices passed down to researchers in different career stages. Despite these reasons, it is challenging to evaluate the studies' methodological choices systematically.

Frequently, the necessity of the modifications was left unreported, and descriptions of scale modifications lacked detail. In such circumstances, understanding what variable is measured and why it is measured in a certain way becomes difficult. As discussed in the limitations section, it is possible that there are studies that modified presence questionnaires that remain undetected due to incomplete reporting.

### 5.2.3 The Impact Problem

Many analyzed papers used modified presence scales without validating them. This raises concerns, as these studies, despite potential inaccuracies, have been widely cited, potentially perpetuating misconceptions about presence within the field. The issue is further amplified by the fact that this practice isn't recent (encompassing studies from 2003 to 2023). Subsequent research has often relied on these inadequately validated works as a foundation for further studies and theoretical development. This creates a snowball effect, perpetuating confusion about the fundamental aspects of presence, which is already vague and complicated. Ultimately, these studies normalize invalid modifications and encourage the non-standardized use of questionnaires in any field of scholarship.

## 5.3 Possible Solutions

In the following, we provide a set of suggestions to formulate a more structured approach to presence research. Following these points, researchers can determine the ways in which they can appropriately use presence questionnaires, improving the validity and reliability of presence assessments.

### 5.3.1 Validation Based on Common Conceptualization

Currently, there are multiple and dissimilar presence questionnaires that are supposed to measure the psychological state of presence. However, definitions of and approaches to presence are also quite diverse in research. Even if we reject the approach to measuring presence based on its factors, there is still uncertainty over the conceptualization of presence, which complicates the development of measurements. As noted in one of the analyzed studies, 'we as researchers need to explicate better and operationalize the concept of presence. While the scale used in this study has been employed by several other studies, it still may not capture the true essence of presence' (Nicovich, 2010, p.617). Once the presence community together agrees upon a general understanding of what presence (not just in VR) is, then it would be possible to develop new measurements that measure the same construct, which may or may not include questionnaires.

### 5.3.2 Multiple And Diverse Measures

As a temporary solution, it is reasonable to combine presence measurements. Currently, presence studies demonstrate that it is impossible to accurately assess presence with a single questionnaire. While using questionnaires is an easy solution, there should be more research that develops other measures and does not rely on self-report scales. Other forms of presence measures, such as interviews or behavioural tasks, may give researchers a different perspective regarding presence experiences and, what is more important, real-time metrics that are necessary to capture the discontinuous state of the sense of presence. Real-time metrics are more reliable and consistent compared to post-test evaluation (Kober et al., 2012).

At the same time, 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 (Turner & Turner, 2006). Thus, while behavioural measures may dominate in testing presence, qualitative research methods can uncover nuanced aspects of presence (Laarni et al., 2015). Ultimately, self-report questionnaires remain irreplaceable for measuring individual factors of presence and control variables.

### 5.3.3 Suggestions to Researchers

For researchers unfamiliar with presence scholarship and its issues, we recommend reviewing multiple studies and approaches, including meta-reviews, to understand better whether there is any agreement regarding constructs and the related measurements. There is no such agreement in the presence studies. While we suggest using multiple measures of presence if using a questionnaire is the only option, it should be used in its validated original form or re-validated if modifications are made.

When reporting modifications to measurement scales, researchers should provide comprehensive details to ensure transparency and rigor. This includes clearly identifying the specific measurement scale(s) used and explaining why those particular scales were chosen in relation to the theoretical constructs and research context. Any modifications to the scale, such as additions, deletions or wording changes, should be detailed, along with the rationale behind these changes. It is essential to assess the reliability and validity of the modified scale by reporting statistics such as Cronbach's alpha for internal consistency, content validity to ensure the changes do not alter the construct's meaning, and exploratory factor analysis (EFA) to evaluate the stability of the scale's structure and dimensions. Confirmatory factor analysis (CFA) should also be used to confirm the stability of the scale's dimensional structure. If possible, researchers should estimate the internal, external and nomological validity of the modified scale to demonstrate its relevance and generalizability. Providing these details will ensure that the modifications are rigorously evaluated and appropriately interpreted.

This will improve the validity of their changes and clarify their research objectives. To enhance clarity, we recommend that the authors consolidate all methodological information in the Methods section to avoid dispersing methodological choices throughout the paper and disclose the details of the questionnaire items in an appendix or open-access repository.

### 5.4 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, widely used 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.

During screening, we did not consider studies as presence studies if they did not conceptualize presence or if they used the scales to measure other constructs than presence—these papers were excluded from the sample. During coding, we had to exclude 64 papers, which happened for a few reasons. While some of the papers were included by technical mistake, since they were not filtered during the search and screening phases, researchers also lost access to some of the papers in the span of 10 months. Finally, we had to exclude papers because they were not screened properly. This probably happened due to the lack of full re-training of the coding team within a 10-month period, which is necessary because coders can forget some of the criteria in the long-term coding process.

### 5.5 Future Directions

In the future, we can evaluate the impact of the studies that improperly use presence questionnaires. We can possibly track the connections between them to see how much they contribute to the ongoing issues. We can also look into how modified measures of presence correspond with the conceptualizations of presence. Many times, in our sample, presence was not properly defined and/or related to other variables. As a result, we urge scholars to look for a more precise conceptualization of presence (across technologies and presence types).

## 6 CONCLUSION

Our study highlights critical methodological issues in the use of presence questionnaires within empirical research. Frequent modifications to these questionnaires include item reduction and paraphrasing. When done without proper re-validation, these changes raise serious concerns about their validity and reliability. These changes are often made without clear justification. This practice undermines the comparability and consistency of presence measurements, which leads to potential inaccuracies in research findings. We emphasize the importance of a unified conceptual understanding of presence, as the current diversity in definitions further complicates the development of standardized measurement tools. Additionally, our findings suggest that combining multiple measures, including quantitative and qualitative approaches, can enhance the robustness of presence assessments. Researchers should prioritize using validated questionnaires in their original forms or ensure rigorous re-validation when modifications are necessary. To improve transparency and replicability, detailed methodological reporting is crucial. Our recommendations aim to guide future research towards more precise and reliable measurements of presence, ultimately contributing to a clearer and more cohesive body of knowledge in the field.

### Acknowledgments

We would like to thank Nima Zargham and Sabrina Sgandurra for help in screening, and the Games Institute for space and overall support. This study was supported by the SSHRC INSIGHT Grant (grant number: 435-2022-0476), NSERC Discovery Grant (grant number: RGPIN-2023-03705), CFI John R. Evans Leaders Fund (grant number: 41844), Meta Research Award, and the Lupina Foundation.

### Competing interests

No competing interest is declared.

## References

- Abbas, A., Choi, M., Seo, J., Cha, S. and Li, H. (2019) Effectiveness of Immersive Virtual Reality-based Communication for Construction Projects. *KSCE J. Civ. Eng.*, **23**, 4972–4983. <https://doi.org/http://link.springer.com/10.1007/s12205-019-0898-0>.
- Adão, T., Pádua, L., Narciso, D., Sousa, J., Agrellos, L., Peres, E. and Magalhães, L. (2019) MixAR: A Multi-Tracking Mixed Reality System to Visualize Virtual Ancient Buildings Aligned Upon Ruins. *J. Inf. Technol. Res.*, **12**, 1–33. <https://doi.org/https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/JITR.2019100101>.
- Aeschbach, L., Perrig, S., Weder, L., Opwis, K. and Brühlmann, F. (2021) Transparency in Measurement Reporting: A Systematic Literature Review of CHI PLAY. *Proc. ACM Hum.-Comput. Interact.*, **5**, 1–21. <https://doi.org/https://dl.acm.org/doi/10.1145/3474660>.
- Affifi, T., Collins, N., Rand, K., Otmar, C., Mazur, A., Dunbar, N., Fujiwara, K., Harrison, K. and Logsdon, R. (2023) Using Virtual Reality to Improve the Quality of Life of Older Adults with Cognitive Impairments and their Family Members who Live at a Distance. *Health Commun.*, **38**, 1904–1915. <https://doi.org/https://www.tandfonline.com/doi/full/10.1080/10410236.2022.2040170>.
- Allman-Farinelli, M., Ijaz, K., Tran, H., Pallotta, H., Ramos, S., Liu, J., Wellard-Cole, L. and Calvo, R. (2019) A Virtual Reality Food Court to Study Meal Choices in Youth: Design and Assessment of Usability. *JMIR Form. Res.*, **3**, e12456. <http://formative.jmir.org/2019/1/e12456/>.
- Baker, S., Kelly, R., Waycott, J., Carrasco, R., Bell, R., Joukhadar, Z., Hoang, T., Ozanne, E. and Vetere, F. (2021) School's Back: Scrolling Reminiscence in Social Virtual Reality with Older Adults. *Proc. ACM Hum.-Comput. Interact.*, **4**, 1–25. <https://doi.org/https://dl.acm.org/doi/10.1145/3434176>.
- Bang, E. and Yildirim, C. (2018) Virtually Empathetic?: Examining the Effects of Virtual Reality Storytelling on Empathy. *Virtual, Augmented And Mixed Reality: Interaction, Navigation, Visualization, Embodiment, And Simulation*, **10909**, 290–298. [https://doi.org/https://link.springer.com/10.1007/978-3-319-91581-4\\_21](https://doi.org/https://link.springer.com/10.1007/978-3-319-91581-4_21).
- Barhoush, Y., Georgiev, G. and Loudon, B. (2020) EMPATHY AND IDEA GENERATION: EXPLORING THE DESIGN OF A VIRTUAL REALITY CONTROLLER FOR REHABILITATION PURPOSES. *Proceedings Of The Sixth International Conference On Design Creativity (ICDC 2020)*, 287–294. <https://www.designsociety.org/publication/43077/EMPATHY+AND+IDEA+GENERATION>.
- Bastiaansen, M., Oosterholt, M., Mitas, O., Han, D. and Lub, X. (2022) An Emotional Roller Coaster: Electrophysiological Evidence of Emotional Engagement during a Roller-Coaster Ride with Virtual Reality Add-On. *J. Hosp. Tour. Res.*, **46**, 29–54. <https://doi.org/http://journals.sagepub.com/doi/10.1177/1096348020944436>.
- Biocca, F., Harms, C. & Burgoon, J. Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria. *Presence*, **12**, 456–480 (2003). <https://direct.mit.edu/pvar/article/12/5/456-480/58921>, <https://doi.org/10.1162/105474603322761270>
- Bisogni, C., Cascone, L., Castiglione, A. & Passero, I. Deep learning for emotion driven user experiences. *Pattern Recognit. Lett.*, **152**, 115–121 (2021), <https://linkinghub.elsevier.com/retrieve/pii/S0167865521003159>, <https://doi.org/10.1016/j.patrec.2021.09.004>
- Bonfert, M., Lemke, S., Porzel, R. and Malaka, R. (2022) Kicking in Virtual Reality: The Influence of Foot Visibility on the Shooting Experience and Accuracy. *2022 IEEE Conference On Virtual Reality And 3D User Interfaces (VR)*, 711–718. <https://ieeexplore.ieee.org/document/9756749/>.
- Bonfert, M., Reinschluessel, A., Putze, S., Lai, Y., Alexandrovsky, D., Malaka, R. and Döring, T. (2023) Seeing the faces is so important—Experiences from online team meetings on commercial virtual reality platforms. *Front. Virtual Real.*, **3**, 945791. <https://doi.org/10.3389/frvir.2022.945791>.
- Bouchard, S., St-Jacques, J., Robillard, G. & Renaud, P. Anxiety Increases the Feeling of Presence in Virtual Reality. *Presence*, **17**, 376–391 (2008), <https://direct.mit.edu/pvar/article/17/4/376-391/18732>
- Breves, P. (2021) Biased by being there: The persuasive impact of spatial presence on cognitive processing. *Comput. Human Behav.*, **119**, 106723. <https://linkinghub.elsevier.com/retrieve/pii/S0747563221000455>.
- Brown, P., Powell, W., Dansey, N., Al-Abbaday, M., Stevens, B. and Powell, V. (2022) Virtual Reality as a Pain Distraction Modality for Experimentally Induced Pain in a Chronic Pain Population: An Exploratory Study. *Cyberpsychol. Behav. Soc. Netw.*, **25**, 66–71. <https://doi.org/https://www.liebertpub.com/doi/10.1089/cyber.2020.0823>.
- Buche, H., Michel, A., Piccoli, C. and Blanc, N. (2021) Contemplating or Acting? Which Immersive Modes Should Be Favored in Virtual Reality During Physiotherapy for Breast Cancer Rehabilitation. *Front. Psychol.*, **12**, 631186. <https://doi.org/https://www.frontiersin.org/articles/10.3389/fpsyg.2021.631186/full>.
- Campelo, A. and Katz, L. (2018) Physical Literacy Promotion in Older Adults Using Active Video Gaming: A Sense of Presence and Attitudes Towards Exercise. *2018 IEEE Games, Entertainment, Media Conference (GEM)*, 163–166. <https://ieeexplore.ieee.org/document/8516546/>.
- Carruth, D. (2017) Virtual reality for education and workforce training. *2017 15th International Conference On Emerging ELearning Technologies And Applications (ICETA)*, 1–6. <http://ieeexplore.ieee.org/document/8102472/>.
- Cavazza, M., Lugrin, J. and Buehner, M. (2007) Causal Perception in Virtual Reality and its Implications for Presence Factors. *Presence*, **16**, 623–642. <https://doi.org/10.1162/pres.16.6.623>. <https://direct.mit.edu/pvar/article/16/6/623-642/18696>.
- Chalil Madathil, K. and Greenstein, J. (2017) An investigation of the efficacy of collaborative virtual reality systems for moderated remote usability testing. *Appl. Ergon.*, **65**, 501–514. <https://doi.org/10.1016/j.apergo.2017.02.011>. <https://linkinghub.elsevier.com/retrieve/pii/S0003687017300376>.
- Chandio, Y., Bashir, N., Interrante, V. and Anwar, F. (2023) Investigating the Correlation Between Presence and Reaction Time in Mixed Reality. *IEEE Trans. Vis. Comput. Graph.*, Early access, 1–13. <https://ieeexplore.ieee.org/document/10264199/>.
- Chessa, M., Maiello, G., Borsari, A. and Bex, P. (2019) The Perceptual Quality of the Oculus Rift for Immersive Virtual Reality. *Hum.-Comput. Interact.*, **34**, 51–82. <https://doi.org/https://www.tandfonline.com/doi/full/10.1080/07370024.2016.1243478>.
- Chicchi Giglioli, I., Bermejo Vidal, C. and Alcañiz Raya, M. (2019) A Virtual Versus an Augmented Reality Cooking Task Based-Tools: A Behavioral and Physiological Study on the Assessment of Executive Functions. *Front. Psychol.*, **10**, 2529. <https://doi.org/https://www.frontiersin.org/article/10.3389/fpsyg.2019.02529/full>.
- Chicchi Giglioli, I., Chirico, A., Cipresso, P., Serino, S., Pedrolì, E., Pallavicini, F. and Riva, G. (2016) Feeling Ghost Food as Real One: Psychometric Assessment of Presence Engagement Exposing to Food in Augmented Reality. *Pervas. Comput. Paradigms Mental Health*, **604**, 99–109. [https://doi.org/http://link.springer.com/10.1007/978-3-319-32270-4\\_10](https://doi.org/http://link.springer.com/10.1007/978-3-319-32270-4_10).
- Chirico, A., Clewis, R., Yaden, D. and Gaggioli, A. (2021) Nature versus art as elicitors of the sublime: A virtual reality study. *PLoS One*, **16**, e0233628. <https://doi.org/https://dx.plos.org/10.1371/journal.pone.0233628>.

- Chirico, A., Ferrise, F., Cordella, L. and Gaggioli, A. (2018) Designing Awe in Virtual Reality: An Experimental Study. *Front. Psychol.*, **8**, 2351. <https://doi.org/http://journal.frontiersin.org/article/10.3389/fpsyg.2017.02351/full>.
- Choi, C., Jun, J., Heo, J. and Kim, K. (2019) Effects of virtual-avatar motion-synchrony levels on full-body interaction. *Proceedings Of The 34th ACM/SIGAPP Symposium On Applied Computing*, 701–708. <https://doi.org/https://dl.acm.org/doi/10.1145/3297280.3297346>.
- Chrysanthakopoulou, A., Kalatzis, K. and Moustakas, K. (2021) Immersive Virtual Reality Experience of Historical Events Using Haptics and Locomotion Simulation. *Appl. Sci.*, **11**, 11613. <https://www.mdpi.com/2076-3417/11/24/11613>.
- Clark, L. and Watson, D. (2019) Constructing validity: New developments in creating objective measuring instruments. *Psychol. Assess.*, **31**, 1412–1427. <https://doi.org/http://doi.apa.org/getdoi.cfm?doi=10.1037/pas0000626>.
- Collins, N., Vaughan, B. and Cullen, C. (2020) Motivation in Situated Immersive Games for Irish Language Learning, a DBR Approach. *ECGBL 2020 14th European Conference On Game-Based Learning*, **96**.
- Cummings, J. and Wertz, E. (2022) Capturing social presence: concept explication through an empirical analysis of social presence measures. *J. Comput.-Mediat. Commun.*, **28**. <https://doi.org/10.1093/jcmc/zmac027>.
- Cummings, J., Tsay-Vogel, M., Cahill, T. and Zhang, L. (2022) Effects of immersive storytelling on affective, cognitive, and associative empathy: The mediating role of presence. *New Media Soc.*, **24**, 2003–2026. <https://doi.org/http://journals.sagepub.com/doi/10.1177/1461444820986816>.
- De Paolis, L. and De Luca, V. (2022) The effects of touchless interaction on usability and sense of presence in a virtual environment. *Virtual Real.*, **26**, 1551–1571. <https://doi.org/https://link.springer.com/10.1007/s10055-022-00647-1>.
- De Paolis, L. and De Luca, V. (2020) The impact of the input interface in a virtual environment: the Vive controller and the Myo armband. *Virtual Real.*, **24**, 483–502. <https://doi.org/http://link.springer.com/10.1007/s10055-019-00409-6>.
- Deb, S., Carruth, D., Sween, R., Strawderman, L. and Garrison, T. (2017) Efficacy of virtual reality in pedestrian safety research. *Appl. Ergon.*, **65**, 449–460. <https://doi.org/10.1016/j.apergo.2017.03.007>. <https://linkinghub.elsevier.com/retrieve/pii/S0003687017300662>.
- Deutsch, J., Myslinski, M., Kafri, M., Ranky, R., Sivak, M., Mavroidis, C. and Lewis, J. (2013) Feasibility of Virtual Reality Augmented Cycling for Health Promotion of People Poststroke. *J. Neurol. Phys. Ther.*, **37**, 118–124. <https://doi.org/10.1097/NPT.0b013e3182a0a078>. <https://journals.lww.com/01253086-201309000-00005>.
- Dienes, Z. (2004) Assumptions of Subjective Measures of Unconscious Mental States: Higher Order Thoughts and Bias. *J. Conscious. Stud.*, **11**, 25–45.
- Difede, J. et al. (2019) Enhanced exposure therapy for combat-related Posttraumatic Stress Disorder (PTSD): Study protocol for a randomized controlled trial. *Contemp. Clin. Trials*, **87**, 105857. <https://doi.org/10.1016/j.cct.2019.105857>. <https://linkinghub.elsevier.com/retrieve/pii/S1551714419305725>.
- Dorado, J. and Figueroa, P. (2015) Methods to reduce cybersickness and enhance presence for in-place navigation techniques. *2015 IEEE Symposium On 3D User Interfaces (3DUI)*, 145–146. <http://ieeexplore.ieee.org/document/7131742/>.
- Dos Santos, S., Da Silva, B. and De Oliveira, J. (2009) Camera control based on rigid body dynamics for virtual environments. *2009 IEEE International Conference On Virtual Environments, Human-Computer Interfaces And Measurements Systems*, 344–349. <https://ieeexplore.ieee.org/document/5068922/>.
- Faas, D., Bao, Q., Frey, D. and Yang, M. (2014) The influence of immersion and presence in early stage engineering designing and building. *Artif. Intell. Eng. Des. Anal. Manuf.*, **28**, 139–151. <https://doi.org/10.1017/S0890060414000055>. [https://www.cambridge.org/core/product/identifier/S0890060414000055/type/journal\\_article](https://www.cambridge.org/core/product/identifier/S0890060414000055/type/journal_article).
- Farmani, Y. and Teather, R. (2018) Viewpoint Snapping to Reduce Cybersickness in Virtual Reality. *Proceedings Of Graphics Interface 2018*, 8 pages, 418.29 KB. <http://graphicsinterface.org/proceedings/gi2018/gi2018-23/>.
- Fauville, G., Queiroz, A., Woolsey, E., Kelly, J. and Bailenson, J. (2021) The effect of water immersion on vection in virtual reality. *Sci. Rep.*, **11**. <https://doi.org/10.1038/s41598-020-80100-y>. <https://www.nature.com/articles/s41598-020-80100-y>.
- Felton, W. and Jackson, R. (2022) Presence: A Review. *Int. J. Hum.-Comput. Interact.*, **38**, 1–18. <https://doi.org/https://www.tandfonline.com/doi/full/10.1080/10447318.2021.1921368>.
- Finn, A. and Kayande, U. (2004) Scale modification: alternative approaches and their consequences. *J. Retail.*, **80**, 37–52. <https://linkinghub.elsevier.com/retrieve/pii/S0022435904000053>.
- Fox, J., Bailenson, J. and Binney, J. V. (2009) Experiences, Physical Behaviors: The Effect of Presence on Imitation of an Eating Avatar. *Presence*, **18**, 294–303. <https://direct.mit.edu/pvar/article/18/4/294-303/18752>.
- Freeman, J., Avons, S., Meddis, R., Pearson, D. and Ijsselstein, W. (2000) Using Behavioral Realism to Estimate Presence: A Study of the Utility of Postural Responses to Motion Stimuli. *Presence*, **9**, 149–164. <https://direct.mit.edu/pvar/article/9/2/149-164/18278>.
- Freeman, J., Avons, S., Pearson, D. and Ijsselstein, W. (1999) Effects of Sensory Information and Prior Experience on Direct Subjective Ratings of Presence. *Presence*, **8**, 1–13. <https://doi.org/10.1162/105474699566017>. <https://direct.mit.edu/pvar/article/8/1/1-13/18209>.
- Garau, M., Friedman, D., Widenfeld, H., Antley, A., Brogni, A. and Slater, M. (2008) Temporal and Spatial Variations in Presence: Qualitative Analysis of Interviews from an Experiment on Breaks in Presence. *Presence*, **17**, 293–309. <https://direct.mit.edu/pvar/article/17/3/293-309/18718>.
- Gonzalez-Morin, D., Gonzalez-Sosa, E., Perez-Garcia, P. and Villegas, A. (2022) Bringing Real Body as Self-Avatar into Mixed Reality: A Gamified Volcano Experience. *2022 IEEE Conference On Virtual Reality And 3D User Interfaces Abstracts And Workshops (VRW)*, 794–795. <https://ieeexplore.ieee.org/document/9757562/>.
- Gorini, A., Griez, E., Petrova, A. and Riva, G. (2010) Assessment of the emotional responses produced by exposure to real food, virtual food and photographs of food in patients affected by eating disorders. *Ann. Gen. Psychiatry*, **9**. <https://doi.org/https://annals-general-psychiatry.biomedcentral.com/articles/10.1186/1744-859X-9-30>.
- Graf, S. and Schwind, V. (2020) Inconsistencies of Presence Questionnaires in Virtual Reality. *26th ACM Symposium On Virtual Reality Software And Technology*, 1–3. <https://doi.org/https://dl.acm.org/doi/10.1145/3385956.3422105>.
- Grassini, S. and Laumann, K. (2020) Questionnaire Measures and Physiological Correlates of Presence: A Systematic Review. *Front. Psychol.*, **11**, 1–21. <https://doi.org/https://www.frontiersin.org/articles/10.3389/fpsyg.2020.00349>.
- Grewe, P., Kohsik, A., Flentge, D., Dyck, E., Botsch, M., Winter, Y., Markowitsch, H., Bien, C. and Piefke, M. (2013) Learning real-life cognitive abilities in a novel 360-virtual reality supermarket: a neuropsychological study of healthy participants and patients

- with epilepsy. *J. Neuroeng. Rehabil.*, **10**, 42. <https://doi.org/http://neuroengrehab.biomedcentral.com/articles/10.1186/1743-0003-10-42>.
- Griol, D., Molina, J. and Callejas, Z. (2014) An approach to develop intelligent learning environments by means of immersive virtual worlds. *J. Ambient Intell. Smart Environ.*, **6**, 237–255. <https://doi.org/https://www.medra.org/servelet/aliasResolver?alias=iospress&doi=10.3233/AIS-140255>.
- Günther, S., Schön, D., Müller, F., Mühlhäuser, M. and Schmitz, M. (2020) PnemoVolley: Pressure-based Haptic Feedback on the Head through Pneumatic Actuation. *Extended Abstracts Of The 2020 CHI Conference On Human Factors In Computing Systems*, 1–10. <https://doi.org/https://dl.acm.org/doi/10.1145/3334480.3382916>.
- Hameed, A. and Perkis, A. (2020) Affects of Perceived-Actions within Virtual Environments on User Behavior on the Outside. *2020 Twelfth International Conference On Quality Of Multimedia Experience (QoMEX)*, 1–6. <https://ieeexplore.ieee.org/document/9123099/>.
- Herrera, F. and Bailenson, J. (2021) Virtual reality perspective-taking at scale: Effect of avatar representation, choice, and head movement on prosocial behaviors. *New Media Soc.*, **23**, 2189–2209. <https://doi.org/http://journals.sagepub.com/doi/10.1177/1461444821993121>.
- Hooks, K., Ferguson, W., Morillo, P. and Cruz-Neira, C. (2020) Evaluating the user experience of omnidirectional VR walking simulators. *Entertain. Comput.*, **34**, 100352. <https://doi.org/10.1016/j.entcom.2020.100352>. <https://linkinghub.elsevier.com/retrieve/pii/S1875952119301284>.
- Hoppe, M., Knierim, P., Kosch, T., Funk, M., Futami, L., Schneegass, S., Henze, N., Schmidt, A. and Machulla, T. (2018) VRHapticDrones: Providing Haptics in Virtual Reality through Quadcopters. *Proceedings Of The 17th International Conference On Mobile And Ubiquitous Multimedia*, 7–18. <https://doi.org/https://dl.acm.org/doi/10.1145/3282894.3282898>.
- Ijsselstein, W., Ridder, H., Freeman, J., Avons, S. and Bouwhuis, D. (2001) Effects of Stereoscopic Presentation, Image Motion, and Screen Size on Subjective and Objective Corroborative Measures of Presence. *Presence*, **10**, 298–311. <https://direct.mit.edu/pvar/article/10/3/298-311/18308>.
- Jacob, Habgood, M., Moore, D., Wilson, D. and Alapont, S. (2018) Rapid, Continuous Movement Between Nodes as an Accessible Virtual Reality Locomotion Technique. *2018 IEEE Conference On Virtual Reality And 3D User Interfaces (VR)*, 371–378. <https://ieeexplore.ieee.org/document/8446130/>.
- Jang, Y. and Hsieh, P. (2021) Understanding consumer behavior in the multimedia context: incorporating gamification in VR-enhanced web system for tourism e-commerce. *Multimed. Tools Appl.*, **80**, 29339–29365. <https://doi.org/https://link.springer.com/10.1007/s11042-021-11149-8>.
- Johns, C., Nuñez, D., Daya, M., Sellars, D., Casanueva, J. and Blake, E. (2000) The Interaction Between Individuals' Immersive Tendencies and the Sensation of Presence in a Virtual Environment. *Virtual Environments 2000*, 65–74. [https://doi.org/http://link.springer.com/10.1007/978-3-7091-6785-4\\_8](https://doi.org/http://link.springer.com/10.1007/978-3-7091-6785-4_8).
- Juniper, E. (2009) Validated questionnaires should not be modified. *Eur. Respir. J.*, **34**, 1015–1017. <https://doi.org/http://erj.ersjournals.com/cgi/doi/10.1183/09031936.00110209>.
- Kern, A., Ellermeier, W. and Wojtusich, J. (2016) Noise-cancelling, steps and soundscapes: the effect of auditory stimulation on presence in virtual realities while walking. *Proceedings Of The 22nd ACM Conference On Virtual Reality Software And Technology*, 87–90. <https://doi.org/https://dl.acm.org/doi/10.1145/2993369.2993398>.
- Khenak, N., Vezien, J., Thery, D. and Bourdot, P. (2019) Spatial Presence in Real and Remote Immersive Environments. *2019 IEEE Conference On Virtual Reality And 3D User Interfaces (VR)*, 1016–1017. <https://ieeexplore.ieee.org/document/8797801/>.
- Kim, K., Kim, C., Cha, K., Park, J., Han, K., Kim, Y., Kim, J., Kim, I. and Kim, S. (2008) Anxiety Provocation and Measurement Using Virtual Reality in Patients with Obsessive-Compulsive Disorder. *Cyberpsychol. Behav.*, **11**, 637–641. <https://doi.org/http://www.liebertpub.com/doi/10.1089/cpb.2008.0003>.
- Kim, H., Oh, Y., Lee, Y., Jung, S., Hwang, H. and Park, J. (2020) Using a VR Training-Evaluation Tool to Evaluate a VR-Simulation Assembly Process. *ICIC International*. <https://doi.org/10.24507/icicelb.11.02.159>.
- Kirya, M., Debattista, K. and Chalmers, A. (2023) Using virtual environments to facilitate refugee integration in third countries. *Virtual Real.*, **27**, 97–107. <https://doi.org/https://link.springer.com/10.1007/s10055-022-00659-x>.
- Kober, S., Kurzmann, J. and Neuper, C. (2012) Cortical correlate of spatial presence in 2D and 3D interactive virtual reality: An EEG study. *Int. J. Psychophysiol.*, **83**, 365–374. <https://doi.org/10.1016/j.ijpsycho.2011.12.003>. <https://linkinghub.elsevier.com/retrieve/pii/S0167876011003825>.
- Kreijns, K., Xu, K. and Weidlich, J. (2022) Social Presence: Conceptualization and Measurement. *Educ. Psychol. Rev.*, **34**, 139–170. <https://doi.org/https://link.springer.com/10.1007/s10648-021-09623-8>.
- Kukshinov, E. (2024) Presence, (re)focused and explicated. *The Communication Review*, [Online first], 1–22. <https://doi.org/10.1080/10714421.2024.2366684>.
- Laarni, J., Ravaja, N., Saari, T., Böcking, S., Hartmann, T. and Schramm, H. (2015) Ways to Measure Spatial Presence: Review and Future Directions. *Immersed In Media: Telepresence Theory, Measurement & Technology*, 139–185. [https://doi.org/10.1007/978-3-319-10190-3\\_8](https://doi.org/10.1007/978-3-319-10190-3_8).
- Lee, S. and Lee, M. (2022) Can Haptic Feedback on One Virtual Object Increase the Presence of Another Virtual Object? *Proceedings Of The 28th ACM Symposium On Virtual Reality Software And Technology*, 1–2. <https://doi.org/https://dl.acm.org/doi/10.1145/3562939.3565658>.
- Lee, C., Rincon, G., Meyer, G., Hollerer, T. and Bowman, D. (2013) The Effects of Visual Realism on Search Tasks in Mixed Reality Simulation. *IEEE Trans. Vis. Comput. Graph.*, **19**, 547–556. <http://ieeexplore.ieee.org/document/6479181/>.
- Lessiter, J., Freeman, J., Keogh, E. and Davidoff, J. (2001) A Cross-Media Presence Questionnaire: The ITC-Sense of Presence Inventory. *Presence*, **10**, 282–297. <https://direct.mit.edu/pvar/article/10/3/282-297/18311>.
- Li, Y., Chn'g, E., Cobb, S. and See, S. (2021a) Presence and Communication in Hybrid Virtual and Augmented Reality Environments. *Presence*, 29–52. [https://doi.org/https://direct.mit.edu/pvar/article/doi/10.1162/pres\\_a\\_00340/107509/Presence-and-Communication-in-Hybrid-Virtual-and](https://doi.org/https://direct.mit.edu/pvar/article/doi/10.1162/pres_a_00340/107509/Presence-and-Communication-in-Hybrid-Virtual-and).
- Li, S., Cui, J., Hao, A., Zhang, S. and Zhao, Q. (2021b) Design and Evaluation of Personalized Percutaneous Coronary Intervention Surgery Simulation System. *IEEE Trans. Vis. Comput. Graph.*, **27**, 4150–4160. <https://doi.org/10.1109/TVCG.2021.3106478>. <https://ieeexplore.ieee.org/document/9523834/>.
- Lombard, M., Biocca, F., Freeman, J., Ijsselstein, W. & Schaevitz, R. *Immersed in media: Telepresence theory, measurement & technology*. Springer International Publishing, 2015, <https://doi.org/http://link.springer.com/10.1007/978-3-319-10190-3>
- Lorenz, M., Brade, J., Diamond, L., Sjölie, D., Busch, M., Tscheligi, M., Klimant, P., Heyde, C. and Hammer, N. (2018) Presence and User Experience in a Virtual Environment under the Influence of Ethanol: An Explorative Study. *Sci. Rep.*, **8**,

6407. <https://doi.org/10.1038/s41598-018-24453-5>. <https://www.nature.com/articles/s41598-018-24453-5>.
- Lorenz, M., Neupetsch, C., Rotsch, C., Klimant, P. and Hammer, N. (2019) Early Virtual Reality User Experience and Usability Assessment of a Surgical Shape Memory Alloy Aspiration/Irrigation Instrument. *2019 IEEE Conference On Virtual Reality And 3D User Interfaces (VR)*, 1056–1057. <https://ieeexplore.ieee.org/document/8798144/>.
- Lull, R. and Bushman, B. (2016) Immersed in violence: Presence mediates the effect of 3D violent video gameplay on angry feelings. *Psychol. Pop. Media Cult.*, **5**, 133–144. <https://doi.org/https://doi.apa.org/doi/10.1037/ppm0000062>.
- Lyu, L., Xu, Q., Lan, C., Shi, Q., Lu, W., Zhou, Y. and Zhao, Y. (2018) Sino-InSpace: A Digital Simulation Platform for Virtual Space Environments. *ISPRS Int. J. Geo-Inf.*, **7**, 373. <http://www.mdpi.com/2220-9964/7/9/373>.
- Ma, X., Cackett, M., Park, L., Chien, E. and Naaman, M. (2018) Web-Based VR Experiments Powered by the Crowd. *Proceedings Of The 2018 World Wide Web Conference On World Wide Web - WWW '18*, 33–43. <http://dl.acm.org/citation.cfm?doi=3178876.3186034>.
- Madden, J., Pandita, S., Schuldt, J., Kim, B., Won, A. S. and Holmes, N. (2020) Ready student one: Exploring the predictors of student learning in virtual reality. *PLoS One*, **15**, e0229788. <https://doi.org/https://dx.plos.org/10.1371/journal.pone.0229788>.
- Mashiba, Y., Salih, H., Wakatsuki, N., Mizutani, K. and Zempo, K. (2022) [Paper] Image Sonification System Based on Auditory Icons and Stereophonic Presentation Considering Image Scenes and Human Action. *ITE Trans. Media Technol. Appl.*, **10**, 216–224. <https://doi.org/10.3169/mta.10.216>. [https://www.jstage.jst.go.jp/article/mta/10/4/10\\_216/\\_article](https://www.jstage.jst.go.jp/article/mta/10/4/10_216/_article).
- Minsky, M. (1980) Telepresence. *OMNI Mag.*, **2**, 45–51. <https://cir.nii.ac.jp/crid/1570854175773549312>.
- Mondellini, M., Arlati, S., Gapeyeva, H., Lees, K., Märitz, I., Pizzagalli, S., Otto, T., Sacco, M. and Teder-Braschinsky, A. (2022) User Experience during an Immersive Virtual Reality-Based Cognitive Task: A Comparison between Estonian and Italian Older Adults with MCI. *Sensors*, **22**, 8249. <https://www.mdpi.com/1424-8220/22/21/8249>.
- Nannipieri, O. (2022) Do Presence Questionnaires Actually Measure Presence? A Content Analysis of Presence Measurement Scales. *Extend. Real.*, **13445**, 273–295. [https://doi.org/https://link.springer.com/10.1007/978-3-031-15546-8\\_24](https://doi.org/https://link.springer.com/10.1007/978-3-031-15546-8_24).
- Nicovich, S. (2010) The effect of involvement on ad judgement in a computermediated environment: The mediating role of presence. *Int. J. Advert.*, **29**, 597–620. <https://doi.org/https://www.tandfonline.com/doi/full/10.2501/S0265048710201361>.
- Nisbett, R. and Wilson, T. (1977) The halo effect: Evidence for unconscious alteration of judgments. *J. Pers. Soc. Psychol.*, **35**, 250–256. <https://doi.org/http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.35.4.250>.
- Nowak, K. and Biocca, F. (2003) The Effect of the Agency and Anthropomorphism on Users' Sense of Telepresence, Copresence, and Social Presence in Virtual Environments. *Presence*, **12**, 481–494. <https://direct.mit.edu/pvar/article/12/5/481-494/58920>.
- Nunez, D. and Blake, E. (2003) A direct comparison of presence levels in text-based and graphics-based virtual environments. *Proceedings Of The 2nd International Conference On Computer Graphics, Virtual Reality, Visualisation And Interaction In Africa*, 53–56. <https://doi.org/https://dl.acm.org/doi/10.1145/602330.602340>.
- Nunez, E., Hirokawa, M., Hautasaari, A. and Suzuki, K. (2022) Remote Communication via Huggable Interfaces - Behavior Synchronization and Social Presence. *CHI Conference On Human Factors In Computing Systems Extended Abstracts*, 1–7. <https://doi.org/https://dl.acm.org/doi/10.1145/3491101.3519699>.
- Oh, S., Park, J. and Cho, S. (2022) Effectiveness of the VR Cognitive Training for Symptom Relief in Patients with ADHD. *J. Web Eng.*, **21**, 767–788. <https://doi.org/10.13052/jwe1540-9589.21310>. <https://journals.riverpublishers.com/index.php/JWE/article/view/11997>.
- Oh, J., Robinson, H. and Lee, J. (2013) Page flipping vs. clicking: The impact of naturally mapped interaction technique on user learning and attitudes. *Comput. Hum. Behav.*, **29**, 1334–1341. <https://doi.org/10.1016/j.chb.2013.01.011>. <https://linkinghub.elsevier.com/retrieve/pii/S0747563213000149>.
- Orduna, M., Perez, P., Diaz, C. and Garcia, N. (2020) Evaluating the Influence of the HMD, Usability, and Fatigue in 360VR Video Quality Assessments. *2020 IEEE Conference On Virtual Reality And 3D User Interfaces Abstracts And Workshops (VRW)*, 682–683. <https://ieeexplore.ieee.org/document/9090529/>.
- Page, M. et al. (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, **372**, n71. <https://doi.org/10.1136/bmj.n71>. <https://www.bmj.com/content/372/bmj.n71>.
- Pai, Y. and Kunze, K. (2017) Armswing: using arm swings for accessible and immersive navigation in AR/VR spaces. *Proceedings Of The 16th International Conference On Mobile And Ubiquitous Multimedia*, 198, 189. <https://doi.org/https://dl.acm.org/doi/10.1145/3152832.3152864>.
- Peiris, R., Chan, L. and Minamizawa, K. (2018) LiquidReality: Wetness Sensations on the Face for Virtual Reality. *Haptics Sci. Technol. Appl.*, **10894**, 366–378. [https://doi.org/http://link.springer.com/10.1007/978-3-319-93399-3\\_32](https://doi.org/http://link.springer.com/10.1007/978-3-319-93399-3_32).
- Perpiñá, C., Roncero, M., Fernández-Aranda, F., Jiménez-Murcia, S., Forcano, L. and Sánchez, I. (2013) Clinical validation of a virtual environment for normalizing eating patterns in eating disorders. *Compr. Psychiatry*, **54**, 680–686. <https://doi.org/10.1016/j.comppsy.2013.01.007>. <https://linkinghub.elsevier.com/retrieve/pii/S0010440X13000345>.
- Perrig, S., Scharowski, N. and Brühlmann, F. (2023) Trust Issues with Trust Scales: Examining the Psychometric Quality of Trust Measures in the Context of AI. *Extended Abstracts Of The 2023 CHI Conference On Human Factors In Computing Systems*, 1–7. <https://doi.org/https://dl.acm.org/doi/10.1145/3544549.3585808>.
- Piccione, J., Collett, J. and De Foe, A. (2019) Virtual skills training: the role of presence and agency. *Heliyon*, **5**, e02583. <https://linkinghub.elsevier.com/retrieve/pii/S2405844019362437>.
- Piercey, C., Charlton, K. and Callewaert, C. (2012) Reducing Anxiety Using Self-Help Virtual Reality Cognitive Behavioral Therapy. *Games Health J.*, **1**, 124–128. <https://doi.org/http://www.liebertpub.com/doi/10.1089/g4h.2012.0008>.
- Pisoni, G., Daniel, F., Casati, F., Callaway, C. and Stock, O. (2019) Interactive Remote Museum Visits for Older Adults: An Evaluation of Feelings of Presence, Social Closeness, Engagement, and Enjoyment in an Social Visit. *2019 IEEE International Symposium On Multimedia (ISM)*, 99–993. <https://ieeexplore.ieee.org/document/8959064/>.
- Podkosova, I. and Kaufmann, H. (2018) Mutual collision avoidance during walking in real and collaborative virtual environments. *Proceedings Of The ACM SIGGRAPH Symposium On Interactive 3D Graphics And Games*, 1–9. <https://doi.org/https://dl.acm.org/doi/10.1145/3190834.3190845>.
- Pontonnier, C., Samani, A., Badawi, M., Madeleine, P. and Dumont, G. (2014) Assessing the Ability of a VR-Based Assembly Task Simulation to Evaluate Physical Risk Factors. *IEEE Trans. Vis. Comput. Graph.*, **20**, 664–674. <https://doi.org/10.1109/TVCG.2013.252>. <http://ieeexplore.ieee.org/document/6654129/>.

- Price, M. and Anderson, P. (2007) The role of presence in virtual reality exposure therapy. *J. Anxiety Disord.*, **21**, 742–751. <https://doi.org/10.1016/j.janxdis.2006.11.002>. <https://linkinghub.elsevier.com/retrieve/pii/S088761850600171X>.
- Ranasinghe, N., Jain, P., Karwita, S., Tolley, D. and Do, E. (2017) Ambiotherm: Enhancing Sense of Presence in Virtual Reality by Simulating Real-World Environmental Conditions. *Proceedings Of The 2017 CHI Conference On Human Factors In Computing Systems*, 1731–1742. <https://doi.org/https://dl.acm.org/doi/10.1145/3025453.3025723>.
- Rietzler, M., Geiselhart, F., Gugenheimer, J. and Rukzio, E. (2018) Breaking the Tracking: Enabling Weight Perception using Perceivable Tracking Offsets. *Proceedings Of The 2018 CHI Conference On Human Factors In Computing Systems*, 1–12. <https://doi.org/https://dl.acm.org/doi/10.1145/3173574.3173702>.
- Ritter, K. and Chambers, T. (2022) Three-dimensional modeled environments versus 360 degree panoramas for mobile virtual reality training. *Virtual Real.*, **26**, 571–581. <https://doi.org/https://link.springer.com/10.1007/s10055-021-00502-9>.
- Riva, G., Mantovani, F., Capideville, C., Preziosa, A., Morganti, F., Villani, D., Gaggioli, A., Botella, C. and Alcañiz, M. (2007) Affective Interactions Using Virtual Reality: The Link between Presence and Emotions. *Cyberpsychol Behav.*, **10**, 45–56. <https://doi.org/http://www.liebertpub.com/doi/10.1089/cpb.2006.9993>.
- Riva, G., Waterworth, J. and Murray, D. (2014) Interacting with Presence: HCI and the Sense of Presence in Computer-mediated Environments. *DE GRUYTER OPEN*. <https://doi.org/https://www.degruyter.com/document/doi/10.2478/9783110409697/html>.
- Robillard, G., Bouchard, S., Renaud, P. and Cournoyer, L. (2002) Validation canadienne-française de deux mesures importantes en réalité virtuelle: l'Immersive Tendencies Questionnaire et le Presence Questionnaire. Communication par affiche présentée au 25e congrès de la Société Québécoise pour la Recherche en Psychologie (SQRP), Trois-Rivières, 1-3 novembre. *Communication Par Affiche Présentée*.
- Rothe, S., Tran, K. and Hußmann, H. (2018) Dynamic Subtitles in Cinematic Virtual Reality. *Proceedings Of The 2018 ACM International Conference On Interactive Experiences For TV And Online Video*, 209–214. <https://doi.org/https://dl.acm.org/doi/10.1145/3210825.3213556>.
- Schrader, C. and Bastiaens, T. (2012) The influence of virtual presence: Effects on experienced cognitive load and learning outcomes in educational computer games. *Comput. Human Behav.*, **28**, 648–658. <https://doi.org/10.1016/j.chb.2011.11.011>. <https://linkinghub.elsevier.com/retrieve/pii/S074756321100255X>.
- Schubert, T., Friedmann, F. and Regenbrecht, H. (2001) The Experience of Presence: Factor Analytic Insights. *Presence*, **10**, 266–281. <https://doi.org/10.1162/105474601300343603>. <https://direct.mit.edu/pvar/article/10/3/266-281/18310>.
- Shafer, D., Carbonara, C. and Korpi, M. (2019) Factors Affecting Enjoyment of Virtual Reality Games: A Comparison Involving Consumer-Grade Virtual Reality Technology. *Games Health J.*, **8**, 15–23. <https://doi.org/https://www.liebertpub.com/doi/10.1089/g4h.2017.0190>.
- Sharma, S. and Ali, S. (2022) Multi-agent Crowd Simulation in an Active Shooter Environment. *Virt. Augment. Mix. Real. Appl. Educ. Aviat. Ind.*, **13318**, 108–120. [https://doi.org/https://link.springer.com/10.1007/978-3-031-06015-1\\_8](https://doi.org/https://link.springer.com/10.1007/978-3-031-06015-1_8).
- Shin, Y., Kim, J., Kim, M., Kyeong, S., Jung, Y., Eom, H. and Kim, E. (2018) Development of an effective virtual environment in eliciting craving in adolescents and young adults with internet gaming disorder. *PLoS One*, **13**. <https://doi.org/https://dx.plos.org/10.1371/journal.pone.0195677>.
- Skarbez, R., Brooks, F. and Whitton, M. (2021) Immersion and Coherence: Research Agenda and Early Results. *IEEE Trans. Vis. Comput. Graph.*, **27**, 3839–3850. <https://doi.org/10.1109/TVCG.2020.2983701>. <https://ieeexplore.ieee.org/document/9055149/>.
- Skarbez, R., Brooks, F. and Whitton, M. (2018) A Survey of Presence and Related Concepts. *ACM Comput. Surv.*, **50**, 1–39. <https://doi.org/https://dl.acm.org/doi/10.1145/3134301>.
- Slater, M., Guger, C., Edlinger, G., Leeb, R., Pfurtscheller, G., Antley, A., Garau, M., Brogni, A. and Friedman, D. (2006) Analysis of Physiological Responses to a Social Situation in an Immersive Virtual Environment. *Presence*, **15**, 553–569. <https://direct.mit.edu/pvar/article/15/5/553-569/18637>.
- Slater, M. (2004) How Colorful Was Your Day? Why Questionnaires Cannot Assess Presence in Virtual Environments. *Presence*, **13**, 484–493. <https://direct.mit.edu/pvar/article/13/4/484-493/18524>.
- Smolentsev, A., Cornick, J. and Blascovich, J. (2017) Using a preamble to increase presence in digital virtual environments. *Virtual Real.*, **21**, 153–164. <https://doi.org/http://link.springer.com/10.1007/s10055-017-0305-4>.
- Sylaïou, S., Mania, K., Karoulis, A. and White, M. (2010) Exploring the relationship between presence and enjoyment in a virtual museum. *Int. J. Hum.-Comput. Stud.*, **68**, 243–253. <https://doi.org/10.1016/j.ijhcs.2009.11.002>. <https://linkinghub.elsevier.com/retrieve/pii/S1071581909001761>.
- Tcha-Tokey, K., Christmann, O., Loup-Escande, E., Loup, G. and Richir, S. (2018) Towards a Model of User Experience in Immersive Virtual Environments. *Adu. Hum.-Comput. Interact.*, **2018**, 1–10. <https://www.hindawi.com/journals/ahci/2018/7827286/>.
- Tuena, C., Serino, S., Maestri, S., Pedroli, E., Stramba-Badiale, C., Brizzi, G., Goulene, K., Cipresso, P., Stramba-Badiale, M. and Riva, G. (2022) Usability of an embodied CAVE system for spatial navigation rehabilitation in mild cognitive impairment. *2022 IEEE International Conference On Metrology For Extended Reality, Artificial Intelligence And Neural Engineering (MetroXRaine)*, 205–210. <https://ieeexplore.ieee.org/document/9967623/>.
- Turner, P. and Turner, S. P. (2006) Sense of Place, and Presence. *Presence*, **15**, 204–217. <https://direct.mit.edu/pvar/article/15/2/204-217/18648>.
- Usoh, M., Catena, E., Arman, S. and Slater, M. (2000) Using Presence Questionnaires in Reality. *Presence*, **9**, 497–503. <https://direct.mit.edu/pvar/article/9/5/497-503/18368>.
- Van De Laar, B., Gurkok, H., Plass-Oude Bos, D., Poel, M. and Nijholt, A. (2013) Experiencing BCI Control in a Popular Computer Game. *IEEE Trans. Comput. Intell. AI Games*, **5**, 176–184. <http://ieeexplore.ieee.org/document/6484110/>.
- Vogt, T., Herpers, R., Askew, C., Scherfgen, D., Strüder, H. and Schneider, S. (2015) Effects of Exercise in Immersive Virtual Environments on Cortical Neural Oscillations and Mental State. *Neural Plast.*, **2015**, 1–9. <https://doi.org/10.1155/2015/523250>. <http://www.hindawi.com/journals/np/2015/523250/>.
- Wirth, M., Gradl, S., Poimann, D., Schaeffe, H., Matlok, J., Koerger, H. and Eskofier, B. (2018) Assessment of Perceptual-Cognitive Abilities among Athletes in Virtual Environments: Exploring Interaction Concepts for Soccer Players. *Proceedings Of The 2018 Designing Interactive Systems Conference*, 1013–1023. <https://doi.org/https://dl.acm.org/doi/10.1145/3196709.3196780>.
- Wissmath, B., Weibel, D. & Stricker, D. When and how to assess subjective overall judgments of presence?. *Wissmath, Bartholomäus; Weibel, David; Stricker, D. (2008). When And How To Assess Subjective Overall Judgments Of Presence? In: Spagnolli, A., Gamberini, L. (eds.) Proceedings Of The 11th Annual International Workshop On Presence, Padova, 16-18 October 2008 (pp. 238–243). Padova: CLEUP. pp. 238–243 (2008), https://boris.unibe.ch/33931/*

- Witmer, B. and Singer, M. (1998) Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence*, **7**, 225–240. <https://direct.mit.edu/pvar/article/7/3/225-240/92643>.
- Witmer, B., Jerome, C. and Singer, M. (2005) The Factor Structure of the Presence Questionnaire. *Presence*, **14**, 298. <https://doi.org/10.1162/105474605323384654>. <https://direct.mit.edu/pvar/article/14/3/298-312/18566>.
- Witmer, B., Sadowski, W. and Finkelstein, N. (2002) VE-Based Training Strategies for Acquiring Survey Knowledge. *Presence*, **11**, 1–18. <https://doi.org/10.1162/105474602317343622>. <https://direct.mit.edu/pvar/article/11/1/1-18/18330>.
- Ye, Y., Wong, S., Li, Y. and Lau, Y. (2020) Risks to pedestrians in traffic systems with unfamiliar driving rules: a virtual reality approach. *Accid. Anal. Prev.*, **142**, 105565. <https://doi.org/10.1016/j.aap.2020.105565>. <https://linkinghub.elsevier.com/retrieve/pii/S0001457519315404>.
- Yildirim, Ç., Bostan, B. and Berkman, M. (2019) Impact of different immersive techniques on the perceived sense of presence measured via subjective scales. *Entertain. Comput.*, **31**, 100308. <https://doi.org/10.1016/j.entcom.2019.100308>. <https://linkinghub.elsevier.com/retrieve/pii/S187595211830096X>.
- Zhang, S., Feng, X. and Shen, Y. (2021) Quantifying Auditory Presence Using Electroencephalography. *Appl. Sci.*, **11**, 10461. <https://www.mdpi.com/2076-3417/11/21/10461>.
- Zhang, K., Kim, L., Guo, Y. and Follmer, S. (2020) Automatic Generation of Spatial Tactile Effects by Analyzing Cross-modality Features of a Video. *Symposium On Spatial User Interaction*, 1–10. <https://doi.org/https://dl.acm.org/doi/10.1145/3385959.3418459>.
- e) To what degree were you able to anticipate what would happen next in response to the actions that you performed?
  - f) How completely were you able to actively survey or search the environment using vision?
  - g) How compelling was your sense of moving around inside the virtual environment?
  - h) How involved were you in the virtual environment experience?
    - i) How much delay did you experience between your actions and expected outcomes?
    - j) How quickly did you adjust to the virtual environment experience?
  - k) How proficient in moving and interacting with the virtual environment did you feel at the end of the experience?
    - l) Were you involved in the experimental task to the extent that you lost track of time?
- 2) Natural Sub-scale
    - a) How natural did your interactions with the environment seem?
    - b) How natural was the mechanism which controlled movement through the environment?
    - c) To what degree did your experiences in the virtual environment seem consistent with your real world experiences?
  - 3) Auditory Sub-scale
    - a) How much did the auditory aspects of the environment involve you?
    - b) How well could you identify sounds?
    - c) How well could you localize sounds?
  - 4) Haptic Sub-scale
    - a) To what extent were you able to actively survey or search the virtual environment using touch?
    - b) Were you able to move or manipulate objects in the virtual environment?
  - 5) Resolution Sub-scale
    - a) How closely were you able to examine objects?
    - b) Were you able to examine objects from multiple viewpoints?
  - 6) Interface Quality Sub-scale
    - a) To what extent did the visual display quality interfere or distract you from performing assigned tasks or required activities?
    - b) To what extent did the control devices interfere with the performance of assigned tasks or with other activities?
    - c) How well could you concentrate on the assigned tasks or required activities rather than on the mechanisms used to perform those tasks or activities?
  - 7) Undetermined
    - a) How completely were all of your senses engaged?
    - b) How aware were you of events occurring in the real world around you?
    - c) How aware were you of your display and control devices?
    - d) To what degree did you experience disconnects or inconsistencies between the information coming from your various senses?
    - e) To what degree did you feel confused or disoriented at the beginning of breaks or at the end of the experimental session?
    - f) Was the control mechanism distracting?

## APPENDIX

### Excluded from search subject areas

Mathematics, dentistry, business, economics and natural sciences, such as physics, astronomy, materials science, environmental science, agricultural and biological sciences, biochemistry, genetics and molecular biology, energy, chemistry, earth and planetary sciences, chemistry and pharmacology, toxicology, pharmaceuticals, immunology and microbiology.

### Three Scales items

The three scales we employed are those developed by Witmer and Singer (Witmer & Singer, 1998), Nowak and Biocca (Nowak & Biocca, 2003) and Lessiter et al. (Lessiter et al., 2001). We include the items and/or subscales within those scales used in our analysis (original, not modified/adapted questionnaires in English).

### Witmer and Singer, 1998 (Witmer & Singer, 1998)

This questionnaire is administered on a 7-point Likert scale (1 = not at all, 7 = very much).

- 1) Involvement/Control Sub-scale
  - a) To what degree do you feel that you were able to control events?
  - b) How responsive was the environment to actions that you initiated (or performed)?
  - c) How much did the visual aspects of the environment involve you?
  - d) How compelling was your sense of objects moving through space?

- g) Did you learn new techniques that enabled you to improve your performance?

### Nowak and Biocca, 2003 (Nowak & Biocca, 2003)

This questionnaire is administered on 5-point Likert scale (1 = not at all, 5 = very much)

#### Self-reported copresence

- 1) I did not want a deeper relationship with my interaction partner.
- 2) I wanted to maintain a sense of distance between us.
- 3) I was unwilling to share personal information with my interaction partner.
- 4) I wanted to make the conversation more intimate.
- 5) I tried to create a sense of closeness between us.
- 6) I was interested in talking to my interaction partner.

#### Perceived Other's copresence

- 1) My interaction partner was intensely involved in our interaction.
- 2) My interaction partner seemed to and our interaction stimulating.
- 3) My interaction partner communicated coldness rather than warmth.
- 4) My interaction partner created a sense of distance between us.
- 5) My interaction partner seemed detached during our interaction.
- 6) My interaction partner was unwilling to share personal information with me.
- 7) My interaction partner made our conversation seem intimate.
- 8) My interaction partner created a sense of distance between us.
- 9) My interaction partner created a sense of closeness between us.
- 10) My interaction partner acted bored by our conversation.

- 11) My interaction partner was interested in talking to me.
- 12) My interaction partner showed enthusiasm while talking to me.

#### Telepresence scale

- 1) How involving was the experience?
- 2) How intense was the experience?
- 3) To what extent did you feel like you were inside the environment you saw/heard?
- 4) To what extent did you feel immersed in the environment you saw/heard?
- 5) To what extent did you feel surrounded by the environment you saw/heard?

#### Social presence

- 1) To what extent did you feel able to assess your partner's reactions to what you said?—Able to assess reactions, not able to assess reactions.
- 2) To what extent was this like a face-to-face meeting? —A lot like face to face, not like face to face at all.
- 3) Sliding scale To what extent was this like you were in the same room with your partner?—A lot like being in the same room, not like being in the same room at all.
- 4) To what extent did your partner seem 'real'?—Very real, not real at all.
- 5) How likely is it that you would choose to use this system of interaction for a meeting in which you wanted to persuade others of something? — Very likely, not likely at all.
- 6) To what extent did you feel you could get to know someone that you met only through this system? —Very well, not at all.

#### Lessiter et al., 2001 (Lessiter et al., 2001)

ITC SOPI is not publicly available, and we cannot distribute its items due to the copyright. Hence we report the subscales of ITC SOPI:

- Sense of Physical Space (19 items)
- Engagement (13 items)
- Ecological Validity / Naturalness (5 items)
- Negative effects (6 items)