

The Fidelity-based Presence Scale (FPS): Modeling the Effects of Fidelity on Sense of Presence

Jacob Belga
University of Central Florida
Orlando, Florida, USA
jacob.belga@ucf.edu

Richard Skarbez
Department of Computer Science &
IT
La Trobe University
Melbourne, VIC, Australia
r.skarbez@latrobe.edu.au

Yahya Hmaiti
University of Central Florida
Orlando, Florida, USA
yohan.hmaiti@ucf.edu

Eric J Chen
Seminole High School
Sanford, Florida, USA
ezhu2006@yahoo.com

Ryan P. McMahan
Computer Science
Virginia Tech
Blacksburg, Virginia, USA
rpm@vt.edu

Joseph J. LaViola Jr.
University of Central Florida
Orlando, Florida, USA
jlaviola@ucf.edu

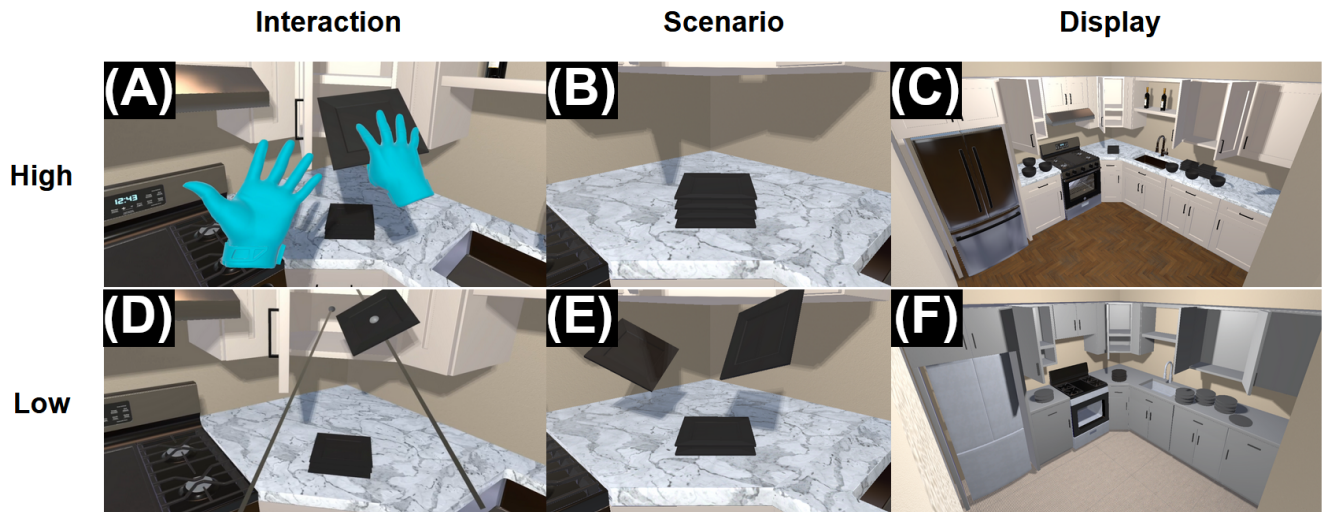


Figure 1: The system fidelity aspects we implemented in our validation study. The top row (A-C) shows the high-fidelity versions of the interaction, scenario, and display components of our virtual reality (VR) study. The bottom row (D-F) shows the low-fidelity versions: D) The Fishing Reel technique [12] is used for interactions instead of the Virtual Hand technique [61]; E) Gravity is not consistent within the scenario; and F) Low-poly models and basic lighting are used instead of high-poly models and advanced lighting.

Abstract

Within the virtual reality (VR) research community, there have been several efforts to develop questionnaires with the aim of better understanding the sense of presence. Despite having numerous

surveys, the community does not have a questionnaire that informs which components of a VR application contributed to the sense of presence. Furthermore, previous literature notes the absence of consensus on which questionnaire or questions should be used. Therefore, we conducted a Delphi study, engaging presence experts to establish a consensus on the most important presence questions and their respective verbiage. We then conducted a validation study with an exploratory factor analysis (EFA). The efforts between our two studies led to the creation of the Fidelity-based Presence Scale (FPS). With our consensus-driven approach and fidelity-based factoring, we hope the FPS will enable better communication within the research community and yield important future results regarding the relationship between VR system fidelity and presence.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHI '25, Yokohama, Japan

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-1394-1/25/04

<https://doi.org/10.1145/3706598.3713566>

CCS Concepts

• **Human-centered computing** → **Virtual reality**.

Keywords

Virtual reality, presence, fidelity, Delphi method

ACM Reference Format:

Jacob Belga, Richard Skarbez, Yahya Hmaiti, Eric J Chen, Ryan P. McMahan, and Joseph J. LaViola Jr.. 2025. The Fidelity-based Presence Scale (FPS): Modeling the Effects of Fidelity on Sense of Presence. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26–May 01, 2025, Yokohama, Japan*. ACM, New York, NY, USA, 15 pages. <https://doi.org/10.1145/3706598.3713566>

1 Introduction

Presence, sometimes noted as “spatial presence”, is a widely investigated sensation in virtual reality (VR) research and is defined as the sense of “being there” in a VR environment [69]. Since its conception in the 1980s [53], numerous techniques have been produced in an effort to better understand and measure presence. Physiological measures, behavioral measures, semi-structured interviews, and questionnaires are some of the common employed techniques found throughout the literature [75]. Of these techniques, questionnaires are the most widely used in user studies. One potential reason for the popularity of questionnaires is that they afford researchers the ability to replicate studies [23]. Previous literature reviews [69, 72, 77] have shown that a multitude of presence questionnaires have been developed and used. Despite the numerous questionnaire options, many researchers still find it necessary to modify and customize these questionnaires [75]. This is concerning as altering the structure of a validated presence survey can lead to adverse results and diminishes the ability to generalize results among multiple studies [39]. This indicates that there is no single presence questionnaire that addresses the broader needs of the research community.

One possible reason for the absence of a presence instrument that meets the needs of the community is that most presence questionnaires do not account for the realism or fidelity of a VR application [69]. Instead, current presence questionnaires provide insight to the user’s general response of being in a virtual environment. For example, the Igroup Presence Questionnaire (IPQ) employs the sub-scales of Presence, Spatial Presence, Involvement, and Realness [67]. As another example, the Spatial Presence Experience Scale (SPES) provides self-location and possible actions as their sub-scales [26]. We believe that VR researchers and developers would greatly benefit from knowing which aspects of a VR application and their respective fidelity impact a user’s sense of presence.

In this paper, we present the development and validation of a presence questionnaire, co-designed by the broader community, to assess how different aspects of fidelity impact presence. We used McMahan et al.’s [48, 51] *system fidelity* framework, which encompasses interaction fidelity (e.g. object manipulation, locomotion technique), scenario fidelity (e.g. gravity of objects, behavior of agents/objects) and display fidelity (e.g. polygon count, audio quality), as our theoretical foundation to address the different aspects of fidelity and selected preexisting presence questionnaires encompassing these aspects. System fidelity is defined as the objective

degree in which real-world interactions or experiences are replicated by an interactive system [48, 51]. We then conducted a Delphi study [29, 35] with 16 presence experts from the broader research community to identify the most important presence questions and what aspects of fidelity they are most affected by. Our Delphi study yielded an initial presence survey comprising of 11 items.

We also present a $2 \times 2 \times 2$ within-subject study that controlled interaction, scenario, and display fidelity at low and high levels (see Figure 1). We used the results of this study ($n=55$) to conduct an exploratory factor analysis (EFA) to identify the new structure of our presence survey, which yielded 10 items across three factors: Interaction Presence, Scenario Presence, and Display Presence. We then present the results of our experiment, which confirm that our new presence survey is capable of identifying how each fidelity component affects presence. Between our Delphi study and validation study, our efforts have culminated in the development of the Fidelity-based Presence Scale (FPS). In summary, our contributions are:

- Co-designing a presence questionnaire with perspectives of expert presence researchers from the broader community.
- Conducting one of the first studies to investigate the effects of interaction, scenario, and display fidelity on one’s sense of presence.
- Using EFA to validate the reliability of our new presence questionnaire and to identify its three subscales: Interaction Presence, Scenario Presence, and Display Presence
- Providing results that demonstrate the validity of the FPS and its ability to indicate how different aspects of VR system fidelity affect presence.
- Providing open access to FPS in common survey formats (e.g., PDF, Word, Qualtrics): [FPS Survey](#).

2 Related Work

From our review of presence literature, which builds upon systematic surveys from Van Baren [77], Skarbez et al. [69], and Souza et al. [72], we have learned of the broader perspectives of presence and the current measurement tools used to evaluate presence. In our Delphi study, we include a series of questionnaires that are predominantly utilized in VR contexts across the presence research community and capture nuances from the past 40 years of research. Table 1 presents a series of presence questionnaires along with their respective item counts, measurement factors, whether they conducted an EFA, and whether they conducted a Delphi study.

2.1 Past Presence Measurement Methods

As highlighted in our introduction, questionnaires and surveys are one of the most prominent forms of measuring the sense of presence. This is further supported by the sheer number of efforts led to create presence surveys. In 2004, Van Baren [77] conducted a systematic review and compiled a list of 28 surveys. In 2017, Skarbez et al. [69] compared and contrasted 14 presence/telepresence surveys. In 2021, Souza et al. conducted a similar review and curated a list of 29 presence surveys [72]. However, in recent literature, it is apparent that there still exists some discourse on which questions are the most important to ask when measuring presence. Particularly, in 2024, Tran et al. [75] conducted a literature review of

Citation	Questionnaire	Items	Year	Measurement Factors	Delphi Study	Exploratory Factor Analysis
[6]	Barfield et al. Questionnaire	3	1993	N/A		
[36]	Kim & Biocca Questionnaire	8	1997	Departure, Arrival		✓
[58]	Questionnaire on Presence and Realism	10	1998	N/A		
[81]	Witmer-Singer Presence Questionnaire	32	1998	Involved/Control, Natural, Auditory, Haptic, Resolution, Interface Quality		
[19]	Dinh et al. Questionnaire	14	1999	N/A		
[43]	Lombard & Ditton Questionnaire	103	2000	Immersion, Parasocial interaction, Parasocial relationships, Physiological responses, Social reality, Interpersonal social richness, General social richness		✓
[54]	Murray et al. Questionnaire	5	2000	N/A		
[55]	Nichols et al. Questionnaire	9	2000	N/A		
[5]	Reality Judgment and Presence Questionnaire	18	2000	Reality Judgement, Internal/External Correspondence, Attention/Absorption		✓
[76]	Slater-Usch-Steed Presence Questionnaire	6	2000	N/A		
[24]	Gerhard et al. Questionnaire	19	2001	Immersion, Communication, Involvement, Awareness		
[67]	Igroup Presence Questionnaire	14	2001	Presence (PRES), Spatial Presence (SP), Involvement (INV), Realness (REAL)		✓
[41]	ITC Sense of Presence Inventory	63	2001	Sense of Physical Space, Engagement, Ecological Validity, Negative Effects		✓
[37]	Krauss et al. Questionnaire	42	2001	Emotionally involved, Degree of Involvement		✓
[66]	Schroeder et al. Questionnaire	10	2001	N/A		
[38]	Swedish Viewer-User Presence	150	2001	N/A		
[73]	Experimental Virtual Environment-Experience Questionnaire	124	2002	Spatial, Attention, Real, Action, Exploration, Skill, Challenge, Personal relevance, Control, Arousal, Valence, Flow, Being there, Impressed, Pleasant, Anxiety, VE distracted, Nausea, Tiredness		✓
[42]	E^2I Scale	9	2002	Sensory Factor, Distraction Factor, Realism Factor, Control Factor		
[15]	Cho et al. Questionnaire	4	2003	N/A		
[56]	Nowak & Biocca Questionnaire	29	2003	Self-reported copresence, Perceived other's copresence, Telepresence scale, Social presence		
[65]	Sas and O'Hare Questionnaire	34	2003	Being there, Not being there, Reflective consciousness		
[10]	Bouchard et al. Questionnaire	1	2004	N/A		
[78]	MEC Spatial Presence Questionnaire	16	2004	Self Location, Possible Actions		
[74]	Presence-Involvement-Flow Framework	15	2004	Physical presence, Emotional Involvement, Situational Involvement, Performance Competence		
[44]	Temple Presence Inventory	42	2009	Spatial Presence, Social presence-actor, Passive social presence, Active social presence, Presence as engagement, Presence as social richness, Presence as social realism, Presence as perceptual realism		✓
[14]	Virtual Experience Test	24	2010	Story Telling, Haptics, Sensory Content, Task Completion, Active		
[26]	Spatial Presence Experience Scale	8	2015	Self-location (SL) and Possible action (PA)		✓
[79]	Self-Assessment-Manikins	6	2015	Attention Allocation, Spatial Situation Model, Self-location, Possible actions, High cognitive involvement, suspension of disbelief		✓
[4]	Self Presence and Spatial Presence	5	2016	N/A		
[46]	Multimodal Presence Scale	15	2017	Physical Realism (PR), Not paying attention to real environment (NARE), Sense of being in the virtual environment (SBVE), Not aware of physical mediation (NAPM), Sense of coexistence (SC), Human realism (HR), Not aware of artificiality of social interaction (NAASI), Not aware of social mediation (NASM), Sense of bodily connectivity (SBC), Sense of bodily extension (SBE)		
Fidelity-based Presence Scale		11		Interaction Fidelity Scenario Fidelity Display Fidelity	✓	✓

Table 1: Summary of presence questionnaires developed in the past and how their measurement factors compare to the FPS. We also include if the presence questionnaire employed a Delphi Study or conducted an EFA in their development. Presence questionnaires in bold are questionnaires we utilized in our Delphi study.

presence studies and found that 201 of their 320 retrieved articles leveraged previously created questionnaires with customizations applied to each survey. The sheer number of surveys reviewed plus the added concern of research performed with customized presence surveys showcases that there is no single presence survey or series of questions that address the broader needs of the research community.

Given these concerns, we conducted a Delphi study in which we solicited a panel of expert researchers in presence. For our Delphi study, we asked our experts to review the entire contents of the Witmer-Singer Presence Questionnaire (WS-PQ) [81], Igroup Presence Questionnaire (IPQ) [67], Slater-Usch-Steed (SUS) presence questionnaire [76], Multi-modal Presence Scale (MPS) [46], and Spatial Presence Experience Scale (SPES) [26]. These surveys were chosen for our Delphi study as we believe each of these surveys

encompass all of the aspects of the system fidelity framework. We choose to focus on spatial presence through our work, which these surveys aim to measure. Co-presence (the sense of being with others [9]), Social presence (the awareness of co-presence and ability to engage with others [9]), and Self-presence (the effect of embodiment in a virtual environment [8]) are other types of presence that have been investigated throughout the literature as highlighted in Table 1. With system fidelity as our guiding framework, we focus on spatial presence as it pertains to the events and interactions an individual experiences in a virtual environment (VE).

Witmer-Singer Presence Questionnaire (WS-PQ): The WS-PQ presence measurement tool is the longest measurement tool that we considered with 32 items. The WS-PQ loads each item into four core factors: control factors (CF), sensory factors (SF), distraction factors (DF), and realism factors (RF) [81]. Subsequently, items in the WS-PQ can also load into these sub-factors: involvement/control (INV/C), natural (NAT), auditory (AUD), haptic (HAPTC), resolution (RES), and interface quality (IFQUAL) [81]. With WS-PQ having the largest inventory in our selected surveys and a wide breadth of loading factors, the questionnaire is aimed to account for a wide range of technological and psychological facets. However, not all virtual experiences employ, for example, auditory or haptic features, which may cause participants to respond negatively to questions explicitly regarding those factors (e.g. WS-PQ question 15 and 16).

Igroup Presence Questionnaire (IPQ): The structure of IPQ is derived from an exploratory factor analysis (EFA) and a confirmatory factor analysis (CFA). Items in IPQ's inventory include questions created by the original researchers and also questions from previous surveys such as the WS-PQ [81]. However, IPQ incorporates differing factor loads onto each of its items. Presence (PRES), spatial presence (SP), involvement (INV), and realism (REAL) are the primary factors a question can load into [67]. An interesting note is that for an item to contribute as a PRES factor, it also needs to collectively load into SP, INV, and REAL [67]. A limitation of IPQ is that it does not explicitly highlight technological factors that compose the virtual experience and impact users' sense of presence. The IPQ focuses on the disconnect users experience when immersed into a virtual environment. Subsequently, IPQ prompts users to evaluate the perceived realism of the virtual environment and their level of disconnection from the real world.

Slater-Usch-Steed (SUS): The SUS presence questionnaire is one of the most recognized presence measurement tools. Items in this survey were developed through a series of studies conducted by Usch et. al [76]. This survey was created on the premise that a presence measurement tool should be capable of discriminating between real-world and virtual experiences [76]. While SUS is a widely utilized presence questionnaire, it does not employ a factor structure, unlike the other questionnaires we consider. The benefit of employing a factor structure is that it enables researchers to draw conclusions regarding the factors as well as the overall sense of presence.

Multimodal Presence Scale (MPS): The MPS was developed with an emphasis on Lee's [40] theoretical framework of presence [47]. This premise led to the construction of a questionnaire that aims to address three types of presence: Physical presence, social presence, and self-presence [46]. As noted in Table 1, the MPS utilizes a

wide range of factors to ensure it covers the three aforementioned presence types. As we have seen with IPQ, we witness the same utilization of CFA to compose the final items in the survey. The MPS excels in combining spatial presence, co-presence, and embodiment into a singular survey. However, it is difficult to conclude which specific components of the virtual experience contributed to the users' senses of physical, social, and self presence.

Spatial Presence Experience Scale (SPES): The SPES survey is based on the spatial presence model proposed by Wirth et al. [26, 80]. Wirth et al.'s [80] model states when users are presented a virtual or media stimulus, they are expected to exhibit a certain level of focus and subsequently develop a perception of possible actions they can perform. With this grounding, the SPES was developed through the composition of questionnaire items and performing an EFA and a CFA to include items in the finalized survey. Similarly to IPQ, the SPES focuses on the individual experience and perceptions of a virtual environment. As with previously discussed surveys, it is difficult to identify the technological components that allowed users to experience high levels of self location and personal agency.

As we have seen, most of the aforementioned presence surveys are grounded within psychological or behavioral constructs [69, 72]. While this can yield productive insights, the results are not particularly easy to translate to future design decisions. Given this gap in the literature, we specifically aimed to design a survey that would yield insights to inform key design decisions. To that end, our work introduces the FPS, a survey grounded in the system fidelity framework. The grounding in system fidelity affords us the ability to report and discuss the components of a VR system that affected sense of presence. This key feature aids in improving result interpretation and comparisons between virtual experiences which, in turn, can yield broader discussions regarding VR system design.

2.2 The Effects of Fidelity on Presence

In our growing understanding of presence, we note that there is a growing body of literature that highlights the idea that increased fidelity leads to higher levels of presence. McMahan et al. [50] conducted a study evaluating display and interaction fidelity. In their study, high display fidelity and high interaction fidelity led to significantly higher presence scores. In another study, Shafer et al. [68] conducted a study evaluating off-the-shelf VR experiences and the games' effects on cybersickness and sense of presence. They presented an initial model showcasing how realism and interactivity affect spatial presence. Adkins et al. [1] conducted a study comparing grasping techniques between a tracked glove and a standard VR controller. In their work, a key outcome was that grabbing objects with the glove led to significantly higher presence scores than the controller. As we can see across multiple studies, there is a strong implication for higher fidelity environments and experiences may yield higher levels of presence.

In addition to the aforementioned studies, many of the previous questionnaires we reviewed, see Table 1, also contain components that allude to the idea that level of fidelity influences sense of presence. For instance, the WSPQ [81] employs auditory and haptic as factors that impact sense of presence. The MPS [46] notes physical realism as a factor that influences presence. Across previous work and questionnaire development, we consider the case that higher

fidelity representations of virtual experiences can positively impact sense of presence. Therefore, the FPS explicitly follows the System Fidelity framework and is validated on its core components.

2.3 The Delphi Method

The Delphi method is a study technique utilized to obtain consensus of opinion through administering a series of questionnaires to experts in a given area of research [13, 18, 29, 35]. While the definition of an “expert” can be debated, Adler [2] defines an expert as an individual that has both extensive knowledge and practical experience with investigating the topic. By design, the Delphi method places a high emphasis on anonymous and iterative review wherein experts can provide their insight and feedback without potential peer pressure or conforming to a dominant view [3, 52, 57, 64].

Previous research highlights the fact that there is no single predominant presence measurement tool or presence measurement theory [40, 69, 72]. This posit is further supported by the number of presence questionnaires that have been developed over the past 40 years. By employing the Delphi Method in its development, the FPS represents a significant step forward in presence measurement as it is, to the best of our knowledge, the first consensus-driven presence measurement tool.

3 Delphi Study Methods

As described in section 2.3, the Delphi method is a study approach designed to gain the insight and feedback of experts from a given field to form consensus on a specific topic [29, 35]. Our Delphi study methodology is adapted from Pandor et al.’s [57] Delphi study on Rapid Reviews. In our Delphi study, our participants completed two rounds of reviewing presence questionnaire items from WS-PQ, IPQ, SUS, MPS, and SPES. Between each round, we modified the questionnaire items based on our expert participants’ feedback which were reviewed by our experts in a subsequent round of the Delphi study. Once our participants provided minimal or no feedback across the questionnaire items, we concluded our Delphi study. Figure 2 summarizes our Delphi process.

3.1 Participant Recruitment

The target participants for our Delphi study were expert presence researchers. Our expert criterion included researchers that have earned their Ph.D., have a thorough history of conducting research in VR contexts, and published three or more articles that employed a standardized presence questionnaire (e.g., SUS or WS-PQ) in a VR study. This follows Adler et al.’s [2] definition of an expert as our expert researchers had extensive knowledge along with practical experience of applying these questionnaires. We compiled an initial list of candidate experts through systematically searching electronic databases (e.g., ACM Digital Library, IEEEExplore, Web of Science, Google Scholar). For databases that allowed for enhanced search parameters, we used the following query to narrow down our searches as well: $TI = (presence \text{ AND } virtual) \text{ OR } AB = (presence \text{ AND } virtual) \text{ OR } AK = (presence \text{ AND } virtual)$. We then gathered our expert candidates’ contact information through publicly available sources (e.g., university websites) and sent personalized emails containing the definition of a Delphi study, the purpose of our study, and a Qualtrics survey link to participate in Round 1 of

our Delphi study. Table 2 shows the demographics of our expert participants.

In previous Delphi studies, the appropriate number of participants varied greatly based on the domain being investigated. Broader topic Delphi studies can result in participant pools from 15 to 60 participants [16, 27, 29]. Conversely, Delphi studies in more specific domains of expertise may contain 5 to 15 participants [16, 29]. To ensure a response rate of at least 15 participants, given previous work, we aimed to invite at least 60 potential researchers to participate in our Delphi study. All participants were allowed to withdraw from the study at any time if needed. For all rounds of the Delphi study, participants responded through Qualtrics surveys.

3.2 Delphi Survey Presentation

Prior to participating in each round of the Delphi Survey, participants were presented with an introductory page that outlined the survey’s purpose, listed the questionnaires included in the survey, and provided definitions for system fidelity, interaction fidelity, scenario fidelity, and display fidelity. Additionally, participants received instructions detailing the information they would encounter for each presence questionnaire item. This included guidance on providing importance ratings, specifying aspects of system fidelity that would affect the response to the item, and providing additional feedback or suggestions to improve a given item.

3.3 Round 1

The Round 1 survey included the entire inventories of WS-PQ, IPQ, SUS, MPS, and SPES [26, 46, 67, 76, 81]. To address item duplication (e.g., IPQ Question 12 draws from WS-PQ question 7), only one instance of each duplicated item was included, resulting in a total inventory of 68 questionnaire items for review. These questionnaires were chosen on the basis that they all are highly cited surveys and we believe they encompass the aspects of the system fidelity framework.

Participants were presented with each questionnaire item in its original form, including the rating scale and anchors. Based on Pandor et al.’s [57] importance rating, we incorporated a Likert

Participant Characteristics	Count	
	Round 1 (N = 16)	Round 2 (N = 14)
Gender		
Male	12 (75%)	10 (71%)
Female	4 (25%)	4 (29%)
Country of employment		
Australia	1 (6%)	1 (7%)
France	1 (6%)	1 (7%)
Germany	2 (13%)	2 (14%)
Italy	2 (13%)	2 (14%)
Portugal	2 (13%)	1 (7%)
United Kingdom	1 (6%)	1 (7%)
USA	7 (44%)	6 (43%)

Table 2: A breakdown of our participant demographics of gender and global location of employment.

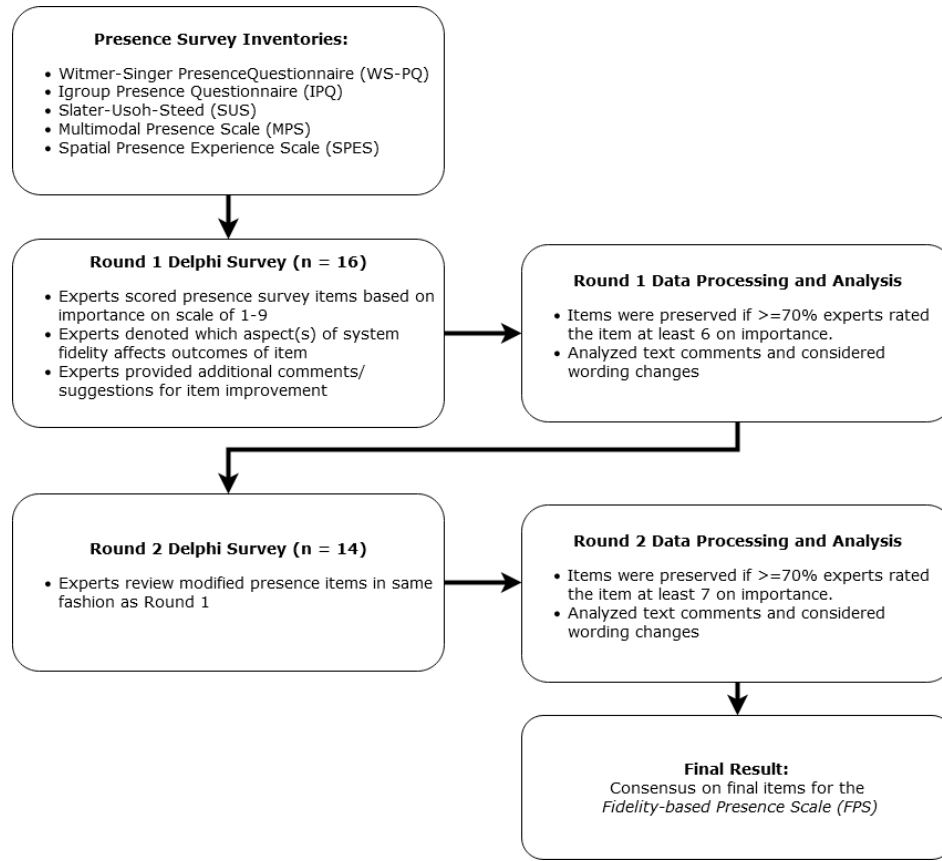


Figure 2: Summary of our Delphi process to create the Fidelity-based Presence Scale.

scale ranging from 1 (not important) to 9 (critically important) where participants rated importance of each questionnaire item. Our participants were also required to specify which component(s) of system fidelity (interaction fidelity, scenario fidelity, display fidelity) would affect the response to the item. If needed, participants were also able to offer additional suggestions to improve the clarity or grammar of the item in the form of a free-response text box. Each presence questionnaire item was accompanied with the definitions of interaction fidelity, scenario fidelity, and display fidelity. This served as a reference in case participants needed a reminder of the definitions during the survey.

After evaluating all 68 presence items, participants were prompted to provide further suggestions or propose additional survey items. They were also asked to indicate the number of points on a Likert scale each item should be rated on and whether items should be presented as questions or statements.

With the conclusion of Round 1, we conducted an analysis of the importance ratings for each item. This analysis was performed to determine which questions would be included in Round 2 of the survey. For items that were included in the second round, we further analyzed our participants' free-text responses and applied necessary changes. We evaluated our participants' comments and re-worded the item to address our participants' concerns while maintaining the core themes and elements of the original questionnaire item. For

example, SPES question 1 was initially phrased as "I felt like I was actually there in the environment of the presentation." Our expert reviewers noted the term 'presentation' was confusing, which led us to modify the item to be phrased as "I felt like I was actually there in the virtual world." Once Round 1 concluded, participants were awarded \$50 USD in the form of an Amazon e-gift card.

3.4 Round 2

Given our applied changes and included items from Round 1, we believed it was necessary to conduct an additional round where our experts evaluated the presentation of the new, reduced questionnaire inventory. All the participants who completed Round 1 of the Delphi survey were invited to participate in Round 2. Along with a new Qualtrics survey link containing the revised questionnaire items, participants were provided a results package containing the overall expert panel consensus distribution for each item and their respective importance ratings as seen in Figure 3.

In the Round 2 survey, experts were once again tasked with rating the importance of each survey item, specifying which component(s) of system fidelity would affect the response, and providing any additional suggestions for clarity or improvement. Each item was presented in the same form and fashion as in Round 1.

Following the conclusion of Round 2, we performed the same analyses on both our participants' perceived importance ratings and

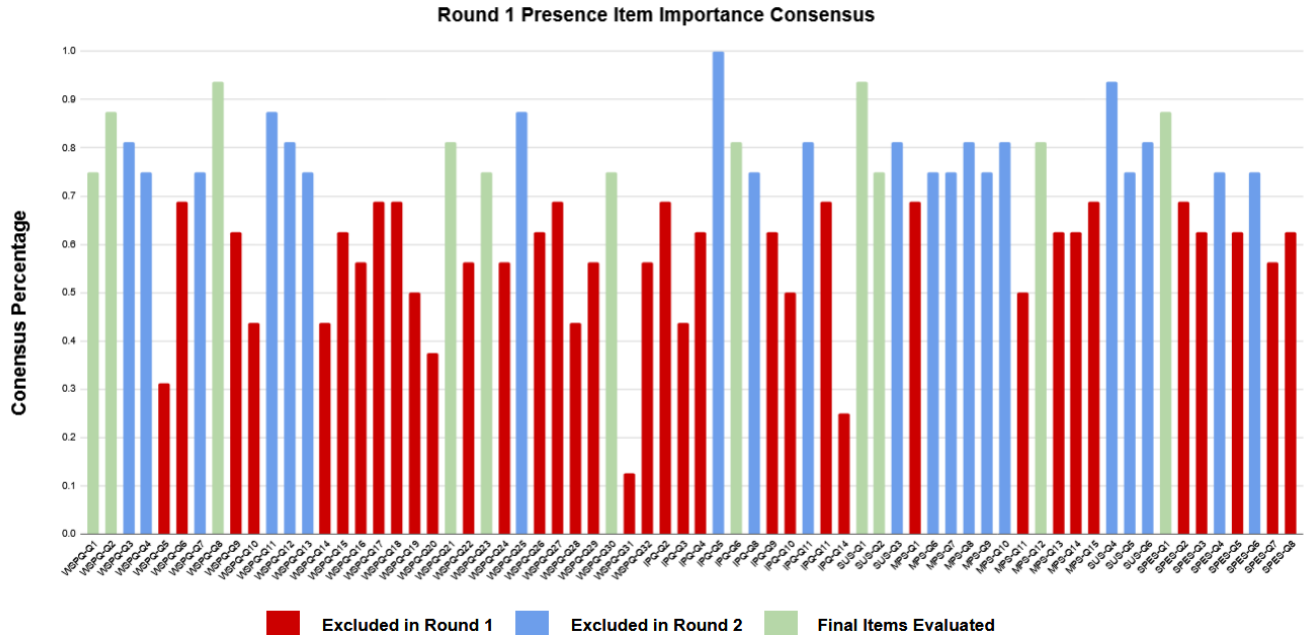


Figure 3: Round 1 summary of presence questionnaire items that met the inclusion criteria to be preserved in Round 2 of our Delphi study. Red indicates that those items were removed after Round 1 analysis. Blue indicates that those items were preserved after Round 1 and removed after Round 2. Green indicates the final items that comprise the FPS after the Delphi study.

free-response comments. Once our inclusion criteria was applied, we found the final set of survey items to be sufficient as there were minimal to no comments warranting further review. Therefore, we concluded our Delphi study after Round 2. Participants were awarded \$50 USD through an Amazon e-gift card for completion of Round 2.

3.5 Inclusion Criteria across Round 1 and Round 2

Initially, we applied the same inclusion criteria as Pandor et al.'s [57] Delphi study where 70% of experts needed to rate a given item 7 or higher in order for that item to be included in a subsequent round. However, in review of our Round 1 results, we found that criteria to drastically drop the questionnaire item count from 68 to 6. While a survey comprising of 6 items could be considered sufficient, we believed that there were numerous items that should be reconsidered. Jorm [31] notes that in cases where there are a high number of items for experts to evaluate, it is common for items to be advocated for reconsideration once during the Delphi study, such as in Yap et al.'s work [83]. Therefore, for Round 1, we included questionnaire items that received a rating of 6 or higher from 70% or more of our experts. In Round 2, we maintained the original inclusion criteria of at least 70% of experts needing to rate an item 7 or higher as our experts' did not present new concerns/changes to any of the items in the survey.

Additionally, our criteria for individual item changes followed the workflow highlighted in section 3.3. For each item, we considered comments and concerns that were raised by *at least two* of our

experts. This allowed for strict changes to be applied across items if needed. We also categorized changes as either being a global change or a local change. For global changes, these were changes that were expressed across multiple items that our experts believed should be modified. For example, SPES #1 was phrased as "I felt like I was actually there in the environment of the presentation." Our experts first noted their concern with the phrasing "environment of the presentation" which was found across multiple items. Given that this concern of phrasing was raised across multiple items, we applied a global change in which we ensured all items would be phrased with "virtual world" as seen in Table 3. A local change is one that we applied specifically to an item. For example, in our Round 1 review, the original phrasing of WSPQ #1 was "How much were you able to control events?". Four of our experts noted the potential confusion of the phrasing "control events", which led us to change the phrasing of this item to "I felt I could control my actions in the virtual world". Our comment/concern count criteria is also the reason we resulted in two rounds for our Delphi study. We conducted additional rounds if two or more experts raised concerns for a given item. In our case, our experts did not raise concerns after reviewing the changed items in round 2, which led to us finalizing our Delphi study.

3.6 Delphi Study Results

In summary, our Delphi study is comprised of two rounds of presence questionnaire item review. Data for Round 1 was collected over the span of 2 months, and data for Round 2 was also collected

over the course of 2 months. In this section, we detail the key results and outcomes of our Delphi study.

3.6.1 Round 1 Consensus. Out of the 68 items presented, 32 items (47%) garnered a consensus of 70% or higher in terms of importance. Participants' comments emphasized maintaining consistency in wording and Likert scale rating range for each item. As detailed in section 3.5, we conducted a thorough analysis of each questionnaire item individually, and the Delphi survey was revised to present each of the 32 items as statements and to utilize a 7-point Likert scale for rating based on our experts' feedback.

3.6.2 Round 2 Consensus. In Round 2, 11 items (28%) out of the 32 items achieved a consensus rating of 70% or higher from our experts. Overall, our participants expressed satisfaction towards the inventory of the survey and no additional suggestions were made for improving the survey within the 11 items. Following a thorough review of our experts' comments and potential suggestions, Round 2 concluded the Delphi study as there were no outstanding suggestions or views that warranted additional rounds. Furthermore, we also wanted to prevent our participants from exhibiting survey fatigue from the iterative process [35]. The resulting version of our presence survey, including the importance rating consensus across both rounds, each item's origin, and which components of system fidelity affect the item, is presented in Table 3.

4 Factor Analysis Validation Study

After the conclusion of our Delphi study, the FPS needed to be validated through a user study. Given FPS' theoretical basis in system fidelity, we conducted a $2 \times 2 \times 2$ (*InteractionFidelity* \times *ScenarioFidelity* \times *DisplayFidelity*) within-subjects VR user study. With participants responding to the FPS within each condition, we applied a Latin square design to ensure conditions were counter-balanced. For all eight conditions, participants were immersed into a kitchen environment with the responsibility of putting dishes away.

4.1 Independent Variables

We investigated three independent variables in our study: interaction fidelity, scenario fidelity, and display fidelity. For each of the system fidelity components, we implemented a low and high fidelity variant to evaluate. Each of our implementations were not intended to comprehensively represent all the aspects of each fidelity construct, but rather focus on specific elements that directly influence user experience in our task context. We further detail which aspects of each fidelity component we include in our study design.

4.1.1 Interaction Fidelity. Since the main task was to grab and put dishes away, we implemented two techniques to represent high interaction fidelity and low interaction fidelity. For high interaction fidelity, we used direct manipulation with virtual hands [61]. This approach included predefined hand gestures for direct manipulation when grabbing objects, allowing for a more intuitive interaction style as seen in Sub-figure A from Figure 1. Conversely, we adapted Bowman et al.'s [12] ray cast "fishing reel" metaphor technique as our low interaction fidelity condition. A ray cast based technique was utilized as it is noted in previous work [11, 12, 60] that ray cast

interaction techniques increase the difficulty of rotating the object in place. Sub-figures A and D from Figure 1 shows the representation of both the interaction technique levels. Between our two selection techniques, direct manipulation with the VR controllers provides a higher level of biomechanical symmetry (i.e. degree with which real-world body movements are reproduced with an interaction [51, 62]) and higher level of control symmetry (i.e. the degree with which control in a task is provided by an interaction [51, 62]) than the ray-cast based technique.

4.1.2 Scenario Fidelity. We implemented two techniques to represent high scenario fidelity and low scenario fidelity. In high scenario fidelity conditions, gravity would always be enabled and dishes would appropriately drop when let go. In the low scenario fidelity conditions, when the participants would let go of a dish, there was a 50% chance that the gravity would be disabled for that single dish. This would lead to instances where some plates would be floating around and some plates would be affected by gravity. Sub-figures B and E from Figure 1 shows the two variations of scenario fidelity. With our scenario fidelity conditions, we were able to vary the physical coherence (i.e. how consistent the physics of the virtual environment are to the real world [49, 70]) and attribute coherence (i.e. how consistent attributes of virtual objects are to their real world counterpart [49]). However, due to our study design not including virtual agents, we cannot evaluate behavioral coherence as it is contingent on the quality and consistency of virtual agent behaviors [49].

4.1.3 Display Fidelity. Throughout the study, participants were immersed into two variants of the kitchen environment. We explored display fidelity by varying visual fidelity (i.e. the degree with which realistic visuals are reproduced in a virtual environment [49, 62]) and auditory fidelity (i.e. the degree with which realistic audio stimuli are reproduced in a virtual environment [49, 62]). In the low display fidelity condition, the textures of the environment were minimal and there was no sound in the environment. Additionally, the dishes were primarily primitive Unity objects (i.e. thin cylinders). In the high display fidelity condition, realistic textures were present, the dishes made noises when colliding with other objects, there was royalty-free background music with spatial audio enabled, and the dishes were actual plate and bowl models. Sub-figures C and F from Figure 1 shows the two display fidelity environments in our study.

4.2 User Study Procedure

Participants recruited for our study were asked to review and complete a pre-screening survey through Qualtrics. In the survey, participants were asked to review an informed consent document, an eligibility document, and provide demographic information, such as age and gender. Following the conclusion of the survey, participants were then asked to schedule a day and time in which they would be able to participate in the in-person VR study.

On arrival, participants were introduced to the Meta Quest Pro, the VR system used to administer the study. This introduction included how to adjust the headset for comfort as well as the controls needed to perform the tasks in the study. We then informed participants on the overall task. We explained to participants that the

Item Origin	Original Item	FPS Item	Round 1 Consensus	Round 2 Consensus
SUS-Q1	Please rate your sense of being in the environment, on the following scale from 1 to 7, where 7 represents your normal experience of being in a place	My sense of being in the virtual world was like being in a real place	15/16 (93%)	11/14 (78%)
SUS-Q2	To what extent were there times during the experience when the virtual environment was reality for you?	During the experience, I felt the virtual world was reality for me.	12/16 (75%)	10/14 (71%)
SPES-Q1	I felt like I was actually there in the environment of the presentation	I felt like I was actually there in the virtual world.	14/16 (87%)	10/14 (71%)
IPQ-Q6	I felt present in the virtual space	I felt present in the virtual world.	13/16 (81%)	11/14 (78%)
WSPQ-Q1	How much were you able to control events?	I felt I could control my actions within the virtual world.	12/16 (75%)	12/14 (86%)
WSPQ-Q2	How responsive was the environment to the actions that you initiated (or performed)?	I felt the virtual world was responsive to my actions.	14/16 (87%)	12/14 (86%)
WSPQ-Q8	How aware were you of events occurring in the real world around you?	I was not aware of events occurring in the real world around me.	15/16 (93%)	10/14 (71%)
WSPQ-Q21	How well could you move or manipulate objects in the virtual environment?	I felt I could move or manipulate objects easily in the virtual world.	13/16 (81%)	10/14 (71%)
WSPQ-Q23	How involved were you in the virtual environment experience?	I felt involved in the virtual world experience.	12/16 (75%)	10/14 (71%)
WSPQ-Q30	How well could you concentrate on the assigned tasks or required activities rather than on the mechanisms used to perform those tasks or activities?	I could concentrate on the virtual activities rather than the controls to perform them.	12/16 (75%)	10/14 (71%)
MPS-Q12	When something happened to my virtual embodiment, it felt like it was happening to my real body.	When something happened to my virtual body, I felt it happened to my real body.	13/16 (81%)	10/14 (71%)

Table 3: Summary of expert importance consensus across Round 1 and Round 2 along with the percentage of experts that identified a component of system fidelity that would affect the outcome of a question after Round 2.

study is comprised of eight trials of putting dishes away. Each trial was comprised of putting away 20 dishes in any of the cabinets within the kitchen environment. We also explained that between each trial, aspects of the virtual world may change (interaction fidelity, scenario fidelity, and display fidelity). Participants would know if all the dishes were put away once the in-VR FPS questionnaire was presented for them to respond to. The environment is based on the designs presented by Hmaiti et al. [28].

After explaining the procedure, participants were immersed into the eight trials. For each trial, participants would respond to the entire inventory of the FPS on a 1 to 7-point Likert scale. After four trials, participants were offered the opportunity to take a break from VR. After their break, participants would complete the remaining four trials of putting dishes away.

Once all eight trials were completed, participants were asked if they had any comments or questions regarding their experience. The overall time required to complete the study was approximately 60 minutes, and participants were compensated \$20 USD via an Amazon e-gift card.

4.3 Participants

We recruited a total of 55 participants (25 female, 30 male) from our local university. Participants were required to be 18 years of age or older, have normal or corrected-to-normal vision, and be able to hear, walk, extend both arms, use both hands, and speak and understand English. Participants with any visual, auditory, neurological, or physical disabilities were excluded. The ages of our participants ranged from 18 to 32 with a mean age of 21.62.

4.4 Exploratory Factor Analysis

In our validation study, 55 participants responded to the FPS under 8 unique conditions, totaling 440 complete responses to the FPS. Following our validation study, we conducted an exploratory factor

analysis (EFA) to reveal further insights into the factors each item loads onto and if there are items that need to be removed.

On the initial data collected, we constructed a correlation matrix to learn if there were items in the FPS that were highly correlated. We found no item pairings from the original FPS inventory with correlation values higher than 0.9. We then verified the sampling adequacy of our data using the Kaiser-Meyer-Olkin (KMO) measure, $KMO = 0.9$, which indicates that our data is suitable for factor analysis [33, 34]. Our Bartlett's test of sphericity ($\chi^2 = 3270.034$, $p < 0.001$) was significant and indicated the presence of correlations among the FPS' survey items [7].

To determine the appropriate number of factors to apply to our survey, we first computed the eigenvalues to learn of the number of factors Kaiser's criterion recommends. We also conducted a parallel analysis to compare the factor valuation Kaiser's criterion recommended. Our eigenvalues, based on Kaiser's criterion of 1, recommends that we maintain two factors while our parallel analysis suggested four factors. Given our experts' consensus on three factors in our Delphi Study, we conducted our initial EFA with three factors, which falls within the suggested ranges from Kaiser's criterion and the parallel analysis. For each of the structure matrices, we applied an inclusion threshold of $> |0.3|$ since our user study resulted in 440 samples of completed FPS surveys. This follows the factor analysis inclusion criteria based on sample size set by Hair et al. [25].

In Table 4, we present the initial structure matrix for the original 11 survey items. Upon running the EFA, we note that WSPQ-Q8 did not sufficiently load into any of the identified factors of interaction fidelity, display fidelity, and scenario fidelity as seen in Table 4. Therefore, we removed WSPQ-Q8 from our final item pool and conducted an additional EFA on the remaining survey items.

In the revised data, we constructed a correlation matrix and found no item pairings to have correlation values higher than 0.9. The KMO mean sampling adequacy maintained a value of 0.9. Our Bartlett's test of sphericity ($\chi^2 = 3178.535$, $p < 0.001$) remained

Item	Delphi Results			EFA Results		
	Interaction Fidelity Consensus	Scenario Fidelity Consensus	Display Fidelity Consensus	Interaction Fidelity	Scenario Fidelity	Display Fidelity
SUS-Q1	7/14 (50%)	13/14 (93%)	11/14 (78%)	0.40	0.16	0.77
SUS-Q2	10/14 (71%)	14/14 (100%)	10/14 (71%)	0.31	0.18	0.84
SPES-Q1	10/14 (71%)	11/14 (78%)	11/14 (78%)	0.24	0.40	0.70
IPQ-Q6	12/14 (86%)	10/14 (71%)	11/14 (78%)	0.34	0.86	0.36
WSPQ-Q1	14/14 (100%)	4/14 (29%)	1/14 (7%)	0.86	0.20	0.23
WSPQ-Q2	14/14 (100%)	9/14 (64%)	3/14 (21%)	0.89	0.19	0.20
WSPQ-Q8	3/14 (21%)	5/14 (36%)	10/14 (71%)	0.15	0.10	0.30
WSPQ-Q21	12/14 (86%)	7/14 (50%)	2/14 (14%)	0.73	0.10	0.35
WSPQ-Q23	9/14 (64%)	11/14 (78%)	6/14 (43%)	0.51	0.37	0.37
WSPQ-Q30	14/14 (100%)	2/14 (14%)	4/14 (29%)	0.58	0.19	0.29
MPS-Q12	8/14 (57%)	5/14 (36%)	11/14 (78%)	0.13	0.1	0.61

Table 4: The Structure Matrix after running EFA after performing an orthogonal rotation on the 11 questions. Values greater than |0.3| are in bold based on criteria from Hair et al. [25]. Alongside the matrix are the experts' rating distributions of which component of System Fidelity would affect each of the FPS items

significant as well. As with the original data, we also computed the eigenvalues and conducted a parallel analysis. Kaiser's criterion recommended 2 factors and the parallel analysis recommended 3 factors are appropriate for the revised data. Table 5 presents the structure matrix on the revised data along with the calculated Cronbach's alpha for each of the factors. None of our factors were found to be violation of the Cronbach's alpha threshold as all values were greater than 0.7 [17].



Figure 4: How an item from the FPS was presented in the virtual environment. Each item is rated on a 7-point Likert scale with the anchors: "Fully disagree" and "Fully agree"

4.5 The Fidelity-based Presence Scale (FPS)

The final version of the FPS consists of 10 items that can be utilized across VR experiments that aim to evaluate participants' sense of presence in virtual environments. Across the Delphi study and our validation study, we have shown the development process of the FPS from theory into practice. Based on feedback from our experts, the FPS should be collected utilizing a 7-point Likert scale ranging from "Fully Disagree" to "Fully Agree". Figure 4 shows how an item from the FPS would be employed in a virtual reality context.

4.6 Calculating the FPS Score

Results of data collected using the FPS will be presented across four scores: Interaction Presence Score, Scenario Presence Score, Display

Presence Score, and Total Presence Score. The purpose of breaking down the results into these four scores to allow future work to provide additional insight into which features of the environment attributed to overall sense of presence.

- Interaction Presence Score: $(FPS-1 + FPS-2 + FPS-4 + FPS-5 + FPS-6 + FPS-7 + FPS-8 + FPS-9) / 8$
- Scenario Presence Score: $(FPS-3 + FPS-4 + FPS-8) / 3$
- Display Presence Score: $(FPS-1 + FPS-2 + FPS-3 + FPS-4 + FPS-7 + FPS-8 + FPS-10) / 7$
- Total Presence Score: $(\sum_{n=1}^{10} FPS_n) / 10$

4.7 Reporting Results of the FPS

After collecting our FPS results across our 55 participants, we calculated each of the presence scores, as highlighted in section 4.6. We found the data to violate normality, therefore we normalized our data through the Aligned Rank Transform (ART) approach [82]. For each presence score, we report the analysis of variance (ANOVA) and their corresponding pairwise t-test(s) for the main and interaction effects.

4.8 Total Presence Score

For our Total Presence Score, our repeated measures ANOVA revealed significant main effects for interaction fidelity $F_{1,54} = 180.42, p < 0.001, \eta^2 = 0.77$ and scenario fidelity $F_{1,54} = 39.26, p < 0.001, \eta^2 = 0.421$. Our analysis also revealed a significant interaction effect between interaction fidelity and display fidelity $F_{1,54} = 7.01, p = 0.008, \eta^2 = 0.115$. Our post-hoc tests for interaction fidelity revealed that virtual hands yielded significantly higher Total Presence Scores than the ray cast interaction technique $t_{54} = 13.43, p < 0.001$. Similarly, post-hoc tests for scenario fidelity revealed that consistent gravity yielded higher total presence scores than inconsistent gravity $t_{54} = 6.27, p < 0.001$. Table 6 shows the post-hoc tests for the interaction effect between interaction fidelity and display fidelity. Figure 5 also highlights the mean Total Presence Scores from our study.

FPS Item	Interaction Fidelity	Scenario Fidelity	Display Fidelity
FPS-1	0.40	0.18	0.77
FPS-2	0.32	0.20	0.84
FPS-3	0.20	0.41	0.69
FPS-4	0.34	0.87	0.35
FPS-5	0.86	0.21	0.23
FPS-6	0.89	0.20	0.19
FPS-7	0.73	0.11	0.34
FPS-8	0.51	0.38	0.36
FPS-9	0.58	0.20	0.28
FPS-10	0.14	0.12	0.60
Cronbach's Alpha	0.91	0.82	0.89

Table 5: The Structure Matrix after running EFA with an orthogonal rotation on the revised data. Values greater than |0.3| are in bold based on criteria from Hair et al. [25]. The final row reveals sufficient Cronbach's alpha values across each of the factors.

Condition Pairing 1	Condition Pairing 2	t-score	Significance
High Interaction x High Display	High Interaction x Low Display	$t_{54} = 3.23$	$p = 0.008$
High Interaction x High Display	Low Interaction x High Display	$t_{54} = 11.62$	$p < 0.001$
High Interaction x High Display	Low Interaction x Low Display	$t_{54} = 10.73$	$p < 0.001$
High Interaction x Low Display	Low Interaction x High Display	$t_{54} = 8.39$	$p < 0.001$
High Interaction x Low Display	Low Interaction x Low Display	$t_{54} = 7.50$	$p < 0.001$

Table 6: Post-hoc tests highlighting the interaction effect for interaction fidelity and display fidelity in regard to the Total Presence Score. Condition pairings in bold are condition parameters that received higher Interaction Presence Scores than their corresponding pairing in the same row.

4.9 Interaction Presence Score

For our Interaction Presence Score, we conducted an ANOVA, which found significance with regard to interaction fidelity level $F_{1,54} = 171.11, p < 0.001, \eta^2 = 0.76$. For high interaction fidelity conditions, we had a mean of 5.28 with standard deviation of 0.995. For low interaction fidelity conditions, we had a mean of 4.20 and a standard deviation of 1.22. Figure 6 highlights the mean Interaction Presence Scores from our validation study across low interaction and high interaction fidelity conditions.

4.10 Scenario Presence Score

Our ANOVA for the Scenario Presence Score revealed significance for scenario fidelity level employed $F_{1,54} = 8.6579, p < 0.001, \eta^2 = 0.138$. For high scenario fidelity conditions, we had a mean Scenario Presence Score of 5.04 with a standard deviation of 1.12. Conversely, the low scenario fidelity conditions had a mean of 4.78 and a standard deviation of 1.19. Figure 6 highlights the mean scores between the low scenario fidelity and high scenario fidelity conditions.

4.11 Display Presence Score

Lastly, in regard to Display Presence Score, our ANOVA revealed significance for level of display fidelity $F_{1,54} = 6.871, p < 0.001, \eta^2 = 0.113$. In high display fidelity conditions, the mean Display Presence Score was 4.58 with a standard deviation of 1.22. In low display fidelity conditions, the mean was 4.37 with a standard deviation of 1.18. Figure 6 shows the mean Display Presence Scores for the low display fidelity conditions and high display fidelity conditions.

5 Discussion

In this paper, we have demonstrated the capabilities that FPS provides with respect to measuring presence. Through our grounding in system fidelity, FPS ventures to strengthen the relationship between the fidelity of a VR system and presence. This is further supported by our reported results of our exploratory factor analysis and validation study. Additionally, the FPS is the result of consensus-driven design. In our Delphi study, we uncovered which questions our experts believe are the critical questions to ask when measuring presence. From our results in our Delphi study and validation study, we also present the implications of our results and future research directions.

5.1 Bridging the Gap between System Fidelity and Presence

In previous literature, there are instances where researchers have noted the disconnect between system fidelity factors and the sense of presence, particularly in how system fidelity can impact presence. Skarbez et al. [69] mentions that a “potential shortcoming of presence as a generalizable measure is that it does not account for the realism of the scenario being presented.” Souza et al. [72] also posits a similar thought wherein they mention a lack of clarity on whether or not there exists a relationship between the sense of presence and different technological factors.

Given this premise, we believe that FPS provides an opportunity to explicitly draw the connection between the fidelity features in a VE and the sense of presence. The system fidelity framework set by McMahan et al. [48, 51] affords us the ability to meaningfully

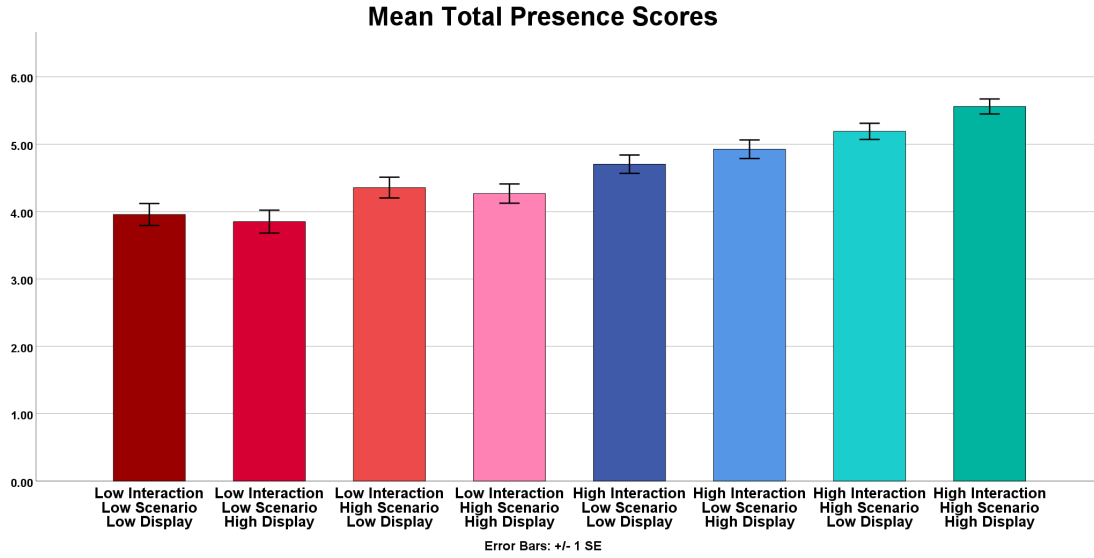


Figure 5: The mean Total Presence Scores across the 8 conditions of our validation study

categorize immersive characteristics of a VR system. As mentioned in section 1, we describe interaction fidelity, scenario fidelity, and display fidelity individually and how they work together to form a cohesive virtual experience. Through the FPS, we can now report on those individual components of the VR system that influence presence as we have shown in section 4.7. This can lead to better operationalization of presence as we can extend our results to discuss what components of the VE affected sense of presence and compare and contrast VEs from other studies.

The ability to discuss and break down presence measurement by system fidelity component is useful for being able to compare presence results from different environments. Jicol et al. [30] conducted a study in which they measured the sense of presence across four environment types. In their study, they modified the environment representation, the actions a participant can engage in, and the audio being played. In another study, McMahan et al. evaluated the display fidelity and interaction fidelity in a first-person shooter

environment [50]. Inherent in their study design, McMahan et al. evaluated the impact of changing fidelity on their participants' sense of presence. In both of these examples, a tool like FPS would have been beneficial as, in addition to measuring presence, they can also report presence in respect to each of the components of system fidelity regardless of the study conditions and research questions. The capability of being able to measure presence ratings within the components of system fidelity can lead to improved future work where researchers and developers can report and compare within their own and across others' VR environments.

5.2 Consensus-driven Survey Design

A key component to the development of FPS is that the finalized items are a result of applying the Delphi method to the presence domain. In previous work, the notion that there is not a single presence measurement tool or theoretical grounding of presence is a reoccurring theme [40, 69, 72]. A possible reason for this phenomenon is that there is no single presence instrument that is developed with consensus in mind. FPS represents a significant step forward in drawing consensus across the presence research community due to our Delphi study's diversity of input and output.

In regard to diversity of input, we note early on that the items in FPS originate from WS-PQ [81], IPQ [67], SUS [76], MPS [46], and SPES [26]. These surveys represent a broad range of intellectual schools of thought and theory in the realm of presence in virtual environments for VR. They each pose their own psychological factoring and theoretical grounding of presence as seen in Table 1. One of the aims of FPS was to draw out the best qualities of these questionnaires in the form of a new presence measurement tool.

By design of the Delphi method, an additional goal of the FPS was to gain diversity of output through our experts evaluating the inventory of each of the aforementioned questionnaires. Having

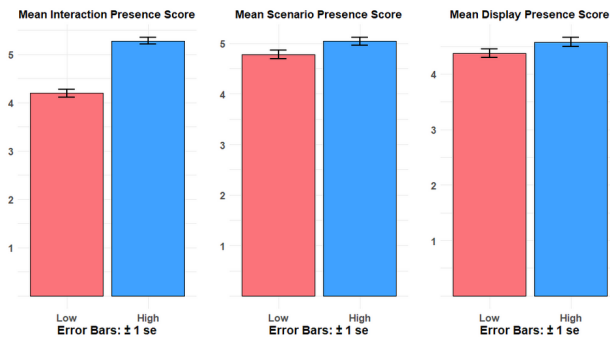


Figure 6: The mean Interaction, Scenario, and Display Presence Scores between low and high fidelity conditions.

multiple experts afforded us the opportunity to gain decades of research experience and differing perspectives. This approach differs greatly from previous questionnaires as most questionnaires are developed under the guidance of a singular lab.

5.3 System Fidelity and Plausibility Illusion

With system fidelity being applied to presence, we consider the possibility that system fidelity has close ties with another construct: Plausibility Illusion. In Slater's work [71], plausibility illusion is defined as the illusion that a scenario being depicted is actually occurring. In comparison, system fidelity is the objective degree in which real-world interactions or experiences are replicated by an interactive system [48, 51]. In our understanding of both of these concepts, plausibility illusion is linked more towards the subjective response a user has to the sensory stimuli present in a virtual experience [39, 71] whereas system fidelity is linked to the objective design decisions a developer/researcher included in a virtual experience. In regard to the FPS, plausibility illusion influences the responses individuals report on the survey. When exposed to a virtual environment, users will determine how believable the environment is to them, which in turn affects plausibility illusion and, subsequently, sense of presence. To further clarify, we consider the idea of a virtual environment set in a fantasy world, such as those highlighted in Rogers et al.'s literature review of realism in video games [63] (e.g. *Skyrim*). In a fantasy world, we have creatures such as centaurs, a creature with a body of a horse paired with the upper torso of a human, that can roam the environment. Inherently, centaurs do not exist in the real world and are therefore unrealistic. However, a user can still experience presence when engaged in a fantasy virtual environment. If the behaviors of the centaur match the real-world counterpart behavior and interaction (e.g. the human torso behaves like a human and the horse body behaves like a horse) it is possible for a user to experience a high level of presence and plausibility illusion. This previous case describes a high scenario fidelity experience as users would experience high physical, attribute, and behavioral coherence due to the centaur being developed based on real-world counterparts. On the converse side, if the centaur began roaming the world with flexible horse legs, similar to an octopus, we would be describing a low scenario fidelity experience as the physical coherence and attribute coherence are systematically changed.

5.4 Limitations and Future Work

One of the limitations of the FPS is the potential biases within our expert participant pool. First, within the 16 expert participants, there exists the possibility of biased perspectives among those who partook in the Delphi study. As highlighted in Table 2, our participant pool predominantly exists within North America and Europe, which may not effectively represent other cultures and perspectives. Initially, we contacted 90 potential experts with the intent of having a broad range of both intellectual and cultural perspectives. Due to factors outside of our control (e.g., scheduling), there was no feasible way to absolutely ensure diversity of culture into our Delphi study.

Another limitation is that the FPS needs to be applied to more studies and broader contexts. We acknowledge that our validation

study is in one task and context. However, in its current state, FPS is validated and ready to be used in future VR studies given our exploratory factor analysis. Future work will involve administering the FPS in a variety of environments and tasks, which will in turn contribute to a comprehensive confirmatory factor analysis in the future. In our study, we intentionally focused on extreme anchors for each fidelity component. For example, physical/attribute coherence within scenario fidelity and visual/auditory fidelity within Display Fidelity, to minimize potential confounds. Hence, we did not examine additional modifications to each of the system fidelity components, paving the path for future work to investigate the influence of these additional aspects to better generalize our findings and the application of FPS to different virtual environments.

An additional limitation is with regard to avatar embodiment in VR environments. Recent work, such as Lugrin et al. [45] examined how VR self-body avatars influence the sense of presence through the enhancement of body ownership and spatial awareness. However, their findings suggest that these effects might not be significant all the time especially for action-based VR games, where task performance and control take precedence over self avatar representations and embodiment. In our work, we emphasize on the sense of presence within the VE itself without the inclusion of avatar embodiment. This was done to ensure that our results generalize to a variety of VR experiences (e.g. ones without self-avatar representations). We also considered the possibility that the addition of self-body avatars could introduce confounding factors and additional complexities. We acknowledge the nuanced contributions of avatar embodiment [20–22, 32, 56, 59] and believe there is an opportunity to apply the FPS to scenarios that leverage self-body avatars.

6 Conclusion

In this paper, we have presented the development and validation of the Fidelity-based Presence Scale. In our developmental stage, we applied iterative, consensus-driven review to learn what questions experts from the VR research community deem most important. Additionally, the FPS was grounded in the system fidelity framework as we believe that future researchers and developers would greatly benefit from understanding how components of their VR applications and their fidelity impacted a user's sense of presence. The results of our Delphi study revealed a subset of 11 questions drawn from five of the most highly cited presence questionnaires for VR experiences. Following our Delphi study, we conducted a validation user study (n=55) and performed an exploratory analysis on the collected data. The results of our validation study showcased FPS' capability of identifying which component affects presence, and from our EFA, we finalized the inventory of the FPS to 10 items with their respective factor loadings. Between the Delphi study and validation methodology, we have shown the transition from theory to validation of the FPS. We also have shown how to utilize the FPS and how each component of system fidelity can be evaluated to provide additional insights and results regarding a user study. Through grounding the FPS in the system fidelity framework, our work represents a significant stride forward in presence measurement as the FPS affords the ability to inform on which components of an environment impacted the sense of presence. This capability

can lead to improved comparisons between VR environments and future design implications for both developers and researchers.

Acknowledgments

The authors would like to thank Delphi Study participants (who agreed to name acknowledgment and contribution. They are listed in alphabetical order): Dr. Sarker Asish, Dr. Sabarish Babu, Dr. Maximino Bessa, Dr. Gerd Bruder, Dr. Marcello Carrozzino, Dr. Nicola Döring, Dr. Regis Kopper, Dr. Ernst Kruijff, Dr. Miguel Melo, Dr. Anthony Steed, and Dr. Pamela Wisniewski.

References

- [1] Alex Adkins, Lorraine Lin, Aline Normoyle, Ryan Canales, Yuting Ye, and Sophie Jörg. 2021. Evaluating grasping visualizations and control modes in a vr game. *ACM Transactions on Applied Perception (TAP)* 18, 4 (2021), 1–14.
- [2] Michael Adler and Erio Ziglio. 1996. *Gazing into the oracle: The Delphi method and its application to social policy and public health*. Jessica Kingsley Publishers.
- [3] Solomon E Asch. 1956. Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological monographs: General and applied* 70, 9 (1956), 1.
- [4] Jakki O Bailey, Jeremy N Bailenson, and Daniel Casasanto. 2016. When does virtual embodiment change our minds? *Presence* 25, 3 (2016), 222–233.
- [5] Rosa María Baños, Cristina Botella, Azucena García-Palacios, Helena Villa, Concepción Perpiñá, and Mariano Alcaniz. 2000. Presence and reality judgment in virtual environments: a unitary construct? *CyberPsychology & Behavior* 3, 3 (2000), 327–335.
- [6] Woodrow Barfield and Suzanne Weghorst. 1993. The sense of presence within virtual environments: A conceptual framework. *Advances in Human Factors Ergonomics* 19 (1993), 699–699.
- [7] Maurice S Bartlett. 1950. Tests of significance in factor analysis. *British journal of psychology* (1950).
- [8] Frank Biocca. 1997. The cyborg's dilemma: Progressive embodiment in virtual environments. *Journal of computer-mediated communication* 3, 2 (1997), JCMC324.
- [9] Frank Biocca, Chad Harms, and Jenn Gregg. 2001. The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity. In *4th annual international workshop on presence*, Philadelphia, PA. 1–9.
- [10] Stéphane Bouchard, Geneviève Robillard, Julie St-Jacques, Stéphanie Dumoulin, Marie-Josée Patry, and Patrice Renaud. 2004. Reliability and validity of a single-item measure of presence in VR. In *The 3rd IEEE international workshop on haptic, audio and visual environments and their applications*. IEEE, 59–61.
- [11] Doug A Bowman and Larry F Hodges. 1995. User interface constraints for immersive virtual environment applications. *Atlanta: Graphics, Visualization and Usability* (1995).
- [12] Doug A Bowman and Larry F Hodges. 1997. An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In *Proceedings of the 1997 symposium on Interactive 3D graphics*. 35–ff.
- [13] Veronica Cateté and Tiffany Barnes. 2017. Application of the Delphi method in computer science principles rubric creation. In *Proceedings of the 2017 ACM conference on innovation and technology in computer science education*. 164–169.
- [14] Dustin B Chertoff, Brian Goldiez, and Joseph J LaViola. 2010. Virtual Experience Test: A virtual environment evaluation questionnaire. In *2010 IEEE virtual reality conference (VR)*. IEEE, 103–110.
- [15] Dongsik Cho, Jihye Park, Gerard Jounghyun Kim, Sangwoo Hong, Sungho Han, and Seungyong Lee. 2003. The dichotomy of presence elements: The where and what. In *IEEE Virtual Reality, 2003. Proceedings*. IEEE, 273–274.
- [16] Youjin Choe, Martin Tomko, and Mohsen Kalantari. 2023. Assessing Mapper Conflict in OpenStreetMap Using the Delphi Survey Method. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.
- [17] Lee J Cronbach. 1951. Coefficient alpha and the internal structure of tests. *psychometrika* 16, 3 (1951), 297–334.
- [18] Norman Dalkey and Olaf Helmer. 1963. An experimental application of the Delphi method to the use of experts. *Management science* 9, 3 (1963), 458–467.
- [19] Huong Q Dinh, Neff Walker, Larry F Hodges, Chang Song, and Akira Kobayashi. 1999. Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments. In *Proceedings IEEE Virtual Reality (Cat. No. 99CB36316)*. IEEE, 222–228.
- [20] Tiffany D Do, Juanita Benjamin, Camille Isabella Protoko, and Ryan P McMahan. 2024. Cultural Reflections in Virtual Reality: The Effects of User Ethnicity in Avatar Matching Experiences on Sense of Embodiment. *IEEE transactions on visualization and computer graphics* (2024).
- [21] Tiffany D Do, Camille Isabella Protoko, and Ryan P McMahan. 2024. Stepping into the Right Shoes: The Effects of User-Matched Avatar Ethnicity and Gender on Sense of Embodiment in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* (2024).
- [22] Tiffany D Do, Steve Zelenty, Mar Gonzalez-Franco, and Ryan P McMahan. 2023. Valid: a perceptually validated virtual avatar library for inclusion and diversity. *Frontiers in Virtual Reality* 4 (2023), 1248915.
- [23] Florian Echtler and Maximilian Häußler. 2018. Open source, open science, and the replication crisis in HCI. In *Extended abstracts of the 2018 CHI conference on human factors in computing systems*. 1–8.
- [24] Michael Gerhard, David J Moore, and Dave J Hobbs. 2001. Continuous presence in collaborative virtual environments: Towards a hybrid avatar-agent model for user representation. In *Intelligent Virtual Agents: Third International Workshop, IVA 2001 Madrid, Spain, September 10–11, 2001 Proceedings* 3. Springer, 137–155.
- [25] J Hair. 2009. Multivariate data analysis. *Exploratory factor analysis* (2009).
- [26] Tilo Hartmann, Werner Wirth, Peter Vorderer, Christoph Klimmt, Holger Schramm, and Saskia Böcking. 2015. Spatial presence theory: State of the art and challenges ahead. *Immersed in media: Telepresence theory, measurement & technology* (2015), 115–135.
- [27] Felicity Hasson, Sinead Keeney, and Hugh McKenna. 2000. Research guidelines for the Delphi survey technique. *Journal of advanced nursing* 32, 4 (2000), 1008–1015.
- [28] Yahya Hmaiti, Mykola Maslych, Amirpouya Ghasemaghahi, Ryan K Ghamandi, and Joseph J LaViola Jr. 2024. Visual Perceptual Confidence: Exploring Discrepancies Between Self-reported and Actual Distance Perception In Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* (2024).
- [29] Chia-Chien Hsu and Brian A Sandford. 2007. The Delphi technique: making sense of consensus. doi: 10.7275. PDZ9-TH90 (2007).
- [30] Crescent Jicol, Chun Hin Wan, Benjamin Doling, Caitlin H Illingworth, Jinha Yoon, Charlotte Headey, Christof Lutteroth, Michael J Proulx, Karin Petrini, and Eamonn O'Neill. 2021. Effects of emotion and agency on presence in virtual reality. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–13.
- [31] Anthony F Jorm. 2015. Using the Delphi expert consensus method in mental health research. *Australian & New Zealand Journal of Psychiatry* 49, 10 (2015), 887–897.
- [32] Sungchul Jung, Gerd Bruder, Pamela J Wisniewski, Christian Sandor, and Charles E Hughes. 2018. Over my hand: Using a personalized hand in vr to improve object size estimation, body ownership, and presence. In *Proceedings of the 2018 ACM Symposium on Spatial User Interaction*. 60–68.
- [33] Henry F Kaiser. 1970. A second generation little jiffy. (1970).
- [34] Henry F Kaiser and John Rice. 1974. Little jiffy, mark IV. *Educational and psychological measurement* 34, 1 (1974), 111–117.
- [35] Sinead Keeney, Felicity Hasson, and Hugh McKenna. 2006. Consulting the oracle: ten lessons from using the Delphi technique in nursing research. *Journal of advanced nursing* 53, 2 (2006), 205–212.
- [36] Taeyong Kim and Frank Biocca. 1997. Telepresence via television: Two dimensions of telepresence may have different connections to memory and persuasion. *Journal of computer-mediated communication* 3, 2 (1997), JCMC325.
- [37] Michael Krauss, Rainer Scheuchenflog, Walter Piechulla, and Alf Zimmer. 2001. Measurement of presence in virtual environments. *Experimentelle Psychologie im Spannungsfeld von Grundlagenforschung und Anwendung Proceedings* (2001), 358–362.
- [38] Pontus Larsson, Daniel Västfjäll, and Mendel Kleiner. 2001. The actor-observer effect in virtual reality presentations. *CyberPsychology & Behavior* 4, 2 (2001), 239–246.
- [39] Joseph J LaViola Jr, Ernst Kruijff, Ryan P McMahan, Doug Bowman, and Ivan P Poupyrev. 2017. *3D user interfaces: theory and practice*. Addison-Wesley Professional.
- [40] Kwan Min Lee. 2004. Presence, explicated. *Communication theory* 14, 1 (2004), 27–50.
- [41] Jane Lessiter, Jonathan Freeman, Edmund Keogh, and Jules Davidoff. 2001. A cross-media presence questionnaire: The ITC-Sense of Presence Inventory. *Presence: Teleoperators & Virtual Environments* 10, 3 (2001), 282–297.
- [42] JJ-W Lin, Henry Been-Lirn Duh, Donald E Parker, Habib Abi-Rached, and Thomas A Furness. 2002. Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. In *Proceedings IEEE virtual reality 2002*. IEEE, 164–171.
- [43] Matthew Lombard, Theresa B Ditton, Daliza Crane, Bill Davis, Gisela Gil-Egui, Karl Horvath, Jessica Rossman, and S Park. 2000. Measuring presence: A literature-based approach to the development of a standardized paper-and-pencil instrument. In *Third international workshop on presence, delft, the netherlands*, Vol. 240. 2–4.
- [44] Matthew Lombard, Theresa B Ditton, and Lisa Weinstein. 2009. Measuring presence: the temple presence inventory. In *Proceedings of the 12th annual international workshop on presence*. 1–15.
- [45] Jean-Luc Lugin, Maximilian Ertl, Philipp Krop, Richard Klüpfel, Sebastian Stierstorfer, Bianka Weisz, Maximilian Rück, Johann Schmitt, Nina Schmidt, and Marc Erich Latoschik. 2018. Any "Body" There? Avatar Visibility Effects in a Virtual Reality Game. In *2018 IEEE Conference on Virtual Reality and 3D User*

- Interfaces (VR)*. 17–24. <https://doi.org/10.1109/VR.2018.8446229>
- [46] Guido Makransky, Lau Lilleholt, and Anders Aaby. 2017. Development and validation of the Multimodal Presence Scale for virtual reality environments: A confirmatory factor analysis and item response theory approach. *Computers in Human Behavior* 72 (2017), 276–285.
- [47] Guido Makransky, Thomas S Terkildsen, and Richard E Mayer. 2019. Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and instruction* 60 (2019), 225–236.
- [48] Ryan P. McMahan. 2018. *Virtual Reality System Fidelity*. Springer International Publishing, Cham, 1–8. https://doi.org/10.1007/978-3-319-08234-9_251-1
- [49] Ryan P McMahan. 2024. Virtual reality system fidelity. In *Encyclopedia of computer graphics and games*. Springer, 2041–2048.
- [50] Ryan P McMahan, Doug A Bowman, David J Zielinski, and Rachael B Brady. 2012. Evaluating display fidelity and interaction fidelity in a virtual reality game. *IEEE transactions on visualization and computer graphics* 18, 4 (2012), 626–633.
- [51] Ryan P McMahan and Nicolas S Herrera. 2016. AFFECT: altered-fidelity framework for enhancing cognition and training. *Frontiers in ICT* 3 (2016), 29.
- [52] Stanley Milgram. 1963. Behavioral study of obedience. *The Journal of abnormal and social psychology* 67, 4 (1963), 371.
- [53] M Minsky. 1980. Telepresence. *Omni* (1980).
- [54] Craig D Murray, Paul Arnold, and Ben Thornton. 2000. Presence accompanying induced hearing loss: Implications for immersive virtual environments. *Presence* 9, 2 (2000), 137–148.
- [55] Sarah Nichols, Clovissa Haldane, and John R Wilson. 2000. Measurement of presence and its consequences in virtual environments. *International Journal of Human-Computer Studies* 52, 3 (2000), 471–491.
- [56] K Nowak. 2003. The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments.
- [57] Abdullah Pandor, Eva Kaltenthaler, Marrissa Martyn-St James, Ruth Wong, Katy Cooper, Munya Dimairo, Alicia O' Cathain, Fiona Campbell, and Andrew Booth. 2019. Delphi consensus reached to produce a decision tool for Selecting Approaches for Rapid Reviews (STARR). *Journal of Clinical Epidemiology* 114 (2019), 22–29.
- [58] Anne Parent. 1998. A virtual environment task analysis workbook for the creation and evaluation of virtual art exhibits. *National Research Council of Canada-Reports-ERB* (1998).
- [59] Tabitha C Peck and Mar Gonzalez-Franco. 2021. Avatar embodiment: a standardized questionnaire. *Frontiers in Virtual Reality* 1 (2021), 575943.
- [60] Ivan Poupyrev, Mark Billinghurst, Suzanne Weghorst, and Tadao Ichikawa. 1996. The go-go interaction technique: non-linear mapping for direct manipulation in VR. In *Proceedings of the 9th annual ACM symposium on User interface software and technology*. 79–80.
- [61] Ivan Poupyrev, Tadao Ichikawa, Suzanne Weghorst, and Mark Billinghurst. 1998. Egocentric object manipulation in virtual environments: empirical evaluation of interaction techniques. In *Computer graphics forum*, Vol. 17. Wiley Online Library, 41–52.
- [62] Eric D Ragan, Doug A Bowman, Regis Kopper, Cheryl Stinson, Siroberto Scerbo, and Ryan P McMahan. 2015. Effects of field of view and visual complexity on virtual reality training effectiveness for a visual scanning task. *IEEE transactions on visualization and computer graphics* 21, 7 (2015), 794–807.
- [63] Katja Rogers, Sukran Karaosmanoglu, Maximilian Altmeyer, Ally Suarez, and Lennart E Nacke. 2022. Much realistic, such wow! a systematic literature review of realism in digital games. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–21.
- [64] Gene Rowe and George Wright. 1999. The Delphi technique as a forecasting tool: issues and analysis. *International journal of forecasting* 15, 4 (1999), 353–375.
- [65] Corina Sas and Gregory MP O'Hare. 2003. Presence equation: An investigation into cognitive factors underlying presence. *Presence* 12, 5 (2003), 523–537.
- [66] Ralph Schroeder, Anthony Steed, Ann-Sofie Axelsson, Ilona Heldal, Åsa Abelin, Josef Wideström, Alexander Nilsson, and Mel Slater. 2001. Collaborating in networked immersive spaces: as good as being there together? *Computers & Graphics* 25, 5 (2001), 781–788.
- [67] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. 2001. The experience of presence: Factor analytic insights. *Presence: Teleoperators & Virtual Environments* 10, 3 (2001), 266–281.
- [68] Daniel M Shafer, Corey P Carbonara, and Michael F Korpi. 2019. Factors affecting enjoyment of virtual reality games: a comparison involving consumer-grade virtual reality technology. *Games for health journal* 8, 1 (2019), 15–23.
- [69] Richard Skarbez, Frederick P Brooks, Jr, and Mary C Whitton. 2017. A survey of presence and related concepts. *ACM Computing Surveys (CSUR)* 50, 6 (2017), 1–39.
- [70] Richard Skarbez, Solene Neyret, Frederick P Brooks, Mel Slater, and Mary C Whitton. 2017. A psychophysical experiment regarding components of the plausibility illusion. *IEEE transactions on visualization and computer graphics* 23, 4 (2017), 1369–1378.
- [71] Mel Slater. 2009. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364, 1535 (2009), 3549–3557.
- [72] Vinicius Souza, Anderson Maciel, Luciana Nedel, and Regis Kopper. 2021. Measuring presence in virtual environments: A survey. *ACM Computing Surveys (CSUR)* 54, 8 (2021), 1–37.
- [73] Jari Takatalo. 2002. Presence and flow in virtual environments: An explorative study. *Master's thesis, University of Helsinki* (2002).
- [74] Jari Takatalo, Jukka Häkkinen, H Särkelä, J Komulainen, and Göte Nyman. 2004. The experiential dimensions of two different digital games. *Proc. PRESENCE 2004* (2004), 274–278.
- [75] Tanh Quang Tran, Tobias Langlotz, and Holger Regenbrecht. 2024. A Survey On Measuring Presence in Mixed Reality. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–38.
- [76] Martin Usoh, Ernest Catena, Sima Arman, and Mel Slater. 2000. Using presence questionnaires in reality. *Presence* 9, 5 (2000), 497–503.
- [77] Joy Van Baren. 2004. Measuring presence: A guide to current measurement approaches. *Deliverable of the OmniPres project IST-2001-39237* (2004).
- [78] Peter Vorderer, Werner Wirth, Feliz Ribeiro Gouveia, Frank Biocca, Timo Saari, Lutz Jäncke, Saskia Böcking, Holger Schramm, Andre Gysbers, Tilo Hartmann, et al. 2004. Mec spatial presence questionnaire. Retrieved Sept 18, 2004 (2004), 2015.
- [79] David Weibel, Jan Schmutz, Olivier Pahud, and Bartholomäus Wissmath. 2015. Measuring spatial presence: Introducing and validating the pictorial presence SAM. *Presence: Teleoperators and Virtual Environments* 24, 1 (2015), 44–61.
- [80] Werner Wirth, Tilo Hartmann, Saskia Böcking, Peter Vorderer, Christoph Klimmt, Holger Schramm, Timo Saari, Jari Laarni, Niklas Ravaja, Feliz Ribeiro Gouveia, et al. 2007. A process model of the formation of spatial presence experiences. *Media psychology* 9, 3 (2007), 493–525.
- [81] Bob G Witmer and Michael J Singer. 1998. Measuring presence in virtual environments: A presence questionnaire. *Presence* 7, 3 (1998), 225–240.
- [82] Jacob O Wobbrock, Leah Findlater, Darren Gergle, and James J Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 143–146.
- [83] Marie BH Yap, Pamela D Pilkington, Siobhan M Ryan, Claire M Kelly, and Anthony F Jorm. 2014. Parenting strategies for reducing the risk of adolescent depression and anxiety disorders: a Delphi consensus study. *Journal of affective disorders* 156 (2014), 67–75.