

# When Do We Feel Present in a Virtual Reality? Towards Sensitivity and User Acceptance of Presence Questionnaires

Annalisa Degenhard

Media informatics

University of Ulm

Ulm, Germany

annalisa.degenhard@uni-ulm.de

Michael Rietzler

Ulm University

Institute of Mediainformatics

Ulm, Germany

michael.rietzler@uni-ulm.de

Ali Askari

Media Informatics

Ulm University

Ulm, Germany

ali.askari@uni-ulm.de

Enrico Rukzio

University of Ulm

Ulm, Germany

enrico.rukzio@uni-ulm.de



Figure 1: The five scenarios used in our study were designed to manipulate the intensity of certain dimensions of presence.

## Abstract

Presence is an important and widely used metric to measure the quality of virtual reality (VR) applications. Given the multifaceted and subjective nature of presence, the most common measures for presence are questionnaires. But there is little research on their validity regarding specific presence dimensions and their responsiveness to differences in perception among users. We investigated four presence questionnaires (SUS, PQ, IPQ, Bouchard) on their responsiveness to intensity variations of known presence dimensions and asked users about their consistency with their experience. Therefore, we created five VR scenarios that were designed to emphasize a specific presence dimension. Our findings showed heterogeneous sensitivity of the questionnaires dependent on the different dimensions of presence. This highlights a context-specific

suitability of presence questionnaires. The questionnaires' sensitivity was further stated as lower than actually perceived. Based on our findings, we offer guidance on selecting these questionnaires based on their suitability for particular use cases.

## CCS Concepts

• Human-centered computing → Empirical studies in HCI; HCI theory, concepts and models.

## Keywords

virtual reality, presence, plausibility, place illusion, embodiment, social presence, involvement, questionnaires



This work is licensed under a Creative Commons Attribution 4.0 International License. CHI '25, Yokohama, Japan

© 2025 Copyright held by the owner/author(s).

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

<https://doi.org/10.1145/3706598.3714204>

## ACM Reference Format:

Annalisa Degenhard, Ali Askari, Michael Rietzler, and Enrico Rukzio. 2025. When Do We Feel Present in a Virtual Reality? Towards Sensitivity and User Acceptance of Presence Questionnaires. In *CHI Conference on Human Factors in Computing Systems (CHI '25)*, April 26–May 01, 2025, Yokohama, Japan. ACM, New York, NY, USA, 29 pages. <https://doi.org/10.1145/3706598.3714204>

## 1 Introduction

The popularity of virtual reality has increased significantly over the last three decades. Allowing users to leave the real world and immerse into a virtual world, Virtual Reality (VR) has become popular not only in gaming but also in the industrial [13, 48] and educational sectors [47]. Many studies observed a positive effect of VR on enjoyment, focus, and learning outcomes. VR allows the filtering of distracting factors and leads the user's focus on essential aspects. To be able to maximize the positive effect of fully immersing in the virtual world, research tried to identify what aspects of a VR experience influence the feeling of being there, which has been defined as *presence* [16]. By clustering the influencing factors, *dimensions of presence* were defined. The number of dimensions and the subjective character of presence make it difficult to measure it objectively. Therefore, questionnaires have become a common way to determine users' perceived presence. The most popular questionnaires regarding the number of citations in recent years [15] as well as in general [12] are the SUS questionnaire [43, 51] by Slater, Usoh, and Steed, the IPQ (IGroup Presence Questionnaire) questionnaire [33, 34] by Schubert and the PQ (Presence Questionnaire) [58] by Witmer & Singer. The questionnaires are widely used today to gain insight into the perceived presence of users in a virtual environment (VE). While the subjective nature of presence does not allow for a cross-study comparison of presence scores, insights on presence are often gathered by comparing different versions of a VR scenario and assigning differences in presence scores to the differences between the scenarios. However, presence questionnaires rarely include guidelines for their applicability to different VR solutions. In fact, they have been developed based on different understandings of presence and for different scenarios. Previous investigations have already drawn attention to the effects of questionnaire choice and the setting of a presence assessment on the outcome [15, 35, 51]. The context and wording, for example, may influence what a particular VR experience is compared to and, hence, influence the quality of a presence assessment. A lack of comprehension regarding the effects caused by design choices for a presence assessment may lead to misinterpretation and erroneous conclusions. With respect to that, we consider two limitations of the use of questionnaires in presence research:

The first one is that presence questionnaires are rarely analyzed for their *sensitivity to certain dimensions of presence*. This makes it unclear if a questionnaire considers all dimensions of presence the researchers want to consider or that influences a user's experience respectively. Especially when it comes to comparative studies, as earlier mentioned, presence researchers would want to exclude the possibility that missing differences in presence scores might simply be caused by the chosen questionnaire not being sensitive to changes in the respective presence dimension. Considering the different underlying concepts of presence questionnaires, we suspect that the choice of the questionnaire could potentially bias the determination of the users' perceived presence by not being sensitive to some included presence dimensions of a virtual world. Single-item presence scales (SIP) like the one by Bouchard et al. [5] seem more neutral. However, in many cases, the item still contains a definition of presence, thus a certain bias on the user's part.

Second, it is rarely investigated whether *respondents approve* to the scores of the questionnaires. Remembering the subjective nature of presence, not only being influenced by technical aspects but also by individual factors like suspension of disbelief [52] or motivation to get immersed [20], and considering the difficulty of assessing such an elusive phenomenon - not only for experienced presence researchers but especially for respondents who are, in many cases, not familiar with the concept - we think that it is important to consider not only sensitivity but also consistency of scores with the respondents' actual experiences.

Our goal was to investigate the ability of different presence questionnaires to assess changes in specific dimensions of presence. Therefore, we investigated (1) the *sensitivity of presence questionnaires to known dimensions of presence* on the one hand and (2) the users' approval of the resulting scores considering their actual experience on the other hand.

*Sensitivity of Presence Questionnaires.* Since there are no guidelines on which presence questionnaire is best suited for a specific VR solution and on their individual strengths and weaknesses, it is still not clear which one to choose to determine users' presence with the highest validity. An adverse choice could cause a dimension of presence relevant to the respective VR solution not to be considered in the questionnaire. Therefore, results could lead to misinterpretations. We investigated the sensitivity of SUS, IPQ, PQ, and SIP to changes in the intensity of specific presence dimensions. To understand the potential bias by the choice of presence questionnaires for a VR solution, we wanted to answer three questions:

- (1) How sensitive are they to changes in specific dimensions of presence? (RQ1)
- (2) Do users agree to the differences in scores affected by the changes of a specific dimension of presence? (RQ2)
- (3) Can questionnaire items be clustered according to their sensitivity to changes in a specific dimension of presence? (RQ3)

We picked five dimensions of presence based on two presence models by Skarbez from 2017 [39] and Hein from 2018 [15]. Skarbez's model combines multiple presence models based on a systematic literature review of existing presence research (331 citations from 2018-2020). We additionally chose Hein's model because it considers further presence dimensions from popular presence models by Schubert (2001, [34], 1075 citations from 2018-2022) and Slater (2003, [40] 619 citations from 2018-2022) that Skarbez chose not to include as stand-alone dimensions. This resulted in the dimensions of place illusion, plausibility, involvement, user embodiment, and social presence. To gain insight into the sensitivity of the four questionnaires to changes in one of the five dimensions, we designed a VR scenario for each dimension that emphasized aspects of that dimension while reducing aspects of other dimensions to a minimum. We then altered its intensity within the scenario, resulting in two conditions per dimension: one for the high-intensity scenario and one for the low-intensity scenario. We asked users to complete all questionnaires for each condition and compared the resulting scores.

Our findings indicate that the sensitivity of the questionnaires to changes in the presence dimension varies. The SUS and IPQ showed high sensitivity to changes in place illusion, whereas the PQ was



most responsive to changes in plausibility. The SIP questionnaire showed the highest sensitivity across all dimensions but with a notable standard deviation in scores. None of the four questionnaires were particularly sensitive to the involvement, social presence, or embodiment dimensions. Our results suggest that researchers should carefully select a questionnaire based on the focus of their study to ensure sensitivity and validity. Thoughtful questionnaire choice is crucial when investigating specific aspects of presence to avoid misinterpretation.

*Approval of Sensitivity.* Since the high sensitivity of questionnaires to changes must not imply consistency with the users' actual perceptions, we further investigated whether users agreed with the scores. Therefore, we asked participants whether the score differences between the two conditions of each scenario complied with their expectations. We found that, in many cases, participants would have expected greater changes. Only in the involvement and embodiment scenarios were the score differences of certain questionnaires higher than users expected. This not only underscores our findings in the previous research question that questionnaires may lack sensitivity when it comes to certain dimensions of presence but also confirms our assumption that greater sensitivity may not imply a realistic reflection of changes in scores.

Our evaluation pointed out the significant effects of the choice of presence questionnaires on the validity of presence assessments and provided various insights about the sensitivity of the questionnaires. The main findings of our investigations are:

- (1) Findings on the sensitivity of presence questionnaires to different presence dimensions that should be considered for the choice of a questionnaire in the future.
- (2) A validation of the questionnaires based on the subjective agreement of participants regarding their sensitivity to changes in individual presence dimensions.
- (3) Findings about the sensitivity of individual questionnaire items to changes in different presence dimensions that should be considered when interpreting scores.

## 2 Related Work

Presence is the phenomenon of experiencing a virtual world and getting so involved in it that one starts to think that they are actually in the virtual world, often described as the "sense of being there" [25]. Definitions of presence in research consider different aspects of presence, such as physical location, the sense of being elsewhere, their concurrent discrepancy, consciousness, awareness, or the users' actions. Common approaches mention the mental location of the user in the other place, meaning the virtual world [25, 27]. Slater et al., for example, defined presence in 1997 as "a state of consciousness, the (psychological) sense of being in the virtual environment" [45]. Other definitions also focus on the aspect of consciousness but with regard to the awareness of the real world. Seth et al. defined presence as "the sense of now being in a VE while transiently unaware of one's real location and the technology delivering the sensory input and recording the motor output." [36]. Witmer and Singer also based their approach on a user's perceived location but refined it by saying that the user has to believe what

he is seeing to be able to be present. In 1994, they defined presence as the "subjective experience of being in one place when one is physically in another" and supplemented their definition by stating that "Presence is the result of a continuously updated interior model of the environment, stressing the necessity for suspension of disbelief" [58]. This concept was one of the first to consider the perception of the VE. In 2004, Lee proposed a concept that focused entirely on the user's perception of the VE. He defined presence as "a psychological state in which virtual objects are experienced as actual objects in either sensory or non-sensory ways" [22]. Sanchez and Flach provided a more functional perspective. They focused their definition on the users' ability to interact with a VE: "[...] the key to the approach is that the sense of 'being there' in a VE is grounded on the ability to 'do there' [32]. Flach et al. considered a more general perspective on the functionality of a VE: "[...] the reality of experience is defined relative to functionality, rather than to appearances" [9]. Concepts of presence are of various nature. Providing a certain definition of presence is likely to influence the user's perspective on the phenomenon. Consequently, definitions in questionnaires are likely to bias the user's rating and should be taken into account when it comes to the choice of questionnaires.

### 2.1 Presence Models

Many aspects of VR experiences have been explored on possible correlations with presence [10, 25, 31, 38, 49, 56, 59]. While factors of immersion seem to influence presence in most VR experiences, it is still unclear which user-specific aspects influence presence and to what extent. For some, auditory cues might be the most relevant aspects of virtual worlds to be able to feel present. For others, it may be visual cues. The facets of an individual that can influence a VR experience are various and may even vary per individual and experience. For this reason, it is difficult to provide an exact model of presence. Different approaches have been published over the years. We will use the term *dimensions of presence* in the following to describe clusters of aspects that are assumed to influence presence according to previous work.

#### *Historical Foundations of Presence Models*

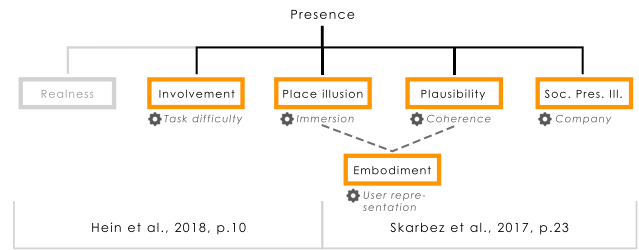
Sheridan et al. (1992) proposed one of the first models. They identified five dimensions that support the feeling of telepresence: the extent of sensory information, the control of sensors relative to the VE, the ability to modify the environment, task difficulty and the degree of automation. Since telepresence was defined to describe the phenomenon of presence in the context of workstations [37], Sheridan et al. focused their approach primarily on factors of immersion. However, the concept has similarities with models of presence in general. Zeltzer proposed a comparable model in 1992 [60]. For him, the quality of a simulation system is influenced by the degree of a user's autonomy, the quality of interaction, and presence. Steuer (1992) was the first to differentiate between technical aspects and the resulting subjective experience [49]. According to Steuer, the ability to perceive telepresence is influenced by the quality of sensory feedback (vividness) and interaction (interactivity). They split up the concept of telepresence into immersion as the technical aspect of a VE and presence as the subjective experience of an

immersive system. Since then, several models of presence dimensions have been proposed that focus more on the user experience of a VE [45]. Lombard and Ditton categorized five dimensions of presence: spatial presence, sensory presence, social realism, social presence, and engagement [25]. Similarly, Heeter et al. [14] stated that presence consists of personal presence, social presence, and environmental presence.

**Psychological Models** With the publication of more perceptual perspectives on presence, psychological approaches began to appear. One was published by Waterworth et al. in 2001 [54]. They considered presence to be influenced by Locus (whether a user's attention lies on the virtual or real world), Focus (the user's state on a continuum scale from being absent to being present), and Sensus (the user's arousal on a continuum scale from unconscious to conscious). Their concept shows similarities to Lombard and Ditton's engagement but focuses entirely on perceptual involvement with the VE.

**Differentiation between Place Illusion and Plausibility** A frequently used model in the last years has been proposed by Slater, who extended his concept of presence [45] by introducing place illusion (PI) and plausibility (Psi) and by noting the relevance of the illusion of body ownership [41, 42]. He criticized the lack of specification of the "there" in the definitions of presence and the consideration of the individual differences of users. PI is defined as "the illusion of being at the place depicted by the VR". Orthogonal to PI, Psi describes "the illusion that the virtual situations and events are really happening" [42]. Comparable to plausibility, Weber et al. proposed a presence model which states that presence occurs "if a mediated environment captures and maintains the attention of users and is perceived as realistic [55]. The difference is that realism is "not just an illusion but rather the result of a conscious evaluation of the credibility and realness of the VE". Lee et al. [22] considered presence to be composed of social presence, physical presence, and self presence. In this context, he defined social presence as "the sense of being together with another and mental models of other intelligences (i.e., people, animals, agents, gods, etc.) that help us simulate other minds", physical presence as "a psychological state in which virtual (para-authentic or artificial) physical objects are experienced as actual physical objects in either sensory or non-sensory ways." and self presence as "a psychological state in which virtual (para-authentic or artificial) self/selves are experienced as the actual self in either sensory or non-sensory ways". Again, physical presence can be compared with place illusion and self-presence with body ownership.

**Combined Presence Models** In their 2017 literature review, Skarbez et al. identified further similarities between prominent models of presence [39] and derived their own model. As this model combines multiple concepts of presence, we decided to consider it for our investigations. The dimensions considered from Skarbez's model can be seen in Figure 2 on the right. However, Skarbez's model does not consider attentional aspects that have been part of several presence models [23, 34, 58]. Skarbez notes that the term is used differently in the literature and needs to be defined more precisely. However, since we consider attentional resources to influence presence, we



**Figure 2: Overview of the presence dimensions that we chose for our investigations split by the two presence models considered. On the right: Dimensions of the presence model by Skarbez et al. (2017, p.23), which summarizes previous models based on a literature review [39]. On the left: Dimensions of the presence model by Hein et al. (2018, p.10), which summarizes the presence models of Slater and Schubert et al. [15].**

decided to include a second model of presence further.

Hein et al. [15] proposed a model for presence that combined Slater's distinction between immersion and presence [46] with Schubert et al.'s model. Based on the work from Witmer and Singer in 1998 and 1999 [57, 59], Slater considered presence to result from immersion and user characteristics [40]. Based on this concept, Schubert, Regenbrecht, and Friedman (2001) conducted a factor analysis. They used items from several presence questionnaires and three first-order presence factors, including involvement, spatial presence, and realness [34]. Involvement describes the user's focus during the mediated experience. Whether they focus more on the virtual or real environment and how captivated they are by the virtual environment. The factor of spatial presence is about the actual sense of being there. The factor loads high when the users perceive the mediated environment as the actual environment in which they are located and with which they can interact. The realness factor summarizes the consistency of events in the virtual environment with real-world experiences and the extent to which the virtual environment is perceived as real. Hein et al. combined these results with Slater's approach to a presence model. The considered presence dimensions of this model can be seen in Figure 2 on the left.

**Comparison of the Presence Models** The models by Skarbez and Schubert share multiple similarities. Slater's place illusion by Slater can be compared with the concept of spatial presence by Schubert. Both concepts describe the illusion of actually being in the environment depicted by the media system. Plausibility and realness also share similarities, but plausibility is more precise. While realness describes the consistency with real-world experiences in general, which leaves room for interpretation, plausibility is defined as the illusion that the depicted scenario is actually occurring. Slater focuses his concept on the believability of events happening in the virtual scenario, opening up the stiff concept of realness to experiences that may not be realistic but are reasonable. Slater states that as long as the users are able to construct a mental model of the virtual environment, they will be able to experience presence. Involvement and social presence may have similarities, but these

concepts should be considered distinct. One can perceive a virtual avatar as a social companion in the VE but not interact with it and thus not necessarily be involved. On the other hand, one can be involved in another task in the VE while being alone. Thus, the user may be involved but not necessarily have the illusion of social presence. Both concepts state the influence of immersion on presence. Schubert further explicitly considers user characteristics as an explicit factor. This differs from Skarbez's concept. It is not explicitly considered, but user characteristics influence the transitions from immersion, coherence, and company to the subsequent concepts. Individual differences will influence whether and how one experiences place illusion when exposed to a certain level of immersion.

*Choice of presence models for this study:* We decided to consider both models of presence in our study. Both are based on prominent models of presence, and we were interested in gaining insight into both involvement and social presence. Therefore, we chose the following dimensions of presence: place illusion, plausibility, social presence, user embodiment, and involvement. We used place illusion and plausibility as defined by Slater et al. [41]. To investigate the effects of changes in place illusion, we manipulated immersion as a direct factor according to the model by Skarbez et al. [39]. Regarding plausibility, we focused on the consistency of events in the VE with expectations. We thus decided to manipulate coherence. Events that were inconsistent with expectations would likely make the user question their plausibility, while expected events should improve plausibility. We decided to further investigate Slater's concept of body ownership as a separate dimension. As previously mentioned, Slater does not consider it a separate dimension but rather a component of place illusion and plausibility. However, as the direct influence of a user representation on presence has been shown by many researchers, e.g. [2, 7, 30, 61], we expected that this separation would provide additional interesting insights. We will refer to this dimension as user embodiment. To investigate the effects of changes in user embodiment, we altered the representation of the user in the VE. We distinguished between social presence and involvement to investigate the effects of demanding tasks and the presence of other individuals on the different presence scores. Since the questionnaires chosen for this work treat involvement and social presence differently, we expected that this distinction would provide more meaningful insights.

## 2.2 Presence Questionnaires

Many approaches to measuring presence have been introduced over the years. Due to the scope of our project, we decided to limit our examination to some of the most significant surveys in terms of their frequency of use. The questionnaires considered are listed in the table 2. We selected the following four presence questionnaires:

- (1) *The Slater-Usch-Steed questionnaire (SUS)* by Slater, Usch, and Steed (2000) measures presence with six items. Unlike other presence questionnaires, the SUS questionnaire focuses on the subjective experience rather than specific aspects of presence [43, 51]. Although published in 2000, the SUS is still one of the most frequently used presence questionnaires.

- (2) *The Presence Questionnaire (PQ)* by Witmer and Singer (1998) [59] consists of 19 items with four subscales: involvement/control, sensory fidelity, naturalness, and interface quality. According to Hein et al., it is still by far the most widely used questionnaire in terms of the number of citations.
- (3) *The IGroup Presence Questionnaire (IPQ)* developed by Schubert (2001) consists of 14 questions with subscales on spatial presence, involvement and realism [33, 34]. The IPQ focuses less on the design aspects of the VR system and the VE but rather on the user's perception of the VE.
- (4) *The single-item presence scale (SIP)* created by Bouchard (2004) [5] was also included in our research. The single-item approach may help to isolate the effects of the VR content itself on presence perceptions, as any variation in presence ratings can be more confidently attributed to the content rather than to variations in the measurement tool.

Our decision was based on literature reviews on the use of presence questionnaires in the years 2016-2017 [15] and 2002-2019 [12]. In addition, we aimed to use questionnaires with distinct dimensions to identify potential advantages in different use scenarios.

To summarize, we decided to use a combination of Skarbez's presence model [39] and by Hein et al.'s [15] (Figure 2). This resulted in a total of five dimensions, hereafter referred to as the *place illusion* dimension (PI), the *plausibility* dimension (PSI), the *involvement* dimension (INV), the *social presence* dimension (SOC), and the *user embodiment* dimension (EMB) in the following. Our goal was to evaluate the sensitivity of our four picked questionnaires for these dimensions. Some of the questionnaires' underlying concepts consider parts of the dimensions, and some are not considered by any of the questionnaires, such as social presence. If we find differences in sensitivity and user approval, this would mean that the choice of the questionnaire would introduce a bias to presence assessments and that presence research should try to raise awareness for that. More refined guidelines for the choice of questionnaires would be needed to support a more user-centered and case-sensitive choice of questionnaires and promote the validity of presence assessments.

Therefore, we designed methods to investigate isolated effects of changes in a specific presence dimension and for the assessment of users' approval. The former is challenging as changes in assessed scores must be attributable to the considered dimension to derive meaningful insights for dimension-specific sensitivity. The latter has the difficulty that it should be avoided to ask users to rate absolute presence scores as the subjective phenomenon is difficult to quantify. The following section explains our approach to assessing dimension-specific sensitivity and user approval concerning the challenges provided by the subjective construct.

## 3 Study Design

The study aimed to evaluate the sensitivity of common presence questionnaires to changes in a specific dimension and how well this corresponds to the actual user experience. The selection of the four questionnaires resulted in a total of 40 questions (see Appendix: SUS: Table 3, PQ: Table 4, IPQ: Table 5, SIP: Table 6). We did not examine the sensitivity of the subscales of the IPQ because we wanted to focus our work on measuring the overall concept of presence. To gain insight into how closely participants agreed with the results

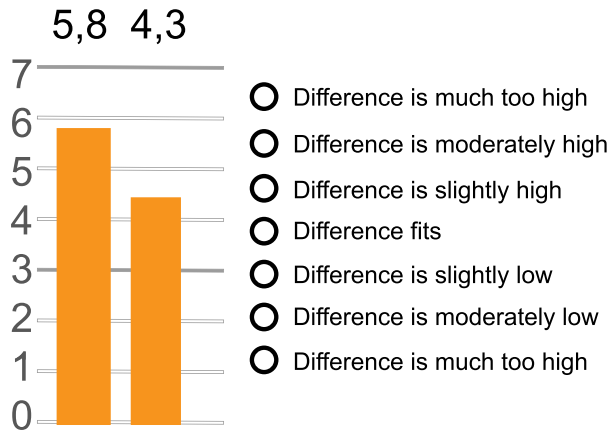


Figure 3: Example of a paired bar chart showing the scores of positive and negative condition. The score of each bar was shown above the bar. A radio button list on the right allowed users to state their approval.

of particular questionnaires, we would have had to introduce additional concepts to the participants to help them understand the terminology. This would have affected the participants' ratings of the overall presence scores, as we would have introduced a bias for certain aspects of presence. These five dimensions and four presence questionnaires formed the basis of our study design to answer our research questions.

### 3.1 Sensitivity of Presence Questionnaires

Our aim was to gain insight into the relevance of the four questionnaires for different use cases in VR. To achieve this, we designed scenarios that allowed us to isolate one of the five chosen dimensions at a time. This allowed us to understand the relationships between changes in a particular presence dimension and changes in overall presence scores. In designing the changes in each scenario, we had to take into account that the boundaries of dimensions for presence are unclear, and modifications to certain aspects of one dimension could also affect others. Therefore, we designed the scenarios to focus the experience on a single dimension with as little impact on the other dimensions as possible. We developed two versions of each scenario - one in which the intensity of the dimension is reduced to a minimum, referred to as the *negative* condition, and another with a high intensity, referred to as the *positive* condition in the following. Minimum, in this case, refers to the reduction of these dimensions to an intensity that is necessary to prevent participants from being distracted by their absence. Aspects of the experience related to dimensions that are not in focus should remain unchanged between conditions. The detailed design decisions for each dimension and scenario are described below.



Figure 4: Left: Dimension of the real room and the object placed in it. Middle & right: Photos of the study room and the interaction space from the right back corner and the right front corner of the room. a) The cabinet that is recreated in the virtual scenario, b) The scented candle that creates the olfactory stimuli, c) The fan that creates the haptic stimuli, d) The doors that are placed in the same position in the virtual scenario.

### 3.2 Approval of Sensitivity

As mentioned earlier, we also wanted to gain insight into the participants' agreement with the presence scores obtained from the various questionnaires. The difficulty with this investigation was that due to the subjective nature of presence, which is influenced by a variety of factors, it can be anticipated that participants will have difficulty quantifying an exact value of the difference in presence that they actually perceived. Consequently, we asked participants to rate the differences in scores on all four questionnaires after each scenario with both conditions. The scores of the negative and positive conditions were presented as bars in a paired bar chart (Figure 3). To counterbalance any effects, the order of the four paired bar charts was randomized, and participants were not told which questionnaire each chart corresponded to. They were then asked to rate how much the difference between the scores corresponded to their actual perception of the differences between the conditions. Therefore, they were asked to rate the difference on a 7-point Likert scale ranging from -3 = "difference is much too low" to 3 = "difference is much too high", with 0 indicating that the difference corresponded to the participant's actual perception.

### 3.3 Scenario Design

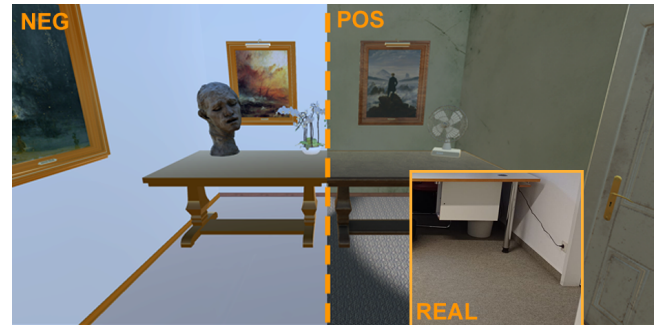
We investigated a questionnaire's dimension-specific sensitivity and users' approval using these two approaches, tailored to the difficult, elusive nature of presence. We restricted our evaluation to the introduced five dimensions and four questionnaires, but in theory, the approach can be applied to any other constellation as well.

The study was conducted in a room measuring 3.40 m × 4.60 m, divided into a control area for the supervisor, separated by two tables from the study area where the participants moved around. The study area was approximately 2.65 m × 3.00 m and had a cabinet and two doors at the back and right side of the room (see Figure 4). The Meta Quest 2 VR HMD, along with the corresponding two

wireless controllers, was chosen. It provides six degrees of freedom, allowing the user to move freely and easily. Head and hand movements were translated one-to-one in the virtual environment. Except for the embodiment scenario, the user could see virtual hands that represented their hand position and gestures made with the controllers. They did not have a full-body representation of the user (except in the *positive* condition of the embodiment scenario). Since we maintained consistency between negative and positive embodiment and solely focused only on score differences, we believe that this limitation did not affect our results. The test scenarios were created using the Unity 3D game engine.

*Place illusion dimension.* Slater defines the dimension of *place illusion* as "[...] the human response to a given level of immersion [...] bounded by the set of SCs possible at that level of immersion" [41]. In this context, sensorimotor contingencies (SCs) refer to learned actions that influence an individual's perception. Thus, we decided to modify the levels of immersion to achieve the desired changes in PI. In accordance with Slaters' concept of immersion ("[...] immersion is bound to a particular set of valid actions that support perception and effectual action within a particular virtual reality [...]") [41]), we designed the changes between the two conditions in terms of sensory breadth and sensory depth, referring to the number and resolution of sensory channels for the user. For this purpose, we created a 3D model of the study room in its true-to-scale size. Real-world objects, such as the cabinet and the tables, were placed in identical positions in the VE to enable the participants to touch them (Figure 5). The fan also existed in the real world and was running during the *positive* condition to create a haptic sensation of wind. Instead of the flower, we placed a magnolia-scented candle on the table. It was uncovered during the *positive* condition. All items were placed on the back of the table, out of the users' reach, to prevent them from trying to touch the objects. Otherwise, the haptic mismatch with the visual appearance of the objects may have irritated them. In addition, four paintings were placed on the wall to provide a visual task for the users. Three spatial audio sources were positioned in the VE to provide auditory stimulation that matched the visual experience.

To achieve changes in presence, we replaced all high-definition photo-textures with simple one-colored textures in the *negative* condition. In addition, the directional spotlights positioned on the ceiling of the virtual room, shadow casting, and the fan were turned off in the *negative* condition. The candle was covered with a glass. All audio sources were turned off. In both conditions, the users had two minutes to explore the room closely. The design challenge was to avoid any co-influences on other presence factor clusters. Therefore, no items were highlighted, and sounds were kept at a constant volume, which is usual for the type of sound, so that it would not affect unintended changes in involvement. No social perception was possible as no social partners were present. All objects that could be suggested to be interactive were placed out of the participants' range. Hence, we expected plausibility not to be affected by the task design. However, we cannot preclude effects of the additional sensory feedback from the *positive* condition on plausibility. The users' virtual representation was equal for both



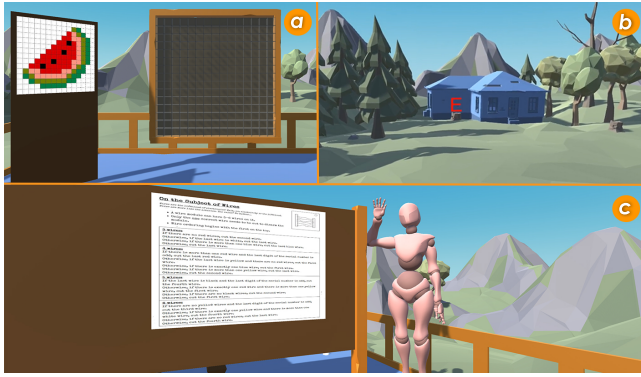
**Figure 5: The virtual environments of the place illusion scenario. On the left is the low-dimension intensity condition, and on the right is the high-dimension intensity condition. On the bottom right is a section of the real-world room from the same perspective.**

conditions. Therefore, we did not expect user embodiment to be affected.

*Involvement dimension.* We based the design of the involvement scenario on the definition of involvement by Witmer et al. [58], who defined it as "[...] focusing one's energy and attention on a coherent set of stimuli or meaningfully related activities and events". Thus, we focused the design on influencing the extent to which users paid attention to the aspects of the VE. We asked users to observe the VE and try to remember symbols that appeared for a short time (2 s). The VE was a landscape, as shown in Figure 6. The users were positioned in a fenced area within the landscape so that they automatically stayed in the play area. In the *negative* condition, exactly one letter appeared once in the virtual environment. In the *positive* condition, five letters appeared one after the other, forming a word. Before the condition started, users were told how many symbols would appear and that the task was to remember them and enter them in an interface that appeared after 100 seconds. In the *negative* condition, the symbol appeared after 10 seconds. Since the only task was to spot this single symbol, the users should focus less on the virtual environment once they recognize it. In the *positive* condition, a symbol appeared after every 15 seconds. Consequently, we expected the users' involvement to be higher in the *positive* condition, as they had to actively scan the VE until the end of the condition in order not to miss a symbol. The only differences between the conditions were the number of displayed symbols and the timing of the symbols' appearance. Hence, we did not expect any effects of the changes on place illusion, plausibility, embodiment, or social presence.

*Embodiment dimension.* To test the effect of embodiment on the perceived presence of users, we used the approach of Zhang et al. [61]. For our investigations, we compared a full-body embodiment condition with a condition in which only the controllers represented the user's hands. In the *positive* condition, interactions with the controllers were translated into gestures of the virtual hands, while the controllers themselves were not visualized. To





**Figure 6: The plausibility scenario (a), the involvement scenario (b), and the social presence scenario (c).**

mimic a real jigsaw puzzle, users could grab pieces, rotate them, and release them at the desired location. If the rotation and shape of the piece matched, the piece would fit into the corresponding slot.

The difficulty of the task remained the same because the interaction was unaffected. In both conditions, there were no changes in environmental detail, acoustic feedback, or task difficulty. Thus, changes in presence should be due to differences in embodiment. According to the results of Zhang et al., users should perceive more presence in the *positive* condition [61].

*Social presence dimension.* We based our definition of the *social dimension* of presence on Lombard’s categorization regarding social realism and social presence [25]. This includes parameters such as the feeling of communicating with an entity as a social partner. Therefore, we designed a basic avatar that resembled a human in stature but was kept minimalistic in detail. To manipulate social realism, we alternated between a computer-generated voice and a real voice and modified the interlocutor’s conformity to social norms. To achieve interaction between user and avatar, we implemented a game similar to “Keep talking, and nobody explodes”<sup>1</sup> (Figure 6). Users had three minutes to solve the task. At the beginning of each condition, the avatar stood to the side of users, greeted them, and then walked behind the board. After the riddle was solved, the avatar returned to the front of the board and said goodbye to the users. For the *positive* condition, the supervisor himself replied to the users’ questions. His voice was transmitted to the headphones to support the illusion of co-presence in the virtual world. The avatar first greeted the users, asked for their names, and responded accordingly. In the *negative* condition, the avatars had a computer-generated voice that read texts entered by the supervisor or predefined responses that were unlikely to match the users’ questions. Responses partially interrupted the users or were presented with delay. Environmental detail was not altered and was reduced to the minimum necessary for task completion. In both scenarios, the interaction was the same with respect to the task, and only the details of the social interaction changed between conditions. Since

the task difficulty is the same in both conditions, we argue that there were also no differences in involvement between the conditions. The embodiment of the users was the same in both conditions.

*Plausibility dimension.* To realize changes in plausibility, we designed a scenario where users were asked to recreate a pixel art painting. An empty canvas was placed next to the painting (Figure 6a). The users had to choose the right-colored cube by pressing left or right on their controller and aiming at the desired spot on the canvas. The task was the same in both scenarios. In the *positive* condition, the interaction worked as explained by the experimenter. The color picker would sequentially go through the list of available colors and repeat when at the end of the list. In the *negative* condition, the placement of the cubes was randomized. In addition, colors were randomly picked by the color picker, which could result in colors being picked multiple times. No social interaction was included, which could have influenced social interaction factors. The task was equivalent in both conditions to avoid possible differences in involvement. The interaction also remained the same in both conditions, with the expectation that identical actions could lead to different outcomes in the *negative* condition. The changes should not have had any impact on the place illusion dimension, as the VE was the same in both conditions. However, we considered it important to provide interaction feedback through a designated sound when a cube was placed in a random location. Otherwise, users might have believed that the misplacement was due to an error. However, since correct placement was also supported by auditory feedback, we do not expect this design choice to affect the place illusion dimension.

### 3.4 Procedure

Overall, the study design resulted in three independent variables: (1) the *presence dimension*, including place illusion, involvement, plausibility, social presence, and embodiment, (2) the *dimension intensity*, which could be either low or high, and (3) the *presence questionnaire* including the SUS, PQ, IPQ, and SIP. We used a within-study design and counterbalanced the order of scenarios and conditions per scenario across users according to the Latin square design. Thus, potential carry-over effects between the scenarios should have been counterbalanced. None of the questionnaires provide a prior explanation of the concepts included, such as presence. Subsequently, we chose not to bias participants’ judgments by defining presence or any related concepts to promote authentic usage. Due to the assumed positive effects on the reliability and consistency of presence assessments [11, 35], participants were asked to answer all questionnaires without leaving the VR. To avoid the presence ratings being influenced by the experience of the questionnaire assessment itself, we chose to use a neutral lobby scenario for it. The SUS and IPQ were rated on a 7-point Likert scale, the PQ was scored on a 5-point Likert scale, and the SIP was scored on a 10-point Likert scale using the response options provided by the questionnaires. To facilitate a comparison of the different scores, we converted the scores to an equal range between 1 and 7. All questions are summarized and roughly categorized in the appendix in Table 3–Table 6.

<sup>1</sup><https://bombmanual.github.io/english/>

At the end of the study, we also asked a few demographic questions.

### 3.5 Participants

Based on a *a priori* power analysis (G\*Power v3.1.9.7, repeated measures within-factors ANOVA,  $\alpha = 0.05$ ,  $power = 0.9$ ,  $r = 0.25$ ), the study was conducted with  $N = 50$  participants (26 male, 24 female). Participants were recruited from university students and staff members and had a mean age of 25.39 years ranging from 18 to 47 years. 78.0 % had prior experience with VR of which 36.0 % only used it occasionally and 34.0 % used it for less than two hours per week.

## 4 Results

To ensure that our results are meaningful, we compared equivalent questions from the different questionnaires. If the participants answered conscientiously, the answers should be the same. We compared the PQ7\_consistency and IPQ12\_consistency (difference  $M = 0.12$ ). The dataset did not show any inconsistencies.

### 4.1 Sensitivity of Presence Questionnaires

To address our first research question, "How sensitive are they to changes of specific dimensions of presence?" we compared the scores of the positive and negative conditions for each scenario and each questionnaire. The results are shown in Figure 7. The average of the presence scores for each scenario and questionnaire evidenced an upward trend from the negative to the positive condition (differences between  $M = 0.02$  and  $M = 1.29$  on average). As shown in Figure 8, all questionnaires result in more substantial score differences between the conditions in the *place illusion* dimension. The SUS, IPQ and SIP have their largest average difference in this scenario (SUS:  $M = 0.80$ , IPQ:  $M = 0.94$ , SIP:  $M = 1.29$ ). The greatest difference of the PQ was found in the *plausibility* scenario. For all questionnaires, the *embodiment* scenario performed the lowest with an average score difference ranging from  $M = 0.02$  to  $0.016$ . Except for the EMB and PI scenarios, the SUS, IPQ, and SIP had comparably similar score differences to the other three scenarios (SUS:  $M = 0.29 \pm 0.07$ , PQ:  $M = 0.31 \pm 0.11$ , SIP:  $M = 0.46 \pm 0.03$ ). The PQ showed comparable results across the PI, INV, and SOC scenarios, with a mean of  $0.3 \pm 0.08$ .

We performed *related-samples Wilcoxon signed-rank* tests to compare the conditions within each scenario for each questionnaire. For further insight, we determined the respective effect sizes (small effect:  $0.1 - \leq 0.3$ , medium:  $0.3 - \leq 0.5$ , large:  $\geq 0.5$ ) and 95 % confidence intervals of the differences between positive and negative scores using the *Hodges-Lehman estimator* for non-parametric related samples [28, 53].

The results are summarized in Figure 7. For further information on the results, refer to Table 12 in the appendix. The results of the dimension-specific evaluation of the PQ and IPQ are presented in table 15.

### 4.2 Approval of Sensitivity

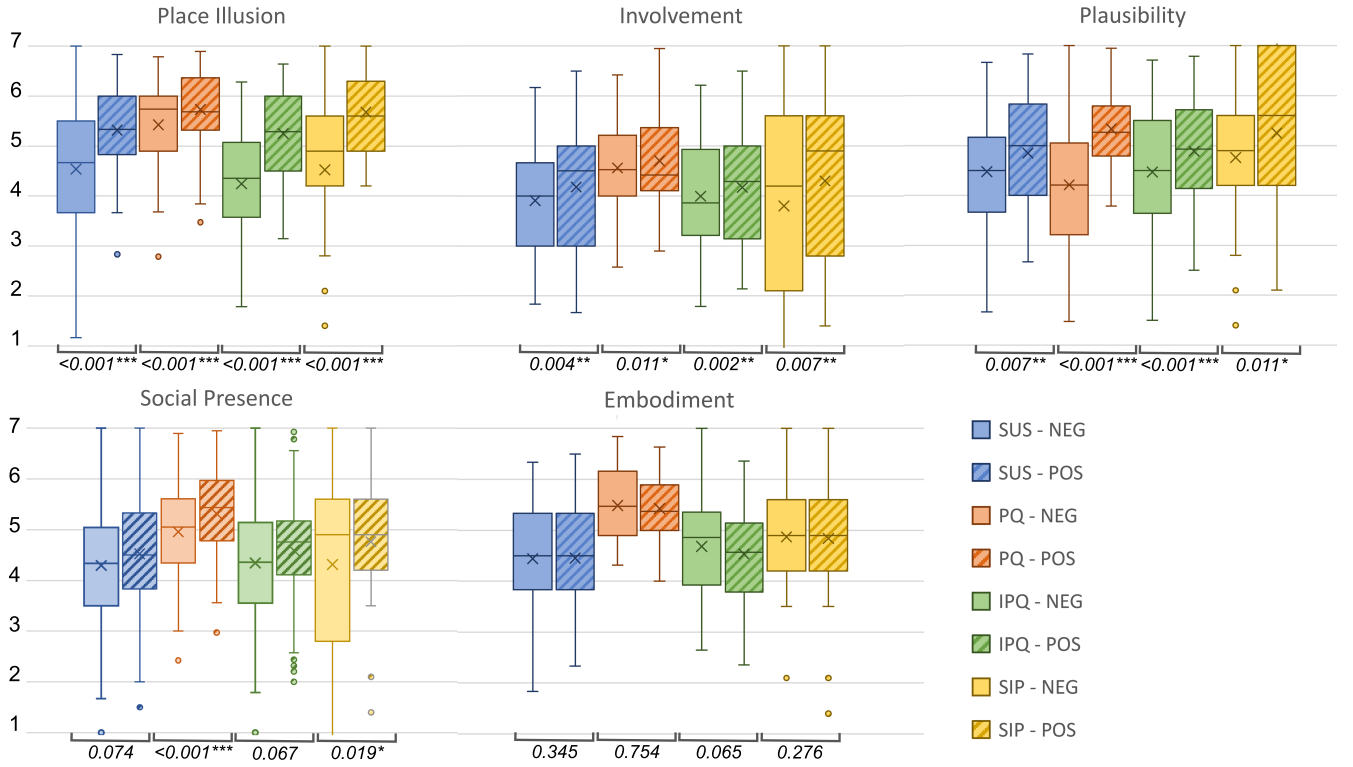
Participants rated the score differences for each questionnaire on a scale from  $-3 =$  "Difference is much too low" to  $3 =$  "Difference is much too high," with  $0 =$  "Agree to the score difference." The results are illustrated in Figure 9.

The results showed that the IPQ scores aligned most closely with users' expectations ( $|M| = 0.964$ ,  $\sigma = 1.21$ ), with 33.6% of participants agreeing with the score differences. The PQ achieved the second-highest agreement rate (25.6%). Across scenarios, participants generally perceived score differences as too low, except for the involvement and plausibility scenarios, where some questionnaires had higher-than-expected scores (e.g., INV: SIP:  $M = 0.26$ ,  $\sigma = 1.87$ ).

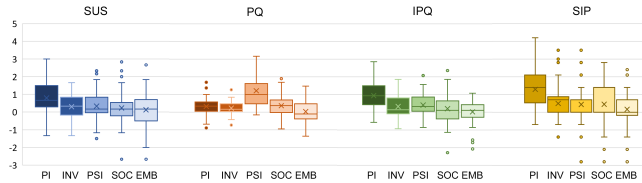
The SUS and PQ had the lowest reflection of changes users agreed with. The SUS aligned most closely with user experience in the embodiment ( $M = -0.08$ ,  $\sigma = 1.49$ ) and involvement scenarios ( $M = -0.06$ ,  $\sigma = 1.52$ ). For the PQ, the highest agreement was in the involvement scenario ( $M = -0.04$ ,  $\sigma = 1.3$ ) and the lowest in the place illusion scenario ( $M = -1.16$ ,  $\sigma = 1.14$ ). The IPQ had the highest agreement in the involvement scenario ( $M = -0.18$ ,  $\sigma = 1.09$ ) and the lowest for plausibility ( $M = -0.98$ ,  $\sigma = 1.05$ ). Variance was highest for SIP scores, particularly in the social presence scenario ( $\sigma = 1.73$ ), while the IPQ showed the lowest variance across all scenarios ( $\sigma = 0.89$ ). The results are summarized in Table 10, Table 9 and Table 11. Non-parametric Friedman tests indicated significant differences in the user agreement for the place illusion ( $\chi^2(2) = 23.74$ ,  $p < .001$ ) and plausibility ( $\chi^2(2) = 30.92$ ,  $p < .001$ ) scenarios. Post-hoc analyses revealed significant differences in agreement with PQ score differences compared to other questionnaires ( $p \leq .018$  for place illusion,  $p < .001$  for plausibility).

## 5 Discussion

The results showed that our approach to investigate dimension-specific sensitivity works. All questionnaires showed an upward trend in scores from negative to positive conditions for all presence dimensions except for the embodiment scenario. In compliance with previous work, we found a significant bias of presence assessments introduced by design. While previous work showed a bias caused by the questionnaire choice [15], the assessments' setting [11] and individual interpretation of questionnaire items [51], we focus our analysis on the effects caused by changes in specific presence dimensions. The reason is that presence questionnaires by design are not considered to be suitable for cross-scenario comparisons but rather for direct comparisons of VR experiences. Accordingly, Schwind et al. assume that the questionnaires do not directly measure the construct of presence but the contrast of VR experiences [35]. Hence, a VR experience could also receive lower presence ratings than another, although the user was comparably present. Simply caused by the fact that an aspect of the experience was perceived as less qualitative than in the other experience, regardless of whether this discrepancy influenced presence or not. Compared to another VR experience, the score might turn out differently. Hence, we not only evaluate the questionnaires' sensitivity but also users' approval of the resulting scores. In fact, not only did the dimension-specific sensitivity analysis provide meaningful insights regarding differing sensitivity to changes in the respective dimension, but the



**Figure 7: Distribution of the measured scores from the SUS, PQ, IPQ, and SIP questionnaires. Each graph shows the scores from all four questionnaires for a given dimension (blue = SUS, red = PQ, green = IPQ, yellow = SIP) paired as negative (solid pattern) and positive (stripe pattern) conditions. The cross indicates the mean, and the line indicates the median.**



**Figure 8: Distribution of the score differences (score of positive condition - score of negative condition per user) of SUS, PQ, IPQ, and SIP clustered by dimension (left = place illusion, second left = involvement, middle = plausibility, second right = social presence, right = embodiment). The line indicates the median value, and the cross indicates the mean.**

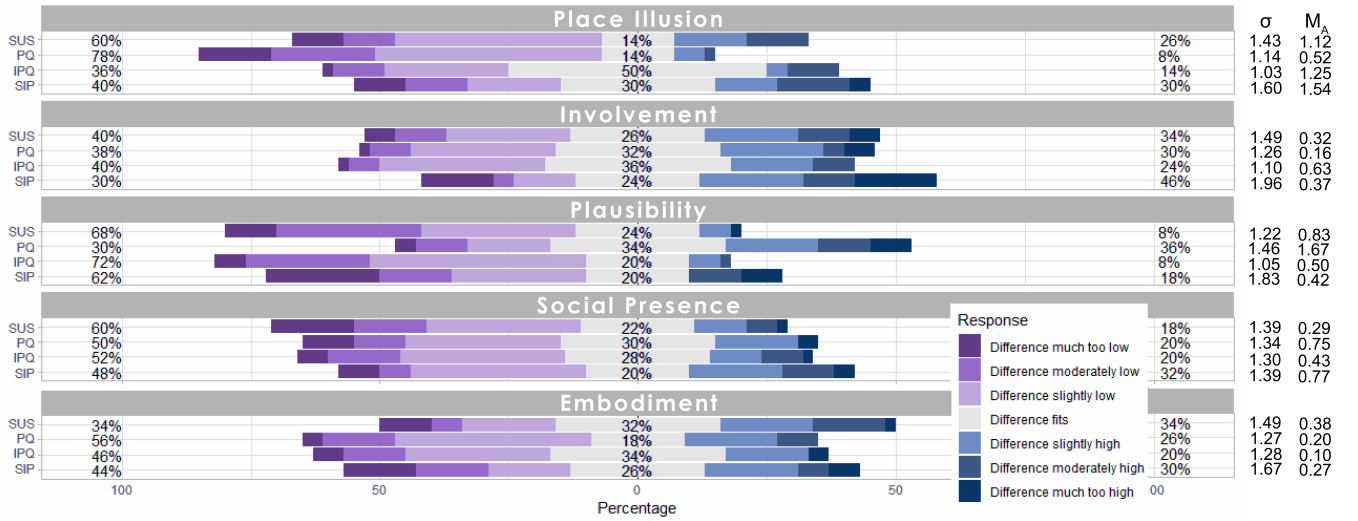
assessment of user approval proved that greater sensitivity does not necessarily mean more authentic presence assessments.

To address our research questions RQ1 ("How sensitive are the questionnaires to changes of specific dimensions of presence?"), RQ2 ("Do participants agree to the differences in scores affected by the changes of a specific dimension of presence?"), and RQ3 ("Can questionnaire items be clustered according to their sensitivity to changes of a specific dimension of presence?"), we conducted a comprehensive analysis of the quantitative results of our study. This included an overall analysis (RQ1), an examination of participants'

approval of scores (RQ2), and an analysis of scores per scenario as well as per question (RQ3). Statistically significant differences were observed between the scores of the positive and negative conditions for certain scenarios related to a particular questionnaire. However, although we designed the embodiment scenario based on previous research, the variations between the respective conditions were not significant enough to achieve measurable differences in presence. Participants agreed that the differences between the two conditions were negligible. Consequently, we decided to omit this dimension from further discussion.

### 5.1 Sensitivity of Presence Questionnaires

Our results demonstrate the influence of a specific presence dimension on presence scores, with higher scores on average in the positive condition. However, there were discrepancies between the presence questionnaires and dimensions. The *SIP* questionnaire showed the greatest sensitivity to change, with a mean difference of 0.57 in scores. The *SUS* had an average change of 0.36, while the *PQ* and *IPQ* had changes of  $M = 0.38$  and  $M = 0.43$ , respectively. The *SIP* scores showed increased differences, possibly due to changes not moderated through averaging multiple questions, some of which may partially not have been triggered by the changes. However, the high standard deviation ( $\sigma = 30.9\%$  higher than average with  $\sigma = 1.25$ ) and the lower user agreement (11.4% below



**Figure 9: Participants' agreement with the score differences of the different questionnaires per scenario. Participants could rate the score differences on a scale from -3 (purple) = "Difference is much too low" to 3 (blue) = "Difference is much too high", where 0 (gray) = "Agree with the score difference".**

average) indicate that users have difficulty quantifying perceived presence.

The *SUS* showed the highest sensitivity to the changes made in the place illusion scenario, probably because it primarily considers the perception of presence in general as well as environmental aspects. The sensitivity to changes in the other presence dimensions was lower and comparably uniform across the scenarios. Apart from the *SIP*, the *SUS* exhibited the second-highest sensitivity to the variations in the involvement scenario.

The *IPQ* demonstrated a comparable sensitivity profile, with the greatest sensitivity in the place illusion scenario, which was slightly higher than that of the *SUS* and lower sensitivity in the other scenarios. The *IPQ* was slightly more sensitive to changes in the *PSI* scenario but less sensitive in the involvement scenario. Both questionnaires appear to be suitable for environmental aspects-focused scenarios but provide moderate sensitivity to changes in the other dimensions. The *IPQ* is preferable for cases with respect to plausibility, while the *SUS* is better able to discern differences in users' involvement.

The *PQ* was less responsive to variations in involvement and place illusion. With an average score difference of  $M = 0.32$ , he was the least sensitive questionnaire in the place illusion scenario. The *PQ* showed the highest sensitivity towards alterations in the plausibility scenario ( $M = 1.21$ ). According to Slaters' definition of plausibility [41], the alterations in plausibility are likely to have had an impact on naturalness evaluated by multiple items. They may also have contributed to the *PQ* being the most responsive questionnaire to the modifications in the social presence scenario since the interaction with the avatar and the reception of real voice replies should enhance naturalness. None of the questionnaires demonstrated significant sensitivity in the social presence scenario, probably because none considered respective aspects. Another questionnaire, such as the one created by Nowak and Biocca [29], which

considers social aspects, may have performed better in this situation. The analysis of the individual dimensions of *PQ* and *IPQ* showed that the changes made to the involvement scenario did not impact the relevant items in the *PQ* as expected. We speculate that the minor effects can be attributed to our focus on involvement with visual aspects of the *VE* as opposed to the various involvement types considered by the *PQ*'s items on involvement. Thus, the individual use case influences the sensitivity of certain dimensions of the *PQ*. Probably due to the involvement items mainly focusing on how *VE* design affects the user's focus rather than on the design itself, as the *PQ* items do, the *IPQ* showed greater sensitivity to changes in the involvement scenario. We anticipated greater sensitivity to the changes in the social presence and embodiment scenario since the positive conditions should be more realistic. We assume the changes to be too specific, as the items primarily concern the *VE*'s realism in general, and the *VE*'s of the respective scenarios were still comparatively unrealistic. In conclusion, the sensitivity of the *PQ*'s and *IPQ*'s items strongly depends on the associated use case and the use of another questionnaire that considers this dimension may already be more or less sensitive.

## 5.2 Approval of Sensitivity

We presented the scores of both conditions of each scenario to the participants, who were then asked to indicate their level of approval regarding the score differences based on their experience. Overall, the participants would have expected a greater score difference in almost all cases. The differences observed in *IPQ* scores were most consistent with their expectations, demonstrating a 12% higher level of approval than the other questionnaire approvals. A closer examination of approval ratings per scenario indicates that participants perceived greater differences between the conditions of the place illusion scenario than reflected by the score differences.

The IPQ and SIP proved most sensitive in this scenario, although participants agreed more with the differences in the IPQ scores. We attribute this effect to the high variance in SIP scores (SIP:  $\sigma = 1.6$ , IPQ:  $\sigma = 1.09$ ). Based on the higher frequency of "much too low" and "much too high" ratings, 54.54 % more frequent than in the other questionnaires, the single item concept may lead to under- or overestimating perceived presence. On average, the SIP score difference received the lowest approval ratings on average ( $|M| = 1.384$ ), and both the score differences ( $\sigma = 1.25$ ) and approval ratings ( $\sigma = 1.73$ ) had a high standard deviation. The approval to the differences in scores for the PQ varied significantly ( $p < .001$ ). The lowest agreement was reported for the PI scenario ( $M = -1.16$ ), where score differences were also the lowest (PI:  $M = 0.84$ , PI\_PQ:  $M = 0.32$ ). In contrast, users mostly agreed on the score difference of the PQ in the plausibility scenario ( $M = 0.14$ ). Alongside the SIP, the SUS was the second most sensitive questionnaire in the INV scenario and received the most user approval. The PI scenario showed the greatest sensitivity. However, users still scored the differences too low in most cases. Regarding the social presence scenario, none of the questionnaires satisfied the users' expectations. On average, all score differences were considered too low ( $M = -0.404$ ). This aligns with our expectations, as none of the questionnaires includes items related to social aspects.

We did not determine the users' approval of the score differences in the individual presence dimensions of PQ and IPQ. Nevertheless, we conducted a brief comparison of their score differences with the differences of the overall score differences to get insights as to whether participants would have potentially concurred more with the score discrepancies of the items linked to the respective presence dimension. Participants could have indeed agreed more on score differences of specific dimensions as, in some cases, these showed a greater difference between the two conditions. The changes in the plausibility scenario affected items on involvement and naturalness. The changes of place illusion had mixed effects. As anticipated, the items on spatial presence, involvement, and realness of the IPQ were affected. However, we would have expected greater effects on the resolution items of the PQ. Moreover, the questions about the involvement of both PQ and IPQ were less impacted by the variations in the involvement rating than anticipated. Therefore, a follow-up study that investigates the sensitivity of questionnaire items or subscales could offer valuable insights. Introducing weighting on items pertinent to a particular usage scenario may potentially enhance the alignment of score variations with user experience.

### 5.3 Sensitivity of Single Items

We also analyzed items that were less affected or less affected by change than anticipated. As illustrated in table 15, items on involvement showed little sensitivity in all scenarios. Most effects on involvement were observable for the place illusion scenario. We assume the limited sensitivity was due to participants' difficulties in interpreting the term involvement, as we were repeatedly asked about its meaning. This assumption is supported by the fact that the IPQ7 and IPQ8, which ask the user whether they are paying attention to the real world, were more affected by the changes.

Another question that received low scores regardless of the scenario was *question 14 of the IPQ*. This question asks the user whether the virtual world is more realistic than the real world. As expected, this question was rated very low for both conditions in each scenario (positive:  $M = 2.32$ , negative:  $M = 2.12$ , difference:  $M = 0.192$ ). We question the relevance of this question, as previous research [6, 18] suggests that presence does not necessarily depend on high realism. Slater and Lee both argued that instead of high realism, it is only necessary for the user to be able to construct a mental model that accounts for what they perceive in the virtual world [22, 41]. Baños proposed that presence and reality judgments are two separate concepts that may or may not be positively correlated with each other but do not have to [3]. The relevance of realism for presence depends on the individual context of use. Furthermore, we believe that some users may not even know how to define a world that is more realistic than what we know as reality.

## 6 Limitations

Since the PQ and IPQ include items on involvement, we would have expected greater effects of our changes in the INV scenario. We designed our scenarios based on Witmers' and Singers' definition of involvement [58]. It would be interesting to examine how different types of involvement affect larger score differences. Our embodiment scenario design was based on previous work. However, the differences between the conditions were too small to achieve perceivable differences. We consider the short task duration or the ease of the task to have potentially caused minor effects. A follow-up study on the sensitivity of the questionnaires to changes in social realism and user embodiment would be interesting to determine if the small score differences were indeed caused by the limited sensitivity of the questionnaires or by the design of our conditions.

## 7 Implications

Our study was the first to quantify the yet only assumed bias of presence assessments by the choice of a questionnaire. Even though our scenarios were designed to maximize dimension-specific differences, the questionnaires' score differences were comparatively small, and users would have expected greater score differences in almost all cases (Table 1). Furthermore, the dimension-specific sensitivity varied significantly among the questionnaires. As VR scenarios are rarely designed to match certain conceptual requirements of presence but concern a variety of presence dimensions, we assume that the bias is usually even larger as it is more likely that the chosen questionnaire does not perfectly cover the presence dimensions affected by the scenario. Arguably, requesting a more concept-focused scenario design for presence investigations is not realistic. To promote the validity of presence assessments, awareness of such choice-dependent biases should be raised on the side of presence researchers, and we strongly encourage questionnaire developers to provide more detailed guidelines on the applicability of questionnaires for certain presence dimensions. Based on the results of our sensitivity evaluation, we were able to provide various insights on the applicability of SUS, PQ, IPQ, and SIP:

### Dimension-specific insights:



**Table 1: Mean score differences (positive minus negative score per user) including significance level and mean approval ratings of questionnaires per presence dimension**

Dim.	Que.	Score Difference	Approval
PI	SUS	0.77** ( $\sigma = 0.90$ )	-0.52 ( $\sigma = 1.19$ )
	PQ	0.31** ( $\sigma = 0.50$ )	-1.16 ( $\sigma = 0.46$ )
	IPQ	0.88** ( $\sigma = 0.88$ )	-0.26 ( $\sigma = 1.37$ )
	SIP	1.15** ( $\sigma = 1.15$ )	-0.20 ( $\sigma = 1.24$ )
INV	SUS	0.32* ( $\sigma = 0.74$ )	-0.06 ( $\sigma = 0.38$ )
	PQ	0.22 ( $\sigma = 0.63$ )	-0.04 ( $\sigma = 0.18$ )
	IPQ	0.29* ( $\sigma = 0.32$ )	-0.18 ( $\sigma = 0.71$ )
	SIP	0.55* ( $\sigma = 0.52$ )	0.26 ( $\sigma = 0.30$ )
PSI	SUS	0.33 ( $\sigma = 0.80$ )	-1.04 ( $\sigma = 0.50$ )
	PQ	1.17** ( $\sigma = 0.88$ )	0.14 ( $\sigma = 0.27$ )
	IPQ	0.40* ( $\sigma = 0.40$ )	-0.98 ( $\sigma = 0.50$ )
	SIP	0.47 ( $\sigma = 0.47$ )	-0.76 ( $\sigma = 0.73$ )
SOC	SUS	0.23 ( $\sigma = 0.93$ )	-0.78 ( $\sigma = 0.71$ )
	PQ	0.38** ( $\sigma = 0.62$ )	-0.52 ( $\sigma = 1.73$ )
	IPQ	0.22 ( $\sigma = 0.04$ )	-0.46 ( $\sigma = 0.45$ )
	SIP	0.45 ( $\sigma = 0.24$ )	-0.20 ( $\sigma = 0.60$ )
EMB	SUS	0.14 ( $\sigma = 0.95$ )	-0.08 ( $\sigma = 0.31$ )
	PQ	-0.01 ( $\sigma = 0.62$ )	-0.44 ( $\sigma = 1.27$ )
	IPQ	0.05 ( $\sigma = 0.02$ )	-0.42 ( $\sigma = 0.59$ )
	SIP	0.09 ( $\sigma = 0.08$ )	-0.38 ( $\sigma = 0.47$ )

- **Place Illusion:** Based on our results, we recommend using the IPQ or SIP. Both exhibited great sensitivity to the changes and received the most approval from the users. However, we recommend interpreting results cautiously as both also showed great uncertainty in terms of high standard deviation.
- **Involvement:** All questionnaires showed to be sensitive to changes in involvement, but the SUS and PQ seemed to reflect users' experience most accurately.
- **Plausibility:** The PQ showed exceptional sensitivity and was most approved by the users.
- **Social Presence:** Users approved most of the SIP that was most sensitive to the dimension changes. However, sensitivity was still rated as too low. We recommend the choice of a questionnaire that explicitly considers social presence for investigations on that dimension.
- **Embodiment:** All questionnaires exhibited little sensitivity to the changes. However, according to user approval, the SUS could reflect the experience accurately.

#### Questionnaire-specific insights:

- **Slater-Usch-Steed Questionnaire (SUS) [43, 51]:** The SUS showed to be a good choice when it comes to investigations on involvement and embodiment in terms of sensitivity and user approval.
- **IGroup Presence Questionnaire (IPQ) [33, 34]:** Based on our results, it seems that the IPQ is the best choice for VR experience that affect multiple presence dimensions as the questionnaire showed sensitivity to all of the dimensions.

Especially if investigations consider place illusion, the IPQ might be a beneficial choice.

- **Presence Questionnaire (PQ) [58]:** The questionnaire was the least sensitive on average. We assume that this was caused by the interface quality subscale of the PQ as technical aspects were not changed between the conditions (e.g., PQ17\_display quality). Furthermore, we observed little sensitivity of the involvement subscale. We assume this to be caused by uncertainties due to the wording of the items. However, the PQ performed exceptionally well for changes in plausibility and was most approved by the users. This was likely caused by the PQ considering task-related aspects. Hence, we recommend the PQ for use cases focusing on task design.
- **Single-Item Presence Questionnaire (SIP) [5]:** The SIP exhibited the greatest sensitivity of all questionnaires. However, we observed uncertainty in the quantification of presence, which was illustrated by very mixed scores with equally mixed approval ratings. The SIP may be beneficial in measuring minor alterations in VR experiences. However, presence researchers should interpret differences cautiously, as they may not accurately represent users' actual experience.

## 8 Conclusion

Presence research is often characterized by studies comparing different versions of a virtual experience to investigate the effects of alterations in design on presence. Hence, measurement tools are needed that are capable of accurately reflecting these effects on presence perception. Questionnaires are the most common way to assess presence. However, the choice of presence questionnaires faces some limitations: there is little research on (1) the sensitivity of questionnaires to changes in specific dimensions of presence and (2) the consistency of the resulting scores with users' actual experience. Our work is the first to quantify the potential bias of presence assessments through the choice of questionnaires. We designed five virtual environments aimed at isolating changes in virtual reality that address a specific dimension of presence (place illusion, involvement, plausibility, social presence, and embodiment). Each scenario was implemented with two conditions (one to induce higher and one to induce lower presence). Analysis of the SUS, IPQ, PQ, and SIP scores from the low-dimension intensity and high-dimension intensity conditions revealed that the four questionnaires have different sensitivities to different dimensions of presence. The SUS and IPQ questionnaires were most sensitive to our place illusion scenario. The PQ showed the highest sensitivity to changes in plausibility.

Our findings are consistent with our initial assumptions that not every questionnaire may be able to measure all aspects of presence with equal sensitivity. While all of them seemed to address aspects of place illusion and plausibility, the questionnaires showed low sensitivity when it came to aspects of user involvement, social presence, and user embodiment, with almost no sensitivity to the differences between the two conditions. Furthermore, we were able to point out that for most of the cases, the sensitivity of the scores

was lower than expected by the participants. In particular, the discrepancies in the SUS scores indicated insufficient sensitivity to the changes. Bouchard's single-item presence scale showed comparably high sensitivity for all dimensions considered. Participants seemed to have difficulty in quantifying their perceived presence, resulting in high score variability and discrepancies with actual perception. As a result, the SIP questionnaire received the least agreement out of all four questionnaires despite being the most sensitive in all scenarios. Some questions in the IPQ appeared to cause difficulties in interpretation, resulting in lower scores, which may have led to lower approval ratings regardless of the altered presence dimension. Regarding the dimensions of the PQ and IPQ, some showed less sensitivity to changes in certain presence dimensions than expected. Nevertheless, the results suggest that in some cases, participants could have approved more of the differences of the questionnaire items of that dimension. In general, our findings confirm our assumptions that the currently commonly used questionnaires are not as universally applicable as it seems to be assumed.

Our approach of evaluating questionnaires' sensitivity to changes in specific dimensions of presence was able to point out a lack of sensitivity of the questionnaires in general but also for specific items that may be responsible for the overall lack of sensitivity. The results underline the need for awareness of the bias in presence assessments caused by the choice of a questionnaire on the side of presence researchers. Results should be interpreted cautiously as they might not reflect the users' actual experience. Second, as VR experiences are mostly not designed to match a certain concept of presence and often include aspects of various presence dimensions, we need more insights into a questionnaire's sensitivity to promote the validity of presence assessments. We encourage questionnaire developers to provide recommendations for the applicability concerning specific dimensions of presence, including insights on sensitivity and users' approval. This way, a more user-centered and case-sensitive presence assessment could be achieved, which would, in turn, increase the assessment's validity.

## Acknowledgments

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through the project "Empirical Assessment of Presence and Immersion in Augmented and Virtual Realities" (project number: 425365472).

## References

- [1] Woodrow Barfield and Claudia Hendrix. 1995. The effect of update rate on the sense of presence within virtual environments. *Virtual Reality* 1, 1 (June 1995), 3–15. <https://doi.org/10.1007/BF02009709>
- [2] Andrea Bartl, Christian Merz, Daniel Roth, and Marc Erich Latoschik. 2022. The Effects of Avatar and Environment Design on Embodiment, Presence, Activation, and Task Load in a Virtual Reality Exercise Application. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, Singapore, 260–269. <https://doi.org/10.1109/ISMAR55827.2022.00041>
- [3] Rosa Baños, Cristina Botella, Azucena Garcia-Palacios, Helena Martin, Conxa Perpiñá, and Mariano Alcañiz Raya. 2000. Presence and reality judgment in virtual environments: A unitary construct? *CyberPsychology & Behavior* 3, 3 (June 2000). <https://doi.org/10.1089/10949310050078760>
- [4] R.M. Baños, C. Botella, A. Garcia-Palacios, H. Villa, C. Perpiñá, and M. Alcañiz. 2000. Presence and Reality Judgment in Virtual Environments: A Unitary Construct? *CyberPsychology & Behavior* 3, 3 (2000), 327–335. <https://doi.org/10.1089/10949310050078760>
- [5] Stéphane Bouchard, Geneviève Robillard, J. St-Jacques, Stéphanie Dumoulin, M.J. Patry, and Patrice Renaud. 2004. Reliability and validity of a single-item measure of presence in VR. In *IEEE International Workshop on Haptic, Audio and Visual Environments and Their Applications*. IEEE, Ottawa, ON, Canada, 59–61. <https://doi.org/10.1109/HAVE.2004.1391882>
- [6] Kevin Cheng and Paul A. Cairns. 2005. Behaviour, Realism and Immersion in Games. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems* (Portland, OR, USA) (CHI EA '05). Association for Computing Machinery, New York, NY, USA, 1272–1275. <https://doi.org/10.1145/1056808.1056894>
- [7] Mark R. Costa, Sung Yeun Kim, and Frank Biocca. 2013. Embodiment and Embodied Cognition. In *Virtual Augmented and Mixed Reality: Designing and Developing Augmented and Virtual Environments*, Randall Shumaker (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 333–342.
- [8] H.Q. Dinh, N. Walker, L.F. Hodges, Chang Song, and A. 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, Houston, Texas, USA, 222–228. <https://doi.org/10.1109/VR.1999.756955>
- [9] J. Flach and J. Holden. 1998. “The reality of experience: Gibson’s way”. *Presence* 7, 1 (1998), 90–95. <https://doi.org/10.1162/105474698565550>
- [10] Alessandra Gorini, Claret S Capideville, Gianluca De Leo, Fabrizia Mantovani, and Giuseppe Riva. 2011. The role of immersion and narrative in mediated presence: the virtual hospital experience. *Cyberpsychology, Behavior, and Social Networking* 14, 3 (2011), 99–105. <https://doi.org/10.1089/cyber.2010.0100> Publisher: Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA.
- [11] Sarah Graf and Valentin Schwind. 2020. Inconsistencies of Presence Questionnaires in Virtual Reality. In *Proceedings of the 26th ACM Symposium on Virtual Reality Software and Technology* (Virtual Event, Canada) (VRST '20). Association for Computing Machinery, New York, NY, USA, Article 60, 3 pages. <https://doi.org/10.1145/3385956.3422105>
- [12] Simone Grassini and Karin Laumann. 2020. Questionnaire Measures and Physiological Correlates of Presence: A Systematic Review. *Frontiers in Psychology* 11 (2020). <https://doi.org/10.3389/fpsyg.2020.00349>
- [13] Mehdi Hafsia, Eric Monacelli, and Hugo Martin. 2018. Virtual Reality Simulator for Construction Workers. In *Proceedings of the Virtual Reality International Conference - Laval Virtual (Laval, France) (VRIC '18)*. Association for Computing Machinery, New York, NY, USA, Article 11, 7 pages. <https://doi.org/10.1145/3234253.3234298>
- [14] Carrie Heeter. 1992. Being There: The Subjective Experience of Presence. *Presence Teleoperators & Virtual Environments* 1, 2 (May 1992), 9. <https://doi.org/10.1162/pres.1992.1.2.262> Presence definition and factors.
- [15] Dimitri Hein, Christian Mai, and Heinrich Hußmann. 2018. The usage of presence measurements in research: a review. The International Society for Presence Research Prague, Proceedings of the International Society for Presence Research Annual Conference (Presence), Prague, 21–22.
- [16] Richard M. Held and Nathaniel I. Durlach. 1992. Telepresence. *Presence: Teleoperators and Virtual Environments* 1, 1 (Feb. 1992), 109–112. <https://doi.org/10.1162/pres.1992.1.1.109>
- [17] Charlene Jennett, Anna L. Cox, Paul Cairns, Samira Dhoparee, Andrew Epps, Tim Tijs, and Alison Walton. 2008. Measuring and defining the experience of immersion in games. *International J. Human-Computer Studies* 66 (May 2008), 641–661. <https://doi.org/10.1016/j.ijhcs.2008.04.004>
- [18] Crescent Jicol, Christopher Clarke, Emilia Tor, Rebecca M Dakin, Tom Charlie Lancaster, Sze Tung Chang, Karin Petrini, Eamonn O'Neill, Michael J Proulx, and Christof Lutteroth. 2023. Realism and Field of View Affect Presence in VR but Not the Way You Think. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 399, 17 pages. <https://doi.org/10.1145/3544548.3581448>
- [19] 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 (Sept. 1997). <https://doi.org/10.1111/j.1083-6101.1997.tb00073.x>
- [20] Christoph Klimmt and Peter Vorderer. 2003. Media Psychology “is not yet there”: Introducing Theories on Media Entertainment to the Presence Debate. *Presence* 12, 4 (Aug. 2003), 346–359. <https://doi.org/10.1162/10547460322391596>
- [21] Pontus Larsson, Daniel Västfjäll, and Mendel Kleiner. 2001. The Actor-Observer Effect in Virtual Reality Presentations. *CyberPsychology & Behavior* 4, 2 (2001), 239–246. <https://doi.org/10.1089/109493101300117929>
- [22] Kwan Min Lee. 2004. Presence, Explicated. *Communication Theory* 14, 1 (Feb. 2004), 27–50. <https://doi.org/10.1111/j.1468-2885.2004.tb00302.x> Publisher: John Wiley & Sons, Ltd.
- [23] Jane Lessiter, Jonathan Freeman, Edmund Keogh, and Jules Davidoff. 2001. A Cross-Media Presence Questionnaire: The ITC-Sense of Presence Inventory. *Presence* 10 (June 2001), 282–297. <https://doi.org/10.1162/105474601300343612>
- [24] James Jeng-Wei Lin, Don E Parker, Henry Been-Lirn Duh, Habib Abi-Rached, and Thomas A. Furness. 2002. Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. *Proceedings - Virtual Reality Annual International Symposium* 171, 2 (Jan. 2002). <https://doi.org/10.1109/VR.2002.996519>
- [25] Matthew Lombard and Theresa Ditton. 1997. At the heart of it all: The concept of presence. *Journal of computer-mediated communication* 3, 2 (1997), JCMC321. <https://doi.org/10.1111/j.1083-6101.1997.tb00072.x> Publisher: Oxford University Press Oxford, UK.
- [26] 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*. International Society for Presence Research Los Angeles, CA, 1–15.
- [27] Thomas Metzinger. 2003. *Being No One: The Self-Model Theory of Subjectivity*. The MIT Press. <https://doi.org/10.7551/mitpress/1551.001.0001>
- [28] Roger Newson. 2002. Parameters behind “Nonparametric” Statistics: Kendall’s tau, Somers’ D and Median Differences. *The Stata Journal* 2, 1 (2002), 45–64. <https://doi.org/10.1177/1536867X0200200103>
- [29] Kristine Nowak and Frank Biocca. 2003. The Effect of the Agency and Anthropomorphism on Users’ Sense of Telepresence, Copresence, and Social Presence in Virtual Environments. *Presence Teleoperators & Virtual Environments* 12 (10 2003), 481–494. <https://doi.org/10.1162/10547460322761289>
- [30] Ye Pan and Anthony Steed. 2019. How Foot Tracking Matters: The Impact of an Animated Self-Avatar on Interaction, Embodiment and Presence in Shared Virtual Environments. *Frontiers in Robotics and AI* 6 (2019). <https://doi.org/10.3389/frobt.2019.00104>
- [31] Daniel Roth, Jean-Luc Lugin, Dmitri Galakhov, Arvid Hofmann, Gary Bente, Marc Erich Latoschik, and Arnulph Fuhrmann. 2016. Avatar realism and social interaction quality in virtual reality. In *2016 IEEE Virtual Reality (VR)*. IEEE, Greenville, South Carolina, 277–278. <https://doi.org/10.1109/VR.2016.7504761>
- [32] Maria Sanchez-Vives and Mel Slater. 2005. From presence to consciousness through virtual reality. *Nature reviews. Neuroscience* 6 (05 2005), 332–9. <https://doi.org/10.1038/nrn1651>
- [33] Thomas Schubert. 2003. The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realism. *Zeitschrift für Medienpsychologie* 15 (April 2003), 69–71. <https://doi.org/10.1026/1617-6383.15.2.69>
- [34] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. 2001. The Experience of Presence: Factor Analytic Insights. *Presence* 10 (June 2001), 266–281. <https://doi.org/10.1162/105474601300343603>
- [35] Valentin Schwind, Pascal Knierim, Nico Haas, and Niels Henze. 2019. Using Presence Questionnaires in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland UK) (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3290605.3300590>
- [36] Anil K Seth, Keisuke Suzuki, and Hugo D Critchley. 2012. An interoceptive predictive coding model of conscious presence. *Frontiers in psychology* 2 (2012), 395. <https://doi.org/10.3389/fpsyg.2011.00395> Publisher: Frontiers Research Foundation.
- [37] Thomas Sheridan. 1992. Musings on Telepresence and Virtual Presence. *Presence* 1 (Jan. 1992), 120–125. <https://doi.org/10.1162/pres.1992.1.1.120>
- [38] Yongxia Xia Skadberg and James R Kimmel. 2004. Visitors’ flow experience while browsing a Web site: its measurement, contributing factors and consequences. *Computers in human behavior* 20, 3 (2004), 403–422.
- [39] Richard Skarbez, Frederick P. Brooks, Jr., and Mary C. Whitton. 2017. A Survey of Presence and Related Concepts. *ACM Comput. Surv.* 50, 6 (Nov. 2017). <https://doi.org/10.1145/3134301>
- [40] Mel Slater. 2003. A Note on Presence Terminology. *Presence Connect* 3 (Jan. 2003). [https://www.researchgate.net/publication/242608507\\_A\\_Note\\_on\\_Presence\\_Terminology](https://www.researchgate.net/publication/242608507_A_Note_on_Presence_Terminology)
- [41] 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 (Dec. 2009), 3549–3557. <https://doi.org/10.1098/rstb.2009.0138>

- [42] Mel Slater, Domna Banakou, Alejandro Beacco, Jaime Gallego, Francisco Macia-Varela, and Ramon Oliva. 2022. A Separate Reality: An Update on Place Illusion and Plausibility in Virtual Reality. *Frontiers in Virtual Reality* 3 (June 2022), 914392. <https://doi.org/10.3389/fvrvir.2022.914392>
- [43] Mel Slater and Anthony Steed. 2000. A Virtual Presence Counter. *Presence* 9 (Oct. 2000), 413–434. <https://doi.org/10.1162/105474600566925> 315 citations (Crossref) [2022-07-15].
- [44] Mel Slater, Martin Usoh, and Anthony Steed. 1994. Depth of Presence in Virtual Environments. *Presence* 3 (Jan. 1994), 130–144. <https://doi.org/10.1162/pres.1994.3.2.130>
- [45] Mel Slater and Sylvia Wilbur. 1997. A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments* 6, 6 (Dec. 1997), 603–616. <https://doi.org/10.1162/pres.1997.6.6.603>
- [46] Mel Slater and Sylvia Wilbur. 1997. A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. *Presence: Teleoperators & Virtual Environments* 6 (1997), 603–616. <https://api.semanticscholar.org/CorpusID:9437981>
- [47] Pavel Smutny. 2022. Learning with virtual reality: a market analysis of educational and training applications. *Interactive Learning Environments* 31, 10 (Feb. 2022), 1–14. <https://doi.org/10.1080/10494820.2022.2028856> Publisher: Routledge.
- [48] Squelch A.P. 2001. Virtual reality for mine safety training in South Africa. *Journal of the Southern African Institute of Mining and Metallurgy* 101, 4 (July 2001), 209–216. [https://doi.org/10.10520/AJA0038223X\\_2672](https://doi.org/10.10520/AJA0038223X_2672) Publisher: Southern African Institute of Mining and Metallurgy.
- [49] Jonathan Steuer. 1992. Defining Virtual Reality: Dimensions Determining Telepresence. *Journal of Communication* 42, 4 (02 1992), 73–93. <https://doi.org/10.1111/j.1460-2466.1992.tb00812.x>
- [50] Jari Takatalo. 2002. *Presence and Flow in Virtual Environments: An Explorative Study*. Ph. D. Dissertation. Aalto University. <https://doi.org/10.1089/cyber.2021.0037>
- [51] Martin Usoh, Ernest Catena, Sima Arman, and Mel Slater. 2000. Using Presence Questionnaires in Reality. *Presence: Teleoperators and Virtual Environments* 9, Presence (April 2000), 497–503. <https://doi.org/10.1162/105474600566989>
- [52] Peter Vorderer, Werner Wirth, Feliz Gouveia, Frank Biocca, Timo Saari, Lutz Jäncke, Saskia Böcking, Holger Schramm, Andre Gysbers, Tilo Hartmann, Christoph Klimmt, Jari Laarni, Niklas Ravaja, Ana Sacau, Thomas Baumgartner, and Petra Jäncke. 2004. MEC spatial presence questionnaire (MEC-SPQ): Short documentation and instructions for application. *Report to the European Community, Project Presence: MEC (IST-2001-37661)* 18 (June 2004). <http://www.ijk.hmt-hannover.de/presence>.
- [53] Ronald Wasserstein and Nicole Lazar. 2016. The ASA’s Statement on p-Values: Context, Process, and Purpose. *The American Statistician* 70 (03 2016), 129–133. <https://doi.org/10.1080/00031305.2016.1154108>
- [54] Eva Waterworth and John Waterworth. 2001. Focus, Locus, and Sensus: The Three Dimensions of Virtual Experience. *Cyberpsychology & behavior: the impact of the Internet, multimedia and virtual reality on behavior and society* 4 (May 2001), 203–13. <https://doi.org/10.1089/109493101300117893>
- [55] Stefan Weber, David Weibel, and Fred W. Mast. 2021. How to Get There When You Are There Already? Defining Presence in Virtual Reality and the Importance of Perceived Realism. *Frontiers in Psychology* 12 (2021), 628298. <https://doi.org/10.3389/fpsyg.2021.628298>
- [56] Robert B. Welch, Theodore T. Blackmon, Andrew Liu, Barbara A. Mellers, and Lawrence W. Stark. 1996. The Effects of Pictorial Realism, Delay of Visual Feedback, and Observer Interactivity on the Subjective Sense of Presence. *Presence: Teleoperators and Virtual Environments* 5, 3 (08 1996), 263–273. <https://doi.org/10.1162/pres.1996.5.3.263> arXiv:<https://direct.mit.edu/pvar/article-pdf/5/3/263/1836794/pres.1996.5.3.263.pdf>
- [57] A Witmer and Mel Slater. 1999. Measuring Presence: A Response to the Witmer and Singer Presence Questionnaire. *Presence (Camb.)* 8 (Dec. 1999), 560–565. <https://doi.org/10.1162/105474699566477>
- [58] Bob G. Witmer and Michael J. Singer. 1994. Measuring Presence in Virtual Environments. *ARI Technical Report* 1014 (Oct. 1994), 47. <https://doi.org/10.1080/10447310802546732>
- [59] Bob G Witmer and Michael J Singer. 1998. Measuring presence in virtual environments: A presence questionnaire. *Presence* 7, 3 (1998), 225–240.
- [60] David Zeltzer. 1992. Autonomy, Interaction, and Presence. *Presence: Teleoperators and Virtual Environments* 1, 1 (02 1992), 127–132. <https://doi.org/10.1162/pres.1992.1.1.127>
- [61] Jingjing Zhang, Mengjie Huang, Lixiang Zhao, Rui Yang, Hai-Ning Liang, Ji Han, Liu Wang, and Wenxin Sun. 2020. Influence of Hand Representation Design on Presence and Embodiment in Virtual Environment. In *2020 13th International Symposium on Computational Intelligence and Design (ISCID)*. IEEE, Hangzhou, China, 364–367. <https://doi.org/10.1109/ISCID51228.2020.00088>

## A Questionnaires

### A.1 Overview

Authors	Year	Items	Factors/Indicators/Dimension
Banos et al. ([4])	1998	77	Reality judgment, presence, emotional involvement, interaction, control, attention, realism, perceptual congruence, anticipation
Barfield & Hendrix ([1])	1995	13	Presence, navigation, real-world dissociation
Bouchard (SIP [5])	2004	1	None
Dinh et al. ([8])	1999	14	Presence, involvement, realism, control, enjoyment
Jennett (IEQ [17])	2008	31	Cognitive involvement, Real World Dissociation, Transportation, Challenge, Emotional involvement, Enjoyment.
Kim & Biocca ([19])	1997	8	Arrival, departure
Larsson et al. (SVUP [21])	2001	150	Quality evaluations, attitudes, presence, realism, provided information, simulator sickness
Lessiter et al. (ITQ-SOPI [23])	2001	44	Sense of Physical Space, Engagement, Naturalness
Lombard et al. (TPI [26])	2009	42	Transportation, Immersion, Realism; Social Actor within a Medium Social Richness
Lin et al. (E <sup>2</sup> I [24])	2002	14	Sensory, distraction, realism, and control
Nowak & Biocca [29]	2003	30	Co-presence of self and others, telepresence, social presence
Regenbrecht et al. (IPQ [34])	2001	14	Spatial Presence, Involvement, Realness
Slater, Usoh, Steed (SUS [43])	2000	6	None
Slater et al. (UCL [44])	1994	3	Subjective presence
Takatalo et al. (EVEQ [50])	2002	124	Spatial presence, action, attention, real[ness], and arousal
Vorderer et al. (MEC-SPQ [52])	2004	L:72 M:54 S:36	Attention allocation; Spatial situation model; Self-location; Possible actions; Cognitive involvement; Suspension of disbelief; Domain-specific interest; Visual/spatial imagery; absorption
Witmer et al. (PQ [59])	1998	32	Involved/ Control, Natural; Interface quality

**Table 2: Overview of presence questionnaires considered for this study with year of publication, number of items and dimensions included.**



## A.2 Items of the Questionnaires Used for Our Investigations

Slater-Usch-Steed Questionnaire (SUS, 2000) [43, 51]		
Question	Description	Tag
1	Rate the sense of being there.	SUS1_presence overall
2	Extent of VE being reality.	SUS2_became reality
3	Does user remember VE more as images or as environment?	SUS3_images
4	Was the sense of being in the VE or in real world stronger?	SUS4_dominant location
5	How detailed is the memory of the VE?	SUS5_memory
6	Frequency of thinking that user is in VE.	SUS6_frequency present

**Table 3: Overview of SUS questions examined in the study [43, 51]. The corresponding tag of each question will be used in the paper as a reference for that question.**

Presence Questionnaire by Witmer and Singer (PQ, 1998) [59]		
Question	Description	Tag
1	Rate the ability to control events. (INV)	PQ1_control events
2	Rate responsiveness to users' actions. (INV)	PQ2_responsiveness
3	How natural was the interaction? (NAT)	PQ3_natural interaction
4	Level involvement through visual aspects. (INV)	PQ4_visual involvement
5	How natural was the control mechanism? (NAT)	PQ5_natural control
6	How compelling was sense of moving objects? (INV)	PQ6_compelling objects
7	Consistency with real world experiences? (NAT)	PQ7_consistency
8	Ability to anticipate what will happen based on users' actions. (INV)	PQ8_anticipation
9	Ability to visually inspect the environment. (INV)	PQ9_visual exploration
10	How compelling was sense of moving around in the VE? (INV)	PQ10_compelling navigation
11	Ability to closely examine objects. (RES)	PQ11_examination detail
12	Ability to investigate objects from multiple viewpoints. (RES)	PQ12_examination viewpoints
13	Involvement by VE experience. (INV)	PQ13_involvement
14	How much delay was between action and outcome? (INV)	PQ14_delay
15	Speed of adjusting to the VE. (INV)	PQ15_adjustment
16	Proficiency of movement and interaction in the VE. (INV)	PQ16_proficiency
17	Extent of interference by display quality. (IQ)	PQ17_display quality
18	How much did the control devices interfere the experience? (IQ)	PQ18_interference control
19	Ability to concentrate on task rather on control mechanism. (IQ)	PQ19_task focus

**Table 4: Overview of the PQ questions examined in the study [59]. The presence dimension of each item is given in parentheses after each question. The dimensions included are Involvement (INV, 11 items), Naturalness (NAT, 3 items), Resolution (RES, 2 items), and Interface Quality (IQ, 3 items). The corresponding tag of each question will be used in the paper as a reference for the respective question.**

<b>The IGroup Presence Questionnaire by Schubert (IPQ, 2001) [33, 34]</b>		
Question	Description	Tag
1	Rate the sense of being there. (PRES)	IPQ1_presence overall
2	Feeling of being surrounded by the VE. (SP)	IPQ2_surrounding
3	Extent of not feeling present in the VE. (SP)	IPQ3_images
4	Sense of acting in the VE instead from outside. (SP)	IPQ4_no presence
5	Feeling of being present in the VE. (SP)	IPQ5_acting inside
6	Awareness of real world when navigating through VE. (SP)	IPQ6_presence
7	Not being aware of the real environment. (INV)	IPQ7_awareness navigation
8	Extent of not being aware of the real environment. (INV)	IPQ8_awareness realworld
9	How much did the user still pay attention to the real world? (INV)	IPQ9_attention realworld
10	Extent of being captivated by the VE. (INV)	IPQ10_captivated
11	Extent to which VE seemed to be real. (REAL)	IPQ11_realism1
12	Consistency of experience with real-world experience. (REAL)	IPQ12_consistency
13	How real seemed VE? (REAL)	IPQ13_realism2
14	Was VE more realistic than real world? (REAL)	IPQ14_realism dominance

**Table 5: Overview of IPQ questions investigated in the study [33, 34]. The presence dimension of each item is given in parentheses after each question. In addition to presence in general (PRES), the dimensions included are spatial presence (SP, 5 items), involvement (INV, 4 items), and realness (REAL, 4 items). The corresponding tag of each question will be used in the paper as a reference for the respective question.**

<b>The single item presence scale by Bouchard (SIP, 2004) [5]</b>		
Question	Description	Tag
1	To what extent do you feel present in the virtual environment, as if you were really there?	SIP_presence overall

**Table 6: Overview of the single-item presence scale by Bouchard (2004) studied in [5]. The corresponding tag will be used in the paper as a reference for the respective question.**

## B Descriptive Analysis

### B.1 Score differences

Questionnaire Presence Dimension Dimension Intensity	SUS									
	PI		INV		PSI		SOC		EMB	
	POS	NEG	POS	NEG	POS	NEG	POS	NEG	POS	NEG
Average Score ( $M$ )	5.27	4.48	4.13	3.82	4.68	4.35	4.52	4.29	4.46	4.32
Median Score ( $Md$ )	5.33	4.50	4.33	3.83	4.58	4.25	4.50	4.33	4.42	4.33
Standard Deviation ( $\sigma$ )	0.88	1.30	1.26	1.21	1.10	1.17	1.15	1.24	1.16	1.15

Questionnaire Presence Dimension Dimension Intensity	PQ									
	PI		INV		PSI		SOC		EMB	
	POS	NEG	POS	NEG	POS	NEG	POS	NEG	POS	NEG
Average Score ( $M$ )	5.69	4.48	4.66	4.44	4.39	4.16	4.55	4.96	5.48	5.45
Median Score ( $Md$ )	5.68	5.55	4.42	4.47	5.00	4.21	5.00	5.05	5.53	5.45
Standard Deviation ( $\sigma$ )	0.83	0.95	1.03	1.10	2.02	1.11	1.98	0.96	0.68	0.70

Questionnaire Presence Dimension Dimension Intensity	IPQ									
	PI		INV		PSI		SOC		EMB	
	POS	NEG	POS	NEG	POS	NEG	POS	NEG	POS	NEG
Average Score ( $M$ )	5.21	4.27	4.12	3.80	4.81	4.41	4.54	4.34	4.59	4.56
Median Score ( $Md$ )	5.21	4.32	4.18	3.86	4.71	4.39	4.64	4.36	4.50	4.64
Standard Deviation ( $\sigma$ )	0.89	1.12	1.15	1.23	1.06	1.13	1.11	1.22	1.02	1.07

Questionnaire Presence Dimension Dimension Intensity	SIP									
	PI		INV		PSI		SOC		EMB	
	POS	NEG	POS	NEG	POS	NEG	POS	NEG	POS	NEG
Average Score ( $M$ )	5.68	4.37	4.21	3.72	4.90	4.45	4.31	4.79	4.45	4.46
Median Score ( $Md$ )	5.60	4.20	3.85	3.85	4.90	4.90	4.90	4.90	4.90	4.90
Standard Deviation ( $\sigma$ )	0.93	1.63	1.80	1.73	1.63	1.79	1.52	1.62	1.56	1.47

**Table 7: Mean, median, and standard deviation of all questionnaire scores split by scenario and condition. Visualization in Figure 7.**

Scenario	Questionnaire	Item	Significance
PI	SUS	1-3,5,6	$p < .004^{**}$
PI	PQ	3,4,6,7,10,11,13,14,17	$p < 0.017^*$
PI	IPQ	1-14	$p < 0.02^*$
PI	SIP	1	$p < .001^{***}$
INV	SUS	2,4	$p < .042^*$
INV	PQ	3,6,8,13	$p < 0.038^*$
INV	IPQ	2,4,5,8-10	$p < 0.046^*$
INV	SIP	1	$p < .007^{**}$
PSI	SUS	1,2,5	$p < .005^{**}$
PSI	PQ	1-11,13-16,18,19	$p < 0.03^*$
PSI	IPQ	1,2,5,6,10-13	$p < 0.05^*$
PSI	SIP	1	$p < .011^*$
SOC	SUS	2,5	$p < .037^*$
SOC	PQ	3,5-9,14-16	$p < 0.05^*$
SOC	IPQ	5,12	$p < 0.035^*$
SOC	SIP	1	$p < .019^*$
EMB	SUS	4,6	$p < .0036^{**}$

**Table 8:** The table shows all items where the *paired-samples Wilcoxon signed rank test* for non-parametric data reported significant differences between the scores of positive and negative condition ( $p < 0.05$ ,  $N = 50$ ).

	Presence dimension				
	PI	INV	PSI	SOC	EMB
<b>SUS</b>	$M = 0.77, Md = 0.67$ $\sigma = 1.43$	$M = 0.32, Md = 0.33$ $\sigma = 1.52$	$M = 0.33, Md = 0.17$ $\sigma = 1.22$	$M = 0.23, Md = 0.17$ $\sigma = 1.49$	$M = 0.14, Md = 0.17$ $\sigma = 1.49$
<b>PQ</b>	$M = 0.31, Md = 0.34$ $\sigma = 1.14$	$M = 0.22, Md = 0.16$ $\sigma = 1.30$	$M = 1.17, Md = 1.00$ $\sigma = 1.48$	$M = 0.38, Md = 0.3$ $\sigma = 1.36$	$M = -0.01, Md = -.11$ $\sigma = 1.27$
<b>IPQ</b>	$M = 0.88, Md = 0.89$ $\sigma = 1.09$	$M = 0.29, Md = 0.14$ $\sigma = 1.09$	$M = 0.4, Md = 0.32$ $\sigma = 1.05$	$M = 0.22, Md = 0.18$ $\sigma = 1.34$	$M = 0.05, Md = 0.07$ $\sigma = 1.28$
<b>SIP</b>	$M = 1.15, Md = 1.40$ $\sigma = 1.6$	$M = 0.55, Md = 0.70$ $\sigma = 1.87$	$M = 0.47, Md = 0.00$ $\sigma = 1.83$	$M = 0.45, Md = 0.00$ $\sigma = 1.48$	$M = 0.09, Md = 0.0$ $\sigma = 1.67$

**Table 9:** Mean and median of the differences of presence scores (score of positive - negative condition) from the questionnaires split by conditions. Visualization in figure 8.

## B.2 Sensitivity of Questionnaire Items

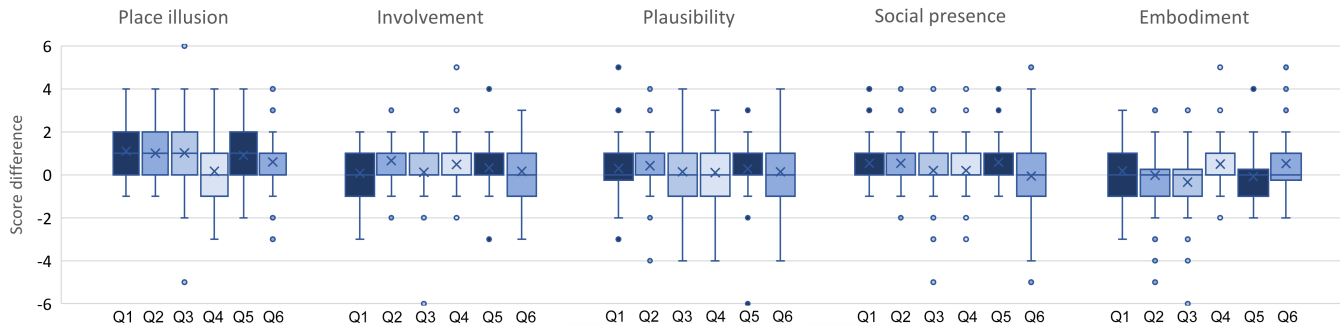


Figure 10: SUS score differences between positive and negative condition grouped by scenario.

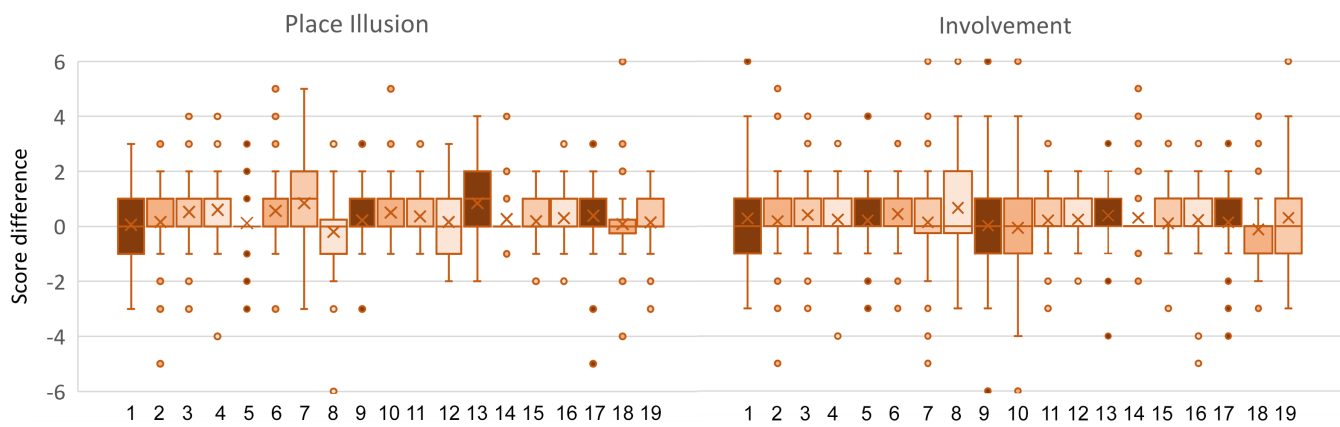


Figure 11: PQ score differences between positive and negative condition of the place illusion and involvement scenario.

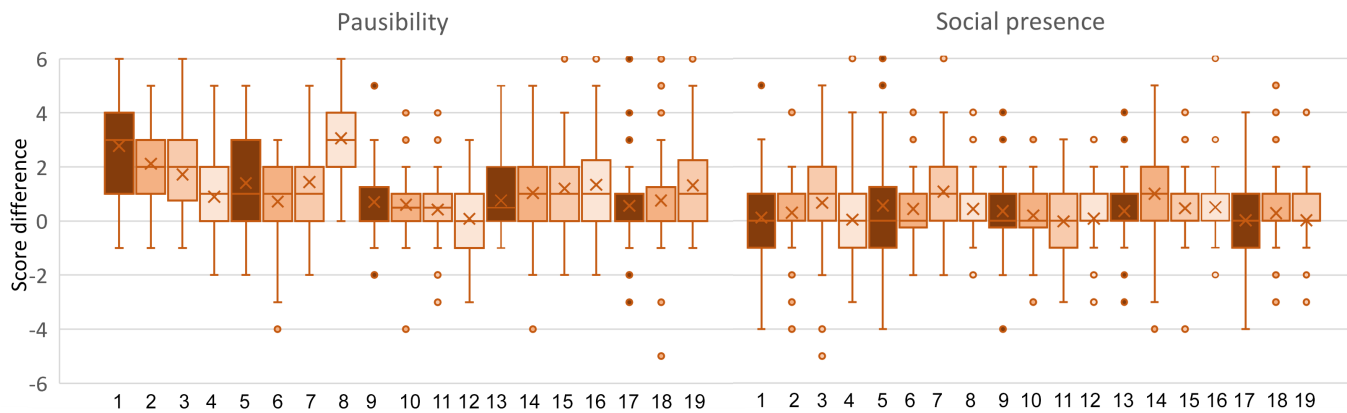


Figure 12: PQ score differences between positive and negative condition of the plausibility and social presence scenario.



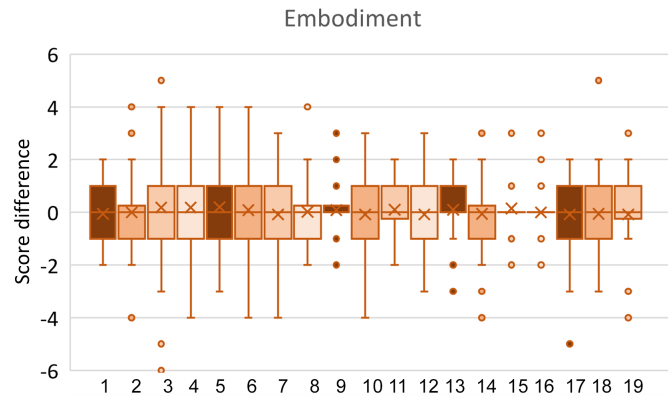


Figure 13: PQ score differences between positive and negative condition of the embodiment scenario.

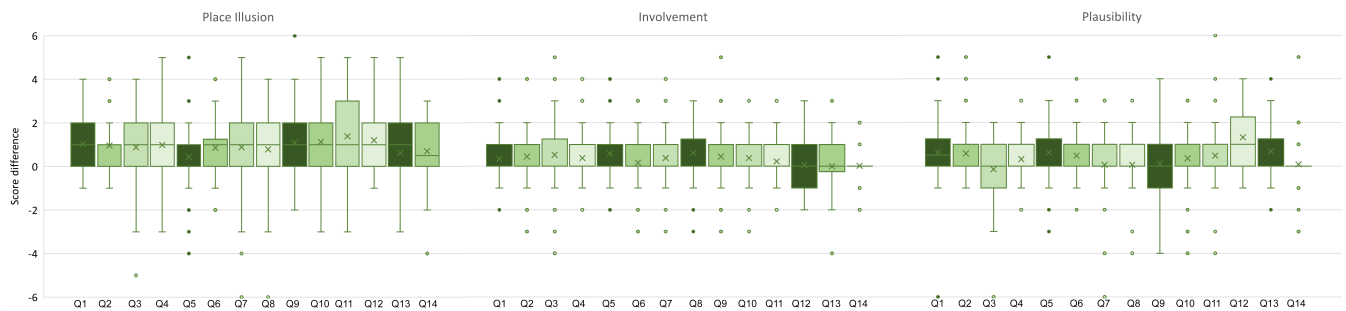


Figure 14: IPQ score differences between positive and negative condition of the place illusion, involvement, and plausibility scenario.

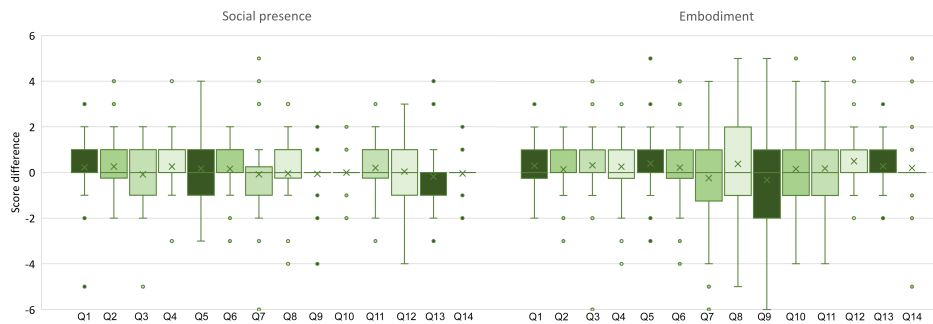


Figure 15: IPQ score differences between positive and negative condition of the social presence and embodiment scenario.

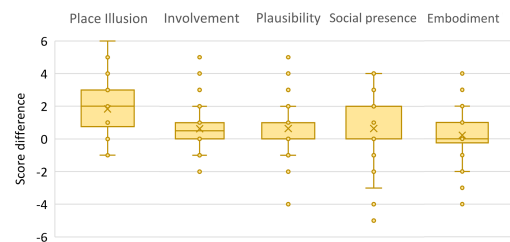


Figure 16: SIP score differences between positive and negative condition grouped by scenario.

### B.3 Approval to Score Differences

Questionnaire	SUS	PQ	IPQ	SIP
Score Difference	$M = 0.36, Md = 0.33$ $\sigma = 0.88$	$M = 0.43, Md = 0.37$ $\sigma = 0.73$	$M = 0.38, Md = 0.27$ $\sigma = 0.77$	$M = 0.57, Md = 0.7$ $\sigma = 1.25$
Approval	$M = -0.50, Md = -1.00$ $\sigma = 1.48$	$M = -0.40, Md = -1.00$ $\sigma = 1.39$	$M = -0.49, Md = 0.00$ $\sigma = 1.21$	$M = -0.26, Md = 0.00$ $\sigma = 1.73$

Table 10: The table shows the average, median, and standard deviation of the score difference between the two conditions, as well as the mean, median, and standard deviation of participants' agreement with this difference and the columns representing the four questionnaires. For agreement with the difference, 0 is to be interpreted as the target value, as it means that the participants agree with the score difference. Positive values mean that the score is higher than anticipated, negative values that the score is lower than expected.

	Presence dimension				
	PI	INV	PSI	SOC	EMB
SUS	$M = -0.52$ $Md = -1.00$ $\sigma = 1.43$	$M = -0.06$ $Md = 0.00$ $\sigma = 1.52$	$M = -1.04$ $Md = -1.00$ $\sigma = 1.22$	$M = -0.78$ $Md = -1.00$ $\sigma = 1.49$	$M = -0.10$ $Md = 0.00$ $\sigma = 1.49$
PQ	$M = -1.16$ $Md = -1.00$ $\sigma = 1.14$	$M = -0.04$ $Md = 0.00$ $\sigma = 1.30$	$M = 0.14$ $Md = 0.00$ $\sigma = 1.48$	$M = -0.52$ $Md = -0.50$ $\sigma = 1.36$	$M = -0.49$ $Md = -1.00$ $\sigma = 1.27$
IPQ	$M = -0.26$ $Md = 0.00$ $\sigma = 1.09$	$M = -0.18$ $Md = 0.00$ $\sigma = 1.09$	$M = -0.98$ $Md = -1.00$ $\sigma = 1.05$	$M = -0.46$ $Md = -1.00$ $\sigma = 1.34$	$M = -0.42$ $Md = 0.00$ $\sigma = 1.28$
SIP	$M = -0.20$ $Md = 0.00$ $\sigma = 1.6$	$M = 0.26$ $Md = 0.00$ $\sigma = 1.87$	$M = -0.76$ $Md = -1.00$ $\sigma = 1.83$	$M = -0.20$ $Md = 0.00$ $\sigma = 1.48$	$M = -0.38$ $Md = 0.0$ $\sigma = 1.67$

Table 11: Mean, median, and standard deviation of the participants' approval to the differences of the presence scores (score of positive - negative condition) split by questionnaire and condition. Positive values mean the score difference is higher than anticipated, and negative values mean the score difference is lower than anticipated. Visualization in figure 9.

## C Statistical Test Results

### C.1 Score Differences per Questionnaire

**Table 12: Results of the related-samples Wilcoxon signed-rank test comparing the score of positive and negative condition.**

Scenario	Questionnaire	Statistic (z)	p-value	CI	Effect Size (r)
Place Illusion (PI)	SUS	5.095	< 0.001	[0.5, 1.0]	0.72
	PQ	4.04	< 0.001	[0.177, 0.435]	0.57
	IPQ	5.797	< 0.001	[0.785, 1.225]	0.82
	SIP	5.33	< 0.001	[1.050, 1.750]	0.75
Involvement (INV)	SUS	2.868	0.004	[0.083, 0.5]	0.41
	PQ	2.542	0.011	[0.056, 0.318]	0.36
	IPQ	3.142	0.002	[0.151, 0.592]	0.44
	SIP	2.687	0.007	[0.0, 0.7]	0.38
Plausibility (PSI)	SUS	2.714	0.007	[0.083, 0.5]	0.38
	PQ	5.652	< 0.001	[0.903, 1.480]	0.8
	IPQ	4.735	0.001	[0.385, 0.759]	0.67
	SIP	2.554	0.011	[0.0, 0.7]	0.36
Social Presence (SOC)	SUS	1.787	0.074	[0.0, 0.417]	0.25
	PQ	3.412	< 0.001	[0.170, 0.556]	0.48
	IPQ	1.829	0.067	[−0.024, 0.507]	0.26
	SIP	2.342	0.019	[0.0, 0.7]	0.33
Embodiment (EMB)	SUS	−0.945	0.345	[−0.083, 0.333]	0.13
	PQ	1.849	0.754	[−0.01, 0.329]	0.26
	IPQ	1.849	0.065	[−0.01, 0.329]	0.26
	SIP	−0.945	0.276	[−0.083, 0.333]	0.13

## **C.2 Score Differences per Item**

Question	Scenario	Sign.	Effect size	Tag
SUS 1	PI	< .001	0.74	SUS1_presence overall
SUS 1	PSI	.004	0.40	
SUS 2	PI	< .001	0.65	SUS2_became reality
SUS 2	INV	< .001	0.51	
SUS 2	PSI	.005	0.39	
SUS 2	SOC	.003	0.30	
SUS 3	PI	< .001	0.55	SUS3_images
SUS 4	INV	.004	0.29	SUS4_dominant location
SUS 4	EMB	.019	0.33	
SUS 5	PI	< .001	0.52	SUS5_memory
SUS 5	PSI	< .001	0.51	
SUS 5	SOC	.037	0.29	
SUS 6	PI	.004	0.41	SUS6_frequency present
SUS 6	EMB	.036	0.30	
PQ 1	PSI	< .001	0.81	PQ1_control events
PQ 2	PSI	< .001	0.79	PQ2_responsiveness
PQ 3	INV	.024	0.32	PQ3_natural interaction
PQ 3	PSI	< .001	0.74	
PQ 3	SOC	.001	0.36	
PQ 4	PI	.002	0.45	PQ4_visual involvement
PQ 4	INV	.072	0.25	
PQ 4	PSI	< .001	0.51	
PQ 5	PSI	.002	0.45	PQ5_natural control
PQ 5	SOC	0.04	0.29	
PQ 6	PI	.008	0.37	PQ6_compelling objects
PQ 6	INV	.002	0.32	
PQ 6	PSI	.004	0.40	
PQ 6	SOC	< .05	0.28	
PQ 7	PI	< .001	0.53	PQ7_consistency
PQ 7	PSI	< .001	0.70	
PQ 7	SOC	< .001	0.59	
PQ 8	INV	.016	0.34	PQ8_anticipation
PQ 8	PSI	< .001	0.85	
PQ 8	SOC	.012	0.35	
PQ 9	PSI	< .001	0.47	PQ9_visual exploration
PQ 10	PI	.001	0.45	PQ10_compelling navigation
PQ 10	PSI	.002	0.43	
PQ 11	PI	.005	0.40	PQ11_examination detail
PQ 11	PSI	.03	0.31	
PQ 13	PI	< .001	0.57	PQ13_involvement
PQ 13	INV	.038	0.29	
PQ 13	PSI	< .001	0.52	
PQ 14	PI	.017	0.34	PQ14_delay
PQ 14	PSI	< .001	0.50	
PQ 14	SOC	< .001	0.52	
PQ 15	PSI	< .001	0.59	PQ15_adjustment
PQ 15	SOC	.007	0.38	
PQ 16	PSI	< .001	0.66	PQ16_proficiency
PQ 16	SOC	.006	0.38	
PQ 17	PI	.017	0.34	PQ17_display quality
PQ 17	PSI	.059	0.27	
PQ 18	PSI	.010	0.37	PQ18_interface control
PQ 19	PSI	< .001	0.64	PQ19_task focus

**Table 13: Results of the related-samples Wilcoxon signed-rank test for SUS and PQ showing a significant difference between the scores of the negative and positive condition per questionnaire item.**

IPQ 1	PI	< .001	0.67	IPQ1_presence overall
IPQ 1	PSI	.001	0.45	
IPQ 2	PI	< .001	0.64	IPQ2_responsiveness
IPQ 2	INV	.021	0.33	
IPQ 2	PSI	.002	0.44	
IPQ 3	PI	< .001	0.49	IPQ3_images
IPQ 4	PI	< .001	0.60	IPQ4_no presence
IPQ 4	INV	.031	0.30	
IPQ 5	PI	.020	0.33	IPQ5_acting outside
IPQ 5	INV	.003	0.43	
IPQ 5	PSI	.004	0.41	
IPQ 5	SOC	.035	0.30	
IPQ 6	PI	< .001	0.58	IPQ6_presence
IPQ 6	PSI	.008	0.38	
IPQ 7	PSI	< .001	0.48	IPQ7_awareness navigation
IPQ 8	PI	< .001	0.48	IPQ8_awareness realworld
IPQ 8	INV	.003	0.30	
IPQ 9	PI	< .001	0.55	IPQ9_attention realworld
IPQ 9	INV	.036	0.30	
IPQ 10	PI	< .001	0.63	IPQ10_captivated
IPQ 10	INV	.046	0.28	
IPQ 10	PSI	.050	0.28	
IPQ 11	PI	< .001	0.66	IPQ11_realism1
IPQ 11	PSI	.044	0.28	
IPQ 12	PI	< .001	0.67	IPQ12_consistency
IPQ 12	PSI	< .001	0.66	
IPQ 12	SOC	.026	0.32	
IPQ 13	PI	.014	0.35	IPQ13_realism2
IPQ 13	PSI	.001	0.45	
IPQ 14	PI	< .001	0.49	IPQ14_realism dominance
SIP	PI	< .001	0.75	SIP_presence overall
SIP	INV	.007	0.38	
SIP	PSI	.011	0.36	
SIP	SOC	.019	0.33	

**Table 14: Results of the related-samples Wilcoxon signed-rank test for IPQ and SIP showing a significant difference between the scores of the negative and positive condition per questionnaire item.**

### C.3 Analysis of Presence Dimensions Included in PQ and IPQ

Scenario/ Dim	PQ: Involvement	PQ: Naturalness
Place illusion	(z=3.605, p<0.001, r=0.51, CI=[0.136,0.445])	(z=3.484, p<0.001, r=0.49, CI=[0.333,0.833])
Involvement	(z=3.527, p<0.001, r=0.50, CI=[0.136,0.409])	(z=1.574, p=0.116, r=0.22, CI=[0.0,0.5])
Plausibility	(z=6.107, p<0.001, r=0.86, CI=[1.091,1.636])	(z=5.793, p<0.001, r=0.82, CI=[1.0,1.833])
Social Presence	(z=3.530, p<0.001, r=0.50, CI=[0.136,0.5])	(z=3.649, p<0.001, r=0.52, CI=[0.333,1.0])
Embodiment	(z=0.19, p=0.849, r=0.03, CI=[-0.182,0.227])	(z=0.234, p=0.815, r=0.03, CI=[-0.333,0.5])
Scenario/ Dim	PQ: Resolution	PQ: Interface quality
Place Illusion	(z=2.235, p=0.025, r=0.32, CI=[0.0,0.5])	(z=1.591, p=0.112, r=0.23, CI=[0.0,0.333])
Involvement	(z=1.990, p=0.047, r=0.28, CI=[0.0,0.5])	(z=0.631, p=0.528, r=0.09, CI=[-0.167,0.333])
Plausibility	(z=1.376, p=0.169, r=0.19, CI=[0.0,0.5])	(z=3.967, p<0.001, r=0.56, CI=[0.333,1.167])
Social Presence	(z=0.21, p=0.834, r=0.03, CI=[-0.25,0.25])	(z=0.299, p=0.765, r=0.04, CI=[-0.167,0.333])
Embodiment	(z=-2.192, p=0.028, r=0.31, CI=[-1.125,-0.125])	(z=-1.182, p=0.237, r=0.17, CI=[-0.75,0.25])
Scenario/ Dim	IPQ: Spatial presence	IPQ: Involvement
Place Illusion	(z=4.718, p<0.001, r=0.67, CI=[0.5,1.1])	(z=4.686, p<0.001, r=0.66, CI=[0.625,1.250])
Involvement	(z=3.183, p=0.001, r=0.45, CI=[0.2,0.6])	(z=3.123, p=0.002, r=0.44, CI=[0.125,0.75])
Plausibility	(z=2.494, p=0.013, r=0.35, CI=[0.1,0.6])	(z=1.647, p=0.1, r=0.23, CI=[0.0,0.375])
Social Presence	(z=1.932, p=0.053, r=0.27, CI=[0.0,0.5])	(z=0.0, p=1.0, r=0.0, CI=[-0.375,0.5])
Embodiment	(z=0.316, p=0.752, r=0.04, CI=[-0.25,0.25])	(z=-0.657, p=0.511, r=0.01, CI=[-0.333,0.167])
Scenario/ Dim	IPQ: Realness	
Place Illusion	(z=5.269, p<0.001, r=0.75, CI=[0.75,1.250])	
Involvement	(z=1.294, p=0.196, r=0.18, CI=[-0.125,0.375])	
Plausibility	(z=3.927, p<0.001, r=0.56, CI=[0.25,0.75])	
Social Presence	(z=2.623, p=0.009, r=0.37, CI=[0.0,0.5])	
Embodiment	(z=-0.098, p=0.922, r=0.01, CI=[-0.25,0.025])	

**Table 15: Summary of the results of the Wilcoxon signed rank test for related samples for the presence dimensions of PQ and IPQ per condition as well as the corresponding 95 %-confidence interval.**