The Construction and Validation of the SP-IE Questionnaire: An Instrument for Measuring Spatial Presence in Immersive Environments

Nawel Khenak, Jean-Marc Vézien, Patrick Bourdot

The Construction and Validation of the SP-IE questionnaire: An instrument for measuring Spatial Presence in Immersive Environments

Nawel Khenak Jeanne Vézien Patrick Bourdot
VENISE Team, LIMSI, CNRS, Paris-Saclay University, Orsay, France

Abstract. The present study describes the construction and validation of an instrument for measuring spatial presence (the sense of "being there") in the context of highly immersive environments: the SP-IE [Spatial Presence for Immersive Environments] questionnaire, for use in the French-speaking community. A first raw version of the questionnaire was submitted to an item selection procedure and reliability tests with 67 participants. An exploratory factor analysis (EFA) with 179 participants was then employed on the resulting version of the questionnaire to explore its underlying scales. Finally, the outcome scale-structure from the EFA was evaluated using confirmatory factor analyses (CFAs). This process resulted in a well-structured 20-item questionnaire, based on seven scales: (i) the sense of spatial presence, (ii) the affordance of the environment, (iii) the user's enjoyment, (iv) the user's attention allocation to the task, (v) the sense of reality attributed to the environment, (vi) the social embodiment with avatars, and (vii) the possible negative effects of the environment (cybersickness). Results showed overall good internal consistency and satisfactory convergent validity of the scales. The fit indexes obtained ($X^2/df = 1.34$, GFI = .95, CFI= .90 , TLI = .87, SRMR = .068 , and RSMEA = .045) demonstrated a good fitness of the structure proposed. However, even though the scale structure proposed in this paper was confirmed, its low discriminant validity encourages further evaluations.

Keywords: Spatial Presence, Questionnaire Validation, Immersive Environments.

1 Introduction

In Virtual Reality (VR), Spatial Presence is defined by the user's sensation of being located in an environment when it is mediated or virtually represented by means of technologies [1]. It represents an important key to enhance the effectiveness of VR applications. For instance, Spatial Presence can facilitate the transfer of information needed for the successful conduct of surgical operations [2], or the transfer of learning during teaching [3]. It can also intensify the positive effects of the applications and their impact on users' emotional reactions such as enjoyment and satisfaction in virtual games [4], and fear and anxiety in virtual therapies [5]. Consequently, researchers focused on evaluating the sense of presence in different contexts [6, 7]. To this end, they developed instruments to assess presence and determine its underlying factors. While multiple physiological and behavioral indicators have been proposed [8], validated questionnaires are still the most common method for measuring this construct [9]. Among them, the most cited questionnaire is the Presence Questionnaire (WS) designed by Witmer and Singer [10], which has been used in hundreds of studies. This also went for the Slater-Usoh-Steed (SUS) questionnaire [11], the Igroup Presence Questionnaire

(IPQ) [12], and the ITC-Sense of Presence Inventory questionnaire [13]. Using different items and subscales, such questionnaires provide scores, which allow highlighting different factors of Spatial Presence such as the user enjoyment [10, 14], the naturalness of the environment [13], and the attention allocation on the task [10, 15] to name a few. Other factors have also been studied which play a role in the emergence of this sense. In particular, the sensorial and behavioral fidelity of the systems (respectively immersion [16, 17] and interaction [18]) has proved to be important criteria in increasing the user's feeling of Spatial Presence [19-22].

Recent technological advances in VR, including visual quality of head-mounted displays (HMDs), sound spatialization, more efficient tracking systems, and overall system latency reduction, allow the creation of environments with higher sensorial and behavioral fidelity. Such environments, in addition to allowing users to experience a higher sense of spatial presence, could increase the sense of reality (experienced realism [23]) attributed to the virtual or mediated space and the affordance (possibility to act [24]) that shape the user's mental representation of what bodily actions are possible in the environment, which in turn activates the sense of presence [17, 21]. Exploring the formation of spatial presence and the impact of its underlying factors in such environments would provide cues to design better immersive VR experiences. To do so, this paper proposes a new questionnaire to assess spatial presence in highly immersive environments, independent of the type of the environment. The questionnaire was developed within the French-speaking population and was accordingly validated in French. Therefore, the aim of the present study is twofold: (a) exploring the underlying factors of spatial presence in immersive environments, mainly affordance of the environment, user's interest and attention on the activity, and the sense of reality, and (b) provide a validated questionnaire for assessing spatial presence and its factors within the French-speaking population in different environments.

The paper is structured as follows: The first section provides an overview of the instruments developed to assess Spatial Presence. In particular, subjective questionnaires for measuring presence are listed in detail. The second section describes the approach followed to construct the questionnaire, and the item reduction procedure to develop an initial questionnaire with satisfactory reliability. The third section describes the validation process based on an Exploratory Factor Analysis (EFA) and Confirmatory Factor Analyses (CFAs) used to establish the construct and discriminant validity of the questionnaire and its psychometric properties. The fourth section reports the results of the validation process with an interpretation of these results. The fifth and last section concludes the paper with recommendations and future perspectives.

2 Related Work

In order to determine the process of formation of Spatial Presence and evaluate its relationship with potential factors, it is important to establish reliable measures for Spatial Presence. To achieve this goal, a large part of studies proposed several methods to assess Spatial Presence. These methods can be divided into objective measures (using behavioral and/or physiological indicators [25, 26]), and subjective measures (using subjective ratings or questionnaires [10-14]).

2.1 Objective measures

Physiological measures. Studies on the reliability of physiological indicators to measure presence such as changes in Heart Rate (HR) [27] and skin temperature and conductance based on Electrodermal Activity (EDA) [28, 29] provided promising results. For example, Meehan et al. [30] showed that in a stressful virtual environment (VE) depicting a pit room, changes in HR correlated positively with self-reported presence. However, these measurements require a baseline comparison for each participant, which means a considerable effort in some study designs. In addition, it has been shown that additional equipment to measure physiological responses (e.g. Electroencephalography (EEG) to measure brain responses [31]) can be a cause of breaks in presence [32]. Moreover, this equipment is more efficient when participants do not move [33], which reduces the scope of possible experiments.

Behavioral indicators. The relationship of behavioral indicators with presence was also studied. These indicators are based on direct observation of users' behavior such as adaptive behaviors evoked by virtual dangers [34] and body movement in response to the context of the VE [35]. For example, Usoh et al. [36] run an experiment in which participants were located in a virtual corridor with a virtual pit. They were interested into what extent people were willing to walk out over the pit. The behavioral measure they used was the path participants actually chose when they navigated to a chair on the other side of the pit. They found that there was a positive correlation between the behavioral measure and subjective presence measured by a questionnaire. More recently, Lepecq et al. [37] studied the correlation between postural adjustment of the body in an experiment in which participants had to walk through either a virtual or a real aperture. Results showed that participants swiveled their body similarly in both real and virtual situations.

Thus, physiological and behavioral indicators exist that could be potentially reliable measures for presence (for more details, see Lombard et al. [38, pp 150-185]). Yet, investigation is still needed to evaluate the correlation between them and the sense of presence [39]. A common approach to achieve this goal is to compare results from this kind of measures with results from presence questionnaires [40, 41].

2.2 Subjective Questionnaires

Presence questionnaires are the most common method for assessing presence as they have been shown to be sensitive enough to find differences in presence [9]. The earliest questionnaire to measure presence in VEs was proposed by Barfield and Weghorst in 1993 as a 6-item one-dimensional questionnaire [42]. Similarly, Slater and Chrysanthou [43] proposed a one-dimensional questionnaire in which the presence score is taken as the number of answers that have a high score. Also, Kim and Biocca [14] designed a questionnaire based on their metaphor of transportation comprising two dimensions: (i) arrival, being present in the mediated environment, and (ii) departure, not being present in the unmediated environment.

Table 1. Overview of most known presence questionnaires in the literature (the most used questionnaires are highlighted in bold).

Authors	Year	Items	Subscales
Witmer & Singer (WS) [10, 45]	**1998**	**32**	**Involvement** **Naturalness** **Concentration**
Usoh et al. (SUS) [11]	2000	6	Spatial presence VE is the dominant reality VE is remembered as a place
Lessiter et al. (ITC-SOPI) [13]	**2001**	**44**	**Spatial Presence** **Engagement** **Naturalness** **Negative effects**
Schubert et al. (IPQ) [12, 21, 23]	**2003**	**20**	**Spatial Presence** **Involvement** **Experienced Realism** **Immersion** **Interaction**
Vorderer et al. (MEC-SPQ) [50]	2007	32-64	Spatial Presence Attention Allocation Possible Actions Involvement Suspension of Disbelief Domain Specific Interest Spatial Situation Model Visual Spatial Imagery
Lombard et al. (TPI) [48]	2009	4-8	Spatial Presence Immersion Engagement perceptual realism social presence-actor passive social presence active social presence

However, as Spatial Presence was early on considered as a multi-dimensional construct [13, 44], researchers quickly focused on developing multi-scale questionnaires rather than one-dimensional questionnaires to take into account the different factors of presence. Table 1**.** summarizes the most known multi-scale questionnaires. In 1998, Witmer and Singer [10, 45] designed a 32-item questionnaire (WS) based on three subscales: (i) involvement, (ii) behavioral fidelity (naturalness), and (iii) user's ability to concentrate on the tasks. The questionnaire was criticized for the low number of items directly assessing presence [23, 46]. In addition, it was not able to discriminate between presence in different environments (real vs. virtual environment) [11]. Nevertheless, the questionnaire has been translated into French by the Cyberpsychology laboratory

of UQO [47]. Usoh et al. [11] developed the Slater-Usoh-Steed (SUS) questionnaire based on three themes: (i) the sense of being in the VE (spatial presence), (ii) the extent to which the VE becomes the dominant reality, and (iii) the extent to which the VE is remembered as a 'place'. The current version of the questionnaire has six items. However, although being a popular instrument, the questionnaire was criticized for measuring only one dimension of presence: "presence as transportation" [48]. In addition, as for the WS questionnaire, the SUS was not able to discriminate between presence in a VE and presence in a physical reality [11].

Later, Lessiter et al. [13] developed the Independent Television Commission Sense of Presence Inventory (ITC-SOPI) consisting of 44 items organized in four subscales: (i) the sense of spatial presence, (ii) the user's engagement, (iii) the ecological validity of the environment (naturalness), and (iv) the negative effects (such as cybersickness). One of the advantages of this instrument is its applicability to several types of environments. In addition, it is easy to administer and score. However, its use is somewhat limited due to the restrictions imposed by its proprietors. Later in 2009, Lombard et al. [48] refined the ITC-SOPI and introduced the Temple Presence Inventory (TPI), which aimed to measure eight subscales: (i) spatial presence (sense of transportation), (ii) social richness (immersion), (iii) engagement, (iv) social realism, (v) perceptual realism, (vi) social presence-actor within medium, (vii) passive social presence, and (viii) active social presence. However, the low number of items (one item per subscale) makes the construct validity of the instrument questionable.

Another common questionnaire is the Igroup Presence Questionnaire (IPQ) created by Schubert et al. [12, 21, 23] by combining both the Slater and Chrysanthou's [43] and Kim and Biocca's [14] questionnaires. It is based on eight factors of 20 items. Three of them, merging 13 items, were found to be directly concerned with presence: (i) the sense of spatial presence, (ii) the involvement into the environment, and (iii) the sense of reality attributed to the virtual space (experienced realism). The others were considered as immersion and interaction variables that may influence presence. The IPQ has been translated into French but has not been subjected to a proper validation procedure [49].

Finally, Vorderer et al. [19, 50] developed the Measurement, Effects, Condition Spatial Presence Questionnaire (MEC-SPQ). It is based on eight factors: (i) spatial presence, (ii) attention allocation, (iii) possible actions, (iv) involvement, (v) suspension of disbelief, (vi) domain specific interest, (vii) spatial situation model, and (viii) visual-spatial imagery. The questionnaire has the advantage to be applicable to different type of environments. However, its varying number of items (between 4 and 8 per scale, 32 and 64 for the overall questionnaire) make the comparison between the environments difficult. In addition, no evaluation of its construct validity was made.

Thus, many questionnaires to assess presence and its factors have been proposed into the literature. However, some factors related to the sensorial and behavioral fidelity that new immersive systems could provide were disregarded. In addition, in the case where these questionnaires were subjected to a validation procedure, it was only in the context of English-speaking population that is no guarantee of their validity in other languages [51]. Thus, no properly validated questionnaires exist within the French-speaking context. Consequently, the current paper develops the Spatial Presence Questionnaire for Immersive Environments (SP-IE), combining the previous questionnaires with other factors that could play a role in the emergence of Spatial Presence. Moreover,

the questionnaire is constructed in French and its reliability, construct validity, and psychometric properties are determined in order to allow thereafter its use in the French-speaking population to compare spatial presence between different environments.

3 Construction of the questionnaire

This section describes the two steps followed to construct the questionnaire, namely the item generation and translation procedure. An item reduction analysis using internal consistency (reliability) test and item-total correlation is then performed on the questionnaire based on data collected from two user studies run in different environments.

3.1 Scale Construction and Item Generation

The aim of this stage is to develop a raw set of items associated with potential scales in the questionnaire. Each scale will represent one principal factor of spatial presence.

Theoretical background. Based on the literature review and following a multi-dimensional approach, six categories of factors were determined for the scale construction of the questionnaire:

1. The *technological factors* related to the ability of the system to be high immersive (depth, breadth, and consistency) [20] and interactive (control and modification) [21].
2. The *content-related factors* related to the degree of naturalness and sense of reality attributed to the environment, and its affordance [52] (user's perception of possible actions [24] and matching with its expectations [53]).
3. The *user factors* related to the users' involvement [13], their satisfaction, and the willing of suspension of disbelief [19].
4. The *activity factors* related to the users' interest in the activity [50], and the attention allocated to the task [10].
5. The *social embodiment factors*, related to the user's feeling of being with other entities, defined as "the sense of social presence" [54, 55], and the influence of embodied avatars [56].
6. The *negative factors* related to latency as perceived by users in an environment (lags and interruption) [57], and cybersickness (nausea, headaches, and dizziness) [58].

These factors were inspired from empirical studies and previous multi-dimensional questionnaires on Presence (see Related Work section), except for the negative factors (perceived latency and cybersickness) that were neglected in previous questionnaires despite of their influence in the formation of spatial presence. Indeed, latency causes breaks in display that disturbs the users and are likely to reduce their feeling of presence [59], while applications with fast update rates (low latency) can create a better illusion of continuous and fast responses of the environment to users actions, and therefore increase their sense of being in this environment [57]. Nevertheless, no questionnaires including latency-related items exist. Therefore, the questionnaire developed in this paper will attempt to associate items with perceived latency issues. By including such

items in a presence questionnaire, evaluation on the consequences of latency on spatial presence will be possible.

Conversely, cybersickness was the concern of many studies that demonstrated its negative correlation with presence [58, 60]. However, this concept was mostly evaluated using specific cybersickness questionnaires independent from presence-assessing questionnaires because of its complexity. Consequently, except for the ITC-SOPI questionnaire [13], no presence questionnaires included items related to cybersickness issues. However, assessing presence and cybersickness within the same questionnaire would be beneficial, on one hand for directly evaluating the correlation of cybersickness with different factors of spatial presence, and on the other hand to save participants' time.

In addition, particular attention was drawn to include in the questionnaire social factors related to the sense of social presence referred to the feeling of "being with others" in virtual or mediated environments [54]. Indeed, many studies demonstrated the positive relationship between social and spatial presence [61, 62]: social presence can provide strong evidence of the existence of the virtual or mediated space, and therefore improve the sense of spatial presence [63]. Furthermore, Social Presence can be experienced with other human or nonhuman entities physically represented or psychologically assumed [64]. Nevertheless, this sense is enhanced by using *avatars* as they promise users the affordance of 'real' bodies by physically representing the whole body or parts of the body (such as projected hands) [65]. More precisely, avatars enable *embodiment* [53, 56] and provide users with new possibilities to interact with themselves and the environment, which in turn enhance social presence [66, 67]. Again, the fidelity of the avatars in representing the actual self of users is of major importance [18, 68]. Different questionnaires to assess Social Presence were developed [69]. However, except for the TPI [48], no spatial presence questionnaire included factors related to social presence as potential subscales. In the current paper, the SP-IE questionnaire will attempt to include these social factors.

Thus, the six categories of factors were considered for the construction of the SP-IE questionnaire. In addition, a scale that aimed to assess Spatial Presence was added to the questionnaire with items from different questionnaires. Consequently, the SP-IE consisted in a seven-scale construct: (i) Spatial Presence, (ii) Fidelity, related to the sensorial and behavioral fidelity of the system, (iii) Affordance, related to the content of environment, (iv) Involvement, related to the users' enjoyment and state of mind, (v) Attention, related to the user's engagement and attention allocated to perform the task/activity, (vi) Social embodiment, related to the sense of social presence with avatars, and (vii) Negative Effects, related to perceived latency and cybersickness.

According to this assumption, items that assessed each scale were generated. The semantic content of the items was based on previous presence questionnaires, mainly the WS [45], the IPQ [12], the ITC-SOPI [13], and the MEC-SPQ [50] questionnaires. These questionnaires have been widely used within the literature and proved to be reliable instruments (cf. Table 1**.**). Consequently, basing the items on these questionnaires ensures a more reliable content of the SPI questionnaire. In addition, some items were proposed that could be potentially relevant to represent issues and factors not taken into

account by the previous questionnaires. Finally, each item had to satisfy a number of criteria in order to obtain an adequate questionnaire as follows (based on [13]):

- No item should directly ask participants how present they feel: the understanding of presence should not be assumed.
- Each item should avoid addressing two or more issues.
- Items should not make reference to specific systems (form) and environments (content).

Therefore, 60 items were initially generated that tapped possible manifestations of the different scales of the SP-IE questionnaire. A discussion was held to reach an agreement to delete redundant items and combine some items into one to eliminate content overlap, and thus, shortening the questionnaire. After this procedure, a set of 50 items remained. Each item represented one of the seven potential subscales as following: five items represented the "Spatial Presence" scale, 12 items represented the "Fidelity" scale, seven items represented the "Affordance", seven items represented the "Involvement" scale, nine items represented the "Attention" scale, four items represented the "Social Embodiment" scale, and six items represented the "Negative" scale.

A 5-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*) was employed for the evaluation of each item in order to reduce the central tendency bias [70].

Translation. Given that the questionnaire will be administered to a French-speaking population, all the items needed to be translated in French in order to fit the cultural (and linguistic) context.

The translation was made based on the back-translation method [71]. Initially, two members fluent in both languages translated independently the English version of the questionnaire to French, and, without consulting the original version, they back-translated the questionnaire from French to English. Then, they met to evaluate both their French and English versions and agreed on final versions. In addition, the items translated were compared to the French version of the IPQ [49] and WS [47] questionnaires. Then, the two members compared their English version with the original version of the questionnaire and made minor modifications to reach a satisfactory semantic and content equivalence in all items. These corrections lead to several item corrections in the French language version until a consensus was reached among the two members that certified that there were no incompatibilities with the original version with respect to the specific terminology and technical terms. This first step resulted in a French language version of the SP-IE questionnaire.

A second step was taken to analyze the form and content of items in terms of clarity and comprehensibility [72]. A committee composed of six persons: three other members of the team that have a good understanding of the presence concept and three external persons with no specific knowledge of the concept, were individually asked to indicate their agreement or disagreement regarding the clarity and relevance of the items in the questionnaire. Based on the comments received from this committee, some items were slightly reworded to be more understandable and to ensure the questionnaire could be completed within a reasonable time frame (10-15 minutes).

Finally, demographic information such as gender, age, and VR experience was added to the questionnaire. In addition, the questionnaire was rendered anonymous to preserve the integrity of responses.

3.2 Item analysis and reduction

In order to reduce the number of items and purify the scales, the fifty items the raw-version of the questionnaire were analyzed. The questionnaire was submitted to participants that were exposed to different environments in order to take into account the variation of systems and contents [13]. Then, an item analysis was performed on the data collected to evaluate the overall reliability and internal consistency of the items. The experimental and statistical procedures, as well as the results, are described below.

Samples and experimental procedures. The questionnaire was used in two experiments. In the following, the sample and experimental procedure of both experiment are briefly explained.

Experiment 1 - Remote vs. Real. (Number of participants: N = 29; location: L = local laboratory). In this experiment, two rooms with a very similar layout were used (see Fig. 1.): (1) an "operating room", representing a rectangular office where 12 tablets were attached to the walls at fixed positions, and (2) a "tele-operating room" where a teleoperation system including an HTC-Vive, leap motion, and binaural audio headset, allowed participants to be *remotely transported* in the operating room.

The participants were seated in the middle of one of the two rooms and had to perform a pointing task (see Appendix 1**.** for an overview of the general settings of the participants). More specifically, the task consisted in pointing as fast as possible a sequence of images that were displayed sequentially (i.e. one image at a time) on the tablets in a time limit of 3 minutes. One person at a time could perform the task. For more information about the experimental design, the reader is reported to [73].

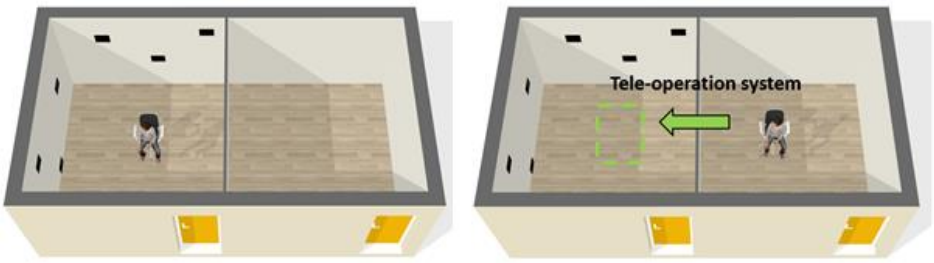

Fig. 1. 3D overview of the rooms. (Left) The operating room. (Right) The tele-operating room.

Experiment 2 - Drone Arena. (N = 40; L = Arena Drone pilot center, Lille, France). Drone Arena (https://www.dronearena.com/) is a pilot center for drones races, open to the general public. During the experiment, the participants sat in front of a tuned car steering wheel that allowed them to control the movements of drones. The drones were located in an immersive remote environment (see Fig. 2**.**). To access this environment,

participants wore an immersive headset that transmitted, in real time, the images filmed by an onboard camera. The task consisted in finishing a circuit as quickly as possible without crashing the drones. Up to six persons could play at the same time. The duration of the task was about 20 minutes. This experiment was selected to ensure that the total sample would be representative of a homogenous population.

Fig. 2. (Left) A drone located in the immersive environment of Drone Arena. (Right - Top) General setting of a participant. (Right-Bottom) The First Person View (FPS).

The completion of the questionnaire took place after the experiments in a calm space, either in isolation or in small groups (never involving more than six people). Before completing the questionnaire, the participants were informed of the research objectives and signed a free and informed consent (IC) agreement, which guaranteed anonymity and confidentiality of all collected data. Table 2. reports the demographics of participants for each experiment. Four participants were withdrawn from the analysis (one participant in the "Real vs. Remote" experiment due to a technical problem, and three participants in "Drone Arena" experiment due to missing values in the questionnaire). All in all, 65 participants provided complete datasets. Those data were used for the process of the item analysis as described in the following subsection.

Table 2. The demographics of the participants.

Environnent		**Sample size**	**Males and Females**	**M*age* and SD**	**% VR experience N / B / I / E**
Real vs. Remote	**Operating condition**	14	13 males 3 females	26.5 +- 4	7% / 64% / 7% / 21%
	Tele-operating condition	14	8 males 6 females	28 +- 5	28% / 42% / 14% / 14%

Drone Arena	37	25 males 12 females	30 +- 7	40% / 49% / 11% / 0%
Total	**65**	**46 males 19 females**	**29 +- 6**	**31% / 50% / 11% / 8%**

N: none, B: beginner, I: Intermediary, E: expert.

Statistical analysis and results. All the analyses were performed with R 3.6.0. First, the descriptive statistics (mean and standard deviation) of the collected responses were calculated for each item and scale (cf. Additional Materials - Material 1).

The internal consistency of the scales was calculated using Cronbach's alpha: Spatial Presence $\alpha = 0.72$, Fidelity $\alpha = 0.63$, Affordance $\alpha = 0.75$, Involvement $\alpha = 0.48$, Attention $\alpha = 0.70$, Social Embodiment $\alpha = 0.47$, Negative Effects $\alpha = 0.80$. The overall Cronbach's alpha coefficient for all the items was $\alpha = 0.87$.

Then, an item analysis was performed using Cronbach's alpha and Total Inter-Item Pearson correlation coefficients as follows: the alpha value of each item was calculated and items that did not contribute to elevating the Cronbach's alpha of their corresponding scales (i.e. that reduce the alpha value) were excluded from the questionnaire. The same went for unsatisfactory items with an item-total correlation coefficient that failed to load above 0.3 [74] (cf. Additional Materials – Material 2). Consequently, 18 items were removed: one item was deleted from the "Spatial Presence" scale, six items from the "Fidelity" scale, three items from the "Affordance" scale, two items from the "Attention" scale, and two items from the "Negative" scale.

Finally, two items were removed because participants had trouble understanding them: one item from the "Attention" scale and one from the "Negative" scale.

Therefore, the item analysis resulted in a modification version of the SP-IE with 30 items ranging over the seven scale of the questionnaire as follows: four items of "Spatial Presence" scale with $\alpha = 0.71$, six items of "Fidelity" scale $\alpha = 0.62$, four items of "Affordance" with $\alpha = 0.71$, five items of "Involvement" with scale $\alpha = 0.62$, four items of "Attention" scale with $\alpha = 0.73$, four items of "Social Embodiment" scale with $\alpha = 0.58$, and three items of "Negative" scale with $\alpha = 0.85$. The Cronbach's alpha of the revised SP-IE was 0.89 indicating overall good reliability (according to [75] alpha values above 0.70 indicate good reliability). No more improvements based on further items removal were considered because minimal gains would be obtained.

4 Validation of the questionnaire

Any measure of presence must be shown to be both reliable and valid in order to be recommended for Presence research [10]. Therefore, the revised version of the SP-IE questionnaire was subjected to a validation process to analyze its reliability, construct validity, and structural adequacy. In the following section, the process is described, followed by the results of the analysis in the next section.

4.1 Sample

In order to evaluate the validity of the questionnaire, an investigation was run on a sample consisting of 179 participants (119 men and 60 women) with ages ranging from18 to 56 years old (M = 30.51; SD = 8.06). Ten participants were removed from the analysis because of missing values in their questionnaires. Of the remaining 169 participants, 42 (25%) had good experience with virtual reality devices, 83 (49%) had some previous experience and 44 (26%) did not have any experience.

4.2 Experimental Design

Environments. All participants were recruited at the ILLUCITY Park for VR highly immersive experiences (Paris, France, https://illucity.fr/en/). This park proposes 20 immersive games divided into different categories: the escape games, the arcade games and the cinematic experiments (VR films). Some of them are multiplayer (up to 6), while others are single-player. It is accessible to all people, from gamer audience to people with no VR experience at all. Depending on the game, the duration of the experiment may vary from 5 minutes to 40 minutes. Of the 20 experiments, 12 games were the most popular and were therefore chosen to run the investigations. Table 3**.** describes each experiment.

Table 3. The description of the experiments of IllUCITY Parc chose to run the investigation.

The Game	Category	Number of players	Duration (min)	Number of participants
Toyland: Crazy Monkey	Arcade	3-6	25	53
Assassin's CREED: The lost Pyramid	Escape	2-4	40	43
The Raft	Arcade	2-4	10	20
Incarna	Escape	3-4	40	15
Eclipse	Escape	2-4	35	13
Space Pirate Trainer	Arcade	1	10	6
The Corsair's Curse	Escape	2-4	35	5
Space Flight	Film	1	7	4
Knightfall	Arcade	1	7	4
Ragnaröck	Arcade	2	15	2
Far Reach	Film	1	5	2
Asteroids	Film	1	11	2

Hardware. In all the experiments, the participants were equipped with HTC Vive Pro headsets and MSI VR One or HP VR backpacks. The backpack-based configuration allowed removing the influence of tethering on physical movements of participants. With this setup, the applications were running at 90/100 frames per second (fps). For arcade and escape games, the interaction with the environments was made possible using the two HTC-Vive controllers, except for Toyland experiment (see Table 3**.**) where the controllers were replaced with haptic rifles. For VR films, the participants were seated in a D-Box simulator for highly sensorial experiences.

Table 4. Cut-off values for the evaluation of a structure during the CFAs.

Fit index	**Cut-off value**
Normed Chi-square (x2/df) [CMIN]	< 3
Normed Fit Index [CFI]	> 0.9
Goodness of Fit Index [GFI]	> 0.9
Trucker-Lewis Index [TLI]	> 0.9
Standardized Root Mean Square Residual [SRMR]	< 0.08
Root Mean Square Error of Approximation [RMSEA]	< 0.06

Procedure. The completion of the questionnaire took place after the experiments. People who participated in one of the 12 experiments were asked if they could volunteer to fill a questionnaire. For the people who accepted, a paper-pencil version of the questionnaire was administered. Before completing the questionnaire, they were informed of the research objectives and signed a free and informed consent (IC) agreement. The duration to fill the questionnaire was about 5 to 10 minutes. After the participants completed the questionnaire, they were asked not to discuss the questionnaire with other people that could potentially participate in the investigation. The data collected were then used for the process of statistical analyses as described in the following subsection.

4.3 Statistical Analyses

An exploratory factor analysis (EFA) was performed to explore the underlying scale structure of the SP-IE, employing a Principal Axis Factoring (PAF) with oblique rotation. In principle, the obtained variables should coincide with the theoretical structure proposed in this paper (see section 3.1).

The resulting version of SP-IE from the EFA was then submitted to confirmatory factor analyses (CFAs), using the weighted least square mean and variance adjusted (WLSMV) estimator in order to confirm the factor construct of the questionnaire: Construct validity was evaluated by examining the convergent and discriminant validity of

each scale, as well as standard factor loadings for each item [76]. The internal consistency (reliability) of the loading factors from the CFAs was then calculated using Cronbach's alpha, in order to examine the stability of the structure scale.

Finally, the structure suitability was evaluated by a set of adjustment indices (summarized in Table 4**.**) as follows:

a. The Normed Chi-Square (x2/df) [CMIN] represents the ratio resulting from the division of the chi-square (x2) by the degree of freedom (df). This chi-square index represents a fit index that indicates when the adjustment value is not significant ($p > 0.05$). According to Byrne, this ratio should not exceed 3 before it cannot be accepted [77].
b. The Comparative Fit Index [CFI], the Goodness Fit Index [GFI], and the Trucker-Lewis Index [TLI] also called the non-normed fit index [NNFI] which produce scores ranging from 0 to 1. According to Bentler and Bonnet, scores above 0.90 indicate a good fit (i.e. an adequate structure) [78].
c. The Standardized Root Mean Square Residual [SRMR] defined as the standardized difference between the observed correlation and the predicted correlation. A value less than .08 is generally considered a good fit [79].
d. The Root Mean Square Error of Approximation [RMSEA], wherein lower values indicate an acceptable adaptation. Hu and Bentler suggested a cutoff point of .06 to demonstrate an acceptable adjustment [79].

5 Results and Discussion

The descriptive statistics (average and standard deviation) of the 169 complete responses to the questionnaire were calculated (cf. Additional Materials – Material 3). Factor analyses have often been reported as large sample techniques. In the present study, the ratio "participants/item" was above (5:1), i.e. 5 participants for each item, which allows performing exploratory and confirmatory factor analyses [80]. All the analyses were performed with R 3.6.0.

5.1 Exploratory Factor Analysis - EFA

An EFA was performed on the dataset using principal axis factoring (PAF) to clarify the structure of the questionnaire. A parallel analysis [81] using MinRes (minimum residual) suggested that the suitable number of factors to be extracted should be seven. This suggestion fitted the number of theoretical scales proposed in section 3. Then, a PAF was performed with Direct Oblimin (oblique) rotation and the fixed number of seven factors. The findings derived from the PAF are reported in Table 5**.** . The items that loaded lower than 0.4 on all factors after the rotation were removed, as loading of 0.4 or greater are conventionally considered acceptable [82]. A Bartlett Sphericity test was statistically significant ($p < .000$), and the overall Kaiser-Meyer-Olkin (KMO) value obtained was 0.76, which confirmed the sampling adequacy of the data for performing factor analysis.

Although all items were generated based on theoretical Presence background, five items: FID3, ATT4, EMB1, EMB2, and NEG3, failed to load significantly (> 0.4) on

any factor, and two items: SP1 and INV4, had the same loading on two factors (see Table 5.). Consequently, they were removed from the questionnaire in order to achieve a simple structure [83]. In addition, one item (INV2) that was expected to load on the "Involvement" scale, loaded instead on the "Negative Effects" scale. As this item was referring to a negative aspect of the involvement ("I paid attention to inconsistencies in the environment"), it was accepted as assessing negative aspects of the environment. The "Involvement" scale was then redefined as the "Enjoyment" scale because the remaining items of this scale (INV1, INV3, and INV5) were mainly referring to the user enjoyment and satisfaction (e.g. INV5: "I had fun during the experiment").

Furthermore, the items that were initially proposed as "Fidelity" items, loaded instead on three different scales: FID1 on the "Spatial" scale, FID 2, FID3, FID5, and

Table 5. Exploratory Factor Analysis Results. Acceptable values are highlighted in bold.

	Item Code	F1	F2	F3	F4	F5	F6	F7
1	SP1	0,485		0,491				
2	SP2	**0,492**						
3	SP3	**0,613**						
4	SP4	**0,665**						
5	FID1	**0,508**				0,374		
6	FID2		**0,569**					
7	FID3	0,306	0,311					0,336
8	FID5		**0,53**					
9	FID6		**0,408**					
10	AFF2		**0,489**		0,36			
11	AFF3		**0,488**					
12	ATT1			**0,622**				
13	ATT2			**0,551**				
14	ATT3			**0,634**				
15	FID4				**0,786**			
16	AFF1				**0,426**			

17	AFF4				**0,404**			
18	INV1					**0,471**		
19	INV3					**0,579**		
20	INV4				0,407	0,422		
21	INV5					**0,804**		
22	EMB3						**0,746**	
23	EMB4						**0,77**	
24	NEG1							**0,712**
25	NEG2							**0,405**
26	INV2			0,369				**0,499**
27	NEG3				-0,351			
28	ACT4	-0,351						
29	EMB1							
30	EMB2							
	Eigenvalues	2.952	2.781	1.8	1.527	1.422	1.065	0.831
	% of variance	10,50%	9,90%	6,40%	5,50%	5,10%	3,80%	3,00%
	Total explained variance	44,20%						

Extraction Method: Principal Axis Factoring (PAF). Rotation Method: Direct Oblimin. Values lower than 0.3 were omitted. SP: Spatial Presence, FID: Fidelity, AFF: Affordance, INV: Involvement, ATT: Attention, EMB: Social Embodiment, NEG: Negative Effects.

FID6 on the "Affordance" scale, and FID4 on a new scale "Reality" (defined below in the next paragraph). These results can be explained by the possible misunderstanding of people between spatial presence and sensorial fidelity items closely related to immersion [10], and between affordance items and behavioral fidelity items. This would explain why a "Fidelity" scale failed to appear in the exploratory analysis.

However, the more interesting finding was the emergence of a scale "Reality" characterized by items that described the extent to which users have the sensation that the environment is real and that their actions have real consequences. This scale has similarities with the "Experienced Realism" proposed in the IPQ [23] and defined as the sense of reality that users could attribute to an environment.

In total, 23 items remained in the questionnaire. The seven-factor structure explained 44.20% of the total variance and was re-defined based on the items loaded as follows: “Spatial” scale defined by four items, counted for 10.50% of the variance, “Affordance” scale (five items, 9.90%), “Enjoyment” scale (three items, 6.40%), “Reality” scale (three items, 5.50%), “Attention” scale (three items, 5.10%), “Social Embodiment” scale (two items, 3.80%), and “Negative” (three items, 3.00%).

5.2 Confirmatory Factor Analysis - CFA

After modifying the scale construct of the SP-IE according to the results of the EFA, a confirmatory factor analysis (CFA) was performed to assess the construct validity, the internal consistency, and the fitness of the revised version as follows:

Table 6. Summative results of second Confirmatory Factor Analysis (SFL for each item, and CR, AVE, Cronbach’s alpha, and Pearson correlation for each scale). Acceptable values are highlighted in bold.

Scales	Item	SFL	CR	AVE	Cronbach’s alpha	Pearson correlation
Spatial Presence	SP1	**0.61**	**0.76**	**0.53**	**0.75**	
	SP2	**0.7**				
	SP3	**0.69**				
	SP4	**0.65**				
Affordance	AFF2	**0.60**	**0.68**	0.38	**0.67**	
	AFF3	**0.57**				
	AFF4	**0.61**				
	AFF5	**0.57**				
Enjoyment	ENJ1	**0.51**	**0.63**	0.37	**0.63**	
	ENJ2	**0.68**				
	ENJ3	**0.61**				
Reality	REAL1	**0.63**	**0.67**	0.49	**0.67**	
	REAL2	**0.73**				
	REAL3	**0.53**				

Attention Allocation	ATT2	**0.56**	**0.63**	**0.51**		**0.44 (p=0)**
	ATT3	**0.78**				
Social Embodiment	EMB1	**0.80**	**0.74**	**0.60**		**0.59 (p=0)**
	EMB2	**0.74**				
Negative Effects	NEG2	**0.66**	0.53	0.35		**0.37 (p=0)**
	NEG3	**0.53**				
TOTAL			**0.86**	0.45	**0.81**	

Construct Validity. Construct validity was evaluated by examining the standard factor loading (SFL) for each item as well as the values of Composite Reliability (CR) and the Average Variance Extracted (AVE) [76]. More precisely, CR and AVE values were employed for evaluating convergent and discriminant validity respectively. Convergent validity is usually recommended to be above .60 [84]. Discriminant validity is considered as sufficient when the value is above .50 [85]. The cut-off for factor loading of each item with its scale was set at .50 [86].

The values obtained by CFA are reported in Additional Materials – Material 4. The SFL was greater than 0.5 for all items, except for three: AFF1 (SFL = 0.43), ATT1 (SFL = 0.45) that were borderlines and NEG1 (SFL = 0.22) that was very low. The CR values of the scales were all above 0.6 indicating a good convergent validity, except for the "Negative" scale (CR = 0.48) which can be explained by the low SFL of NEG1. Conversely, except for the "Embodiment" scale (AVE = 0.60), the AVE values did not exceed the value of 0.5 in indicating unsatisfactory discriminant validity.

Overall, the results showed insufficient construct validity. Consequently, the three items with unacceptable SFL values were removed from the questionnaire resulting in 20 items. A second CFA was then performed on the new structure. The results are shown in Table 6**.** . This second analysis indicated a more satisfactory construct validity: all items showed an acceptable SFL [>0.5]. Concerning the convergent validity, all the CR values were above the threshold [>0.6], except for the CR value of "Negative" scale which increased but remained borderline. Finally, the discriminant validity showed better results with the increase of AVE values for all scales. In particular, the "Spatial Presence", "Attention", and "Embodiment" scales were all above the threshold [>0.5] and the "Reality" scale showed a borderline AVE value. However, the AVE values of the "Affordance", "Enjoyment", and "Negative" scales remained low.

Internal Consistency (Reliability). Internal reliability was examined with Cronbach's alpha values computed for each scales. The results obtained are shown in Table 6**.** . Alpha's values for the "Spatial Presence", "Affordance", "Enjoyment", and "Reality" scales ranged from 0.63 to 0.75, which are acceptable values [75]. The alpha value of the questionnaire was above 0.80 indicating overall good reliability.

Concerning "Attention", "Embodiment", and "Negative" scales, their low number of items (two) made it impossible to correctly calculate their Cronbach's alpha values.

Indeed, Cronbach's alpha is based on several quite restrictive assumptions, i.e., unidimensionality, uncorrelated errors, and essentially tau-equivalence. At least three items are necessary to test these assumptions [87]. Therefore, their internal consistencies were reported instead using Pearson correlation tests with a cutoff at 0.3 [88]. The results showed satisfactory correlations ranging from 0.37 to 0.59. Based on these results, no item was removed because minimal gains would be obtained.

Fitness of the internal structure. The fit statistics for the model are presented in Table 7. . The structure of the SP-IE after item correction had an acceptable model fit, since all recommended fit indices satisfied the cut-off values, except TLI which was slightly below the cut-off value. The sample size of the present study appears to be sufficient for CFA-based analyses [89]. In addition, among the diverse goodness-of-fit indices that were employed in the present study, RMSEA, which is less sensitive to sample size [90], indicated a good fit between the model and the data.

To summarize, the structural statistical analysis supported the internal structure of the final version of the SP-IE proposed. This process, as described, yielded the final, well-defined questionnaire, composed of 20 five-point Likert items.

Table 7. Goodness-of-fit scores after the CFA evaluation (acceptable values are in bold).

Fit index	x2	df	CMIN	CFI	GFI	TLI	SRMR	RMSEA
SP-IE	200.01	149	**1.34 (p < .03)**	**0.90**	**0.95**	0.87	**0.068**	**0.045 [0.027;0.061] (p < .05)**

6 Conclusion

The present study aimed at developing the SP-IE [Spatial Presence in Immersive environments] questionnaire, an instrument for measuring Spatial Presence and its underlying factors in high immersive environments. The questionnaire was developed in the French language for use within the French-Speaking population.

To achieve this goal, the study adopted a multi-stage process to questionnaire construction and validation. The construction stage consisted of determining the different scales of the questionnaire and generating corresponding items for each scale. This stage was based on empirical presence studies and previous most used questionnaires (namely WS [10, 45], ITC-SOPI [13], IPQ [12, 21, 23], and MEC-SPQ [50] questionnaires). Founding the construction scale and item generation on theoretical presence backgrounds allowed to preserve the content validity of the SP-IE questionnaire. In addition, an item-reduction procedure was performed in order to shorten the questionnaire and reach a satisfactory internal consistency. The dataset for this procedure was collected from an investigation in three different controlled environments.

In the validation stage, the construct validity and the fitness of the SP-IE structure were examined. Data collected from a large sample size investigation was processed with EFA to explore the hypothetical structure of the questionnaire and later confirmed by CFA tests. Item correction based on the factor analyses ended with a seven-scale

questionnaire with 20 items. The results supported the final structure scale with good internal consistency and satisfactory convergent validity. However, discriminant validity was shown to be insufficient. In addition, the structure had an acceptable model fit with indices above their respective cut-off values (CMIN, CFI, GFI, SRMR, and RMSEA), expect for Tucker-Lewis Index (TLI), which was slightly below the cut-off value.

This process yielded a well-structured questionnaire that supports the multidimensionality and hierarchical structure of Spatial Presence and indicates that it is related to different factors, namely: the affordance of the environment, the user's enjoyment, the attention allocation on the activity, the sense of reality and awareness of real consequences, the social embodiment, and the cybersickness.

However, even though the factor structure proposed in this paper was confirmed, the low discriminant validity obtained encourages further attention. Thus, another invariance study with a large sample size in different environments is recommended as a follow-up to the present study in order to examine the psychometric properties of the questionnaire. Furthermore, attempts should be made to increase the number of items per scale regarding the low number of some scales of the questionnaire.

In addition, the SP-IE questionnaire is designed for assignment after experiment exposure: the participants complete the questionnaire at the end of their experience, so as not to cause breaks that reduce their sense of presence [66]. Consequently, it does not provide a continuous measurement of presence during the experiment. This limitation is common with all post-intervention questionnaires. To solve this, it is suggested to rely on a multi-measurements approach combining questionnaires and objective non-invasive metrics for assessing spatial presence. The SP-IE questionnaire being a reliable and valid measure of spatial presence, its scores should be associated in a predictable manner with other variables or constructs that in theory are related to spatial presence. Thus, future studies should investigate the relationship between the SP-IE questionnaire and other reliable measurements of presence, such as behavioral observations. Such mixed-method studies will be critical in providing deeper and more reliable insights of the validity of the questionnaire.

To conclude, the present study contributed to the literature by (a) offering a valid questionnaire to assess Spatial Presence in immersive environments for French-speaking community, and (b) verifying the existence of a multi-level, hierarchical nature of Spatial Presence with emphasis on factors neglected in other questionnaires, namely the affordance of the environment, the sense of reality and awareness of consequences, and the social embodiment using avatars.

This questionnaire will aim to compare the sense of Spatial Presence between different highly immersive environments. By providing a theoretically driven validated assessment of Spatial Presence and its underlying factors, the questionnaire will support presence community researchers and designers of such environments.

Acknowledgments

Special thanks are due to Drone Arena (https://www.dronearena.com/) and Illucity La Villette (https://illucity.fr/en/) who accepted to allow us to administer the questionnaire to their participants.

References

1. Sheridan, T. B. (1992). Musings on telepresence and virtual presence. *Presence: Teleoperators & Virtual Environments*, *1*(1), 120-126.
2. Taylor, R. H., Menciassi, A., Fichtinger, G., Fiorini, P., & Dario, P. (2016). Medical robotics and computer-integrated surgery. In *Springer handbook of robotics* (pp. 1657-1684). Springer, Cham.
3. Anderson, T., Liam, R., Garrison, D. R., & Archer, W. (2001). Assessing teaching presence in a computer conferencing context.
4. Tamborini, R., & Skalski, P. (2006). The role of presence in the experience of electronic games. In P. Vorderer & J. Bryant (Eds.), Playing video games: Motives, responses, and consequences (pp. 225–240). Mahwah, NJ: Lawrence Erlbaum Associates.
5. Juan, M. C., Baños, R., Botella, C., Pérez, D., Alcaníiz, M., & Monserrat, C. (2006). An augmented reality system for the treatment of acrophobia: the sense of presence using immersive photography. *Presence: Teleoperators and virtual environments*, *15*(4), 393-402.
6. Brade, J., Lorenz, M., Busch, M., Hammer, N., Tscheligi, M., & Klimant, P. (2017). Being there again–presence in real and virtual environments and its relation to usability and user experience using a mobile navigation task. *International Journal of Human-Computer Studies*, *101*, 76-87.
7. Mania, K. (2001, November). Connections between lighting impressions and presence in real and virtual environments: an experimental study. In *Proceedings of the 1st international conference on Computer graphics, virtual reality and visualisation* (pp. 119-123). ACM.
8. Meehan, M., Insko, B., Whitton, M., & Brooks Jr, F. P. (2002, July). Physiological measures of presence in stressful virtual environments. In *Acm transactions on graphics (tog)* (Vol. 21, No. 3, pp. 645-652). ACM.
9. Insko, B. E. (2003). Measuring presence: Subjective, behavioral and physiological methods.
10. Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence*, *7*(3), 225-240.
11. Usoh, M., Catena, E., Arman, S., & Slater, M. (2000). Using presence questionnaires in reality. *Presence: Teleoperators & Virtual Environments*, *9*(5), 497-503.
12. Schubert, T. W. (2003). The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realness. *Zeitschrift für Medienpsychologie*, *15*(2), 69-71.
13. Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J. (2001). A cross-media presence questionnaire: The ITC-Sense of Presence Inventory. *Presence: Teleoperators & Virtual Environments*, *10*(3), 282-297.
14. Kim, T., & Biocca, F. (1997). Telepresence via television: Two dimensions of telepresence may have different connections to memory and persuasion. *Journal of computer-mediated communication*, *3*(2), JCMC325.

15. Bystrom, K. E., Barfield, W., & Hendrix, C. (1999). A conceptual model of the sense of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, *8*(2), 241-244.
16. Schubert, T. W., Friedmann, F., & Regenbrecht, H. T. (1999b, April). Decomposing the sense of presence: Factor analytic insights. In *2nd international workshop on presence* (Vol. 1999).
17. Sanchez-Vives, M. V., & Slater, M. (2005). From presence to consciousness through virtual reality. *Nature Reviews Neuroscience*, *6*(4), 332.
18. Schultze, U. (2010). Embodiment and presence in virtual worlds: a review. *Journal of Information Technology*, *25*(4), 434-449.
19. Wirth, W., Hartmann, T., Böcking, S., Vorderer, P., Klimmt, C., Schramm, H., ... & Biocca, F. (2007). A process model of the formation of spatial presence experiences. *Media psychology*, *9*(3), 493-525.
20. Bowman, D. A., & McMahan, R. P. (2007). Virtual reality: how much immersion is enough?. *Computer*, *40*(7), 36-43.
21. Regenbrecht, H., & Schubert, T. (2002). Real and illusory interactions enhance presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, *11*(4), 425-434.
22. Lok, B., Naik, S., Whitton, M. and Brooks Jr., F.P. (2003). Effects of Handling Real Objects and Self-Avatar Fidelity on Cognitive Task Performance and Sense of Presence in Virtual Environments, Presence 12(6): 615–628.
23. Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence: Teleoperators & Virtual Environments*, *10*(3), 266-281.
24. Riva, G., Waterworth, J. A., Waterworth, E. L., & Mantovani, F. (2011). From intention to action: The role of presence. *New Ideas in Psychology*, *29*(1), 24-37.
25. Slater, M. (2003). A note on presence terminology. *Presence connect*, *3*(3), 1-5.
26. Blascovich, J. (2002). Social influence within immersive virtual environments. In *The social life of avatars* (pp. 127-145). Springer, London.
27. Meehan, M., Insko, B., Whitton, M., & Brooks Jr, F. P. (2002, July). Physiological measures of presence in stressful virtual environments. In *Acm transactions on graphics (tog)* (Vol. 21, No. 3, pp. 645-652). ACM.
28. Wiederhold, B. K., Gevirtz, R., & Wiederhold, M. D. (1998). Fear of flying: A case report using virtual reality therapy with physiological monitoring. *CyberPsychology & Behavior*, *1*(2), 97-103.
29. Wiederhold, B. K., Jang, D. P., Kaneda, M., Cabral, I., Lurie, Y., May, T., ... & Kim, S. I. (2001). An investigation into physiological responses in virtual environments: an objective measurement of presence. *Towards cyberpsychology: Mind, cognitions and society in the internet age*, *2*.
30. Meehan, M., Razzaque, S., Insko, B., Whitton, M., & Brooks, F. P. (2005). Review of four studies on the use of physiological reaction as a measure of presence in stressfulvirtual environments. *Applied psychophysiology and biofeedback*, *30*(3), 239-258.

31. Baumgartner, T., Valko, L., Esslen, M., & Jäncke, L. (2006). Neural correlate of spatial presence in an arousing and noninteractive virtual reality: an EEG and psychophysiology study. *CyberPsychology & Behavior*, *9*(1), 30-45.
32. Brogni, A., Slater, M., & Steed, A. (2003, October). More breaks less presence. In *Presence 2003: The 6th Annual International Workshop on Presence* (pp. 1-4).
33. Nalivaiko, E., Davis, S. L., Blackmore, K. L., Vakulin, A., & Nesbitt, K. V. (2015). Cybersickness provoked by head-mounted display affects cutaneous vascular tone, heart rate and reaction time. *Physiology & behavior*, *151*, 583-590.
34. Schuemie, M. J., Van Der Straaten, P., Krijn, M., & Van Der Mast, C. A. (2001). Research on presence in virtual reality: A survey. *CyberPsychology & Behavior*, *4*(2), 183-201.
35. Sheridan, T. B. (1996). Further musings on the psychophysics of presence. Presence: Teleoperators & Virtual Environments, 5(2), 241-246.
36. Usoh, M., Arthur, K., Whitton, M. C., Bastos, R., Steed, A., Slater, M., & Brooks Jr, F. P. (1999, July). Walking> walking-in-place> flying, in virtual environments. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques* (pp. 359-364). ACM Press/Addison-Wesley Publishing Co.
37. Lepecq, J. C., Bringoux, L., Pergandi, J. M., Coyle, T., & Mestre, D. (2009). Afforded actions as a behavioral assessment of physical presence in virtual environments. *Virtual reality*, *13*(3), 141-151.
38. Lombard, M., Biocca, F., Freeman, J., IJsselsteijn, W., & Schaevitz, R. J. (Eds.). (2015). *Immersed in media: Telepresence theory, measurement & technology*. Springer.
39. Freeman, J., Lessiter, J., Pugh, K., & Keogh, E. (2005). When presence and emotion are related, and when they are not. In *8th Annual International Workshop on Presence, September* (pp. 21-23).
40. Freeman, J., Avons, S. E., Meddis, R., Pearson, D. E., & IJsselsteijn, W. (2000). Using behavioral realism to estimate presence: A study of the utility of postural responses to motion stimuli. *Presence: Teleoperators & Virtual Environments*, *9*(2), 149-164.
41. Bracken, C. C., Pettey, G., & Wu, M. (2014). Revisiting the use of secondary task reaction time measures in telepresence research: exploring the role of immersion and attention. *AI & society*, *29*(4), 533-538.
42. Barfield, W., & Weghorst, S. (1993). The sense of presence within virtual environments: A conceptual framework. *Advances in Human Factors Ergonomics*, *19*, 699-699.
43. Slater, M., Usoh, M., & Chrysanthou, Y. (1995). The influence of dynamic shadows on presence in immersive virtual environments. In *Virtual Environments' 95* (pp. 8-21). Springer, Vienna.
44. Biocca, F., & Delaney, B. (1995). Immersive virtual reality technology. *Communication in the age of virtual reality*, *15*, 32.
45. Witmer, B. G., Jerome, C. J., & Singer, M. J. (2005). The factor structure of the presence questionnaire. *Presence: Teleoperators & Virtual Environments*, *14*(3), 298-312.

46. Slater, M. (1999). Measuring presence: A response to the Witmer and Singer presence questionnaire. *Presence*, *8*(5), 560-565.
47. UQO Cyberpsychology Lab (2004). Revised WS Questionnaire.
48. Lombard, M., Ditton, T. B., & Weinstein, L. (2009, October). Measuring presence: the temple presence inventory. In *Proceedings of the 12th Annual International Workshop on Presence* (pp. 1-15).
49. Viaud-Delmo, I. (n.d.). Igroup presence questionnaire (IPQ) item download. Retrieved from http://www.igroup.org/pq/ipq/IPQinstructionsFr.doc
50. Vorderer, P., Wirth, W., Gouveia, F. R., Biocca, F., Saari, T., Jäncke, L., ... & Klimmt, C. (2004). MEC Spatial Presence Questionnaire. *Retrieved Sept*, *18*, 2015.
51. Vasconcelos-Raposo, J., Bessa, M., Melo, M., Barbosa, L., Rodrigues, R., Teixeira, C. M., ... & Sousa, A. A. (2016). Adaptation and validation of the Igroup Presence Questionnaire (IPQ) in a Portuguese sample. *Presence: Teleoperators and virtual environments*, *25*(3), 191-203.
52. Gibson, J. J. (2014). The ecological approach to visual perception: classic edition. Psychology Press.
53. Schubert, T., Friedmann, F., & Regenbrecht, H. (1999). Embodied presence in virtual environments. In *Visual representations and interpretations* (pp. 269-278). Springer, London.
54. Biocca, F. (1997). The cyborg's dilemma: Progressive embodiment in virtual environments. *Journal of computer-mediated communication*, *3*(2), JCMC324.
55. Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of computer-mediated communication*, *3*(2), JCMC321.
56. Taylor, T. L. (2002). Living digitally: Embodiment in virtual worlds. In *The social life of avatars* (pp. 40-62). Springer, London.
57. Meehan, M., Razzaque, S., Whitton, M. C., & Brooks, F. P. (2003, March). Effect of latency on presence in stressful virtual environments. In *IEEE Virtual Reality, 2003. Proceedings.* (pp. 141-148). IEEE.
58. Rebenitsch, L., & Owen, C. (2016). Review on cybersickness in applications and visual displays. *Virtual Reality*, *20*(2), 101-125.
59. Welch, R. B., Blackmon, T. T., Liu, A., Mellers, B. A., & Stark, L. W. (1996). The effects of pictorial realism, delay of visual feedback, and observer interactivity on the subjective sense of presence. *Presence: Teleoperators & Virtual Environments*, *5*(3), 263-273.
60. Ling, Y., Nefs, H. T., Brinkman, W. P., Qu, C., & Heynderickx, I. (2013). The relationship between individual characteristics and experienced presence. *Computers in Human Behavior*, *29*(4), 1519-1530.
61. Slater, M., Sadagic, A., Usoh, M., & Schroeder, R. (2000). Small-group behavior in a virtual and real environment: A comparative study. *Presence: Teleoperators & Virtual Environments*, *9*(1), 37-51.
62. Thie, S., & Van Wijk, J. (1998). A general theory on presence. *1st Int. Wkshp. on Presence*.
63. Heeter, C. (1992). Being there: The subjective experience of presence. *Presence: Teleoperators & Virtual Environments*, *1*(2), 262-271.
64. Lee, K. M. (2004). Presence, explicated. *Communication theory*, *14*(1), 27-50.

65. Nowak, K. L., & Biocca, F. (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*(5), 481-494.
66. Schultze, U., & Leahy, M. M. (2009). The avatar-self relationship: Enacting presence in second life. *ICIS 2009 Proceedings*, 12.
67. Wang, X., Laffey, J., Xing, W., Ma, Y., & Stichter, J. (2016). Exploring embodied social presence of youth with Autism in 3D collaborative virtual learning environment: A case study. Computers in Human Behavior, 55, 310-321.
68. Jin, S. A. A. (2009). Avatars mirroring the actual self versus projecting the ideal self: The effects of self-priming on interactivity and immersion in an exergame, Wii Fit. *CyberPsychology & Behavior*, *12*(6), 761-765.
69. Dean, E., Murphy, J., & Cook, S. (2009). Social presence in virtual world surveys. In *Proceedings of The 12th Annual International Workshop on Presence*.
70. Bertram, D. (2007). Likert scales. *Retrieved November*, *2*, 2013.
71. Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of cross-cultural psychology*, *1*(3), 185-216.
72. Hambleton, R. K., & Zenisky, A. L. (2011). Translating and adapting tests for cross-cultural assessments.
73. Khenak N., Vezien J.M., Théry D., Bourdot P (2019). Spatial Presence in Real and Remote Immersive Environments. *In proceedings of the 26th IEEE Conference on Virtual Reality and 3D User Interface*: file:///H:/web/program/datas/Conference%20_%20Posters/1251-doc.pdf
74. Rammstedt, B. & Beierlein, C. (2014). Can't we make it any shorter? The limits of personality assessment and ways to overcome them. Journal of Individual Differences, 35(4), 212-220. http://dx.doi.org/10.1027/1614-0001/a000141
75. Stormer, F., Kline, T., & Goldenberg, S. (1999). Measuring entrepreneurship with the general enterprising tendency (GET) test: criterion-related validity and reliability. *Human Systems Management*, *18*(1), 47-52.
76. Raubenheimer, J. (2004). An item selection procedure to maximize scale reliability and validity. *SA Journal of Industrial Psychology*, *30*(4), 59-64.
77. Byrne, B. M., & Stewart, S. M. (2006). Teacher's corner: The MACS approach to testing for multigroup invariance of a second-order structure: A walk through the process. *Structural Equation Modeling*, *13*(2), 287-321.
78. Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological bulletin*, *88*(3), 588.
79. Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, *6*(1), 1-55.
80. Ding, L., Velicer, W. F., & Harlow, L. L. (1995). Effects of estimation methods, number of indicators per factor, and improper solutions on structural equation modeling fit indices. *Structural Equation Modeling: A Multidisciplinary Journal*, *2*(2), 119-143.
81. Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179-185.
82. Field, A. (2009). Discovering statistics using SPSS. London: SAGE.

83. Thurstone, L. L. (1947). Multiple-factor analysis; a development and expansion of The Vectors of Mind.
84. Carlson & Herdman, 2012. Retrieved Feb 3, 2016 from: www.management.pamplin.vt.edu/directory/Articles/Carlson1.pdf.
85. Fornell, C. & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research, 18, 39–50.
86. Hair, J. F. (2006). *Multivariate data analysis*. Pearson Education India.
87. McDonald R. P. (1999). Test theory: A unified approach. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
88. Rammstedt, B. & Beierlein, C. (2014). Can't we make it any shorter? The limits of personality assessment and ways to overcome them. Journal of Individual Differences, 35(4), 212-220. http://dx.doi.org/10.1027/1614-0001/a000141
89. Gagne, P., & Hancock, G. R. (2006). Measurement model quality, sample size, and solution propriety in confirmatory factor models. *Multivariate Behavioral Research*, *41*(1), 65-83.
90. Brown, T. A., & Moore, M. T. (2012). Confirmatory factor analysis. *Handbook of structural equation modeling*, 361-379.

Appendix

Appendix 1. "Real vs. Remote" experiment: General setting of participants (top) with their corresponding First Person View (bottom). (Left) The operating room. (Right) The tele-operating room.

Thi ere
def *EuroVR International Conference (EuroVR 2019)*, Oct 2019, Tallinn, Estonia. pp.201-225, https://doi.org/10.1007/978-3-030-31908-3_13