Jacinto Vasconcelos-Raposo*
INESC TEC and UTAD—Education & Psychology Dept.

Maximino Bessa
INESC TEC and UTAD

Miguel Melo
INESC TEC

Luis Barbosa
INESC TEC and FEUP

Rui Rodrigues
UTAD—Education & Psychology Dept.

Carla Maria Teixeira
CITCEM and UTAD

Luciana Cabral
INESC TEC and FEUP

António Augusto Sousa
INESC TEC and FEUP

Adaptation and Validation of the Igroup Presence Questionnaire (IPQ) in a Portuguese Sample

Abstract

The present study aims (a) to translate and adapt the Igroup Presence Questionnaire (IPQ) to the Portuguese context (semantic equivalence/ conceptual and content validity) and (b) to examine its psychometric properties (reliability and factorial validity). The sample consisted of 478 subjects (285 males and 193 females). The fidelity of the factors varied between 0.53 and 0.83. The confirmatory factor analysis results produced a 14-item version of IPQ-PT, accepting covariance between residual errors of some items of the instrument, as the best structural representation of the data analyzed. The CFA was conducted based on a three-variable model. The fit indexes obtained were $X^2/df = 2.647$, GFI = .948, CFI = .941, RSMEA = .059, and AIC = 254. These values demonstrate that the proposed Portuguese translation of the IPQ maintains its original validity, demonstrating it to be a robust questionnaire to measure the sense of presence in virtual reality studies. It is therefore recommended for use in Presence research when using Portuguese samples.

1 Introduction

Virtual reality is a fast-growing technology with multiple applications, such as therapeutic interventions (Anderson, et al., 2013; DeAngelis, 2012), military and medical training (Seymour, et al. 2002), entertainment, simulation (Gallagher, et al., 2005), rehabilitation (Saposnik & Levin, 2011), and sports (Bideau et al., 2010). One of the advantages of this technology is that it allows individuals to experience emotions in a controlled fashion (DeAngelis, 2012), and it facilitates the acquisition and training of a great variety of skills (Psotka, 1995; Bliss, Tidwell & Guest, 1997; Ahlberg et al., 2007; Seidel & Chatelier, 2013).

In the context of mental health, for example, psychotherapists have been using virtual reality to treat phobias, post-traumatic stress disorders, and social anxiety (Gerardi, Cukor, Difede, Rizzo, & Rothbaum, 2010), as well as autism spectrum disorders (Stendal & Balandin, 2015). In medical training, this technology has been applied to develop, for example, surgical skills (Grantcharov et al., 2004). In aviation, it has been used to train pilots and aircrews to respond adequately to certain situations (Taylor, Lintern, Hulin, Talleur, & Emanuel, 1997; O'Neil Jr., Andrews, & O'Neil, 2000). Its greatest visibility and fastest growing application, for the general public, is in the entertainment industry. Nevertheless, its expansion, at this point in time, is open to human creativity.

Presence, *Vol. 25, No. 3, Summer 2016, 1–13*
doi:10.1162/PRES_a_00261

*Correspondence to jvraposo@utad.pt.

The concept of presence seems to unify different perspectives among proponents in the search for evidence for the effectiveness of the available multiple applications of virtual reality. Presence is a construct with many levels and dimensions (Biocca & Delaney, 1995), and several questionnaires have been proposed to measure it. To quantify this construct, many concepts have been proposed, including spatial presence, realness, involvement, self-location, cognitive involvement, naturalness, social-actor within medium, passive social presence, active social presence, engagement social richness, social realism, perceptual realism, and others.

For the purpose of measuring presence, according to Rosakranse and Oh (2014), several instruments have been presented in the specialized literature, and among them, some have gained preponderance, namely, the Slater-Usoh-Steed Presence Questionnaire (SUS) (Slater, Usoh, & Steed, 1994), Presence Questionnaire (PQ) (Witmer & Singer, 1998), ITC-Sense of Presence Inventory (ITC-SOPI) (Lessiter, Freeman, Keogh, & Davidoff, 2001), Igroup Presence Questionnaire (IPQ) (Schubert, Friedmann, & Regenbrecht, 2001), and Temple Presence Inventory (Lombard, Ditton, & Weinstein, 2009).

According to Rosakranse and Oh (2014), the SUS was the first questionnaire that attempted to measure presence in virtual reality. The questionnaire underwent changes throughout the years; currently, it consists of six items. SUS is a popular instrument; however, it measures only one dimension of presence: “presence as transportation” (Lombard, Ditton, & Weinstein, 2009).

A second questionnaire, also widely used, is the Presence Questionnaire (PQ), which was developed by Witmer and Singer (1998). It aims to measure involvement and immersion, and it also underwent several changes throughout time, namely in semantic aspects.

A third instrument widely used is the ITC-Sense of Presence Inventory presented by Lessiter et al. (2001). One of the advantages of this instrument is its applicability to several types of interfaces. However, its use is somewhat limited due to the restrictions imposed by its proprietors. Lombard et al. (2009) refined the ITC-SOPI and introduced the Temple Presence Inventory, which aimed to measure eight subscales: spatial presence, social presence-actor within medium, passive social presence, active social presence, presence as engagement, presence as social richness, presence as social realism, and presence as perceptual realism.

The IPQ, created by Schubert et al. (2001), measures mainly the same scales of the SUS and the PQ questionnaires, and in its final format, it allows the quantification of three subscales: spatial-presence, realness, and involvement. This questionnaire seems to be the only one of the described instruments that was subjected to a Confirmatory Factor Analysis (CFA).

The IPQ questionnaire is considered one of the canonical presence questionnaires and has been used in many research studies, including ones on the effects of mediation in a storytelling virtual environment (VE) (Brown, Ladeira, Winterbottom, & Blake, 2003), the effectiveness of VE in the treatment of acrophobia (Krijn, et al., 2004) and social phobia (Price, Mehta, Tone, & Anderson, 2012), the impact of embodied interaction on mixed-reality spaces (Betella, Carvalho, Sanchez-Palencia, Bernader, & Verschure, 2012), the evaluation of diagnostic and therapeutic tools for cynophobia (Suied, Dretakkis, Warusfel, & Viaud-Delmon, 2013), and even to study whether the feeling of presence has an impact on the induced experience of pain (Czub & Piskorz, 2014). However, most of the results are questionable due to the lack of information on data quality and the version of the questionnaire used and its psychometric properties. These aspects contribute to the poor internal validity of the studies reported. For example, Alsina-Jurnet and Gutiérrez-Maldonado (2010) used the IPQ in a study with a Spanish sample but never provided information regarding which version of the questionnaire they used.

Robust scientific methodology requires proper translation and consequent validation of the developed instruments in other languages. The IPQ has been translated into French (Viaud-Delmon, n.d.), German (Schubert et al., 2001), Japanese (Hyun, et al., 2010), English (Krijn et al., 2004), and Persian (Panahi-Shahri, Fathi-Ashtiani, & P. Azad-Fallah., 2009). However, only the German (original version proposed by the authors) and Persian versions were properly validated.

The reliability presented by the German version ranged from 0.64 to 0.80, and the confirmatory analysis demonstrated an adequate model; these values were calculated using the online database made available by the authors. In the original manuscript, the authors reported the following adjustment index values: $X^2(62, N = 296) = 89.840$, $p = 0.012$, RMSEA $= 0.039$. Based on the scores presented, we are able to determine that the recommended ratio X^2/df for the German sample is 1.464, which is clearly below the value of 2, which is presented in the literature as the ideal value to be attained.

The popularity of the IPQ and the fact that among all of the available instruments, it is the only one that was subject to a confirmatory factor analysis justifies the purpose of the present study to translate and validate the IPQ to be used in nomothetic research.

Because there are no properly validated instruments to measure Presence within the Portuguese population, we intend to translate and validate the IPQ questionnaire in the Portuguese language, establishing semantic equivalence for each item. Using both the Exploratory Factor Analysis (EFA) and the confirmatory factor analysis (CFA), we aim to establish its psychometric properties and consequently recommend the IPQ to be used in Portuguese research projects. Based on databases made available by international researchers, we will perform independent CFAs using their data and consequently externally validate the questionnaire.

2 Method

The sample consisted of 478 subjects (285 males and 193 females) between 17 and 56 years old ($M = 24.54$; $SD = 6.7$). Five subjects withdrew due to nausea and vision problems. Of the 478, 16.7% had some previous experience with virtual reality devices and 83.3% did not. No significant differences were found between levels of previous experience with virtual reality. All subjects were recruited in two different locations, both higher education institutions.

For confirmatory factor analysis purposes, Tabachnick, Fidell, and Osterlind (2001) suggested a ratio of 10 subjects for each item (10:1). In the present study, we had a 34:1 ratio.

2.1 Instruments

2.1.1 Questionnaire. The IPQ is a self-report questionnaire and was initially developed by Schubert et al. (2001). In its original version, it aimed to measure the following variables: spatial presence (SP1, SP2, SP3, SP4, and SP5), involvement (INV1, INV2, INV3, and INV4), realness (REAL1, REAL2, REAL3, and REAL4), and global presence (G1). All questions were presented in a five-point Likert scale format using different anchors.

2.1.2 Translation and Cultural Adaptation. The first stage consisted in the translation/back-translation (back-translation method) of the instrument as proposed by Brislin (1970) and Hambleton and Zenisky (2011). This procedure involved the collaboration of four bilingual experts and PhDs: two of them from the field of psychology, with expertise in psychometry, and the other two from the field of computer science, with expertise in virtual reality. Initially, an expert fluent in both languages translated the English instrument to Portuguese; later, another expert back-translated the instrument from Portuguese to English without consulting the original version. The results were the two versions of the instrument, one in each language. Second, experts met to evaluate and make modifications in the draft. The English version was compared with the original version, and the result showed the existence of semantics and content equivalence in most items. Several items in the Portuguese language version were revised, and corrections were made with respect to the specific terminology to adjust to the proper use of technical terms in Portuguese based on the consensus reached among experts that certified that there were no incompatibilities with the original version. This consultation also served to analyze the form and content of items in terms of clarity and comprehensibility (Almeida & Freire, 2003, Hambleton & Zenisky, 2011).

A second step was taken to assess the content validity: a committee composed of two experienced PhD researchers in psychology and two in computer science were individually asked to indicate their agreement or disagreement regarding the inclusion of items in the theoretically proposed factors. Then, researchers were asked to calculate the percentage of concordance and relevance of the items on their respective factors based on a 10-point scale (1 = not relevant/important to 10 = extremely relevant/important). This process allowed the calculation of the content validity index (CVI) (Waltz, Strickland, & Lenz, 1991). The scores obtained showed values above 80% for all items, which supports the suitability/inclusion of the items in their respective factors.

Once the required authorization from the institutional authorities (i.e., ethical committees) was obtained, 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. The completion of the questionnaire took place in a calm and serene environment, either in isolation or in small groups (never involving more than five people).

2.1.3 Virtual Environment. To evaluate the questionnaire, subjects were exposed to virtual environments, which would provide different stimuli to attempt to induce presence, and after which they would need to answer the questionnaire. As for the stimulus itself, we used the virtual reality experience "Don't let go!" (Skydome Studios, 2014), a game-like app that consists of a first-person experience where the user is behind a desk and has to hold both Ctrl keys while they are confronted with a series of events intended to induce fear and cause the user to release the Ctrl keys. This strategy aimed to ensure that participants did not release the Ctrl keys accidentally; nevertheless, they could quit the experiment at any time they wanted by knocking twice, as instructed.

For presenting the stimuli, an Asus N550JK-CN104H laptop computer equipped with an Intel Core i7 4700HQ CPU, a NVIDIA GeForce GTX850M with 4 GB DDR3 graphic card, and 16 GB of RAM with an added SSD drive was used. To simulate the interaction, the subjects had in front of them a Microsoft Wired 600 keyboard that was not attached to the laptop. The headphones used were the Bose QuietComfort 15 model, which provides very effective active acoustic noise cancelling. The HMD used was the Oculus Rift DK2. The resolution used was FHD (1920 × 1080) with an average framerate of 60 FPS.

2.1.4 Experimental Procedure. Before each experiment, all participants were informed about the procedures and briefed about what to expect and how to proceed properly when interacting with the elements presented in the virtual environment. All participants were informed that their participation involved no risk to their well-being and that they could immediately abandon their participation at any time they wanted to do so during the experiment by knocking twice on the table where they stood. To complete the experimental procedure each participant took on average 15 minutes. The virtual experiment had the duration of approximately 3:45 minutes.

A booth was placed in the experimental room to isolate participants from the surrounding environment during the experiments. Inside the booth, there was a table where the user stood, a keyboard, a pair of headphones, and the Oculus Rift. Participants were assisted with putting the equipment in place. Immediately after finishing this phase, subjects were asked to go to another room where they would fill out the questionnaire (IPQ).

2.1.5 Statistical Procedures. First, descriptive statistics such as the average, standard deviation, and minimum and maximum scores were calculated for each item and scale. To verify the asymmetry of the distribution, we used skewness and kurtosis. The results showed that all scores observed were in the range from -1 to $+1$. The internal consistency of the factors was later calculated using Cronbach's alpha.

An exploratory factor analysis was performed to identify a set of latent constructs underlying the measured variables through the IPQ. In principle, the obtained variables should coincide with the different theoretical suggestions that have been advanced in different languages (i.e., French, German, Japanese, and Dutch). The databases available were taken into consideration and consequently submitted to a confirmatory factor analysis (CFA) for comparative purposes.

The CFA (Amos 22.0) was used to test the models, using the maximum likelihood estimation (MLE)

Table 1. *Descriptive and Univariate Normality Analyses*

	M ± *SD*	Skewness	Kurtosis
G1	3.97 ± 0.70	−0.937	2.502
SP1	4.09 ± 0.71	−1.151	3.195
SP2	3.98 ± 0.98	−1.027	0.727
SP3	3.98 ± 0.92	−0.779	0.174
SP4	3.79 ± 0.86	−0.632	0.203
SP5	4.04 ± 0.71	−1.020	2.777
INV1	3.03 ± 1.21	0.005	−1.005
INV2	3.07 ± 1.23	−0.201	−1.008
INV3	3.19 ± 1.30	−0.144	−1.193
INV4	3.82 ± 0.89	−0.696	0.428
REAL1	3.37 ± 0.99	−0.316	−0.393
REAL2	2.70 ± 1.05	0.196	−0.521
REAL3	2.79 ± 1.09	0.045	−0.769
REAL4	2.14 ± 1.06	0.832	0.203
REALNESS	11.00 ± 3.4	0.168	−0.208
INVOLVEMENT	13.10 ± 3.01	−0.059	−0.392
SPATIAL PRESENCE	23.85 ± 3.00	−0.446	1.156

method with 34 subjects per item. This ratio is six times greater than what is recommended by Ding, Velicer, and Harlow (1995). After specifications and model estimation, the model suitability was evaluated by a set of adjustment/adaptation indices. The χ^2 (chi-square) score indicates when the adjustment value is not significant ($p > 0.05$). However, the results of this test per se are questionable because of its vulnerability to the sample size. Jöreskog and Sörbom (1989) suggested a correction procedure which consisted of using the ratio between chi-square and degrees of freedom (df), represented by χ^2/df. In agreement with this suggestion, Ullman (2001) recommended the use of 2.0 as the reference to determine the model's acceptability. In addition to this ratio, other indexes have been recommended, namely:

a. CFI (Comparative Fit Index) and GFI (Goodness Fit Index), which produce scores ranging from 0 to 1. According to Bentler and Bonnet (1980), scores above 0.90 represent an appropriate model. Later, Hu and Bentler (1999) suggested a cutoff of 0.95 as an indicative value of a good fit of the model;
b. RMSEA (Root Mean Square Error of Approximation), wherein lower values indicate an acceptable adaptation. In the opinion of Browne & Cudeck (1993), this score should be equal or below .08. However, Hu & Bentler (1999) suggested a cutoff point of .06.

3 Results and Discussion

The univariate normality scores obtained for each subscale (see Table 1) demonstrated a normal distribution with all scores ranging from −1.151 to 0.832 for skewness and from −1.193 to 2.777 for Kurtosis (see Marôco, 2014). The correlation score between items within the same subscale were: 1− Spatial Presence: Highest = 0.601 − lowest = −0.101; Involvement: Highest = 0.456, lowest = 0.071; Realness: Highest = 0.588, lowest = 0.384.

3.1 Exploratory Factor Analysis

Based on the literature review, we identified no publications with Portuguese speaking samples that used

Table 2. *Exploratory Factor Analysis Results*

	Item code	F 1	F 2	F 3
1	INV1			0.769
2	INV2	0. 212		0.679
3	INV3	−0.175		0.635
4	INV4	.349	0.566	
5	REAL1	0.718	0.257	
6	REAL2	0.844	0.179	
7	REAL3	0.833	0.215	
8	REAL4	0.734		
9	SP1	0.230	0.724	0.147
10	SP2		0.296	0.427
11	SP3		0.533	
12	SP4		0.460	−0.343
13	SP5	0.220	0.787	0.118
14	G1	0.207	0.751	0.166
	Eigenvalues	4.196	1.804	1.44
	% of variance	29.9%	12.88%	10.2%
	Total explained variance	58.69%		

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. Rotation converged in five iterations. Values lower than 0.1 were omitted.

the IPQ. With this in mind, an EFA was performed using principal component analysis (PCA) and a fixed number of factors (3) (see Table 2). Two rotation methods were used: the Varimax, with Kaiser Normalization, and the Direct Oblimin. The results obtained were similar in the number of extracted factors and with the same items loading in the same factors. More precisely, Varimax rotation converged in five interactions and the Oblimin rotation presented the same result in ten interactions. A Bartlett Sphericity test was statistically significant ($p < 0.001$), and the Kaiser-Meyer-Olkin (KMO) value obtained was 0.826, thus confirming the adequacy of the data for performing factor analysis. The results presented a three-factors structure. Two of the items from the initially translated version did not load ($< .40$) within the same factors as suggested in the original version of the questionnaire, namely SP2, which loaded in the Involvement scale, and INV4, which loaded in the Spatial scale.

The item G1, which was initially proposed as an independent item, had a loading value of 0.668 in the Spatial Presence subscale. After a content analysis, the item was accepted as measuring Spatial Presence. The three-factor structure obtained explained 58.69% of the total variance. The first factor (Realness) explained 27.33% of the variance, the second (Spatial) explained 18.03%, and the third (Involvement) explained 10.12%.

3.2 Internal Consistency ("Reliability")

In Table 3, we can see the results obtained for each of the questionnaire subscales. For Spatial Presence $\alpha = 0.66$, for Involvement $\alpha = 0.53$, for Experienced Realism $\alpha = 0.83$, and for the Composite $\alpha = 0.76$. When we compared the alpha Cronbach values obtained in the different language IPQ databases that are available online, Spatial Presence and Involvement emerged as the

Table 3. *Cronbach Values of the French, German, and Portuguese Versions of the IPQ*

Sub-scale	French version	German version	Portuguese version IPQp
Spatial Presence	0.78	0.80	0.66
Involvement	0.75	0.68	0.53
Realness	0.54	0.64	0.83
IPQ Total	0.85	0.83	0.76

least stable variables. These values scores were obtained based on the original theoretical version, which consequently was used in the confirmatory factor analysis.

No improvements based on item removal were considered because minimal gains would be obtained.

Regarding the threats to internal validity, none was observed throughout the experimentation. All subjects were from similar sociocultural backgrounds and with similar experiences, even though 16.7% of them were in some degree acquainted with virtual reality. This familiarity, however, was based on knowledge and not on physical experience with such equipment. When it comes to the dropout rate it did happen with only five subjects (1.035%) representing a fairly low percentage and thus assuming a nonthreat factor.

3.3 Confirmatory Factor Analysis—CFA

In order to perform the CFA two models were considered: the original theoretical proposal and the structure obtained in the EFA. Additionally, based on the comments of IPQ authors, Schubert, Friedmann, and Regenbrecht (2001), we considered both the inclusion and exclusion of the G1 variable in the subscale Spatial Presence. Several indexes and adjustment procedures were used to evaluate the adequacy of the model. The X^2 (Qui-square) value obtained, *per se*, is not a good indicator to base a final decision regarding the adequacy of the model on because its values are influenced by the sample size (the bigger the sample the smaller the significance value obtained). To solve this limitation and based on the suggestions of Jöreskog and Sörbom (1989), we used the ratio resulting from the division of the Qui-square by the degrees of freedom (df), which is represented by X^2/df. Byrne (2010) defined scores ranging from 2.0 to 5.0 as acceptable values. Other indexes generally used in the literature are the GFI (Goodness of Fit Index) and CFI (Comparative Fit Index), whose values should be between 0 and 1. Values greater than 0.90 indicate an adequate model (Bentler and Bonnet, 1980). Finally, the RMSEA (Root Mean Square Error of Approximation) index was used, and according to Li, Harmer, Duncan, T., and Duncan, C. (1998), the values should be smaller than 0.08 to demonstrate an acceptable adjustment. In the same way, the AMOS CFA output usually provides a set of modification values that allow the improvement of the model, namely the Lagrange Multiplier and Wald test.

The factorial validity of the IPQ was determined using a confirmatory factor analysis (AMOS, v.22, an SPSS, an IBM Company, Chicago, IL) based on the procedures recommended by Marôco, (2014). The composite reliability was calculated using the formula suggested by Marôco. The Mahalanobis square distance (D^2) and the skewness (*Sk*) and kurtosis (*Ku*) values were used to determine both uni and multivariate normality of the data ($|\text{SK}| < 3$ and $|\text{KU}| < 10$; see Marôco, 2014). No outliers were identified. The overall quality of the adjustment factorial model was based on the following indicators: χ^2/df, GFI, CFI, RMSEA, P[rmsea $\leq$ 0:05], and MECVI (Marôco 2010, p50). The quality of local adjustment was evaluated through the loading weights and the reliability of individual items, and the adjustment model was made through the modification rates greater than 11; with $p < 0.01$ produced by AMOS and based on theoretical considerations. Best practices usually indicate that indices lower than .08 for RMSEA and greater than .90 for CFI are interpreted as indicating acceptable fit (Bentler, 1990; Bollen, 1989; Browne & Cudeck, 1992; Marôco, 2014). Cutoff values of 0.06 for RMSEA, and 0.95 for CFI have been suggested by Hu and Bentler (1999).

We adopted the Weighted Least Squares Mean and Variance adjusted procedure (WLSMV). This provides for asymptotically unbiased, consistent, and efficient parameter estimates and correct goodness of fit indices for variables that have a non-normal distribution. Several CFA were performed using the online databases available for different nationalities (see Table 4).

Table 4. *CFA Results for the Portuguese, French, and German Versions of the IPQ*

	N	χ^2/df	GFI	CFI	RSMEA	AIC	MECVI
Portuguese versionIPQp with G1	478	2.647	0.948	0.941	0.059	254	0.539
Portuguese three factor model excluding G1	478	2.970	.948	0.931	0.064	237. 304	0.502
Portuguese version re-assigning SP2 and INV 4	478	2.813	0.941	0.930	0.062	270.193	0.571
German version	542	2.437	0.955	0.958	0.052	244.15	0.448
French version	77	1.630	.828	.898	0.091	210.0	2.661

Different three-factor models of the original IPQ_p were tested with our sample of 478 university students from northern Portugal (see Table 4). Showing adequate fit indexes, these versions were: Portuguese version IPQ_p with G1; Portuguese version IPQ_p excluding G1; and Portuguese version re-assigning SP2 and INV4. In Table 4 we also present the data for the German and French version, only the French version does not present adequate fit indexes.

When we tested the IPQ_p, without any corrections results showed an acceptable adjustment quality ($\chi^2/df = 2.717$; GFI = 0.942, CFI = 0.935; RMSEA = 0.060; P [RMSEA = ≤ 0.08]; MECVI = 0.59). Analysis showed that there were no identified outliers, and no need to eliminate any item from the model. When we took into account the correlation between measurement errors, several covariances were identified requiring correction, namely those between items e1–e2 and e1–e3 belonging to the Involvement factor, e6–e7 in the realness factor, and e9–e10 and e9–e12 in the realness factor. Following these corrections, we obtained an improvement in the adjustment quality ($\chi^2/df = 2.647$; GFI = 0.948; CFI = 0.941; RMSEA = 0.059 [0.048 to 0.069]; MECVI = 0.539) and in support of the proposed theoretical model. The model with the adjustments showed better indicators, as seen in Table 4 from the scores obtained. We highlight the MECVI, which was reduced from 0.59 to 0.54. The composite reliability of each factor was as follows: Realness (R) = 0.823, Involvement (I) = 0.314, and Spatial (S) = 0.728. The composite reliability of the three factors demonstrated a low value for the Involvement and acceptable values for the Realness and Spatial Presence variables.

The determination of the validity of the constructs as proposed is key to the validation process of instruments. In the present study the discriminant validity of the factors was evaluated by comparing the VEM (Marôco 2010, p. 62) values with the squares of the correlations between factors. It was found that VEMi = 0.145, VEMr = 0.542 and VEMs = 0.346. The squared correlations obtained were $r^2_{IR} = 0.77$, and $r^2_{SI} = 0.46$, and $r^2_{RS} = 0.46$. When we looked separately at VEM values and the square of the correlation between variables, we found that discriminant validity was obtained in only one of the three factors, namely in VEM realness, whose values were higher and more desirable than those of r^2. In the case of the correlations to obtain r^2_{SI} the highest correlation value was verified between SP1 and INV4 ($r^2 = 0.216$), and for the case of RS the highest correlation was $r^2 = 0.116$. The SI correlation could be anticipated based on the EFA structure proposed whose CFA indexes showed to be poorer than the original proposal suggested.

The CFA based on the EFA presented poorer results than the original theoretical proposal ($\chi^2/df = 2.813$; GFI = 0.941; CFI = 0.930; RMSEA = 0.062; [0.052 to 0.72]; MECVI = 0.539). When we compare the obtained results from the different models tested it became apparent to the authors that the IPQp with G1 was the most appropriate to accept once it better corroborates the Schubert et al.'s (2001) initial proposal (see Figure 1). The results suggested that the original theoretical proposal is indeed theoretically robust and should consequently be preserved in the Portuguese version.

Presence is a key dependent variable in virtual reality research. The requirements for an appropriate assessment of human perception and satisfaction relating to

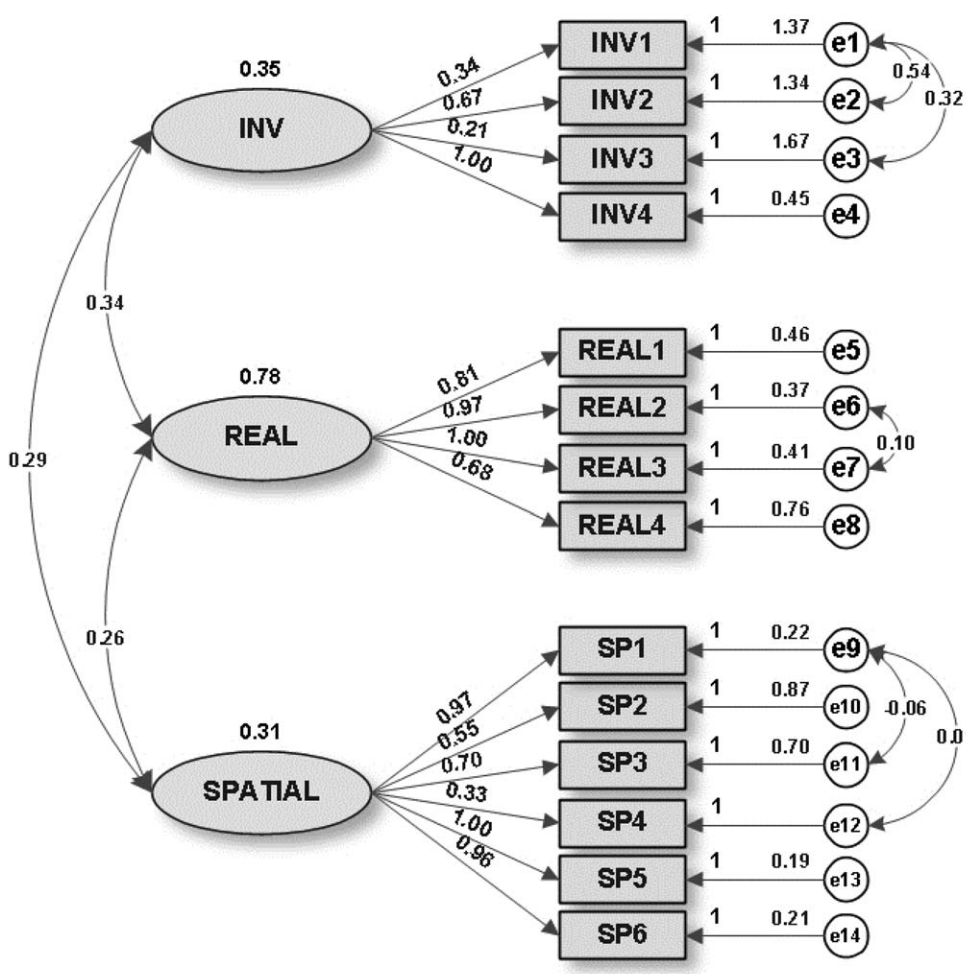

Figure 1. *Theoretical model of the* IPQ_p.

the use of the technology and its use demand extra care regarding the variables and means to evaluate the human experiences elicited. Thus, presently, given the lack of fully validated instruments to measure human experience elicited through virtual reality exposure, it is urgent for researchers to overcome some of the existent limitations to the credibility of research in this domain.

In this study, we created a cross-cultural adaptation of the IPQ, originally developed by Schubert et al. (2001), for a sample of 478 university students in northern Portugal. Several procedures were used, as recommended (Matsumoto & van de Vijver, 2011; Hambleton & Zenisky, 2011), to proceed with cultural adaptation of the IPQ, namely evaluating the semantic and cognitive aspects of the items to improve the construct validity as related to a criterion and the reliability of IPQ subscales. Compared with the original version of the IPQ, no items were removed from any of the subscales, and all items had adequate correlation values and were within the theoretically predicted scales. Good adjustment values were obtained for the discriminant validity for the realness scale. Very high correlations were obtained between the scales of Involvement and Realness.

The final model presented less than desirable VEM values for the Involvement and Spatial variables and desirable values for the Realness variable. Convergent validity of the subscales was established. The IPQ's overall discriminant validity was not accepted because R_s^2 of the factors was always less than the VEM scores obtained in each of the factors taken into consideration. One possible solution to improve the weaknesses identified might be to increase the number of items in each subscale.

Some of the limitations found do not allow the rejection of the test validated. Indeed, some of the problems identified might be overcome in future studies paying particular attention to the type of subjects to be studied.

According to van de Vijver and Leung (2011) theory driven testing requires that modifications are made based on the initial theoretical proposal. For this reason, even though the EFA suggested that item SP2 loaded in the involvement scale and the INV4 in the spatial scale, in order to respect the principles associated with theory validation, both items were maintained in the original factorial structure and consequently submitted to hypothesis testing. Such procedures should be respected to properly validate the theoretical proposal of the original scale. If such practices are not respected one of the key aspects for scientific theoretical consolidation will be violated: the search for the universality of concepts to assure proper interpretation of instruments (theory) and consequently explain human behavior. In the case of our study the factorial structure was confirmed.

4 Conclusions

In the present study, particular attention was given to the multiple levels of equivalence bias. The use of instruments developed in one cultural (and linguistic) context, in order to be applied in another, needs to be translated and validated. However, in order to overcome these limitations, the focus was centered on theory-driven validation procedures, namely the type of statistical analysis used (i.e., CFA).

The different types of adaptation domains considered were (1) Concept, which in turn was subdivided into concept driven and theory driven adaptation; (2) Culture, subdivided into terminology/factual driven and norm driven; and (3) Language, subdivided into linguistic and pragmatics driven, and finally format driven. Of all these aspects, the theory driven validation was assumed as the most important. The remaining domains were taken into consideration in the required processes that preceded the statistical analysis to test the hypothesis of the structural factorial theory validation. These included that the same number of factors and items was maintained, that the factorial loading was equivalent in the different cultural groups (the group used in the original version and the sample used in our study), if the same items explained the same variability level of the construct (Byrne, 2010). The Portuguese version of the IPQ was successful in confirming all these aspects.

Altogether, the results from the reliability and CFA demonstrated that this Portuguese version of the IPQ has excellent psychometric properties and supported the three dimensionality of the scale as theoretically proposed. The confirmation of the three factors reinforces the questionnaire robustness. The analyses show that the original version (German) and our version (Portuguese), both submitted to a CFA, demonstrated adequate indices to validate the theoretical proposal of a three factor structure. The IPQ is, so far has we know, the only instrument properly validated in different cultural context. For that reason, until other instruments are subjected to the same validation procedures, the IPQ should be the instrument of choice for the study of presence. The French version, based on the free access IPQ database, when subjected to the CFA did not prove its structural validity.

The other translations versions of the IPQ, available online, still need to be validated, in order to ensure that the psychometric properties of the questionnaire are adequate for research use. The nonconformity to these procedures produce highly questionable knowledge, because studies using them produce results that lack both internal and external validity.

Notwithstanding the limitations identified, the Portuguese version of the IPQ (and the first, to the best of our knowledge) is an important and useful instrument for assessing presence when conducting research in virtual reality. With this version, research in this topic with Portuguese speaking subjects is facilitated.

However, much research is needed with regard to this version, whether due to its preliminary nature or because of the uniqueness of the variables being studied and the theoretical background on which studies are based. Future research should try to recruit a greater number of individuals with previous experience with virtual reality equipment, and with different sociocultural backgrounds. In addition, attempts should be made to increase the number of items per subscale.

Even though the factor structure originally proposed was confirmed, the obtained discriminant validity results encourage further attention, namely when it comes to the possibility of increasing the number of items per factor. An invariance study is recommended as a follow-up to the analysis as described in this study.

Acknowledgments

This work was partially supported by the project REC I/EEI-SII/0360/2012 entitled "MASSIVE—Multimodal Acknowledgeable multiSenSorial Immersive Virtual Enviroments." This work was also partially supported by the project "TEC4-Growth—Pervasive Intelligence, Enhancers and Proofs of Concept with Industrial Impact/NORTE-01-0145-FEDER-000020" and is financed by the North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, and through the European Regional Development Fund (ERDF).

References

Ahlberg, G., Enochsson, L., Gallagher, A. G., Hedman, L., Hogman, C., McClusky, D. A., Ramel, S., Smith, C. D., & Arvidsson, D. (2007). Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies. *The American Journal of Surgery, 193*(6), 797–804.

Almeida, L. S., & Freire, T. (2003). Metodologia da investigação em psicologia e educação. *Cadernos de Consulta Psicológica, 2*: 143–154.

Alsina-Jurnet, I., & Gutiérrez-Maldonado, J. (2010). Influence of personality and individual abilities on the sense of presence experienced in anxiety triggering virtual environments. *International Journal of Human-Computer Studies, 68*(10), 788–801.

Anderson, P. L., Price, M., Edwards, S. M., Obasaju, M. A., Schmertz, S. K., Zimand, E., & Calamaras, M. R. (2013). Virtual reality exposure therapy for social anxiety disorder: A randomized controlled trial. *Journal of Consulting and Clinical Psychology, 81*(5), 751.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*(3), 588.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238.

Betella, A., Carvalho, R., Sanchez-Palencia, J., Bernardet, U., & Verschure, P. F. (2012). Embodied interaction with complex neuronal data in mixed-reality. *Proceedings of the 2012 Virtual Reality International Conference,* 3.

Bideau, B., Kulpa, R., Vignais, N., Brault, S., Multon, F., & Craig, C. (2010). Using virtual reality to analyze sports performance. *Computer Graphics and Applications, IEEE, 30*(2), 14–21.

Biocca, F., & Delaney, B. (1995). Immersive virtual reality technology. *Communication in the Age of Virtual Reality,* 57–124.

Bliss, J. P., Tidwell, P. D., & Guest, M. A. (1997). The effectiveness of virtual reality for administering spatial navigation training to firefighters. *Presence: Teleoperators and Virtual Environments, 6*(1), 73–86.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: John Wiley & Sons.

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology, 1*(3), 185–216.

Brown, S., Ladeira, I, Winterbottom, C, & Blake, E (2003). The effects of mediation in a storytelling virtual environment. O. Balet, G. Subsol, & P. Torguet (Eds.), in *Virtual storytelling: Using virtual reality technologies for storytelling* (pp. 102–111). Berlin, Heidelberg: Springer-Verlag.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research, 21*(2), 230–258.

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming.* New York: Routledge.

Czub, M., & Piskorz, J. (2014). How body movement influences virtual reality analgesia. *2014 International Conference on Interactive Technologies and Games (iTAG)* 13–19.

DeAngelis, T. (2012). A second life for practice. *Monitor on Psychology, 43*(3), 48–51.

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.

Gallagher, A. G., Ritter, E. M., Champion, H., Higgins, G., Fried, M. P., Moses, G., Smith, C, D., & Satava, R. M.

(2005). Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. *Annals of Surgery, 241*(2), 364–372.

Gerardi, M. Cukor, J., Difede, J. , Rizzo, A. , & Rothbaum, B. (2010). Virtual reality exposure therapy for post-traumatic stress disorder and other anxiety disorders. *Current Psychiatry Reports, 12*, 298–305. doi: 10.1007/s11920-010-0128-4

Grantcharov, T. P., Kristiansen, V. B., Bendix, J., Bardram, L., Rosenberg, J., & Funch-Jensen, P. (2004). Randomized clinical trial of virtual reality simulation for laparoscopic skills training. *British Journal of Surgery, 91*(2), 146–150.

Hambleton, R. K., & Zenisky, A. L. (2011). Translating and adapting tests for cross cultural assessments. In Em D. Matsumoto & F. J. R. van de Vijver (Eds.), *Cross-cultural research methods in psychology* (pp. 46–74). New York: Cambridge University Press.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.

Hyun, J., Habuchi, Y., Park, A., Ishikawa, T., Kourogi, M., & Kurata, T. (2010). Service-field simulator using MR techniques: Behavior comparison in real and virtual environments. *Proceedings of the 20th International Conference on Artificial Reality and Telexistence (ICAT2010),* 14–21.

Jöreskog, K. G., & Sörbom, D. (1989). *LISREL 7: A guide to the program and applications* (2nd ed.). Chicago: SPSS.

Krijn, M., Emmelkamp, P. M., Biemond, R., de Ligny, C. D. W., Schuemie, M. J., & van der Mast, C. A. (2004). Treatment of acrophobia in virtual reality: The role of immersion and presence. *Behaviour Research and Therapy, 42*(2), 229–239.

Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J. (2001). A cross-media presence questionnaire: The ITC-Sense of Presence Inventory. *Presence: Teleoperators and Virtual Environments, 10*(3), 282–297.

Li, F., Harmer, P., Duncan, T. E., Duncan, S. C., Acock, A., & Boles, S. (1998). Approaches to testing interaction effects using structural equation modeling methodology. *Multivariate Behavioral Research, 33*(1), 1–39.

Lombard, M., Ditton, T. B., & Weinstein, L. (2009). Measuring presence: The temple presence inventory. *Proceedings of the 12th Annual International Workshop on Presence,* 1–15.

Marôco, J. (2010). *Análise de Equações Estruturais: Fundamentos teóricos, software & aplicações.* Pêro Pinheiro: Report Number.

Marôco, J. (2014). *Análise estatística com o SPSS statistics* (6^{th} ed.). Pêro Pinheiro: Report Number.

Matsumoto, D., & van de Vijver, F. J. (Eds.). (2011). *Cross-cultural research methods in psychology.* New York: Cambridge University Press.

O'Neil Jr, H. F., Andrews, D. H., & O'Neil, H. F. (Eds.). (2000). *Aircrew training and assessment.* Boca Raton: CRC Press.

Panahi-Shahri, M., Fathi-Ashtiani, A., & Azad-Fallah, P. (2009). Reliability and validity of Igroup Presence Questionnaire (IPQ). *Journal of Behavioral Sciences, 3*(1), 27–34.

Price, M., Mehta, N., Tone, E. B., & Anderson, P. L. (2011). Does engagement with exposure yield better outcomes? Components of presence as a predictor of treatment response for virtual reality exposure therapy for social phobia. *Journal of Anxiety Disorders, 25*(6), 763–770. doi: 10.1016/j.janxdis.2011.03.004

Psotka, J. (1995). Immersive training systems: Virtual reality and education and training. *Instructional Science, 23*(5–6), 405–431.

Rosakranse, C. & Oh, S. Y. (2014). "Measuring presence: The use trends of five canonical presence questionnaires from 1998–2012." In A. Felnhofer & O. D. Kothgassner (Eds.), *Challenging presence: Proceedings of the 15th international conference on presence,* 25–30.

Saposnik, G., & Levin, M. (2011). Virtual reality in stroke rehabilitation a meta-analysis and implications for clinicians. *Stroke, 42*(5), 1380–1386.

Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence: Teleoperators and Virtual Environments, 10*(3), 266–281.

Seidel, R. J., & Chatelier, P. R. (Eds.). (2013). *Virtual reality, Training's future? Perspectives on virtual reality and related emerging technologies* (Vol. 6). New York: Springer Science & Business Media.

Seymour, N. E., Gallagher, A. G., Roman, S. A., O'Brien, M. K., Bansal, V. K., Andersen, D. K., & Satava, R. M. (2002). Virtual reality training improves operating room performance: Results of a randomized, double-blinded study. *Annals of Surgery, 236*(4), 458–464.

Skydome Studios. (2014). Don't let Go! (Version 1.2). Retrieved from https://share.oculus.com/app/dont-let-go

Slater, M., Usoh, M., & Steed, A. (1994). Depth of presence in virtual environments. *Presence: Teleoperators and Virtual Environments, 3*(2), 130–144.

Stendal, K., & Balandin, S. (2015). Virtual worlds for people with autism spectrum disorder: A case study in second life. *Disability & Rehabilitation,* (0), 1–8.

Suied, C., Drettakis, G., Warusfel, W., & Viaud-Delmon, I. (2013). Auditory-visual virtual reality as a diagnostic and therapeutic tool for cynophobia. *Journal of Cybertherapy and Rehabilitation, iACTOR, 16*(2), 145–152.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics*. Needham Heights, MA: Allyn & Bacon.

Taylor, H. L., Lintern, G., Hulin, C. L., Talleur, D., & Emanuel, T. (1997). *Transfer of Training Effectiveness of Personal Computer-Based Aviation Training Devices*. University of Illinois at Urbana-Champaign Savoy, Institute of Aviation.

Ullman, J. B. (2001). Structural equation modeling. In B. G. Tabachnick & L. S. Fidell (Eds.), *Using multivariate statistics* (4th ed., pp. 653–771). Needham Heights, MA: Allyn & Bacon.

Viaud-Delmo, I. (n.d.). Igroup presence questionnaire (IPQ) item download. Retrieved from http://www.igroup.org/pq/ipq/IPQinstructionsFr.doc

Vijver, F., & Leung, K. (2011). Equivalence and bias: A review of concepts, models, and data analitics procedures. In D. Matsumoto & F. Vijver (Eds.), *Cross-cultural research methods in psychology* (pp. 17–45). New york: Cambridge University Press.

Waltz, C., Strickland, O., & Lenz, E. (1991). *Measurement in nursing research*. Philadelphia: F. A. Davis.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments, 7*(3), 225–240.