Presence measurement revisited: Developing presence scales from self-reports and behavioral observations using the Rasch model

Antal Haans & Wijnand A. IJsselsteijn

Eindhoven University of Technology

Author Note

Correspondence concerning this article should be addressed to Antal Haans, Eindhoven University of Technology (IPO 1.35), P.O. Box 513, NL-5600 MB Eindhoven, The Netherlands; e-mail: a.haans@tue.nl

Abstract

In this contribution we provide empirical evidence for the validity and reliability of a novel method for measuring presence (defined as the illusion of non-mediation). Implemented within the Rasch family of models, the proposed method measures presence on a unidimensional scale using wide variety of perceptual, visceral, cognitive and behavioral responses toward the virtual environment as indicators. Results from an experiment demonstrated the validity and reliability of the proposed method, as well as its sensitivity in differentiating between the effects of real and simulated walking. In addition, moderate to high correlations were found with all main dimensions (i.e., spatial presence, involvement, and realism) of the existing Igroup Presence Questionnaire (IPQ). Our research illustrates that both self-reports and behavioral observations—as long as reflecting clearly that a person did or did not fall for the illusion of non-mediation—are valid indicators of presence. Further implications and limitations of the research are discussed.

Keywords: Virtual environments; Presence; Measurement; Navigation Mode; Rasch model

Presence measurement revisited: Developing presence scales from self-reports and behavioral observations using the Rasch model

It is almost forty years ago that Marvin Minsky (1980) coined the term telepresence, but despite decades of research the measurement of presence remains challenging (see, e.g., Laarni, Ravaja, Saari, Böcking, Hartmann, & Schramm, 2015; Slater, Lotto, Arnold, & Sanchez-Vives, 2009). This results, in part, from different scholars entertaining different conceptions of what presence is (see, e.g., Lombard & Jones, 2015), leading necessarily to differences in operationalization of the concept. At the same time, it results from a misconception of what constitutes measurement in psychology (Michell, 1999; for a discussion specifically on presence, see Haans, 2014). As a result, there now exist a multitude of presence instruments whose outcomes do not always appear to overlap (see e.g., Laarni et al., 2015; Slater et al., 2009). Recently, we proposed yet another method of measuring presence in interactive virtual environments (Haans, 2014). A first empirical test of this method is reported in this contribution.

Before introducing the method and its empirical validation, it is inevitable that we first explain, in brief, how we conceive of presence and its underlying mechanisms. Many other theoretical explanations are, of course, found in the presence literature, including so-called “inner presence” theories (e.g., Riva, Waterworth, Waterworth, & Mantovani, 2011) or Slater’s (2009) distinction between place and plausibility illusion. An elaborate discussion of these alternative theories is outside the scope of the present contribution. Instead, we provide the theoretical background required to understand the measurement method that we propose.

Presence as the illusion of non-mediation

To us, presence in interactive virtual environments is best defined as the illusion of non-mediation (Lombard & Ditton, 1997). In this conception, presence occurs when the mediating technology becomes transparent to the user, as a result of which the user will respond to the virtual environment as if it were real (IJsselsteijn, 2005). Our explanation of the processes underlying presence relies heavily on Biocca (1997) who has argued that our biological receptors and actuators too are communication media between us and the immediate environment; media that are transparent to such an extent that we never think about them as such. That is, we never think about the retinal image when visually searching for our lost keys, or about which specific patterns of muscles to contract when walking. Instead, we usually “simply” see the world and act on it directly.

Following Biocca (1997), we thus have argued that any explanation of presence should first and foremost consider how we as humans are embodied (Haans & IJsselsteijn, 2012). This includes the capacity of our body schema and body image to include tools and other artefacts as functional and phenomenological extensions of the body. These processes of incorporation or adaptation allow media technologies to become as transparent as our own biological sensors and actuators; allowing us to interact with a simulated environment as if the mediating technology were not there (i.e., the illusion of non-mediation).

An extensive discussion of how the body schema and image can adapt to and / or incorporate tools and artifacts is outside the scope of the present contribution (for a more detailed discussion, see Haans & IJsselsteijn, 2012). For the present discussion, however, it is important to stress the importance of sensorimotor integration and establishing existing or new sensorimotor contingencies in the process. Every time we look towards or away from an object, the information received by the retina will change in a manner that corresponds lawfully and consistently to our head and eye movements. Similarly, when approaching an object, the visual angle of its retinal image will increase lawfully and consistently as well. Registering such sensorimotor contingencies appear crucial in rendering the workings of the body and its parts transparent to its user; allowing it to be used and perceived of as a coherent functional unity. Interactive media technologies, such as a head-mounted display, can become part of that same functional unity, as these too allow for sensorimotor contingencies to be established (or enacted).

Based on this theoretical conception, the strength of the illusion of non-mediation thus primarily depends on the extent to which the mediating technology allows sensorimotor contingencies to be enacted. This includes both interactivity (the extent to which the user can influence the stream of mediated information) and vividness (the number of sensory receptors that are addressed by the medium; Steuer, 1992). Of course, the extent to which presence occurs, even with the same technology, can be different for different people (Coxon, Kelly, & Page, 2016; Lombard & Ditton, 1997; Steuer, 1992). Such individual differences are also found in the capacity to incorporate foreign objects (e.g., a virtual hand) in the body image (e.g., Haans, Kaiser, Bouwhuis, & IJsselsteijn, 2012; IJsselsteijn, de Kort, & Haans, 2006).

In contrast to Lombard and Ditton (1997), who defined presence as the perceptual illusion of non-mediation, we choose not to include the term perceptual. It is not only our perceptual apparatus that is susceptible to the illusion of mediation—other systems too can be fooled by media technology. These include our visceral and cognitive systems as well as those systems responsible for the more automatic behavioral responses to the environment. For example, when looking down into a virtual pit, most people will experience, not just a perception of depth, but a visceral reaction as well (e.g., Meehan, Razzaque, Insko, Whitton, & Brooks, 2005). Cognitive responses, such as not challenging the reality of the simulated environment, may occur when media technologies become more advanced, as, for example, in the Star Trek: Next Generation episode “Ship in a Bottle”. In other words, we must consider all systems of the organism as being potentially fooled into treating the virtual environment as if it were real (Haans, 2014; also Slater et al., 2009).

Measuring presence: a proposal

Any response to the virtual environment that reflects the illusion of non-mediation can thus be considered as an indicator of presence (Haans, 2014). These include self-reported cognitions and experiences, visceral and automatic responses, and volitional behaviors (as for example, refraining to jump into a virtual hole in the floor; Usoh, Arthur, Whitton, Bastos, Steed, Slater, & Brooks, 1999). This conclusion has been raised before by multiple authors, most notably by Slater and colleagues (2009; p. 206):

> *"We are interested across the board—in what people say, in what thoughts they report they had, how their emotions were effected, how their attention was distributed, in addition to observable behaviour, physiological responses, eye-movement patterns, brain activity and so on—. We do not give a higher priority to any of these, and results from these different levels may indeed be contradictory (e.g., people saying that they felt nothing but the physiological and observable behaviourable responses show otherwise)."*

This quote from Slater and colleagues (2009) also points to a mistake often made in the presence literature: When a self-report contradicts with a behavioral response, then either the self-report or the behavioral response must be an invalid indicator. To appreciate this common mistake, one should consider that a particular self-report, or a certain observed behavior, is not a measure of presence, but merely one of the many possible indicators on the basis of which a presence measure could be obtained (see also Haans, 2014). Only measures of the same attribute are required to be correlated with each other, not the nominal or ordinal observations on which these measures are based. Mathematical ability measurements obtained with one test should correlate strongly with measurements of the same persons obtained with another, but the nominal observation that a person can solve a problem of simple addition does not tell us anything, and does not need to tell us anything, about his or her performance on a higher order differentiation. Yet the first is just as valid as an indicator of mathematical ability as the latter. What they differ on, instead, is item difficulty.

Indicators of presence can be expected to differ in item difficulty as well (Haans, 2014). Consider the following example adapted from Dennett (1996): One may not feel consciously present in the VR environment, but the optical flow and virtual objects in our peripheral vision may nonetheless affect the length of our steps when navigating through the simulated world. Similarly, and in all likelihood, more people will experience a visceral response when looking down into a virtual pit, than there are people that do not dare to jump in it (see Figure 1).

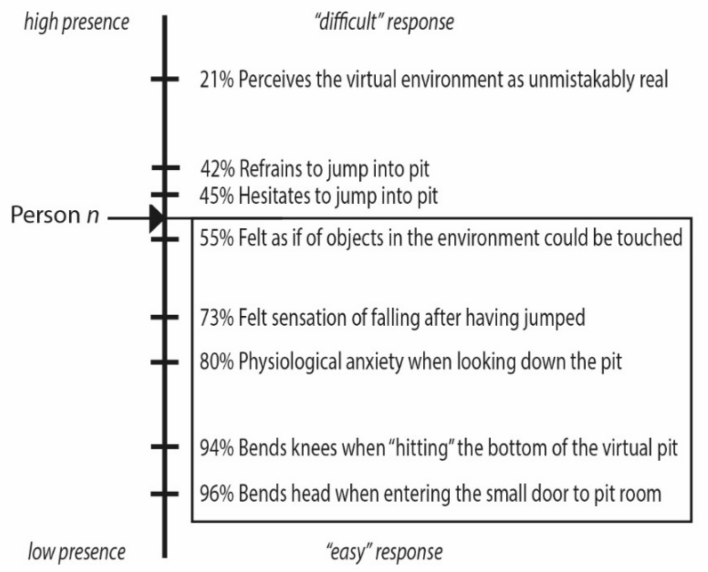

Figure 1. Hypothetical position of eight indicators along the presence scale for a pit environment, and the presence of a single individual *n*. Percentages reflect the probabilities for any individual with *n*’s level of presence to display each response. Figure obtained from Haans (2014).

Correlational attempts to combine these different observations into a presence measure, as for example by means of factor analytical models, are most likely to fail (also Haans, 2014). It has, for example, been well-documented that the number of extracted factors is artificially inflated when indicators vary in item difficulty (see, e.g., Gorsuch, 1997). Instead, we propose a presence measurement approach based on the Rasch model (Rasch, 1960; Bond & Fox, 2007); in particular, the many-facet Rasch model (Linacre, 2002).

A Rasch model implementation

Let us consider an experiment in which each of *n* individuals visits *j* different virtual environments. Presence is measured with a series of *i* indicators (e.g., self-reported experiences and behavioral observations). As mentioned briefly above, individuals are expected to differ in the extent to which they experience presence. Let us call this the individual's propensity to experience presence (indicated with a θ_n). The higher a person's score the more likely he or she is to fall for the illusion of non-mediation, and thus to endorse or demonstrate the presence indicators. The various virtual environments also will differ in the extent to which they elicit presence (e.g., through the specific mode of navigating through the environment; Slater, Usoh, Steed, 1995). Let us call this the constraints upon presence imposed by the technology (indicated with a λ_j). The higher the environment's score the more the experience of presence will be constrained. Finally, and as explained above, the indicators (or items) will differ in item difficulty (indicated with a δ_i). The more difficult the indicator is, the less likely it is to be endorsed. The Rasch model, then, becomes:

$$LN\left(\frac{P(x_{nij}=1)}{1-\ P(x_{nij}=1)}\right)=\theta_n-(\lambda_j+\delta_i)$$

In this model, the natural logarithm of the odds that person *n* shows or endorses a certain indicator *i* in virtual environment *j* is governed by three factors: *n*'s propensity to experience presence (θ_n), the specific constraints imposed by the characteristics of virtual environment *j* (λ_j) and the difficulty of the specific item or indicator *i* (δ_i). All parameters in this equation are estimated by means of maximum-likelihood estimation, and are expressed in log odd units (also called logits; for a similar example in a different domain, see Haans et al., 2012).

The Rasch model, however, is not only descriptive, but also prescriptive. It requires items to be ordered according to their difficulty invariantly and transitively across persons and environments, and vice versa (also Haans et al., 2012). Thus, if a person endorses, for example, four out of 10 indicators, the Rasch model also prescribes which four are most likely be endorsed: They should be among the least difficult ones. And for the most difficult to endorse item (e.g., agreeing to know for sure that virtual environment was real and not a simulation), we expect to exclusively find such reports from the most susceptible persons and / or in the environments that provide the most immersion.

This invariance assumption should be sufficiently met in order to map persons, virtual environments, and indicators on a single (unidimensional) scale of equal additive units, and thus to compare individuals and virtual environments in a meaningful manner with respect to their relation with presence. These formal Rasch model expectations can be tested empirically against the observed data (see, e.g., Bond & Fox, 2007). Typically, Mean Square (MS) fit statistics are used to test such data-to-model fit.

Research Aims

In this experiment, we test the proposed Rasch-based method for measuring presence in immersive virtual environments. We expect that participants' perceptual, visceral, cognitive, and behavioral responses to the simulated content can all be scaled onto a unidimensional scale as evidenced by an adequate data-to-model fit. In addition, we test the method’s sensitivity in differentiating between individual propensities to experience presence, and between the constraints imposed by technological features of the virtual environment; in particular, the effect of real versus simulated walking. We expect that presence will be less constrained with real walking than with navigating by means of a controller (cf. Slater et al., 1995; Usoh et al., 1999), as the former allows participants to enact more sensorimotor contingencies. As a test of convergent and/or divergent validity, we compare measures obtained with the proposed Rasch-based method to those obtained with an established multi-dimensional presence instrument: The Igroup Presence Questionnaire (IPQ; Schubert, Friedmann, & Regenbrecht, 2001).

Method

Participants

A convenience sample of 50 people participated in the experiment, of which 24 (48%) were men and 26% (52%) were women. Average age of the participants was M = 22.4 (SD = 7.8; range 17 to 45). All participants were recruited from the J.F. Schouten database of Eindhoven University of Technology, Eindhoven, The Netherlands. All received € 5,- euro as a compensation for their participation.

Design

We conducted a two condition (simulated versus real walking) within-subject experiment with presence as the dependent variable. In the real walking condition, participant could navigate through the virtual environment through natural locomotion. In the simulated walking condition, the participant remained at a fixed position and navigated through the environment by means of a controller. The order of the two conditions was counterbalanced across participants.

Materials and apparatus

The virtual environment covered 6 by 7 meters and consisted of two rooms connected by a small door of 1.7 meter in height (see Figure 2). The larger of the two rooms contained a 12-meter-deep pit with a grate on top that could be opened and closed. An avatar holding a white ball was situated in the pit room, and, when activated by the experimenter, would throw the ball in the direction of the participant's head. Various objects, including a book case, tables, luminaires, and a picture frame, where situated in the virtual environment. The virtual environment was created in Autodesk 3D Studio Max. The avatar and all animations and interactions were implemented using the WorldViz Vizard 4.0 Virtual Reality toolkit.

Figure 2. Virtual environment used in the present experiment, showing the two rooms connected by the small door, the avatar holding the white ball, and the pit enclosed by the grate.

The virtual environment was displayed on an nVisor SX111 head-mounted display. Head orientations were tracked in both the real and the simulated walking condition using the Intersense InertiaCube 3. In the real-walking condition, the participant’s location was tracked with a PhaseSpace Impulse camera positioning system. In the simulated walking condition, participants used a Nintendo WiiMote controller to navigate through the virtual environment. The directional pad (D-pad) was used for moving forward and side-ways as referenced on the gaze direction of the participant. In both the real and the simulated walking condition, participants used the A-button on the WiiMote controller to open the grate, causing them to fall into the pit.

Finally, a high definition video-camera was used to record the participants’ behavior in the virtual environment. This allowed for behavioral indicators, such as bending when passing through the small door or a reflex to the approaching ball, to be coded and confirmed afterwards by multiple observers.

Procedure

After being welcomed in the laboratory room, participants were asked to read and sign an informed consent form. Next, the experimenter explained that the participant was to visit a virtual environment twice, and that he or she, each time, would be asked to conduct various tasks. It was explicitly mentioned that the participant could decide not to comply with the requests when he or she did not want to.

The first experimental session started with an explanation of how to navigate through the virtual environment. Depending on the experimental condition, the experimenter either explained natural movement (in the real walking condition) or the use of the D-pad of the WiiMote controller (in the simulated condition). In the simulated walking condition, it was explained that while participants needed to turn and / or move their head to navigate, they were otherwise requested to remain physically in the same spot. Finally, and for both conditions, the experimenter explained the use of the A-button. Finally, the participant was assisted with putting on the head-mounted display.

In both conditions, the participant started in the small room. Here the participant was instructed to look around in order to get used to the head-mounted display, before exploring the room through movement. After about a minute, the experimenter opened the small room and the audio instructions (played on headphones) asked the participant to proceed to the second room. After entering the room, the participant was asked to look at the avatar until further instructions. Next the experimenter raised the hand of the avatar, and, after about two seconds, the ball was thrown into the direction of the participant’s head. Next, the participant was instructed to move to the edge of the pit and look down. After doing so, the experimenter would close the grate, and the participant was requested to step onto the middle of it. Finally, the participant was instructed to look down through the grate towards the bottom of the pit, and to press the A-button on the WiiMote controller in order to open the grate and fall down into the pit. Subsequently, the experimenter assisted the participant with taking off the head-mounted display, and asked the participant to fill out a questionnaire.

When the participant had completed the questionnaire, the next session would start but this time with the other means of navigating through the virtual environment. This second session was otherwise similar to the first, except that the questionnaire also asked for demographics, body length, and fear of heights. Finally, the participants where debriefed about the aim and purpose of the experiment and were thanked for their participation.

Measures

Two instruments were used to measure presence: The proposed Rasch-based measurement method, and the Igroup Presence Questionnaire (IPQ). In addition, we asked participant if they suffered from fear of heights.

The Rasch-based presence instrument.

The proposed presence measurement method is based on Haans (2014) and used a combination of behavioral observations and self-reported experiences and cognitions as items (see Table 1 for an overview). Observed behavioral indicators included, for example, bending when passing through the small door, a reflex when the virtual ball was thrown in the direction of the head, or whether participants refrained from opening the grate that covered the pit, thus preventing a fall into the virtual pit. In addition, we recorded the time (in seconds) between the start of the audio file instructing the participant to press the button to open the grate (which had duration of 15 sec) and the actual button press. We hypothesized that a hesitation in pressing the button would result from an internal conflict between the perception of depth and a cognitive understanding of the environment as being virtual; an internal conflict often observed during the pilot experiments.

One observational item “participant braced for impact when hitting the bottom of the virtual pit” was dropped from the analysis. We observed different types of behavioral responses, from bending knees to re-finding one’s balance, and it was unclear whether and if so which of these behaviors would indicate presence, or in fact the opposite. Since no participants refrained from stepping up onto the grate and from pushing the button to fall into the pit in any of the two conditions, these two items were also excluded from the analyses (as these are hence not informative in differentiating between individuals and experimental manipulations). Time recordings were missing for 3 participants in the real walking condition as they did not have the WiiMote controller with them when requested to open the grate.

All observed behaviors were coded into a dichotomous format: “no endorsement” (e.g., did not refrain from pressing the button) versus “endorsement” of the behavioral indicator (e.g., did refrain from pressing the button). For participant whose body length was smaller than or equal to 1.65 meter, and thus for whom refraining to bend when entering through the small doorway cannot be regarded as an indicator for a lack of presence, a missing value was inserted in the data. This was the case for 12 out of 50 participants. A median split (median = 16.5 sec) was used for dichotomizing the elapsed time before pressing the button into a “did not hesitate” and a “hesitated” format.

The video recordings of one person were lost due to human error. The remaining videos were coded by two independent observers for the “bending for the door” and “reflex on the ball” indicator. Any type of bending or reflex (from, for example, a minimal twitch with the head, to trying to protect the head with the hands) were regarded as endorsement of the behaviors. The two coders were in disagreement on 14 (7.1%) out of 196 observations. These cases were discussed amongst the two coders. When the two coders remained in disagreement, then the behavior was coded as “no endorsement”. In case of the lost video, we relied on the notes of the experimenter made during the sessions.

A total of 15 self-report items were included in the questionnaire that participant completed after each session (see Table 1). These items asked participants to confirm or disconfirm a variety of perceptual effects (e.g., having the feeling to be able to touch the various virtual objects that they saw; an indicator of distal attribution, see Loomis, 1992), experienced visceral reactions (e.g., when looking down into the pit), and cognitions (e.g., knowing for sure that the two rooms just visited were real and not a simulation). For persons whose body length was smaller or equal to 1.65 meters, responses to all items pertaining to the small door were recoded as missing responses, for the same reason as outlined above.

One control item was included to check whether or not participant saw the virtual ball being thrown at them. Participant could respond to each item using a dichotomous yes/no response format. Only one participant indicated not to have seen the ball being thrown. For this participant all self-report and observation items related to the ball were coded as a missing response.

The Igroup Presence Questionnaire (IPQ).

The IPQ contains 14 items which tap into four sub-dimensions of presence: Spatial presence (5 items), involvement (4 items), realism (4 items), and the single item general presence dimension (Schubert et al., 2001). We used the Dutch translation of the IPQ (available at: http://www.igroup.org/pq/ipq/download.php#Dutch). These IPQ items were included in the questionnaire that participant completed after each session.

Fear of heights.

Fear of heights was measured with a single question: “Do you suffer from vertigo?” (Leitenberg & Callahan, 1973). Participants could answer this question using a 6-point response format ranging from “no fear of heights” (coded with 0) to “very severe fear of heights” (coded with 5). This item was included in the survey that participants completed after the last session.

Table 1

Item difficulty (*δ*) and standard error of estimate (SE), mean square fit statistic (MS), and the probability of endorsement for an averagely susceptible person (P) in the simulated and real walking conditions

		difficulty		MS fit		P	
	Item	*δ*	SE	infit	outfit	simulated waling	real walking
1	I am sure that the two rooms were real and not a simulation.	3.48	0.52	1.06	0.84	.01	.04
2	I was sure that the ball that was thrown would hit me physically.	1.54	0.28	1.02	1.13	.09	.23
3	*When asked to fall into the pit, I did so without hesitation.*	0.77	0.24	0.84	0.71	.17	.39
4	I briefly thought that a real ball was thrown at me.	0.75	0.24	0.89	0.82	.18	.40
5	After the grate opened, it really felt as if I was falling.	0.60	0.23	0.98	0.92	.20	.44
6	I had the feeling that I needed to evade the ball.	0.42	0.23	1.06	1.01	.23	.48
7	There were moments during the session in which I forgot that I was in a simulation.	-0.01	0.23	0.88	0.93	.31	.59
8	I felt a bodily reaction when hitting the bottom of the pit.	-0.03	0.23	0.99	0.98	.32	.59
9	**Participant bent when passing the small door.**	-0.05	0.26	1.15	1.15	.32	.60
10	**Participant waited longer before opening the grate (above median reaction time).**	-0.18	0.23	1.20	1.27	.35	.63
11	I was sure that I could bump my head against the doorpost.	-0.18	0.26	0.89	0.80	.35	.63
12	**Participant showed a reflex when the ball was thrown.**	-0.33	0.22	1.12	1.11	.38	.66
13	At several moments during the session, I had the feeling that I could have extended my arm to touch the objects in the virtual environment.	-0.42	0.23	1.09	1.08	.41	.68
14	I constantly had the feeling that the objects I saw were situated in that specific location in the real world around me.	-0.58	0.23	1.17	1.22	.45	.72
15	When looking into the pit, I had the thought that I could really fall.	-0.74	0.23	0.93	0.93	.49	.75
16	I had the feeling that I needed to bend as not to bump my head against the doorpost.	-1.30	0.28	1.04	1.20	.62	.84
17	I felt a bodily reaction when looking down to the bottom of the pit.	-1.66	0.26	0.85	0.70	.70	.88
18	I felt somewhat uncomfortable when looking down to the bottom of the pit.	-2.09	0.28	0.84	0.63	.78	.92

Note: Items are translated from their original Dutch version. Items are ordered and numbered according to item difficulty. Items in **bold** are observations, rather than self-reports. Items in *italic* are negatively worded, and were reverse-coded before they were analyzed. P represents the probability of a positive response for an averagely susceptible person ($\theta = -0.23$).

Results

Item difficulties

Estimated item difficulties of the Rasch-based presence measure ranged from δ = -2.1 to 3.5 logits (M = 0.0; SD = 1.21; see Table 1). These item difficulties were estimated with a separation reliability of .95. All items fitted the Rasch model with mean square in- and outfit values of MS ≤ 1.27. An MS-value of 1.27 stands for a 27% excess in variation between the observed and expected responses. For assessing item fit, MS-values up to 1.20 are considered excellent (e.g., for high stake testing), MS-values between 1.20 and 1.30 as sufficient, and MS-values between 1.30 and 1.50 as acceptable (i.e., these do not degrade measurement; Wright & Linacre, 1994).

The fit statistics thus demonstrate that items could be ordered invariantly and transitively according to their difficulty, and thus in a manner that is similar across persons and experimental conditions. As an additional test of the latter, we estimated the interaction between experimental condition (e.g., simulated versus real walking) and the item difficulties. This interaction was not found to be statistically significant, with $\chi^2(36) = 35.3$ and $p = .50$. In other words, we could not reject the null-hypothesis that the item difficulties were the same for the two experimental conditions.

In addition, we estimated item difficulties separately for the real and simulated walking condition. The two sets of estimates were found to correlate with $r = .83$ and $p < .001$ (see Figure 3). After correction for measurement error attenuation (e.g., Charles 2005), the correlation became $r = .94$. Clearly, numerical differences between items with respect to their item difficulty were invariant across the two experimental conditions.

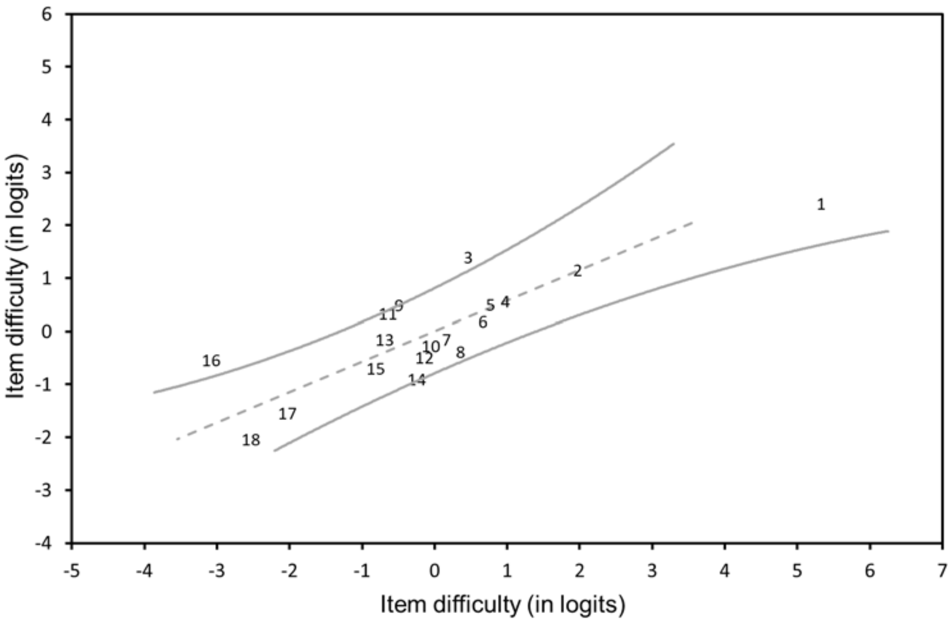

Figure 3. Item difficulties, and 95% confidence intervals, for the 18 items as estimated for the real walking (y-axis) and the simulated walking condition (x-axis) separately. The dotted line represents the line of equivalence.

As another, and more stringent test of the invariance of item difficulties, we estimated the difficulties of the 18 items separately for two extreme groups of participants: The 25 participants with the lowest estimated propensity to experience presence, and the 25 participants with the highest such propensity (i.e., Wright's challenge; see Bond & Fox, 2007). Again, the two sets of item difficulties correlated strongly with each other, with $r = .86$ and $p < .001$ (see Figure 4). After correction for measurement error attenuation, the correlation became $r = .98$. Numerical differences between items thus are independent, not only of the specific experimental conditions used, but of the propensity of the participants to experience presence as well.

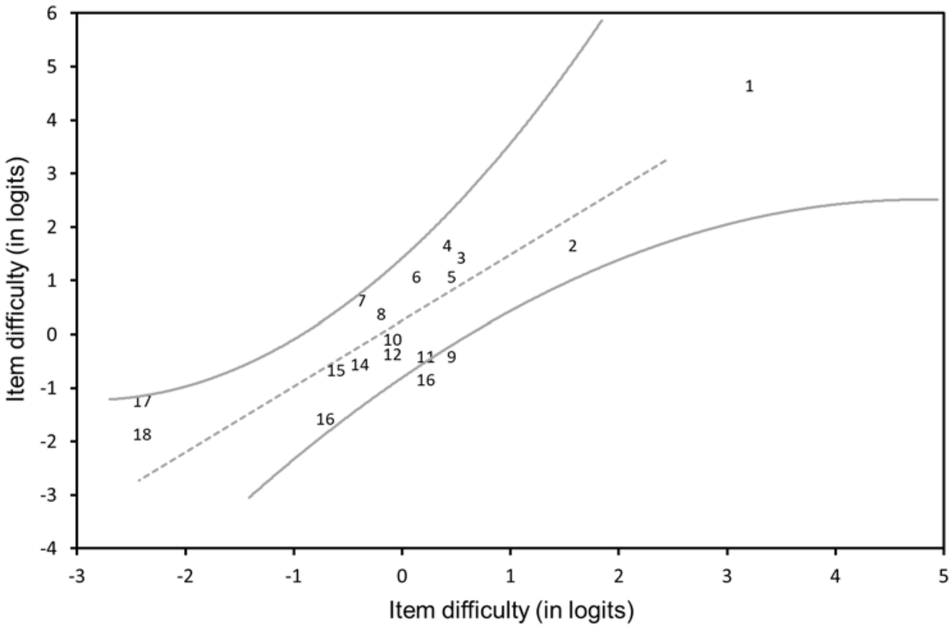

Figure 4. Item difficulties, and 95% confidence intervals, for the 18 items as estimated for participants with a low propensity (y-axis) and a high propensity to experience presence (x-axis) separately. The dotted line represents the line of equivalence.

Effect of simulated versus real walking

The estimates of the constraints imposed by the walking manipulation were $\lambda = 0.57$ logits (SE = 0.08) for simulated walking, and $\lambda = -0.57$ logits (SE = 0.08) for real walking. These constraints were estimated with a separation reliability of .98. Our participant's responses to these two conditions fitted the Rasch model prediction with means square in- and outfit values of MS ≤ 1.05. The difference in the constraint imposed by the simulated and real walking condition was statistically significant with $\chi^2(1) = 94.6$, and $p < .01$. As expected, simulated walking affected presence more negatively than real walking. For example, the estimated probability for an average person (with an individual propensity to experience presence of $\theta = -0.23$) to feel uncomfortable when looking down into the pit reduced from .92 in the real walking condition to .78 in the simulated walking condition (for changes in the probabilities of endorsement on all other items, see Table 1).

A similar effect of the walking manipulation was found on all main dimensions of the IPQ. Spatial presence was found to be higher in the real walking (M = 1.53; SE = 0.09) than in the simulated walking condition (M = 0.64; SE = 0.13), with $t(49) = 7.5$ and $p < .001$. Involvement was found to be higher in the real walking (M = 0.87; SE = .09) than in the simulated walking condition (M = 0.43; SE = 0.16), with $t(49) = 3.2$ and $p = .002$. Realism also was found to be higher in the real walking (M = -0.02; SE = .14) than in the simulated walking condition (M = -0.64 SE = 0.15), with $t(49) = 5.5$ and $p < .001$. Based on the Wilcoxon Singed Ranks Test, no statistically significant difference, with Z = -0.69 and $p = .492$, was found on the single item general presence dimension (M = 1.42 and SE = 0.16, and M = 1.24 and SE = 0.09 for the real and simulated walking condition respectively).

Individual differences in propensity to experience presence

Estimated individual differences in the propensity to experience presence as assessed with the Rasch-based method ranged from $\theta = -2.06$ to 2.06 logits (M = -0.23; SD = 0.96). These individual differences were estimated with a separation reliability of .81. Responses of 7 (14%) out of 50 participants did not sufficiently fit the Rasch model with means square in- and / or outfit values of MS > 1.30. For 4 (8%) of the participants, the responses did not match the model predictions with outfit MS > 1.50. Given that more misfit is expected to occur with persons as compared to items and experimental conditions, these results are nonetheless acceptable.

To compare individual differences on the Rasch-based measure with those obtained with the IPQ instrument, we averaged, for each person, the responses to the IPQ items across the two experimental conditions; thus creating a single score for each person on each of the four IPQ dimensions (i.e., spatial presence, involvement, realism, general presence). After correction for measurement error attenuation, we found medium to large correlations between our Rasch-based estimate and all of the IPQ-dimensions, except for the single item general presence dimension, with $r \geq .44$ (see Table 2). The correlation with the single item general presence dimension was not found to be statistically significant with *rho* = -.02 and $p = .91$. Clearly, individual differences obtained with the Rasch-based instrument overlapped with all main dimensions of the IPQ.

Table 2

Item difficulty (δ) and standard error of estimate (SE), mean square fit statistic (MS), and the probability of endorsement for an averagely susceptible person (P) in the simulated and real walking conditions

		IPQ dimensions			
	Rasch	spatial presence	involvement	realism	general presence
Rasch	**.81**	.43**	.38**	.55**	-.06
spatial presence	.54	**.79**	.53**	.51**	.21
involvement	.46	.64	**.86**	.34*	-.12
realism	.66	.62	.40	**.86**	.12

Note: The diagonal shows reliabilities (in **bold**). Reliabilities are Cronbach's *α* for the IPQ dimensions and the separation reliability for the Rasch-based estimate. Pearson correlations are shown above the diagonal, expect for correlations involving general presence which are Spearman's *rho*. Correlations corrected for measurement error attenuation are shown below the diagonal. Being a single item dimension, no reliability and no corrected correlations are reported for general presence. *$p < .05$. **$p < .01$.

Given the nature of the virtual pit environment, we anticipated that the experience of presence would be correlated with self-reported fear of heights. Indeed, we found a small, and statistically significant, correlation between fear of heights and the Rasch-based individual differences in the propensity to experience presence, with *rho* = .29, and *p* = .04. A similar correlation of *rho* = .25 was found with the involvement dimension of the IPQ, although it did not reach statistical significance with *p* = .08. Apparently, people with higher self-reported fear of heights had higher levels of presence and involvement in our pit environment.

Discussion

In our experiment, we successfully predicted 18 perceptual, visceral, cognitive, and behavioral responses to a virtual pit environment based on estimates of a person's predisposition to experience presence, the constraints imposed by the mode of navigation (real versus simulated walking), and the difficulty of the specific presence indicator. Our model test confirmed our expectation that the various responses to the virtual environment differed in their difficulty of being endorsed. Moreover, the order of the various responses by and large matched our expectations. For example, perceptual (e.g., perceiving the objects in the virtual environment as existing in the external world) and visceral responses (i.e., feeling uncomfortable when looking into the pit) were much easier to endorse, than letting oneself fall into the pit without hesitation (although all participant eventually did so). As expected, cognitive responses, such as knowing for sure that the two rooms were real and not a simulation, were amongst the hardest to endorse.

Nevertheless, there were also responses that were easier or more difficult than expected. For example, participants with an average predisposition to experience presence had only a probability of 66% (in the real walking condition) of showing a minimal reflex to the ball. This and participants' responses to the related self-reports indicated that perhaps the ball throw was not realistically simulated. Similarly, we did not expect any of the participants to confirm being sure about the two rooms not being a simulation. Still, three participants responded affirmative to that item. Most important, however, is that the differences between items, with respect to their item difficulty, was found to be invariant across persons and experimental conditions (as indicated by excellent item fit statistics, and by comparing item difficulties estimated on various subsets of the data; see Figures 3 and 4). As result, the 18 indicators allowed for a meaningful comparison of participants and experimental conditions.

As expected, and in line with existing research findings (e.g., Slater et al, 1995; Usoh et al., 1999), we found that simulated walking by means of the controller resulted in less presence than walking by means of natural locomotion. Interestingly, this effect of walking mode was found, not just on the Rasch-based instrument, but on all three of the main dimensions of the IPQ: Spatial presence, involvement, and realism. Although more research is needed to substantiate this, we believe that the difference between real and simulated walking is due to the former allowing for more sensorimotor contingencies to be enacted.

In contrast to items and experimental conditions, responses of individual participants did not always fit the model’s predictions sufficiently. Although more misfit is expected with persons than with items, because of individual idiosyncrasies, the relatively high number of MS-values above 1.30 deserves further investigation. First, it may reflect differences in response styles. For example, the three participants with outfit MS > 1.50 all responded positive to the question of being sure that the two rooms were real and not a simulation. Second, some of the sessions were disrupted by hardware problems or the participant not having the controller with them. Such idiosyncratic disturbances may have constrained these participants in their presence at specific times during the session (e.g., when asked to open the grate), as a result of which their responses to the respective items (e.g., opening the grate without hesitation) may have been different than anticipated by the model.

Our experiment thus provides first evidence for the validity and reliability of using the Rasch model to construct a unidimensional presence scale based on a wide variety of perceptual, visceral, cognitive and behavioral responses toward the virtual environment as indicators. As such, our research illustrates that both self-reports and behavioral observations—as long as reflecting clearly that a person did or did not fall for the illusion of non-mediation—are valid indicators of presence. Although future research is needed for confirmation, it seems likely that other such indicators, including physiological responses, can be scaled in a similar fashion. The challenge here is to find indicators that reflect unmistakably that some aspect of the organism has accepted or refrained from accepting the virtual content as being real. This may be relatively easy for a pit environment—no sane person would voluntarily step into a 12 meters deep pit—, but what behavioral responses, or the absence thereof, would indicate, for example, that a virtual fire in the home was responded to as being real (see Spanlang, Fröhlich, Descalzo, Antley, & Slater, 2007). Our results also confirm that it is possible to construct a unidimensional presence scale. Although additional analysis is required to confirm whether perhaps a second dimension can be constructed from the data, the moderate to high correlations with all of the three main dimension of the IPQ (spatial presence, involvement, and realism) speak in favor of a unidimensional concept of presence.

There were several noteworthy limitations. First, we have tested the measurement method on a single, specific type of virtual environment: A virtual pit environment. This environment allowed us to observe the required range of perceptual, visceral, cognitive, and behavioral responses, but more research is needed to demonstrate the effectiveness of the method with other, perhaps less engaging, virtual environments.

A second limitation is that the proposed method, for now, has been designed for measuring presence in interactive virtual environments. More research is needed to investigate how the method may be extended for use with non-interactive media such as televisions or books.

A final limitation of the proposed measurement method is that choice of indicators depends on the specific virtual environment in which presence is measured. It would, for example, make little sense to ask about one’s experiences toward a virtual pit, when no such pit is present in the simulated scene. This may sound problematic, but only because, in psychology, measures are typically defined by the particular indicators used (Kaiser, Merten, & Wetzel, 2018). This has resulted in a quest for the gold standard of presence measurement preferably in the form of an instrument that is applicable across multiple situations and populations without any alterations to the indicators; alterations that would affect the concept being measured. However, even with instruments designed to be applicable across a range of media technologies, such as the ITC-Sense of Presence Inventory (ICT-Sopi; Lessiter, Freeman, Keogh, & Davidoff, 2001), it often remains necessary to drop items (e.g., when avatars are not present in the simulated world). The proposed Rasch-based method can be used for making comparisons even between different types of virtual environments (e.g., a pit environment with a virtual class room) provided that there is some overlap in the items used to enable scale calibration and test equating. More work is needed in order to develop such general items, but several of the self-report items included in the present experiment are promising candidates.

From a strict measurement perspective, however, estimates of presence should be independent of the specific set of items used (a principle called specific objectivity; see, e.g., Bond & Fox, 2007; Rasch, 1960). In other words, the difference, or numerical relation between two persons should remain the same (i.e., invariant) regardless of the specific set of indicators used to measure presence. It has been argued that demonstrating specific objectivity provides the most convincing evidence for the validity of the measurement of a psychological variable (also Kaiser, Merten, & Wetzel, 2018); the most convincing evidence, even, for the existence or reality of a latent psychological variable such as presence (cf. Slater, 2004). By being implemented within the Rasch family of models, the proposed measurement method allows for testing specific objectivity against the empirical data. The present contribution is but a first step in that direction, and more research is needed to confirm whether presence meets this criterion. We have demonstrated that item difficulties are independent of the specifics of the experimental condition (i.e., the means of navigating through the environment) and of the participants’ level of propensity to experience presence. Additional research, however, is needed to demonstrate that estimated differences between experimental conditions and between individuals are, in turn, indeed independent of the specific set of indicators used. To do so requires a larger set of items than included in the present study, so that differences between persons and / or experimental conditions can be estimated with sufficient reliability even when subsets of indicators are used.

Despite these limitations, our experiment provides a first and successful test of the proposed Rasch-based measurement of presence. Our results demonstrate that presence indicators are invariant across persons and experimental conditions. As a result, we have demonstrated that it is possible to construct a unidimensional presence scale from a wide variety of perceptual, visceral, cognitive, and behavioral responses toward the virtual environment as indicators. As such, it provides a promising method for exploring the origins behind individual differences in the propensity to experience presence, for testing the effect of manipulations of the virtual environment and / or the mediating technology, and for uncovering the nature of presence itself.

Acknowledgements

We thank Peter Frumau for his assistance in the data collection, and Sima Ipakchian Askari for her assistance in video coding. We also gratefully acknowledge Aart van der Spank for providing technical support.

References

Biocca, F. (1997). The cyborg's dilemma: Progressive embodiment in virtual environments. *Journal of Computer-Mediated Communications, 3*(2).

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah, NJ: Erlbaum.

Charles, E. P. (2005). The correction for attenuation due to measurement error: Clarifying concepts and creating confidence sets. *Psychological Methods, 10,* 206–226.

Coxon, M., Kelly, N., & Page, S. (2016). Individual differences in virtual reality: Are spatial presence and spatial ability linked? *Virtual Reality, 20*(4), 203–212.

Dennett, D. C. (1996). *Kinds of minds: The origins of consciousness.* New York: Basic Books.

Gorsuch, R. L. (1997). Exploratory factor analysis: Its role in item analysis. *Journal of Personality Assessment, 68*, 532–560.

Haans, A. (2014). In search of the fixed points on the presence scale. In A. Felnhofer, & O. D. Kothgassner (Eds.), *Challenging Presence: Proceedings of the International Society for Presence Research 2014* (pp. 17-21). Vienna, Austria: Facultas.wuv.

Haans, A., & IJsselsteijn, W. A. (2012). Embodiment and telepresence: Toward a comprehensive theoretical framework. *Interacting with Computers, 24*, 211-218.

Haans, A., Kaiser, F. G., Bouwhuis, D. G. & IJsselsteijn, W. A. (2012). Individual differences in the rubber-hand illusion: Predicting self-reports of people's personal experiences. *Acta Psychologica, 141,* 169-177.

IJsselsteijn, W. A. (2005). Towards a neuropsychological basis of presence. *Annual Review of CyberTherapy and Telemedicine, 3,* 25-30.

IJsselsteijn, W. A., de Kort, Y. A. W., & Haans, A. (2006). Is this my hand I see before me? The rubber hand illusion in reality, virtual reality, and mixed reality. *Presence: Teleoperators and Virtual Environments, 15,* 455–464.

Kaiser, F. G., Merten, M., & Wetzel, E. (2018). How do we know we are measuring environmental attitude? Specific objectivity as the formal validation criterion for measures of latent attributes. *Journal of Environmental Psychology, 55,* 139-146.

Laarni, J., Ravaja, N., Saari, T., Böcking, S., Hartmann, T., & Schramm, H. (2015). Ways to measure spatial presence: Review and future directions. In: M. Lombard, F. Biocca, J. Freeman, W. IJsselsteijn, & R. J. Schaevitz (Eds.), *Immersed in Media: Telepresence Theory, Measurement & Technology* (pp. 139-185). Cham, Switzerland: Springer.

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.

Linacre, J. M. (2002). Construction of measures from many-facet data. *Journal of Applied Measurement, 3,* 486-512.

Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. J*ournal of Computer-mediated Communication, 3*(2).

Lombard, M., & Jones, M. T. (2015). Defining presence In: M. Lombard, F. Biocca, J. Freeman, W. IJsselsteijn, & R. J. Schaevitz (Eds.), *Immersed in Media: Telepresence Theory, Measurement & Technology* (pp. 13-34). Cham, Switzerland: Springer.

Loomis, J. M. (1992). Distal attribution and presence. *Presence: Teleoperators and Virtual Environments, 1,* 113– 119.

Meehan, M., Razzaque, S., Insko, B., Whitton, M., & Brooks Jr, F. P. (2005). Review of four studies on the use of physiological reaction as a measure of presence in stressful virtual environments. *Applied Psychophysiology and Biofeedback, 30*(3), 239-258.

Michell, J. (1999). *Measurement in psychology: A critical history of a methodological concept.* Cambridge: Cambridge University Press.

Minsky, M. (1980). Telepresence. *Omni, 21,* 45–51.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Copenhagen, Denmark: Danish National Institute for Educational Research.

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.

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

Slater, M. (2004). How colorful was your day? Why questionnaires cannot assess presence in virtual environments. *Presence: Teleoperators and Virtual Environments, 13*(4), 484-493.

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1535), 3549–3557.

Slater, M., Lotto, B., Arnold, M. M., & Sanchez-Vives, M. V. (2009). How we experience immersive virtual environments: the concept of presence and its measurement. *Anuario de Psicología, 40*(2), 193-210.

Slater, M., Usoh, M., & Steed, A. (1995). Taking steps: The influence of a walking technique on presence in virtual reality. *ACM Transactions on Computer-Human Interaction, 2*(3), 201-219.

Spanlang, B., Fröhlich, T., Descalzo, V. F., Antley, A., Slater, M. (2007). The making of a presence experiment: Responses to virtual fire. In L. Moreno & Starlab Barcelona S. L. (Eds.), P*roceedings of the 10th Annual International Workshop on Presence* (pp. 303-307). Barcelona, Spain: Starlab Barcelona S.

Steuer, J. (1992). Defining virtual reality: dimensions determining telepresence. *Journal of Communication, 42*(4), 73–93.

Usoh, M., Arthur, K., Whitton, M. C., Bastos, R., Steed, A., Slater, M. & Brooks Jr., F. P. (1999). Walking > walking-in-place > flying, in virtual environments. In P*roceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques* (pp. 359–364). New York: ACM Press.

Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions, 8,* 370.