MASTER

The scaling of presence in virtual reality

Frumau, P.F.W.

Award date:
2016

Link to publication
The Scaling of Presence in Virtual Reality

by Peter Frumau

Identity number 0871047

in partial fulfilment of the requirements for the degree of

Master of Science
in Human-Technology Interaction

Department of Industrial Engineering & Innovation Sciences
Eindhoven University of Technology (TU/e)

Supervisors:
Dr. Ir. A. (Antal) Haans  IE & IS, TU/e
Prof. Dr. W.A. (Wijnand) IJsselsteijn  IE & IS, TU/e
Keywords: Media technology, virtual reality, presence, and measurement.
Abstract

Media technologies allow people to experience an environment that might only exist in the medium itself. Virtual reality is such a technology, allowing people to actually experience and walk around in a virtual environment. When immersed in virtual reality, a user can experience a sense of “being there”, a phenomenon commonly referred to as presence (Sheridan, 1992). Presence is an interesting phenomenon to research, since it may influence how people behave and react to the media technology. However, defining presence is problematic, since the concept of presence is such a fuzzy one. Also, measuring presence has proven to be problematic. Although the general consensus is that presence can actually be measured (i.e. suggesting an intermediate states theory of presence; e.g. Witmer and Singer, 1998), there is no clear consensus about what method is best at measuring presence.

Firstly, we aim to address the problem in defining presence. Because this study will focus on presence in media technologies, this paper uses a definition based on the notion of transparency of those technologies, i.e. that what occurs when a user forgets that a technology is being used to create the media that is perceived (Haans & IJsselsteijn, 2012). Therefore the following definition of presence is used throughout this paper: “the perceptual illusion of non-mediation through transparency of media technology”, specifically in relation to immersive and interactive virtual reality (also in Haans, 2014). The use of this definition creates possibilities in the measurement of presence. When presence is achieved in a user of virtual reality, reactions of that user to the virtual environment are expected to be similar to reactions to the real world. Plenty of subjective measures (e.g. presence questionnaires) have been proposed in order to capture presence through self-reports. However, subjective measures introduce problems such as invariant interpretations of questionnaire items. Instead, objective methods have been proposed as alternatives such as behavioral observations and measurement of visceral reactions. Attempts have been made to compare these different methods. However, different presence measures (e.g. self-reports and behavioral observations) hardly correlate empirically, suggesting none fully captures what is means to be present in a virtual environment (e.g. Haans, 2014; Slater et al., 2009). We argue that most methods proposed might still be valid ways of measuring presence. However, in order to capture the full range of values on a presence scale (assuming a single dimension in presence), it is important to not compare, but to combine these different measures.

Hence in the current study, various cognitive, visceral and behavioral responses were combined into a presence scale using Rasch-type models (e.g. Linacre, 2002). The Rasch model assumes a trade-off between the items' difficulties and the respondents’ ability to answer to certain items. In addition, this model can be extended to include an effect of context. To find out whether the Rasch model is indeed the key to a reliable, sensitive, and valid method for scaling presence, a virtual reality experiment was conducted. Based on an earlier study by Haans (2014), a virtual environment was designed in which participants were requested to engage in various activities as to proceed to a second room through a small door, watch an avatar throw a ball (in their face), look down into a deep pit, step onto a grille that covered the pit, and finally to press a button on a controller that made them (visually) experience a fall in the pit. A total of fifty participants completed the experiment, that visited the virtual environment twice (completing the same tasks each session). A different navigation style was experienced by participants in each session: Real walking (i.e. tracked in the real environment) or indirect walking (i.e. using a controller to navigate). After each session, they completed a short questionnaire (16 items created for the purposes of this experiment) that asked about their experiences during the session (e.g., with respect to the thrown ball, the falling experience, or the extent to which they felt they would be able to pick up objects present in the virtual room). During the sessions the participants were observed for automatic behaviors (e.g. retreat head for the ball, bending knees when hitting the bottom of the virtual pit) or latencies in volitional behaviors (e.g. when hesitating to jump). Additionally, an existing questionnaire that attempts to measure presence, the Igroup Presence Questionnaire (IPQ; Schubert, Friedmann & Regenbrecht,
1999), was completed by participants each session, used to cross-validate our own. In addition, an open question allowed for a more qualitative analysis of participants’ experiences. Finally, demographics questions (age, gender and length), one question addressing fear of heights (obtained from Leitenberg & Callahan, 1973), and one question addressing experience with VR were completed by participants. The self-reports and behavioral observations obtained in the experiment were tested against the Rasch model for invariant ordering of items, ability to distinguish between participants, and ability to detect differences in the immersiveness of the experience.

The proposed many-facets Rasch model turned out to be functional in its capability to reliably predict the difficulty of items (of both self-reports and behavioral observations) and reliably assess a person’s susceptibility to presence. Additionally, it could reliably differentiate between a different navigation style (where real walking elicited a higher sense of presence in participants compared to indirect walking). Moreover, significant (small to moderate) correlations were found between the proposed presence scale and three out of four different subscales of the IPQ. The lack of high correlation might have resulted from a difference in definition of presence and how questions were formulated across instruments. However, even though the correlations were not that high, they were still interpreted as support for convergent validity, because various presence-related responses are not expected to correlate empirically (e.g. Haans, 2014; Slater et al., 2009). It is the way the different methods are combined, not compared, that make up the most reliable presence score. The IPQ might only be able to distinguish a small range of presence scores. Because of the full range of cognitions, behaviors and visceral responses that are captured by our Rasch-based presence measure, this measure might more fully capture what it means to be present in a virtual environment.

Despite the high reliability of the scale’s ability to differentiate between participants, a high percentage of participants (18%) did not fit the model well. Several limitations could be causing these aberrant results, including a potential error in filling in the questionnaire, or a misinterpretation of an item in the questionnaire. However, a more likely explanation is found in the research design. Since we tested participants twice, assumptions regarding invariance of parameter values across sessions could have been wrong. By treating presence as a trait in the many-facets Rasch model, we did not allow changes in person susceptibility to presence across sessions or conditions. However, it may likely be the case that susceptibility to presence is very dependent on the moment (i.e. invariant across sessions or conditions). For example, attention might have been focused on different aspects of the virtual environment in the second session (e.g. due to expectation of events, knowledge of the objects in the environment or memory of what was said during audio instructions). This is a limitation of the current study that should be addressed in future research. Perhaps a different Rasch-based model should be used, one that can account for invariant person scores across sessions.

Other limitations that pertained to the virtual reality setup might have influenced results, such as the absence of a virtual body, the task design (e.g. the height of the door was independent of participant length) or length of audio instructions. Despite these limitations, due to the model’s capabilities (i.e. reliably predict the difficulty of items, reliably assess a person’s susceptibility to presence, and the ability to detect differences in the immersiveness of the media), we argue that instead of comparing different presence measures through correlation, future research should focus on combining different methods that attempt to capture presence. Applying the proposed method in presence research regarding media technologies, will create a better understanding of what the phenomenon of presence is, and of the way people react to media technologies. The Rasch model, if adapted to the likely invariance of person susceptibility to presence, will allow for a presence scale that captures the full range of intermediate states in presence.
# Contents

1. Introduction ................................................................................................................................. 1  
   1.1. Presence .................................................................................................................................. 2  
   1.2. Measuring presence .................................................................................................................... 3  
   1.3. The Rasch model ...................................................................................................................... 4  
   1.4. Research aims ........................................................................................................................... 7  

2. Methods ........................................................................................................................................... 8  
   2.1. Participants ............................................................................................................................... 8  
   2.2. Research design .......................................................................................................................... 8  
   2.3. Virtual environment design ...................................................................................................... 8  
   2.4. Apparatus .................................................................................................................................. 9  
   2.5. Procedure ................................................................................................................................... 9  
   2.6. Measures .................................................................................................................................. 10  

3. Results ............................................................................................................................................ 13  
   3.1. Model test .................................................................................................................................. 13  
      3.1.1. Item difficulties ................................................................................................................... 14  
      3.1.2. Person scores ...................................................................................................................... 15  
      3.1.3. Effect of navigation style .................................................................................................. 16  
   3.2. Convergent and discriminative validity ..................................................................................... 17  
      3.2.1. Correlation with IPQ instrument ......................................................................................... 17  
      3.2.2. Correlation with Fear of heights ....................................................................................... 18  

4. Discussion ....................................................................................................................................... 18  
   4.1. Quality of the instrument .......................................................................................................... 18  
   4.2. Merits to differentiate individuals ............................................................................................ 19  
   4.3. What is being measured (convergent and discriminative validity) ........................................... 19  
   4.4. Detection of differences in the immersiveness of the media ....................................................... 20  
   4.5. Limitations ................................................................................................................................ 21  
   4.6. Conclusion ................................................................................................................................ 22  

References ......................................................................................................................................... 23
A. Informed consent form ........................................................................................................................... 25
B. Audio instructions.................................................................................................................................... 27
C. Questionnaires ......................................................................................................................................... 28
   C.1. Main questionnaire ........................................................................................................................... 28
   C.2. Additional questions .......................................................................................................................... 32
D. Specifications Regarding the Virtual Reality System ........................................................................... 34
E. Code of Scientific Conduct ....................................................................................................................... 35
1. Introduction

Creation of media has always been intriguing to humans. From photography to movies, media technologies have the capability to show its users a world that only exist in the medium itself. In a classical movie, a director shows some story from a certain point of view. However, the director has all control over what environment and story is shown to the user. Some more recent media technologies tried passing over control to the user through virtual environments (VEs). A VE can be defined as a “digital space in which a user’s movements are tracked and his or her surroundings rendered or digitally composed and displayed to the senses, in accordance with those movements” (Fox et al., 2009, p95). Frequently used in video games, VEs introduced a large increase in interactivity with computer generated content, because now users can control what is shown and what happens in the VE. Immersive Virtual Environments (IVEs) brought the quality of interactivity to the next level. In IVEs, users are immersed in the environment by wearing some form of computer equipment. Immersion can be defined as losing oneself in the digital environment and shutting out cues from the physical world (Fox et al., 2009; Witmer & Singer, 1998). In Virtual Reality (VR) for example, users can wear a Head-Mounted Display (HMD) that displays the computer generated content in front of the user’s eyes, excluding the real world from vision. This allows the user not only to control what happens and what is shown in the VE, but to exert this control through a first person view. In other words, the user now actually is the main character of the story. Virtual reality is often praised for its capability to create an experience that would not have been possible or feasible without the qualities of this technology. Consequently, VR is a useful tool in conducting research. Experimental settings and stimuli can be replicated exactly, whilst retaining a large degree of control the parameters.

Application areas for VR are for example exposure therapy, cognitive behavioral therapy, physical rehabilitation, training environment and networked collaboration (Fox et al., 2009). One strong capability of VR that makes these applications possible is the feeling that a user can get of “being there” in the virtual environment, a concept known as presence (Sheridan, 1992). Presence in a mediated or virtual environment is often defined as “the subjective experience of being in one place or environment, even when one is physically situated in another” (Witmer & Singer, 1998, p. 225). However, these definitions are only two of many used in the existing literature which explicates the complexity of the phenomenon itself. Furthermore, measuring presence has proven to be problematic. Not only numerous different questionnaires exist, they often ask about experiences that are very abstract or ambiguous (e.g. “In the computer generated world I had a sense of “being there””, an item in the SUS presence questionnaire; Slater, Usoh & Steed, 1994). If scientists cannot even agree on a sole definition of presence, how can they expect participants in VR studies to understand presence the same way they do? Therefore other methods, such as behavioral observation (e.g. Fox et al., 2009), and visceral measurements (e.g. Meehan et al., 2005) have been used in an attempt to measure presence. However, what method is ‘best’ at measuring presence remains an unanswered question.

In the upcoming sections, firstly the scope of this research is delimited, by giving a definition of presence that is focused on dealing with interactive immersive media. Thereafter, current methods that attempt to measure presence are discussed and criticized. Finally, a proposal for a method that creates the possibility to form a single, unidimensional measure of presence is presented, accompanied by the current research aim.
1.1. Presence

Because of the problems that arise in defining presence, the aim of the current study is not to find the best definition of presence possible, but to give a definition of presence that is usable for technologies that simulate mediated media. The focus will therefore be placed on a more comprehensive definition of presence, proposed by Lombard and Ditton (1997): presence as “the perceptual illusion of non-mediation”. An example of such an illusion is a person wearing glasses. Although these glasses are not part of the person’s biological body, after continued use they can become “transparent” (i.e. a functional extension of the body; Haans & IJsselsteijn, 2012). That person basically forgets the glasses are mediating what his or her eyes receive, and accepts that what is seen is an unmediated view on the world. Presence related to media technologies could be defined as that what occurs when a user forgets that a technology is being used to create the media that is perceived, or when a media technology becomes a functional or phenomenological extension of the body (Haans & IJsselsteijn, 2012). As a consequence, the user is left to perceive and act in the mediated or simulated environment as if he or she were physically there (IJsselsteijn, 2005). Because this study will focus on presence in media technologies, the definition for presence used throughout this paper is based on exactly that notion: presence as “the perceptual illusion of non-mediation through transparency of media technology”, specifically in relation to immersive and interactive virtual reality (also in Haans, 2014). In the current study, presence will be mostly related to the experience that virtual objects are experienced as actual physical objects, a concept known as spatial presence (Lee, 2004). In addition to spatial presence, Lee (2004) subdivided the concept of presence into two more types: self-presence and social presence. Self-presence is achieved when the virtual self is experienced as the actual self, and social presence is achieved when other virtual social actors are experienced as actual social actors. For the sake of not complicating things even further, this study focusses on spatial presence and not on self or social presence.

According to Slater and colleagues (2009), three things are required for spatial presence: a consistent low latency sensorimotor loop between sensory data and proprioception, statistical plausibility (e.g. natural laws such as gravity should be consistent for objects in the virtual environment), and behavior-response correlations that are in line with real world experiences. Vision and movement are an essential aspect of these behavior-response correlations in VR. When a user moves his or her head around, the virtual environment (and the position of the objects within it) should realistically change and shown to the user accordingly. When the virtual world is experienced in the same way that the real world is experienced, presence can be achieved. As a result, the user should act in the virtual environment just like he or she would in the real world. Unfortunately, due to the fact that media technologies are simply not powerful enough (yet) to immerse a user in a virtual world that is indistinguishable from the real world, it is hard to fool that user into thinking that virtual world is real. However, it might be possible to fool some processes of the mind into treating the virtual environment as if it were real.

Behavioral and visceral reactions are operated largely outside conscious awareness, and could therefore be easier to fool than the conscious mind. The body schema is a distributed network of procedures that is aimed at guiding behavior (Haans & IJsselsteijn, 2012). It governs proprioception and guides automatic tasks such as maintaining balance, selecting actions, and controlling the muscles. It also helps the body to use functional extensions (e.g. a pen) with little conscious effort. Due to this unconscious nature, the body schema might produce some behavioral reactions that are not in line with higher-order cognitive beliefs. For example, while persons in a VE might not walk through walls, they will not necessarily belief that the virtual environment was actually a real environment and not a simulation. The same applies to visceral reactions, or “gut feelings”. Visceral reactions like fear are often generated unconsciously and automatically, serving as a protection mechanism (Rosen & Schulkin, 1998). For example, a picture of a spider can induce fear, while cognitively the observant should know it is not a real spider (i.e. cannot harm him or her). The same principles apply to the experience of presence in virtual reality. Many people will
experience a feeling of unease when looking down from a great height into a depth (i.e. fear of height). So where most people are expected to experience a feeling of unease when they look down into a virtual pit (i.e. an automatic visceral reaction), statements such as the belief that the virtual environment was actually a real one (i.e. conscious cognitive reaction), are harder to agree to, since this requires a person to be fooled into the conscious belief that there would be no technology that is mediating his or her experience of the virtual content.

1.2. Measuring presence

There has been a lot of debate on the topic of measuring presence. One could argue that there are no measurable differences in presence levels at all. Accordingly, Lombard and Ditton (1997) conveyed opinions that presence is an all-or-nothing phenomenon: One feels either present in the real world, or present in the virtual world. If this is true, differences in presence levels could still be distinguished as the time someone felt present in the virtual environment, or in terms of breaks-in-presence (BIP): the amount of time someone was no longer present in the virtual environment (Slater & Steed, 2000). Another theory on presence claims there are multiple levels of presence possible within the same person, arguing for the existence of intermediate state(s) in presence, a theory held by Witmer and Singer (1998). Which theory (i.e. all-or-nothing or intermediate states) is accepted in research may likely depend on the way ‘presence’ is defined and measured. However, the amount of presence instruments (e.g. presence questionnaires) that have been proposed suggests a general belief in the theory of intermediate states in presence. Past and current presence research tends to only ask a handful of questions about the participants’ experience of “being there” (e.g. “In the computer generated world I had a sense of "being there" “, an item in the SUS Presence Questionnaire; Slater, Usoh & Steed, 1994). These questions are not only insufficient in number, they are also very hard for a participant to understand. Since feeling present (i.e. “being there”) is a very subjective notion, measuring presence using questionnaires introduces problems on its own due to the way respondents interpret the questions (Slater, 2004). Consequently, there is little agreement over what questionnaire is best at measuring presence. Some of the many different questionnaires that are being used in presence research include the Igroup Presence Questionnaire (IPQ; Schubert, Friedmann & Regenbrecht, 1999), the Body Sensations Questionnaire (BSQ; Chambless et al., 1984), the Witmer & Singer Presence Questionnaire (WS; Witmer & Singer, 1998), and the Slater Usoh Steed Presence Questionnaire (SUS; Slater, Usoh & Steed, 1994). Other questionnaires used are often composed of several subjective statements that ask about presence, adjusted to the experiment at hand (e.g. Yee & Bailenson, 2007; Nowak & Biocca, 2003).

In an attempt to avoid the need for introspection from the user, alternative measures of presence pay more attention to behavioral and visceral reactions to the virtual environment. For example, Fox, Bailenson and Binney (2009) used users’ remembered details or content as a measure for presence, expending the knowledge that when a user is fully engaged in a virtual environment, he or she will remember less specific details or content (Nichols, Haldane, & Wilson, 2000; Bailey et al., 2012). Behavioral data was also used by Fox and colleagues (2009) in addition to cognitive evaluation. They observed participants’ imitation of a virtual avatar as a predictor for presence, utilizing behavioral transfer from a virtual setting to the real world. Others used behavioral observation in presence research. For example, Gonzalez-Franco and colleagues (2010) observed the reaction of users when a virtual ceiling fan approached their heads, and Usoh and colleagues (1999) observed the path a user walked (where a path taken over a virtual pit - floating in the air - was associated with a lower sense of presence than one along a virtual floor). Visceral reactions to a virtual environment have also been utilized. For example, Meehan and colleagues (2005) conducted four studies on the use of visceral reactions as an indication of presence and found that hearth rate (to a high extend) and skin conductance (to a lesser extend) satisfied the requirements for a measure of presence.
Although these methods independently look promising, when compared against each other the various presence-related responses hardly correlate empirically (Haans, 2014). Indeed, Slater and colleagues (2009) did a study that compared subjective reports with physiological measurements regarding presence. They found that subjective reports by participants did not match their physiological responses to a virtual character that walked by. In line with this result, another study showed that even though people said they did not feel present in the virtual environment, behavioral and visceral responses show they did respond to the virtual world like they would in the real world (Haans, 2014). This could mean one of the methods measures a different dimension of presence, but another interpretation could be that different methods all measure a different part (not dimension) of presence. An analogy with a temperature scale explains this second interpretation further. Say there are two thermometers: One can only measure temperatures between 0 and 20 degrees Celsius, while the other can only measure between 20-40 degrees Celsius. This would mean that when a room’s temperature varies between 5 and 15 degrees Celsius, the first thermometer can detect differences, while the second thermometer cannot. As a result, these two temperature scales will correlate insignificantly (for measures within that room), although they both measure temperature. The same could apply to different presence measurements. For example, differences in physiological responses might be observable in participants with a low sense of presence. Acknowledgement of feeling present though, might a much higher sense of presence to be observed. In other words, even though some cognitive processes do not show signs of presence, some of the body schematic procedures might. When a participant answers a question in a presence questionnaire (e.g. “I had a sense of “being there” in the virtual world”), the response may likely only refer to the conscious cognitive processes in which that person evaluated the virtual world, therefore it might not capture some of the more “easy” reactions to the virtual environment (e.g. visceral and behavioral reactions). If the goal is to scale presence, not merely cognitive processes should be measured, since these might simply be insufficient to scale such a complex phenomenon (also Slater et al., 2009).

In line with these thoughts, the current study argues that correlation is not the way the different presence measurement attempts should be compared. All measures can be valid; it is the way the different methods are combined, not compared, that make up the most reliable presence score (e.g. Haans, 2014). Therefore, in the present study, no single method (e.g. behavioral observation, self-reports, or physiological measurement) is hypothesized to be the best one. Cognitive, behavioral and visceral responses to the computer-generated content are different observations that all occur with presence and could all be valid indicators of presence. When the measurement of presence does not only take the conscious experience of “being there” into account, but also the smaller behaviors and visceral responses that build up the intermediate states of being present, the outcome of that measurement should be a better indication of presence.

1.3. The Rasch model

The Rasch model (Bond & Fox, 2007) is a mathematical model that is used for analyzing ordinal data, such as categorical or dichotomous (i.e. yes-no) answers to a questionnaire. This model assumes a trade-off between the items’ difficulties and the respondents’ ability to answer to certain items:

\[
\ln \left( \frac{p(x_{nic} = 1)}{1 - p(x_{nic} = 1)} \right) = \theta_n - \delta_i
\]

(1)

The Rasch model describes the natural logarithm of the odds that a person \( n \) shows a certain reaction to an item \( i \) (i.e. a behavior, cognition, etc.) through the addition of that person \( n \)'s ability \( (\theta_n) \) and the difficulty \( (\delta_i) \) assumed behind that item \( i \). Both parameters in this equation \( (\theta_n, \delta_i) \) are estimated with maximum-likelihood statistics and expressed in log odd units (logits). As a result, different items can be
invariantly ordered based on their difficulty. The pattern of a person’s responses to the items is subsequently matched, with a certain probability, to that person’s ability ($\theta$).

The logic behind this invariant ordering of items can be explained through the items of a math test (see Figure 1). The more difficult the item is, the less likely that participant is to answer that item correctly. However, a person with a high skill is expected to (with a certain probability) answer all items correctly. For example, there are two math questions, a simple one ($1+1=?$) and a more difficult one ($2/5 \times 1/2=?$). When someone correctly answers the simple question, that person is not expected to also correctly answer the difficult question. However, when someone correctly answers to the difficult question, that person is highly likely to also answer correctly to the simple math question.

![Figure 1. Example of item difficulties of a number of math questions. A problem higher up the scale requires a higher ability in math problem solving compared to easy problems. Therefore, a person with a high ability is expected to answer both the difficult problems and the easier problem correctly, while a person with a low ability is expected to only answer the easy problems correctly. The normal curve at the left side indicates the expected distribution of the population’s math problem solving ability.](image)

The current study hypothesizes that the Rasch model can be applied to the various cognitive, visceral and behavioral responses a person can show in a virtual environment. If so, the Rasch model enables us to assess the quality of these different observations and find individual differences in susceptibility to presence. The current study will test users in a virtual pit environment. This study is based on an earlier study by Haans (2014) that tested participants in a virtual environment that consisted of two rooms, where participants had to walk from the first to the second room through a small door, after which they had to walk into a deep virtual pit, located in the second room. The goal was to find out whether the combinations of cognitive, behavioral and visceral reactions to an interactive virtual environment can be scaled according to the Rasch model, in order to find an accurate measure of presence. Although the results looked promising regarding the scaling of presence, definite conclusions were not adequate due to, among other things, a small amount of participants. The current study has a similar experimental setup, but eliminates the shortcomings, by increasing the number of participants, using an improved questionnaire and extending the task design with a new task: a virtual avatar throwing a ball. Participants
were requested to engage in various activities as to proceed to a second room through a small door, watch a virtual avatar throw a ball (in their face), look down into a deep pit, step onto a grille that covered the pit, and finally to press a button on a controller that made them (visually) experience a fall in the pit. After this session, they completed a short questionnaire that asks about their experiences during the session (e.g., with respect to the thrown ball, the falling experience, or the extent to which they felt they would be able to pick up objects present in the virtual room). During the sessions the participants were observed for automatic behaviors (e.g. retreat head for the ball, bending knees when hitting the bottom of the virtual pit) and latencies in volitional behaviors (e.g. when hesitating to jump).

In order to test whether the Rasch model can differentiate between different virtual reality experiences, two virtual environments with different levels of immersiveness were required. In several studies, subjective ratings of presence were found to be dependent on the walking method provided for the virtual body (e.g. Slater, Usoh & Steed, 1995; Usoh et al., 1999). Therefore, in two different sessions a different navigation style (i.e. control over the virtual body) was experienced by participants in two VR systems. The first system let participants physically walk through the real environment (real walking), tracking his or her movements and matching these to the virtual body movements. The second system only allowed participants to move their virtual body through the use of a controller (indirect walking), while physically staying on the same position in the real world. To account for this contextual constraint, a many-facet Rasch-type model is used throughout this experiment (see Linacre, 2002; also Bond & Fox, 2007):

$$\ln \left( \frac{p(x_{nic} = 1)}{1 - p(x_{nic} = 1)} \right) = \theta_n - (\delta_i + \lambda_c)$$

In this model, presence is governed by three factors: person susceptibility to presence ($\theta_n$), item difficulty ($\delta_i$) and navigation style ($\lambda_c$). The different parameter values (e.g. item difficulty) can be used in a presence scale, where item difficulty represents the level of presence that is required for an item to be observed – with a certain probability - in a participant (see Figure 2 for such a presence scale).

![Figure 2](image)

Figure 2. Cognitive, behavioral and visceral reactions to an interactive virtual (pit) environment and their relation to presence (Haans, 2014). A statement higher up the scale requires a higher sense of presence to be affirmatively indicated by participants.
In the context of the Rasch model, an example of an easy item in a VE would be a bodily reflex such as bending the head when moving through a door opening with a lowered frame. Moving through a doorway usually does not require a lot of conscious attention, since such a behavior is managed by the body schema. As mentioned before, the body schema operates largely outside our conscious awareness, and is highly able to allow for transparency in the use of tools such as an HMD (Haans & IJsselsteijn, 2012). Therefore, even when a participant indicates that he or she is not feeling very present in the virtual environment, it is likely that a person will duck in order to avoid any potential collision with a lowered doorframe. A difficult item on the other hand would be the cognitive appraisal that the virtual world is indistinguishable from the real world. Due to limitations in current virtual reality systems, there are cues in the virtual world that will not fool cognition, such as the absence of a virtual body. While all participants are expected to bend their heads, only participants that truly feel present in the virtual world are expected to answer affirmatively to the statement that they could no longer differentiate between the real and the virtual world.

1.4. Research aims

To find out whether the Rasch model is indeed the key to a reliable, sensitive, and valid method for scaling presence, the model will be tested for its invariant ordering of items, ability to distinguish between participants that experience a virtual reality experience, and ability to detect differences in the immersiveness of the media. The main research question that will be answered in this study is: Can presence be scaled through an invariant item order of automatic and volitional behaviors, cognitions and environmentally induced visceral responses? Using the many-facet Rasch-type model from Eq. (2), the expected invariant order of the various observations is tested against the observed responses with fit statistics. Based on the results of a study by Haans (2014), we expect that different observations (automatic and volitional behaviors, cognitions and environmentally induced visceral responses alike) can be ordered in a transitive manner that is more or less similar to all people.

To test whether individual differences in presence susceptibility can be reliably distinguished, and whether the proposed presence scale can differentiate between different virtual reality experiences, three additional questions will be addressed. Firstly, can the proposed method for measuring presence distinguish between participants that experience the same media technology and/or experience? We will investigate how much people differ, and how well they can be distinguished. Since we want to find a scale for presence, we expect our Rasch-based presence measure will find a range of individual differences in presence estimates between participants. Secondly: can the proposed method for measuring presence distinguish within participants that are subjected to a different virtual reality system (real walking versus walking with joystick)? Since indirect walking (in contrast to the real walking) will elicit a mismatch between vision (moving) and bodily action (not moving), significant lower presence scores are on average to be expected in the indirect walking VR experience. We hypothesize our Rasch-based presence measure can detect such differences in immersiveness. Thirdly: Does our presence questionnaire actually measure presence, and not something else? Since the IPQ has been translated in Dutch and has been proven a valid and reliable questionnaire for measuring (spatial) presence, this questionnaire was used to cross-validate our own. We expect a significant correlation with the spatial presence (SP) subscale of the IPQ. Additionally, our presence scale will be correlated with a fear of height measure to test discriminative validity. This correlation is expected to be insignificant, since we want to measure presence, and not fear of heights.
2. Methods

2.1. Participants

A total of fifty participants were recruited (24 male and 26 female). Ages ranged from 17 to 27, with two older participants aged 55 and 62 (M=22.36, SD=7.80). Most participants (90%) were students studying at the Eindhoven University of Technology, Eindhoven, the Netherlands. As can be seen in Table 1, the sample included a range of individual differences regarding experience with VR, fear of heights and length in cm. All participants were recruited through the J.F. Schouten database, and received 7.50 or 9.50 euro (depending on whether they studied at TU/e) as compensation.

<table>
<thead>
<tr>
<th>Time experienced</th>
<th>VR</th>
<th>Fear of height</th>
<th>Length (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>17</td>
<td>0: none</td>
<td>&lt;160</td>
</tr>
<tr>
<td>2nd-3rd</td>
<td>22</td>
<td>1</td>
<td>160-170</td>
</tr>
<tr>
<td>4th-5th</td>
<td>7</td>
<td>2</td>
<td>170-180</td>
</tr>
<tr>
<td>6th - more</td>
<td>4</td>
<td>3</td>
<td>180-190</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>&gt;190</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5: very severe</td>
<td>1</td>
</tr>
</tbody>
</table>

2.2. Research design

A within-subjects design was employed to assess a potential invariant ordering of various cognitive, behavioral, and visceral responses to a virtual environment. Participants visited the same virtual environment twice, with a different navigation style in each session.

1. Real walking
   Participants were tracked for both position and rotation, allowing them to walk and look around in the virtual environment.

2. Indirect walking
   Participants were asked to stay on the same spot in the real environment, but could turn and look around in the virtual environment. They were still being tracked for position and rotation, but only rotational movements were physically allowed. Positional movement was facilitated with a WiiMote controller, controlled by the participant.

The sequence of these conditions was counterbalanced across participants. The total duration of the experiment was around 35 minutes.

2.3. Virtual environment design

A partially similar virtual pit environment was used as the one in a study by Haans (2014). Two rooms were connected by a doorway with a lowered frame set to a height of 1.70 meters. An avatar was located in the corner of the room, holding a ball. A virtual pit (12 meters deep) was located in the center of the room, with a grille that could be closed and opened by the experiment leader and additionally be opened by the participant only when he/she was standing on top of it. This grill enabled participants to step on top of the pit, without floating in the air (because there was no pit in the real world). At the beginning of a trial, the door was closed and the grille was opened. Additional objects (e.g. a lamp and a small table) were placed in the room to make the environment a more realistic one (see Figure 3 for the complete environment). Participants had no virtual body in the virtual environment.
Figure 3. Three views of the virtual room with door and grille opened. 3.a shows a view from the left into the first room, 3.b shows a top view (first room on the left-side, second room on the right-side), 3.c shows a view from the right into the second room (pit room), 3.d shows the depth of the pit in perspective. Three different tasks were carried out: 1= proceeding to the second room through a door opening with a lowered doorframe, 2= watching an avatar throw a ball, 3= looking into a virtual pit, and falling into that pit).

2.4. Apparatus

Hardware used for the virtual environment and the tracking of the user included a camera positioning system (PhaseSpace IMPULSE) and a head mounted display with orientation tracker (nVisor SX111 HMD with InterSense InertiaCube3; for more information see Appendix D). Participants could move in a 7 by 6 m space. The virtual environment itself (the rooms and the various objects in the room) and the interactivity in the virtual environment (including the opening of the door, the animations of the man and the ball, the closing and opening of the grille, and triggering of audio instructions) were implemented using the WorldViz Vizard Virtual Reality Toolkit 4.0 running on a Dell XPS desktop PC.

A Nintendo WiiMote controller was used by participants in the indirect walking condition. They used the directional pad (D-pad) for moving forward and sideways, depending on the direction the HMD was facing. In both the real and the indirect walking condition, participants used the A-button on the controller to open the grille on top of the virtual pit, thus making that participant fall down in the pit.

A high definition video-camera filmed the participants in order to observe behaviors elicited during the sessions in VR.

2.5. Procedure

At the beginning of the experiment, participants were informed about the general purpose of the study and consented to their willingness to participate in this research and to their actions being recorded on camera (Appendix A). In a short introduction participants were told they were going to wear an HMD that would show them a virtual environment, in which they had to complete some tasks. Depending on the condition participants started in, an explanation was provided about how the controller should be used. In the indirect walking condition, participants were told that in order to navigate through the virtual environment, they had to use a controller. The d-pad on the controller was explained as a means of moving around. They were asked to remain physically on (roughly) the same position in the real world, but were allowed to turn and look around. In the real and indirect walking conditions the A-button on the controller was instructed as being the button participants had to press after this was asked of them through the audio instructions (its function was left unexplained until the actual audio instruction). Participants were told that all further instructions would be given through the headphones. Finally, the HMD (with headphones) was put on the participant’s head.
Next, participants completed two sessions in a virtual pit environment. The instructions during the experiment were prerecorded. Audio instructions were used because during pretesting of the virtual environment, verbal instructions given directly by the experiment leader were criticized for increasing awareness for the real world (hearing but not seeing the experiment leader in the virtual environment). The experiment leader triggered each instruction at the right time, depending on the participants’ progress in the virtual environment (see Appendix B for all audio instructions).

At the beginning of each session participants were requested to look and walk around for about one minute in the first room, to get used to the virtual environment and the navigation style (either real walking or using a controller). Afterwards participants were requested to engage in various activities for about two minutes. Firstly, the door opened and participants had to walk to the next room (pit room) through the lowered doorframe, at a height of 1.70 meters. The doorframe was lowered, so that participants had to bend their heads or duck when they wanted to walk to the next room without “hitting their heads”. Afterwards, participants were asked to look at the avatar, and to keep looking at him until further instructions. Only when participants kept looking at the avatar, the experiment leader started the avatar animation, where he would raise his arm and - after around 2 seconds - throw a virtual ball at the participant’s face. Next, participants had to move toward the virtual pit, and were requested to look to its bottom for a couple of seconds. Hereafter, the grille on top of the pit closed, after which participants were asked to step onto the middle of the grille. They were then requested to look down, through the grille, into the pit and press a button on the controller that made the grille open and drop the participant approximately twelve meters down the virtual pit.

After each session, a questionnaire was completed assessing the participants’ experiences in the VE (see Appendix C.1). Directly after completing this questionnaire, participants started the second session, with a different navigation style. Except for a different instruction concerning the use of the controller, the procedure was similar to the first session. After the second session, after completing the presence questionnaires, participants answered some control and demographic questions (regarding age, gender, length, VR experience, and fear of heights; see Appendix C.2).

Finally, participants were debriefed about the aim of the experiment, paid, and thanked for their efforts.

2.6. Measures

2.6.1. Rasch-based presence measure

The first instrument used was one created for this experiment specifically. The items consisted of self-reports and behavioral observations, capturing cognitive evaluation, visceral and behavioral reactions to the virtual environment.

2.6.1.1 Self-reports (questionnaire)

After each session, participants completed a short questionnaire. The first 16 questions consisted of self-reports, asking about their cognitive (e.g. “There were moments during the session where I forgot that I was situated in a simulated environment.”) and visceral (e.g. “It felt somewhat uncomfortable when I looked at the bottom of the pit.”) experiences in the virtual environment with respect to the virtual door, thrown ball, the virtual pit, or the objects present in the virtual rooms (see Table 2 items 5 to 19 or Appendix C.1 items 1 to 16). All items could only be answered with “yes”, “no”, or “not of application”.

One control question (Appendix C.1, item 8) asked whether participants had seen the ball when it was flying towards them. Only one participant indicated he had not seen the ball coming towards him. However, that participant was actually looking at the ball (controlled by the experiment leader) and gave
a physical reaction to the ball. Therefore, this participant’s answer was disregarded and all participants were treated as having seen the ball.

2.6.1.2. Behavioral observations
During the sessions participants were observed for behavioral reactions to the virtual environment (e.g., ducks for the lowered doorpost, avoids the virtual ball, stepped on the grille, pressed the button to fall down the pit, and bended the knees when hitting the bottom of the virtual pit) and for latencies in volitional behaviors (e.g., hesitated to press the button to fall; see Table 2 items 1 to 4). Video recordings were used to review behavioral observations.

Table 2. Item number and observation description for the Rasch-based presence measure. The first four items (1-4) were behavioral observations; the other items (5-19) were statements in a questionnaire.

<table>
<thead>
<tr>
<th>Item</th>
<th>Observation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ducks or bends head for the virtual doorpost</td>
</tr>
<tr>
<td>2</td>
<td>Head moved at the impact moment of the virtual ball</td>
</tr>
<tr>
<td>3</td>
<td>Hesitated to press the button to fall down (&gt;3 seconds delay after instruction)</td>
</tr>
</tbody>
</table>
| 4    | 0= did not step on the grille  
1= stepped on the grille  
2= pressed the button to fall down the pit  
3= bended knees when falling down or hitting the bottom of the virtual pit |
| 5    | I had the feeling I had to evade the ball. |
| 6    | I knew for sure the ball that was thrown was going to hit me physically. |
| 7    | I thought for a moment a real ball was thrown to me. |
| 8    | There were moments during the session where I forgot that I was situated in a simulated environment. |
| 9    | I kept having the feeling that the different objects I saw were situated in a specific location in the real environment around me. |
| 10   | At some moments during the session I felt as if I could stretch my arm to touch the objects in the virtual environment. |
| 11   | I know for sure the two rooms were real and not a simulation. |
| 12   | I knew for sure I could hit my head against the doorpost. |
| 13   | I had the feeling I needed to duck to avoid hitting my head against the doorpost. |
| 14   | *When I was asked to fall in the pit, I did so without hesitation.* |
| 15   | After the grilles opened, it really felt like I fell. |
| 16   | I felt a bodily response when I hit the bottom of the pit. |
| 17   | I had the thought that I could really fall when I looked down the pit. |
| 18   | It felt somewhat uncomfortable when I looked at the bottom of the pit. |
| 19   | I felt a bodily response when I looked down the pit. |

Note that items are translated from Dutch, for the Dutch version see Appendix C.1. For item 4, all participants were at least in group 2.

Since the doorpost in the VE was simply too high (1.70m) for some participants to elicit a ducking behavior in the real world, item 1 (“Ducks or bends head for the virtual doorpost”) was reported as missing for participants smaller than 1.65 m (N=12). Item 3 (“Hesitated to press the button to fall down”) was measured in Vizard as the time between the start of the instruction (telling the participant to press the button to fall down) and the moment the participant pressed the button to fall down. In analyses, this time was recoded to a dichotomous item, where a button press that took longer than three seconds after the instruction was coded as ‘1’. This cutoff point was decided subjectively after an exploratory look at the data, based on the median time participants waited to press the button (2.1 seconds, SD=1.36). Because three participants forgot to bring the controller in the real walking condition, they couldn’t press the button to fall down when this was asked of them. Therefore, item 3 (“Time until button was pressed to fall down”) was reported as missing in these participants. These participants still experienced the fall, after the controller was handed to them by the experiment leader.

Three observations were recoded into one super item: the observations “did not step on the grille”, “stepped on the grille”, “pressed the button to fall down the pit”, and “bended knees when falling down
or hitting the bottom of the virtual pit” were recoded as 0, 1, 2 and 3 respectively in item 4. All participants elicited the behavior of “stepping on the grille”. Therefore, item 4 only distinguishes between “pressed the button to fall down the pit”, and/or “bended knees when falling down or hitting the bottom of the virtual pit”.

The observed cognitive, behavioral and visceral responses obtained through self-reports (i.e. the questionnaire) and behavioral observations were analyzed and scaled with Rasch-type model tests using Facets (Linacre, 2015). The Facets software employs a joint maximum likelihood procedure to estimate each parameter of the factors in the defined Rasch model. Three factors were investigated: Item i’s difficulty ($\delta$), effect of navigation style $c$ ($\lambda$) and person n’s susceptibility ($\theta$; see Eq. 2). The mean-squared statistic was used to assess the match between the model-predicted and observed response patterns. For a good fit, items require MS-values ≤ 1.30, for an acceptable fit, items require MS-values ≤ 1.50, in accordance with Wright and Linacre (1994). Similar MS-values were used for person score tests, however Wright and Linacre (1994) indicate that people should not being treated as strictly as items: Since we are measuring people, some misfit in persons scores is expected. Therefore, a margin of 5% poorly fitting persons was considered acceptable as a rule of thumb (e.g. Haans, 2014). In addition, item invariance across conditions was tested in Winsteps (Linacre, 2016) by comparing the item difficulty parameter ($\delta$) values between navigation styles, with a 95% confidence interval determined by Winsteps.

2.6.2. Igroup Presence Questionnaire (IPQ)

In addition to self-reports used for the Rasch-based presence measure, a remaining 14 questions were obtained from the Dutch translation of the Igroup Presence Questionnaire (IPQ; Schubert, Friedmann & Regenbrecht, 1999), used to test convergent validity (see Appendix C.1 items 17 to 30). The IPQ consists of three subscales: Spatial presence (SP), Involvement (INV) and Experienced Realism (REAL). According to Igroup, SP measures the sense of being physically present (e.g. “Somehow I felt that the virtual world surrounded me”), INV measures the attention devoted to the virtual environment and involvement experienced (e.g. “How aware were you of the real world surrounding while navigating in the virtual world?”) and REAL measures the subjective experience of realism (e.g. “How real did the virtual world seem to you?”; Schubert, Friedmann & Regenbrecht, 1999). Additionally, one general item can be identified that measures a general sense of “being there” (PRES; “In the computer generated world I had a sense of ‘being there’”). The average reliability (Cronbach’s $\alpha$) from two studies done by Igroup (Schubert, Friedmann & Regenbrecht, n.d.) and the average reliability from two sessions in the current study are reported in Table 3. The subscale reliabilities from the current study were .687 (SP), .760 (INV) and .763 (REAL). The Cronbach’s $\alpha$ of the PRES subscale could not be determined because it consisted of only one item. There were no missing values for the IPQ questions in the current study’s data.

Table 3. Reliability ($\alpha$) of three subscales (SP, INV and REAL) of the IPQ estimated through Cronbach’s $\alpha$. The left column contains the average reliabilities obtained in two studies performed by Igroup; the right column contains the average reliabilities obtained in the two sessions of the current study.

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Studies done by Igroup $\alpha$</th>
<th>Both sessions of current experiment $\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SP</td>
<td>.785</td>
<td>.687</td>
</tr>
<tr>
<td>INV</td>
<td>.760</td>
<td>.796</td>
</tr>
<tr>
<td>REAL</td>
<td>.690</td>
<td>.763</td>
</tr>
</tbody>
</table>

Note. The Cronbach’s $\alpha$ of the PRES subscale cannot be determined because it consisted of only one item.
2.6.3. Additional questions

One additional empty page could be used by participants to describe their experiences each session (i.e. an open question), to allow for a more qualitative analysis of their experiences. Additionally, since the virtual pit is a large element in the current study, it could be the case that the Rasch-based presence measure measures fear of heights instead of presence. Therefore an existing fear of heights question (obtained from Leitenberg & Callahan, 1973) was added at the end of the experiment to test discriminative validity through a correlation with obtained person score (i.e. susceptibility to presence) of the Rasch-based presence measure. The fear of heights question stated “Do you suffer from vertigo? Assess the severity of your fear of heights as you experience it in everyday situations.” and was answered on a 6-point scale (coded at the extremes as 0= “no fear of heights”, 5= “very severe fear of heights”). In addition, one question assessed participants’ previous experience with VR, stating “What is your experience with this kind of virtual environments?” This question had four possible answers: “1st time”, “2nd or 3rd time”, “4th or 5th time”, or “6th time or more often” (stated in a full sentence: “This was my ___ time in a virtual environment”). Three demographic questions regarding the participant’s length, age and gender concluded the experiment. For all additional questions in Dutch see Appendix C.2.

3. Results

In the following section, firstly the observations’ fit to the Rasch model will be presented. Thereafter the individual factors’ statistics (item difficulties, navigation style effect and person scores) in the analysis will be reported more in depth. Since one item introduced some problems, both analyses with and without this item were carried out. Finally, convergent and discriminative validity are reported.

3.1. Model test

The data-to-model fitted presence scale is shown in Figure 4. A high position on the scale corresponds to a higher sense of presence (defined as the perceptual illusion of non-mediation through transparency of media technology). For the item parameter, a high parameter value reflects that an item (whether an observed behavior, or self-reported cognition) was more difficult to be endorsed, and thus required a more presence-susceptible participant or presence-inducing condition in order for that cognition or behavior to be observed. For participants with a higher person score, this means he or she was more likely to respond affirmatively to even the more difficult items. For navigation style, a higher position on the presence scale means that navigation style elicited higher presence scores in participants.
3.1.1. Item difficulties

Item difficulties ($\delta$) of 19 items ranged from -2.11 to 3.23 logits (centered at $M=0$; $SD=1.24$) and covered a wide range of values in that range (see $\delta$ in Figure 4 and Table 4). Responses to all but one item fitted the Rasch model, with $MS$-values $\leq 1.30$; only item 14 (“When I was asked to fall in the pit, I did so without hesitation.”) had aberrant fit values of $MS=1.72$ (infit) and $MS=2.43$ (outfit). Item difficulties were estimated with a reliability of .96.

Removal of item 14 from analysis resulted in some slight changes in item difficulties (maximum changes of .19 logits) and $MS$-values (maximum changes of .16 logits), but the order of the items with respect to their difficulties remained largely the same. In this new analysis, all items except one were a good fit to the Rasch model, with $MS$-values $\leq 1.30$. Now, item 3 emerged as having a slightly insufficient outfit with $MS=1.36$, but since that value was below the 1.50 criterion, it was left in all analyses.

A potential effect of session number on item difficulty (excluding item 14) was tested, because participants indicated they “expected what was about to happen” in the second session.
However, no significant effect was found for session number on item difficulty ($\chi^2 = 11.3$ (36), $p > .99$). Therefore session sequence was not considered in other analyses.

Table 4. Item difficulties ($\delta$) and standard error of estimate (SE), mean-squared (MS) infit and outfit statistics, and the probability an averagely susceptible person ($p$) will respond affirmatively to an item (condition effect not included).

<table>
<thead>
<tr>
<th>Observation</th>
<th>(\delta) (SE)</th>
<th>Infit (MS)</th>
<th>Outfit (MS)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>I know for sure the two rooms were real and not a simulation.</td>
<td>3.23 (.52)</td>
<td>1.02</td>
<td>0.70</td>
</tr>
<tr>
<td>3</td>
<td>Hesitated to press the button to fall down (&gt;3 seconds delay after instruction).</td>
<td>1.83 (.31)</td>
<td>1.14</td>
<td>1.27</td>
</tr>
<tr>
<td>4</td>
<td>Bended the knees during fall or on impact with the pit's floor.</td>
<td>1.34 (.27)</td>
<td>0.88</td>
<td>0.77</td>
</tr>
<tr>
<td>6</td>
<td>I knew for sure the ball that was thrown was going to hit me physically.</td>
<td>1.34 (.27)</td>
<td>0.93</td>
<td>0.92</td>
</tr>
<tr>
<td>7</td>
<td>I thought for a moment a real ball was thrown to me.</td>
<td>0.59 (.23)</td>
<td>0.84</td>
<td>0.80</td>
</tr>
<tr>
<td>15</td>
<td>After the grilles opened, it really felt like I fell.</td>
<td>0.43 (.23)</td>
<td>0.96</td>
<td>0.90</td>
</tr>
<tr>
<td>5</td>
<td>I had the feeling I had to evade the ball.</td>
<td>0.28 (.23)</td>
<td>0.98</td>
<td>0.93</td>
</tr>
<tr>
<td>12</td>
<td>I knew for sure I could hit my head against the doorpost.</td>
<td>-0.09 (.22)</td>
<td>0.87</td>
<td>0.82</td>
</tr>
<tr>
<td>8</td>
<td>There were moments during the session where I forgot that I was situated in a simulated environment.</td>
<td>-0.15 (.22)</td>
<td>0.86</td>
<td>0.90</td>
</tr>
<tr>
<td>16</td>
<td>I felt a bodily response when I hit the bottom of the pit.</td>
<td>-0.17 (.22)</td>
<td>0.96</td>
<td>0.94</td>
</tr>
<tr>
<td>1</td>
<td>Ducks or bends head for the virtual doorpost.</td>
<td>-0.34 (.25)</td>
<td>0.98</td>
<td>0.95</td>
</tr>
<tr>
<td>2</td>
<td>Head moved at the impact moment of the virtual ball.</td>
<td>-0.36 (.22)</td>
<td>1.08</td>
<td>1.07</td>
</tr>
<tr>
<td>10</td>
<td>At some moments during the session I felt as if I could stretch my arm to touch the objects in the virtual environment.</td>
<td>-0.53 (.22)</td>
<td>1.06</td>
<td>1.05</td>
</tr>
<tr>
<td>9</td>
<td>I kept having the feeling that the different objects I saw were situated in a specific location in the real environment around me.</td>
<td>-0.69 (.22)</td>
<td>1.06</td>
<td>1.05</td>
</tr>
<tr>
<td>13</td>
<td>I had the feeling I needed to duck to avoid hitting my head against the doorpost.</td>
<td>-0.82 (.22)</td>
<td>1.01</td>
<td>1.00</td>
</tr>
<tr>
<td>17</td>
<td>I had the thought that I could really fall when I looked down the pit.</td>
<td>-0.83 (.22)</td>
<td>0.92</td>
<td>0.89</td>
</tr>
<tr>
<td>14</td>
<td>When I was asked to fall in the pit, I did so without hesitation.</td>
<td>-1.30 (.23)</td>
<td>1.72</td>
<td>2.42</td>
</tr>
<tr>
<td>19</td>
<td>I felt a bodily response when I looked down the pit.</td>
<td>-1.70 (.25)</td>
<td>0.82</td>
<td>0.68</td>
</tr>
<tr>
<td>18</td>
<td>It felt somewhat uncomfortable when I looked at the bottom of the pit.</td>
<td>-2.11 (.28)</td>
<td>0.85</td>
<td>0.66</td>
</tr>
</tbody>
</table>

Note that items are translated from Dutch, for the Dutch version see Appendix C.1. Item 14 was negatively worded (in italic), and thus reversed in their coding before analysis. Items 1-4 were behavioral observations, items 5-19 were self-reports assessed in questionnaires. All participants showed observation 2 of super item 4: “pressed the button to fall down the pit” (see Table 2). Therefore, item 4 only discriminates between participants that did or did not bend their knees.

3.1.2. Person scores

The distribution of person scores ($\theta$) is displayed in Figure 4. Person scores ranged from -1.93 to 1.70 logits ($M = -0.35$, $SD = 0.82$) and were estimated with a reliability of .77. Some individual differences besides the experimental behavioral observations were observed: For example, some participants immediately tried to (rebelliously) walk through the walls, while others moved around very carefully to not bump into objects. Additionally, while some participants commented in the open question about the importance of the environment’s realism, others talked about feelings and emotions that the experience elicited. Nine out of fifty participants did not meet requirements for a good fit with MS-values $\leq 1.30$. This was a total of 18% which, according to the 5% margin of poorly fitting persons, is not considered acceptable. Especially outfit values were found to be insufficient. Of these nine participants, two stood out with insufficient fit MS-values $> 1.50$.

After removing item 14, person scores and their order on the Rasch scale changed a bit. Persons scores now ranged from -2.38 to 1.66 logits ($M = -0.44$, $SD = 0.98$) and reliability of person scores slightly grew to .81. Again, nine out of fifty participants (18%) did not meet requirements for a good fit with MS-values $\leq 1.30$. However, from these participants only five were the same as in the analysis that included item 14. Again, these misfit values were mostly outfit statistics. The total of 18% poorly fitting persons, according to the 5% margin of poorly fitting persons, cannot be considered acceptable. Of these nine participants, three stood out with insufficient fit MS-values $> 1.50$ (two persons with only insufficient outfit statistics and one
A person with both insufficient outfit and infit statistics. For more infit and outfit statistics, and demographics of poorly fitting persons, see Table 5.

Table 5. Person scores ($\theta$) and standard error of estimate (SE), mean-squared (MS) infit and outfit statistics, and demographic information of that person (gender, age, VR experience, fear of heights and length in centimeters), for the data-to-model analysis excluding item 14.

<table>
<thead>
<tr>
<th>$\theta$ (SE)</th>
<th>Infit (MS)</th>
<th>Outfit (MS)</th>
<th>Gender</th>
<th>Age</th>
<th>Times experienced VR</th>
<th>Fear of heights</th>
<th>length (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.09 (.39)</td>
<td>0.99</td>
<td>1.91**</td>
<td>Male</td>
<td>21</td>
<td>≥ 6th time</td>
<td>0</td>
<td>192</td>
</tr>
<tr>
<td>0.09 (.39)</td>
<td>1.54**</td>
<td>1.68**</td>
<td>Female</td>
<td>19</td>
<td>1st time</td>
<td>4</td>
<td>168</td>
</tr>
<tr>
<td>0.89 (41)</td>
<td>1.18</td>
<td>1.54**</td>
<td>Male</td>
<td>55</td>
<td>1st time</td>
<td>5</td>
<td>194</td>
</tr>
<tr>
<td>-0.68 (41)</td>
<td>1.12</td>
<td>1.50*</td>
<td>Female</td>
<td>18</td>
<td>1st time</td>
<td>3</td>
<td>162</td>
</tr>
<tr>
<td>-1.30 (42)</td>
<td>1.02</td>
<td>1.46*</td>
<td>Male</td>
<td>22</td>
<td>2nd or 3rd time</td>
<td>0</td>
<td>180</td>
</tr>
<tr>
<td>-2.30 (53)</td>
<td>1.42*</td>
<td>1.19</td>
<td>Female</td>
<td>20</td>
<td>1st time</td>
<td>0</td>
<td>164</td>
</tr>
<tr>
<td>-1.31 (42)</td>
<td>1.38*</td>
<td>1.24</td>
<td>Male</td>
<td>18</td>
<td>1st time</td>
<td>1</td>
<td>192</td>
</tr>
<tr>
<td>-0.52 (40)</td>
<td>1.22</td>
<td>1.34*</td>
<td>Female</td>
<td>19</td>
<td>2nd or 3rd time</td>
<td>1</td>
<td>160</td>
</tr>
<tr>
<td>-1.69 (45)</td>
<td>0.96</td>
<td>1.31*</td>
<td>Male</td>
<td>22</td>
<td>1st time</td>
<td>3</td>
<td>186</td>
</tr>
</tbody>
</table>

Note: * . Fit statistic is higher than 1.30.
** . Fit statistic is higher than 1.50.
Fear of heights values range from 0-5 (0: none, 5: very severe).

3.1.3. Effect of navigation style

The effect of navigation style ($\lambda$) on presence score is displayed in Figure 4 in the middle column. According to our expectation, real walking (.45 logits) elicited more presence than indirect walking (-.45 logits) to a statistically significant extent, $\chi^2(1, N=50)=64.8, and p<.001$. This means that in the real walking condition, in general participants showed more observations from Table 4. The effect of navigation style was estimated with a reliability of .97. Some differences between conditions besides the experimental behavioral observations were observed. Participants moved around differently in the real walking condition compared to the indirect walking condition: for example, people approached the virtual pit more carefully in the real walking condition. Additionally, in the open question participants described experiencing “more feelings” and “more realism”, as well as a “more natural way of moving around” in the real walking condition. The indirect walking condition was described as a “more static experience”, where more “distance” to the virtual environment was experienced. As one participant commented: “I had the feeling I was controlling a character [in the indirect walking condition], instead of actually being the character myself [like in the real walking condition]”.

When item 14 was removed from analysis, the effect of navigation style remained significant: $\chi^2(1, N=50)=88.4, and p<.001$. The effect of navigation style increased slightly from .9 to 1.1 logits (again centered at $M=0$) and reliability remained about the same with .98.

We performed a test of differential item functioning to find out whether the item difficulties were invariant across the two navigation styles. The item difficulty calibration for the real walking condition was plotted against the item difficulty calibration for the indirect walking condition, along with 95% confidence intervals. This test showed that participants answered very different to expectation across conditions on item 14 (see number 14 in Figure 5). Although a difference of .90 logits was expected between navigation styles, this effect was estimated at 2.99 logits for item 14.
Figure 5. Item difficulty estimates ($\delta$) compared across conditions. The solid curved lines form the 95% confidence interval. Items that show an interaction effect with navigation style are located outside the curved lines (relative to the dotted line). Item 14 elicited very variant item difficulty estimates across conditions (encircled in grey).

Because of the misfit of item 14 to the model and the interaction of its item difficulty with navigation style, person scores obtained from the analysis without item 14 will be used for the convergent and discriminative validity tests below.

3.2. Convergent and discriminative validity

3.2.1. Correlation with IPQ instrument

Convergent validity was tested by correlating the presence measure with an existing one: the IPQ. Since participants completed the IPQ twice (once in the indirect and once in the real walking condition), but only a single person score was obtained per participant, the two IPQ scores were averaged. All scales had reliabilities over .75. Our Rasch-based presence measure correlated significantly with the spatial presence (SP; $r=.403$, $p=.004$), involvement (INV; $r=.297$, $p=.036$) and experienced realism (REAL; $r=.570$, $p<.001$) subscales of the IPQ. Some of the different subscales of the IPQ also correlated with each other. For example the SP subscale correlated significantly with all other subscales (INV: $r=.456$, $p=.001$; REAL: $r=.513$, $p<.001$; PRES: $r=.290$, $p=.041$). For all correlations and reliabilities of the different subscales, see Table 6.

Table 6. Correlations between persons scores ($\theta$) and subscales of the IPQ averaged over both conditions of the current experiment. SP=Spatial presence, INV=involvement, REAL=experienced realism, PRES=one item measuring general presence. Reliabilities (Cronbach’s $\alpha$) of the different scales are on the diagonal.

<table>
<thead>
<tr>
<th></th>
<th>$\theta$</th>
<th>SP</th>
<th>INV</th>
<th>REAL</th>
<th>PRES</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta$</td>
<td>.81</td>
<td>.406</td>
<td>.297</td>
<td>.570</td>
<td>.010</td>
</tr>
<tr>
<td>SP</td>
<td>.69</td>
<td>.456**</td>
<td>.513**</td>
<td>.290*</td>
<td></td>
</tr>
<tr>
<td>INV</td>
<td>.80</td>
<td>.267</td>
<td>-1.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>REAL</td>
<td>.76</td>
<td>.126</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRES</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: *. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).
The person scores were used from the analysis excluding item 14.
The Cronbach’s $\alpha$ of the PRES subscale cannot be determined because it consisted of only one item.
3.2.2. Correlation with Fear of heights

Since many of the tasks that participants had to conduct in the virtual environment involved the deep virtual pit (e.g. looking down in it, or falling in the pit), the question remains whether we are measuring a person’s fear of heights, rather than the intended presence in the virtual environment. Therefore, we tested for discriminative validity by correlating the Rasch-based person scores ($\theta$) with self-reported fear of heights. We found only a small, and statistically insignificant correlation of $r=0.250$ ($p=0.80$) between persons scores and fear of heights. In other words, our presence scale measured something else than fear of heights, despite the large role of the virtual pit in the virtual environment.

4. Discussion

Based on the definition of presence as the perceptual illusion of non-mediation through transparency of media technology (e.g. Haans, 2014), we anticipated that a variety of different responses (automatic and volitional behaviors, cognitions and environmentally induced visceral responses) to computer-generated content could be scaled to a single, unidimensional, presence scale. Therefore we assessed these responses through self-reports and behavioral observations, and fitted the acquired data to a many-facet Rasch model.

4.1. Quality of the instrument

As expected, the scale calibration and fit statistics revealed that we were successful in using a combination of behavioral observations and self-reports about people’s behavioral, cognitive, and affective responses. Of the 19 items, responses to all but one item fitted the model’s expectation sufficiently. Item 14 (“When I was asked to fall in the pit, I did so without hesitation”) did not fit the model with very high infit and outfit MS-values. One possible cause of this misfit was that this item varied across conditions. The item’s difficulty was estimated much lower than expected in the indirect walking condition compared to the real walking condition. In other words, people indicated it was much easier to press a button to fall down in the indirect walking condition compared to the real walking condition. One possible explanation for item 14’s misfit - that matches the invariance in item difficulty - is that in the real walking navigation style participants actually had to move their real body on the (virtual) grid, making it a more plausible scenario that that person is really standing on top of a pit in the real world. Since participants stood on the same spot in the real world during the indirect walking navigation style, the idea of a potentially real pit might simply have been too implausible.

Another explanation for the invariance in this item’s difficulty may be due to how the related question was formulated in the questionnaire, stating “When I was asked to fall in the pit, I did so without hesitation”. Different people - or the same person across conditions - may have read and interpreted the question differently. For example, one can read this item related to behavior (e.g. “I waited 3 seconds before pressing the button”), or to cognition (e.g. “I did not want to press the button but did so anyway”). A potential difference in response caused by a different interpretation is supported by participants in the open question, indicating that the virtual pit - although it made them doubt in the decision to fall down or not - did not keep them from actually pressing the button to fall down. Indeed, the behavioral observations of “stepped on the grille” and “pressed the button to fall down the pit” were elicited by all participants, and a hesitation of longer than 3 seconds to press the button to fall down was observed in only in few participants.
4.2. Merits to differentiate individuals

As expected, the Rasch-based presence measure was found to reliably distinguish between participants ($r = .81$). Since people should not be treated as strict as items, 5% of participants were expected to misfit the model with $MS$-values $> 1.30$. This rule of thumb was not met with 18% of poorly fitting participants. Since no poorly fitting items were left in this analysis, an explanation is not likely to result from problems with a single item. Which begs the question, why did so many participants misfit the model?

Several possible explanations are possible. Firstly, some poorly fitting participants might have simply misread a question, or accidentally ticked the wrong box during completion of the questionnaire. For example, a participant indicated he had not seen the ball, yet that participant was clearly looking at the ball when it was thrown (controlled through video recordings), and he gave a physical reaction to the task involving the ball. However, this is the reason a 5% margin for poorly fitting persons was governed and does not explain the much higher percentage of poorly fitting persons observed (18%). Therefore, another explanation might be found in the way participants interpreted the questionnaire’s items. It might be the case that some participants interpreted certain questions differently than the majority of participants. For example, there were three participants that answered affirmatively to item 11, stating “I know for sure the two rooms were real and not a simulation”. This item was not expected to be answered affirmatively by any participant at all, especially not in participants that did not answer affirmatively to most other items (which was the case in two insufficiently fitting participants with $MS$-values $> 1.50$). Another possible explanation for the large number of poorly fitting participants is related to the experimental design, in which we tested participants twice. Participants might have focused their attention differently in each session. For example in the first session, participants did not yet know what was about to happen, or what questions were going to be asked in the questionnaire. Therefore they might have focused attention on different aspects of the virtual world in the second session. Indeed, in the second session, several participants indicated they were already preparing for certain events (e.g. the thrown ball or falling down the virtual pit). This even led to some participants simply waiting in front of the (virtual) door to the second room, until it was opened by the experiment leader. Accordingly, how participants behaved might have differed between sessions. For example, often in the first session, people were walking around a lot, performing reality checks such as walking through the walls, looking at their (virtual) hands (that were not there), or trying to touch objects in the room. Such behaviors were seldom observed in the second session. Future research should therefore focus on the effects of repeated measures (e.g. expectation of events) on behavior in a virtual environment and a person’s interpretation of items in a presence questionnaire.

4.3. What is being measured (convergent and discriminative validity)

Convergent validity was tested through cross-validation of person score of the Rasch-based presence measure with subscale score on the Igroup Presence Questionnaire (an existing questionnaire that attempts to measure presence; IPQ; Schubert, Friedmann & Regenbrecht, 1999). The insignificant correlation with the general item ($PRES; r = .010$) was expected, since we did not expect a single item can fully capture what it means to be present in a virtual environment. Significant correlations were found between persons score and three IPQ subscales: spatial presence ($SP; r = .406$), involvement ($INV; r = .297$) and experienced realism ($REAL; r = .570$). Since these correlations were not very high, apparently our presence scale is not measuring exactly the same thing as the different subscales of the IPQ (about 8-33% explained variance). This might be due to a difference in what was measured: For example the $SP$ subscale only asks about “feelings” (e.g. “Somehow I felt that the virtual world surrounded me”, or “I felt present in the virtual space”). In comparison, our Rasch-based presence measure consists of cognitions (e.g. “I had the thought that I could really fall when I looked down the pit”), behaviors (e.g. ducks or bends head for the virtual doorpost) and visceral reactions (e.g. “I felt a bodily response when I looked down the pit”).
The items of the REAL subscale (e.g. “How much did your experience in the virtual environment seem consistent with your real world experience”) might represent most how we defined presence related to transparency (hence the moderate correlation): When a virtual reality experience closely resembles a real one, we might conclude the VR equipment has become transparent for the user. The small correlation with the involvement (INV) subscale can also be explained, since the items of this subscale ask about awareness of the real environment (e.g. “I was not aware of my real environment”). We argue that awareness of the real world is not necessarily interrelated with presence in the virtual environment. Indeed, some participants indicated that even though they sometimes still paid attention to the real world, they would still experience a strong sense of “being there” in the virtual environment.

Nevertheless, a moderate to low correlation with different presence measures (i.e. the SP and INV subscales respectively) is not necessarily a bad result. Actually, it is in line with the finding that various presence-related responses hardly correlate empirically (e.g. Slater et al., 2009; Haans, 2014). To recall from the introduction, a low correlation needs not be a cause for deeming these measurement attempts unsuccessful. All measures can be valid; it is the way the different methods are combined, not compared, that make up the most reliable presence score (Haans, 2014). Presence in virtual reality was hypothesized to be measured through the combination of behavioral, cognitive and visceral responses to a virtual environment. The IPQ however does not address the full range of responses possible in a virtual environment. The different subscales of the IPQ might thus only be sensitive to a small part of the full range of possible presence-related responses to the media technology. In line with this argument, we argue that the SP subscale mostly addresses feelings (e.g. “Somehow I felt that the virtual world surrounded me”, or “I felt present in the virtual space”) and the questions in the REAL subscale ask mostly about the realism of the experience as a whole (e.g. “How much did your experience in the virtual environment seem consistent with your real world experience?”). The IPQ also neglects behavioral observations completely. Therefore we argue our Rasch-based presence measure more fully captures what it means to be present in a virtual environment. In addition, although Igroup divided the items of its presence questionnaire into different subscales, we argue these subscales might all be part of the same scale, namely presence. An interesting direction of future research could therefore be to try and fit the items of an existing presence questionnaire such as the IPQ to the Rasch model.

Of course, the presence scale should measure presence, and not something else. Therefore, due to the large role of the virtual pit in this experiment, we correlated estimated person scores with a measure of fear of height to test discriminative validity. The small and statistically insignificant correlation \( r = .254 \) indicates that we did not differentiate participants based on their fear of heights, but that the observed differences in automatic and volitional behaviors, cognitions, and environmentally induced visceral responses to the virtual environment, were indeed reflective of differences in the sense of presence.

### 4.4. Detection of differences in the immersiveness of the media

The hypothesis that different conditional setups can be accurately distinguished through the Rasch model was confirmed, regardless of the high percentage of poorly fitting participants. Real walking elicited significantly higher presence scores compared to indirect walking with a WiiMote controller. This supports the argument that our Rasch-based presence measure can determine differences in the immersiveness of the media experienced by participants. Most differences in reactions to the VE were also very salient depending on participants’ navigation style. For example in the real walking condition participants approached the pit much slower and with more care compared to the indirect walking condition.
4.5. Limitations

There were several limitations to the present experiment that pertained to the virtual reality setup, and how it affected participants’ behavior and presence in the virtual environment. One limitation of this study was the design of the virtual environment itself. With a fixed height of the doorpost, participants smaller than 1.65 meters did not even have to bend their heads in order to avoid collision and these behaviors were reported as missing in the analysis. For future research, the environments’ tasks should be adapted to the user to avoid missing values in the data (e.g. a doorpost height that is set for each user to be on his or her eye-level).

Another limitation was the absence of a virtual body during the experiment. Participant that looked down could not see their own (virtual) body, thereby likely reducing presence during the tasks where people had to look down (i.e. involving the virtual pit). This might also have affected the way participants executed those tasks. For example moving towards the edge of the virtual pit might have been a different experience when one can actually see how far his or her feet are from the edge. The absence of a body could lead to more problems: in a pilot study by Haans (2014), participants accidentally fell in the pit, because they couldn’t see whether they had already stepped over the edge. Therefore in the current study a grille was introduced that covered the pit. This allowed participants to stand in the middle of the grille before falling down into the pit (after the press of a button). A downside of this approach is that the grille made participants less anxious to step on the pit and fall down into it. In fact, none of the participants refused to step on the grille or fall down the pit. Some participants indicated they might have refused if they would have been instructed to step or jump into the virtual pit. Indeed, in the study by Haans (2014) mentioned earlier, some participants refused to step in the virtual pit.

A third limitation was introduced by the audio instructions. One behavioral observation addressed hesitation to press a button to fall down in the virtual pit. The function of the button was left unexplained until the actual audio instruction preceding this task (see Appendix B, audio instruction 6). However, since this audio instruction was about fifteen seconds long, participants were already given some time to think about the upcoming decision to fall down. Even though this item fitted the model sufficiently, the instruction’s length might have influenced this item’s difficulty. Additionally, this item might have led to invariant results across sessions, because in the second session, the instruction was already known to participants. This gave them more time to prepare for the fall (an experience they had already experienced).

The impact of the repeated measures research design is an aspect of this study that needs to be further investigated. By treating presence as a trait in the many-facets Rasch model, we did not allow changes in presence susceptibility across sessions or conditions. However, a salient result of the current study is the large amount of participants that did not fit the Rasch model well. As discussed before, participants might have experienced the virtual environment or interpreted questions in the questionnaire differently between sessions. Due to the large amount of poorly fitting participants, one could argue susceptibility to presence is not simply a trait, or in terms of the Rasch model: a number that we can add to effect of navigation style and item difficulty. Perhaps the many-facets Rasch model is not the right model for measuring presence across sessions within the same person. Presence might be dependent on more than simply one trait (susceptibility to presence) alone. It is plausible that an estimate of presence is very dependent on how attention was focused on the different aspects in the virtual environment. For example, Rand and colleagues (2005) found that changes in the environment and interactivity can help focus participants’ attention and keep them more engaged, influencing a user’s experience of presence. Therefore, it is interesting to investigate presence using a Rasch-type model that accounts for invariant person score estimates.
4.6. Conclusion

Despite these limitations, the present study confirms that observing people’s behavioral, cognitive and visceral responses to a virtual environment are all valid methods for measuring presence. Existing instruments that attempt to measure presence - such as the IPQ - could indeed be valid. However at the same time, such a questionnaire delimits the scope of presence because it ignores behaviors and visceral reactions to the virtual environment that too are intrinsically linked to presence. This study shows a better way of turning the various behavioral, cognitive, and visceral responses to a virtual environment into a presence measure by fitting these responses to a type of Rasch model. When the invariance in susceptibility to presence is accounted for in future work, we might just have the key to a single and unidimensional presence measure, allowing for a presence scale that captures the full range of intermediate states in presence.
References


A. Informed consent form

Informed consent form

This document gives you information about the study “ScalingPresence3”. Before the study begins, it is important that you learn about the procedure followed in this study and that you give your informed consent for voluntary participation. Please read this document carefully.

Aim and benefit of the study
The aim of this study is to find an accurate method of measuring “presence” (often defined as “the sense of being there”) in a virtual environment. The benefit is to improve and combine the current methods in measuring presence, leading to the ability to differentiate between and within users of media technologies, as well as to differentiate between different media technologies and/or experiences. This study is done by Peter Frumau, a student under the supervision of Antal Haans of the Human-Technology Interaction group.

Procedure
You will visit a virtual environment twice for about 10 minutes each. In each visit, there will be a different condition: real walking or walking with a joystick. During each visit you will be requested to engage in various activities. After each visit, you will complete a questionnaire that asks about your experiences during the sessions. Additionally, you will be video recorded during each session. Video data, in particular regarding bodily movements, are necessary for this experiment.

Risks
The study does not involve any risks or detrimental side effects.

Duration
The study will last approximately 40 minutes.

Participants
You were selected because you were registered as participant in the participant database of the Human Technology Interaction group of the Eindhoven University of Technology.

Voluntary
Your participation is completely voluntary. You can refuse to participate without giving any reasons and you can stop your participation at any time during the study. You can also withdraw your permission to use your experimental data up to 24 hours after the study is finished. All this will have no negative consequences whatsoever.
Compensation
You will be paid 7,50 euros (plus 2,00 euros extra if you do not study or work at the TU/e or Fontys Eindhoven).

Confidentiality
All research conducted at the Human-Technology Interaction Group adheres to the Code of Ethics of the NIP (Nederlands Instituut voor Psychologen – Dutch Institute for Psychologists). We will not be sharing personal information about you to anyone outside of the research team. Video recordings that are made will be used only for scientific analysis. Video recordings—facial features not visible—will not be used for public presentation purposes without your explicit consent. The information that we collect from this study is used for writing scientific publications and will be reported at group level. It will be completely anonymous and it cannot be traced back to you. Only the researchers will know your identity and we will lock that information up with a lock and key.

Further information
If you want more information about this study you can ask Peter Frumau (contact email: p.f.w.frumau@student.tue.nl).
If you have any complaints about this study, please contact the supervisor, Antal Haans (A.Haans@tue.nl)

Certificate of Consent
I, (NAME)……………………………………….. have read and understood this consent form and have been given the opportunity to ask questions. I agree to voluntary participate in this research study carried by the research group Human Technology Interaction of the Eindhoven University of Technology.

I CONSENT / DO NOT CONSENT (encircle your preference) with the video recordings of my experimental sessions being used for public presentation purposes.

_________________________    ______________________
Participant’s Signature      Date

Participant’s paraph _____
B. Audio instructions

At start of experiment:
1. “Welkom in de virtuele omgeving, om aan deze omgeving te wennen, mag u nu eerst een aantal minuten rondlopen en rondkijken in deze eerste ruimte. Over een aantal minuten krijgt u verdere instructies.”

After around 2 minutes:
2. “Het experiment begint nu: Ik zal de deur voor u openen, loopt u daarna naar de tweede ruimte, en wacht op verdere instructies.”

After the participant walked through the door opening:
3. “Kijk nu naar de man aan de overkant van de ruimte. Blijf naar deze man kijken, tot verdere instructies volgen.”

After the ball was thrown:
4. “Kijk nu naar het gat in het midden van de ruimte. Loop nu naar de rand van het gat, en kijk in het gat.”

After around 3 seconds of looking down the pit:
5. “Ik zal nu het rooster voor u sluiten. Als het rooster gesloten is, mag u in het midden van het rooster gaan staan.”

After standing on top of the grille:

If participants were not looking down but pressed the button, a reminder to looking down was played:
7. “Kijkt u naar de bodem van het gat, en druk op de knop als u klaar bent om te vallen.”
Vragenlijst 1e sessie

De volgende stellingen gaan over mogelijke ervaringen tijdens de sessie in de **twee kamers**. Geef aan in welke mate u het eens bent met de gegeven stellingen.

- Aangezien iedereen een andere ervaring heeft tijdens de sessie, zijn er geen goede of foute antwoorden.
- Denk niet te lang na over de antwoorden, maar vul dat antwoord in dat het eerst in u opkomt.
- Let op: Bij sommige stellingen wordt aangegeven of het er om gaat of u de desbetreffende ervaring gedurende de hele sessie (**steeds**) of op zijn minst een keer (**sommige momenten**) hebt ervaren.
- U kiest “niet van toepassing” wanneer u geen antwoord kunt geven op een vraag. U kunt bijvoorbeeld geen antwoord geven op de vraag hoe het voelde om in het gat te stoppen wanneer u dat niet hebt gedaan.

<table>
<thead>
<tr>
<th></th>
<th>Oneens</th>
<th>Eens</th>
<th>Niet van toepassing</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Ik wist zeker dat ik mijn hoofd zou kunnen stoten aan de deurpost.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Ik had het gevoel dat ik de bal moest ontwijken.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>Ik had de gedachte dat ik echt zou kunnen vallen toen ik in het gat keek.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>Ik wist zeker dat de bal die gegooid werd me fysiek zou raken.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>Ik had het gevoel dat ik moest bukken om mijn hoofd niet te stoten aan de deurpost.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>Ik dacht even dat er een echte bal naar me werd gegooid.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>Het voelde enigszins ongemakkelijk toen ik naar de bodem van het gat keek.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td>Ik heb de bal gezien toen deze op mij af kwam.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
De volgende stellingen gaan over mogelijke ervaringen tijdens de sessie in de **twee kamers**. Geef aan in welke mate u het eens bent met de gegeven stellingen.

- Aangetekend iedereen een andere ervaring heeft tijdens de sessie, zijn er geen goede of foute antwoorden.
- Denk niet te lang na over de antwoorden, waarvan dat antwoord in dat het eerste in u opkomt.
- Let op: Bij sommige stellingen wordt aangegeven of het er gaat om of u de deskundige ervaring gedurende de hele sessie (**steeds**) of op zijn minst een keer (**sommige momenten**) hebt ervaren.
- U kiest "niet van toepassing" wanneer u geen antwoord kunt geven op een vraag. U kunt bijvoorbeeld geen antwoord geven op de vraag hoe het voelde om in het gat te stappen wanneer u dat niet hebt gedaan.

<table>
<thead>
<tr>
<th></th>
<th>Oneens</th>
<th>Eens</th>
<th>Niet van toepassing</th>
</tr>
</thead>
<tbody>
<tr>
<td>9. Toen ik gevraagd werd in het gat te vallen heb ik dat zonder aarzeling gedaan.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>10. Nadat de roosters openklapten, voelde het echt alsof ik viel.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>11. Er waren momenten tijdens de sessie waarin ik vergat dat ik me in een gesimuleerde omgeving bevond.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>12. Ik had steeds het gevoel dat de verschillende objecten die ik zag zich op die specifieke locatie bevonden in de echte omgeving om mij heen.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>13. Ik voelde een lichamelijke reactie toen ik in het gat keek.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>14. Op sommige momenten tijdens de sessie voelde het alsof ik mijn arm kon uitstrekken om de objecten in de virtuele omgeving aan te raken.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>15. Ik voelde een lichamelijke reactie bij het neerkomen op de bodem van het gat.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>16. Ik weet zeker dat de twee kamers echt waren en geen simulatie.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

**Sluit de bladzijde om voor de volgende vragen!**
De volgende vragen en stellingen gaan nogmaals over uw ervaringen tijdens de sessie. Bij elke vraag staan steeds zeventien blokjes die u kunt gebruiken om antwoord te geven. *Naast de blokjes* is een beschrijving van uiterste van de antwoordschaal gegeven. Kruis steeds het antwoordblokje aan dat het beste overeenkomt met uw beleving.

- Aangezien iedereen een andere ervaring heeft tijdens de sessie, zijn er geen goede of foute antwoorden.
- Denk niet te lang na over de antwoorden, maar wil dat antwoord in dat het eerste in u opkomt.

<table>
<thead>
<tr>
<th>Stem</th>
<th>Helemaal niet</th>
<th>Helemaal mee eens</th>
<th>Zeer bewust</th>
<th>Helemaal mee eens</th>
<th>Helemaal mee eens</th>
<th>Volledige overeenstemming</th>
</tr>
</thead>
<tbody>
<tr>
<td>17.</td>
<td>Ik had het gevoel aanwezig te zijn in de computerwereld.</td>
<td>Helemaal niet</td>
<td>Helemaal mee eens</td>
<td>Zeer bewust</td>
<td>Helemaal mee eens</td>
<td>Volledige overeenstemming</td>
</tr>
<tr>
<td>18.</td>
<td>Hoe bewust was u zich van de echte omgeving (bv. geluiden van buiten, kamertemperatuur), terwijl u zich bevond in de virtuele ruimte?</td>
<td>Helemaal niet bewust</td>
<td>Helemaal mee eens</td>
<td>Zeer bewust</td>
<td>Helemaal mee eens</td>
<td>Volledige overeenstemming</td>
</tr>
<tr>
<td>19.</td>
<td>Ik voelde me aanwezig in de virtuele ruimte.</td>
<td>Helemaal mee eens</td>
<td>Helemaal mee eens</td>
<td>Zeer bewust</td>
<td>Helemaal mee eens</td>
<td>Volledige overeenstemming</td>
</tr>
<tr>
<td>20.</td>
<td>Ik had het gevoel slechts speelstukjes te aanschouwen.</td>
<td>Helemaal mee eens</td>
<td>Helemaal mee eens</td>
<td>Zeer bewust</td>
<td>Helemaal mee eens</td>
<td>Volledige overeenstemming</td>
</tr>
<tr>
<td>21.</td>
<td>Ik lette nog op de echte omgeving.</td>
<td>Helemaal mee eens</td>
<td>Helemaal mee eens</td>
<td>Zeer bewust</td>
<td>Helemaal mee eens</td>
<td>Volledige overeenstemming</td>
</tr>
<tr>
<td>22.</td>
<td>In hoeverre kwam uw ervaring in de virtuele omgeving overeen met uw ervaringen in de echte wereld?</td>
<td>Geen overeenstemming</td>
<td>Helemaal mee eens</td>
<td>Zeer bewust</td>
<td>Helemaal mee eens</td>
<td>Volledige overeenstemming</td>
</tr>
<tr>
<td>23.</td>
<td>Ik had meer het gevoel bezig te zijn in de virtuele ruimte, dan dat ik het gevoel had iets van buitenaf te bedienen.</td>
<td>Helemaal mee eens</td>
<td>Helemaal mee eens</td>
<td>Zeer bewust</td>
<td>Helemaal mee eens</td>
<td>Volledige overeenstemming</td>
</tr>
</tbody>
</table>

Sla de bladzijde om voor de volgende vragen!
<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>24. Hoe echt kwam de virtuele omgeving op u over?</td>
<td>helemaal niet echt</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>25. Ik had het gevoel omgeven te zijn door de virtuele wereld.</td>
<td>helemaal mee oneens</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>26. Ik was me niet bewust van mijn echte omgeving.</td>
<td>helemaal mee oneens</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>27. Hoe werkelijk kwam de virtuele wereld op u over? zoals een denkbeeldige wereld</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>niet te onderscheiden van de echte wereld</td>
</tr>
<tr>
<td>28. Ik had niet het gevoel in de virtuele ruimte aanwezig te zijn.</td>
<td>helemaal mee oneens</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>29. De virtuele wereld kwam echter op mij over dan de werkelijke wereld.</td>
<td>helemaal mee oneens</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>30. Ik ging volledig op in de virtuele wereld.</td>
<td>helemaal mee oneens</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

Sla de bladzijde om voor de volgende vragen!
C.2. Additional questions
Lengte: ………………… cm

<table>
<thead>
<tr>
<th>Ik ben...</th>
<th>…een man</th>
<th>…een vrouw</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>□</td>
<td>□</td>
</tr>
</tbody>
</table>

Ik ben ………………… jaar oud

<table>
<thead>
<tr>
<th>Wat is uw ervaring met dit soort van virtuele omgevingen?</th>
<th>□ Dit was de 1ste keer in een virtuele omgeving</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>□ Dit was de 2de of 3de keer in een virtuele omgeving</td>
</tr>
<tr>
<td></td>
<td>□ Dit was de 4de of 5de keer in een virtuele omgeving</td>
</tr>
<tr>
<td></td>
<td>□ Ik ben 6 keer of vaker in een virtuele omgeving geweest</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hebt u last van hoogtevrees? Beoordeel de ernst van uw hoogtevrees zoals u deze ervaart in alledaagse omstandigheden.</th>
<th>geen hoogtevrees</th>
<th>□</th>
<th>□</th>
<th>□</th>
<th>□</th>
<th>□</th>
<th>zeer hoogtevrees</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>
D. Specifications Regarding the Virtual Reality System

VirTU/e lab - Virtual Reality Lab

The VirTU/e Lab is situated in room 0.54 of the IPO-building. It is a virtual reality facility with a camera positioning system and head mounted display (HMD). Participants may move around in a 7x6 m² space. The interaction with virtual environments is controlled with the WorldViz Vizard software platform.

Available hardware

The lab consists of the following hardware components:

- nVisor SX111 head-mounted display: A display with separate LCD screens in front of both participant's eyes for stereoscopic viewing of 3D virtual environments. The nVisor SX111 is a wide field-of-view (FOV) head-mounted display featuring dual SXGA (1280×1024 pixels) displays with 76Hx64V degrees FOV per eye, total of 111 degree FOV. The HMDs video control unit is attached to a backpack. Replaceable battery packs and a separate battery loader enable participants to move around without cable connections. The SX111 is equipped with a sound option, i.e. a set of head phones and an electret condenser microphone that can be connected to the notebook.

- InterSense InertiaCube3 orientation tracker: The InertiaCube3 is a small inertial orientation reference system. Providing full 360° sourceless tracking in all axes, the InertiaCube3 integrates nine discrete, miniature sensing elements with advanced Kalman filtering algorithms. It is connected to the notebook with a USB interface. An InterSense plugin is available for the Vizard VR toolkit software (see below).

- PhaseSpace IMPULSE position tracker: An optical position tracker consisting of 8 linear detector based cameras (3600 x 3600 resolution at 480 Hz) in combination with active LED markers with a unique ID each. Typically one or more LED markers are attached to the HMD in order to track the participant's head position. When multiple LED markers are used on the HMD, a rigid body tracker can be defined for which it is also possible to track the orientation. The IMPULSE server machine connects to the cameras via an integrated HUB and wirelessly connects to the LED driver(s). Up to 48 LED markers can be tracked simultaneously. The position data is streamed in real-time to multiple clients over the network. A PhaseSpace plugin is available for the Vizard VR toolkit software (see below).

- Dell Alienware gaming notebook: A powerful notebook to be attached to the participant's backpack. Vizard VR applications run on this notebook. Position and orientation data are used to update the visual display.

- Dell XPS desktop PC: This PC is used by the experimenter for remote controlling the Alienware notebook via TeamViewer (see below).

Available software

In the VirTU/e lab the following software components are used:

- WorldViz Vizard virtual reality toolkit: Vizard is an easy-to-learn 3D development platform complete with everything you need to build interactive and immersive 3D content. Students may download a demo Lite version of Vizard at the WorldViz site and use it for free during 90 days. Staff members should contact the lab support team for a license. There are 2 kinds of license available: a couple of Lite licenses and one Enterprise license which is installed on the Dell Alien notebook that is used in the VirTU/e lab in combination with the VR hardware components. The Lite license has some restrictions: Full screen displaying is not possible and a Vizard logo is permanently visible. It is however a perfect tool for developing VR applications. Beginning Vizard users are encouraged to follow the free downloadable tutorial Vizard Teacher in a Book. Since in Vizard the Python scripting language is used it may also be wise to get more familiar with it. Numerous Python tutorials are available on the internet.

- TeamViewer: This tool connects the experimenter's PC to the participant's notebook via the wireless network. You can remote control the notebook while the participant is carrying it in the backpack.
Declaration concerning the TU/e Code of Scientific Conduct for the Master's/PDEng/PhD thesis

I have read the TU/e Code of Scientific Conduct.

I hereby declare that my Master's/PDEng/PhD-thesis has been carried out in accordance with the rules of the TU/e Code of Scientific Conduct.

Date 10/03/2016

Name Peter Fruman

Signature

See: http://www.tue.nl/en/university/about-the-university/integrity/scientific-integrity/
The Netherlands Code of Conduct for Academic Practice of the VSNU can be found here also. More information about scientific integrity is published on the websites of TU/e and VSNU.

July 2 2015