Computers in Human Behavior 72 (2017) 276–285

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Full length article

Development and validation of the Multimodal Presence Scale for virtual reality environments: A confirmatory factor analysis and item response theory approach

Guido Makransky*, Lau Lilleholt, Anders Aaby

Department of Psychology, University of Southern Denmark, Denmark

ARTICLE INFO

Article history:
Received 16 December 2016
Received in revised form
17 February 2017
Accepted 27 February 2017
Available online 28 February 2017

Keywords:
Presence
Virtual reality
Confirmatory factor analysis
Item response theory
Virtual simulations

ABSTRACT

Presence is one of the most important psychological constructs for understanding human-computer interaction. However, different terminology and operationalizations of presence across fields have plagued the comparability and generalizability of results across studies. Lee's (2004) unified understanding of presence as a multidimensional construct made up of physical, social, and self-presence, has created a unified theory of presence; nevertheless, there are still no psychometrically valid measurement instruments based on the theory. Two studies were conducted that describe the development of a standardized multidimensional measure of presence (the MPS) for a VR learning context based on this theory, and its validation using confirmatory factor analysis and item response theory. The results from Study 1 which included 161 medical students from Denmark indicated that the items used in the MPS measure a three dimensional theoretical model of presence: physical, social, and self-presence. Furthermore, IRT analyses indicated that it was possible to limit the number of items in the MPS to 15 (five items per sub-dimension) while maintaining the construct validity and reliability of the measure. The results of Study 2, which included 118 biology students from Scotland, supported the validity and generalizability of the MPS in a new context.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

The construct of presence has great practical relevance for the design and evaluation of media products, especially in education, entertainment, telecommunications, psychology, and health care (Klimmt & Vorderer, 2003; Lee, 2004). Consequently, several scholars have attempted to define the concept of presence in order to measure and operationalize it as a psychological construct. There is some consensus regarding defining presence as the experience or feeling of being present in a mediated environment, rather than the immediate physical environment wherein one is currently bodily present (Steuer, 1992; Witmer & Singer, 1998). However, different fields of research have typically used different terminology (Lee, 2004), which has made it difficult for researchers to create a unified theory of presence, thereby creating difficulties in comparing and generalizing results across studies. However, through an extensive explication process, Lee (2004) made a thorough attempt to provide a unified understanding and general terminology of presence which had heretofore been lacking.

Although there has been great progress in further understanding presence as a psychological construct since this development, one area where the progress has stagnated is on the operationalization and measurement of presence according to Lee's (2004) definition of the construct. This is surprising, considering Lee's emphasis on the importance of creating standard validated measures of presence so that comparison between and among studies, and generalization across studies, would be possible. With the increasing availability of sophisticated technologies for simulating interactions between people and places such as Virtual Reality (VR), the need among researchers and media content developers for a validated standardized measure of presence, using a unified theory such as Lee's (2004) to create such measures seems crucial.

The objective of this study was therefore to develop a standardized measure of presence for VR environments based on Lee's (2004) explication of presence, and to validate this measure using modern test theory; that is Confirmatory Factor Analysis (CFA; Brown, 2015) and Item Response Theory (IRT; Embertson & Reise, 2000). Accordingly, the scope of this study was not to develop a

* Corresponding author. Campusvej 55, DK-5230, Odense M, Denmark.
E-mail address: gmakransky@health.sdu.dk (G. Makransky).

http://dx.doi.org/10.1016/j.chb.2017.02.066
0747-5632/© 2017 Elsevier Ltd. All rights reserved.

measure of presence to be used when actual people and places that are physically separated are brought together through technology, but rather to develop a presence scale pertaining to computer simulations and other programs employing VR. Furthermore, the aim of this study was not to validate Lee's (2004) theory of presence, but to validate a standardized measure of presence based on this theory.

This paper describes the development of the Multimodal Presence Scale (MPS), and the evaluation of its psychometric properties with CFA and IRT. First, a brief introduction to Lee's (2004) explication of presence and an outline of previous measures of presence is presented, followed by a description of the development of the MPS based on existing acknowledged measures of presence. Next, CFA and IRT are explained. This is followed by a description of two studies wherein these methodologies were used to test the psychometric properties of the MPS. Finally, the strengths and weaknesses of the MPS and practical considerations involving its use are discussed.

1.1. Lee's (2004) explication of presence and previous operationalization of presence

According to Lee (2004), both real and virtual experience can be divided into three domains; physical, social, and self. Physical experience involves the experience of physical objects and environments; social experience refers to the experience of social actors; and self-experience describe the experiences people have of their own selves.

In contrast to real experience, there are two ways in which an experience can become virtual (Lee, 2004). Firstly, an experience can become virtual when the act of experience is mediated by, or made possible by, human-made technology and the technology enables users to experience mediated versions of actual entities which hold some kind of valid connection with the actual entities that they represent (para-authentic entities). Secondly, an experience can also become virtual when experienced entities are artificially created or simulated by human-made technology. In this latter case the entities do not actually exist in the real world, but are experienced as if they would exist in the real world due to human-made technology (artificial entities). In sum, an experience can become virtual when the act of experiencing physical objects or environments, social actors and/or the self is mediated, or artificially constructed, by human-made technology.

In addition to the before-mentioned aspects of real and virtual experience, Lee (2004) also distinguishes between sensory and non-sensory experiences. By including this distinction in his definition of presence Lee (2004) is able to incorporate the possibility of experiencing presence during the use of low-tech non-sensory media, such as books.

Based on the outlined theoretical framework, Lee (2004) defines physical, social and self-presence as a psychological state in which virtual (para-authentic or artificial) physical objects, social actors, and the self, respectively, are experienced as actual entities in either sensory or non-sensory ways. However, since the focus of this paper is on VR environments, the distinction between sensory and non-sensory experiences is not relevant here, as VR environments are sensory in nature.

The development of the MPS is based on Lee's (2004) theory of presence, specifically in the division of presence into three sub-dimensions; physical, social and self. However, since Lee (2004) does not provide a specific operationalization of physical, social, or self-presence, it is necessary to review the available literature of some of the most acknowledged previous measures of presence to develop adequate content validity for these three constructs.

Through a review of the literature, it is evident that there are various acknowledged measures of presence available; and many of these capture some aspects of each of the three constructs: physical, social, and self-presence. However, no attempts have been made to measure all three together and none of these measures have been reported in the literature to have been validated using both CFA and IRT. The Igroup Presence Questionnaire (IPQ) developed by Schubert, Friedmannm, and Regenbrecht (2001) did use CFA to identify a general structure, but without the utilization of IRT they lack information on how well the specific items function within each sub-dimension. More importantly, the IPQ fits entirely within the realm of physical presence, and is therefore unable to measure either social or self-presence as opposed to the MPS.

Considering the vast amount of acknowledged measures of presence, the MPS will consist of previously developed items that are selected through an extensive extraction process and validated through psychometric testing with CFA and IRT.

2. Method

The first step in the process of developing the MPS was to review numerous existing presence scales. The databases Scopus, PsycInfo and PsycTest were searched with the keywords “presence”, “virtual environment” and “virtual reality”. The search was limited to include only English written journal articles. Of the 16 articles selected, 13 were identified through the database search while three articles were identified by searching through reference lists. Studies with items that were not easily converted to fit a VR context or where items did not fit this specific context were excluded. Furthermore, studies measuring presence directly by asking respondents how present they felt on a single numerical analogue scale were not included, as this kind of measurement scale only measures the overlying construct and not the sub-dimensions relevant to this study. Lastly, studies using presence scales or subscales of previously identified scales were excluded from this study. A list of the scales that were identified and used in this study is presented in Appendix 1. Key aspects of physical, social, and self-presence were identified in these existing instruments.

2.1. Key aspects of presence

In order to provide adequate content validity for the MPS all of the available items from the existing instruments were considered (see Appendix 1). A thorough review of all these items identified common themes by grouping similar themed items and giving this group a label (referred to as area attributes throughout this paper). A small group of items were excluded through this review as they measured antecedents of experiencing presence (e.g. time spent creating an avatar or personality traits) and not the actual experience of presence. Through this extensive extraction process, 13 key aspects of presence relevant to a VR context were identified and categorized into physical, social and self-presence. This hypothesized model was not set in stone, but was an initial theoretically based categorization needed to perform the CFA. The key aspects of presence are summarized in Table 1.

2.1.1. Key aspects of physical presence

The first physical area attribute is “physical realism”, which describes the extent to which users experience the virtual environment as mimicking the physical appearance and causal relationships of the real world. The second area attribute is “not paying attention to the real world environment”. This refers to the extent to which users are completely captivated by the virtual world and, therefore, become less aware of the real world in which they actually exist. The third area attribute is “sense of control in the virtual environment”, which describes the extent to which

Table 1
Content validity – Key aspects of presence.

Sub-dimension	Area attribute
Physical presence	Physical realism (PR)
	Not paying attention to real environment (NARE)
	Control/act in the virtual environment (CA)
	Sense of being in the virtual environment (SBVE)
	Not aware of the physical mediation (NAPM)
Social presence	Sense of coexistence (SC)
	Human realism (HR)
	Not aware of the artificiality of social interaction (NAASI)
	Not aware of the social mediation (NASM)
Self-presence	Sense of bodily connectivity (SBC)
	Sense of bodily extension (SBE)
	Emotional connectivity (EC)
	Sense of self being in the virtual environment (SSBVE)

users have a sense of being able to act in the virtual environment, to thereby control events and actively search for and manipulate objects. The fourth area attribute is "sense of being in the virtual environment", which describes the extent to which users experience a general and intuitive sense of being in the virtual environment. The fifth area attribute is "not being aware of the physical mediation". This is defined as the extent to which users are unaware of the process by which the physical environment is mediated. Attention is focused on manipulating and navigating the environment and not on the interface for doing the manipulation and navigation.

2.1.2. Key aspects of social presence

The first social area attribute is "sense of coexistence", which refers to the extent to which the user experiences a general and intuitive sense of being in the presence of another person also in the virtual environment. The second area attribute is "human realism", which describes the extent to which the user experiences the avatar representations of humans in the virtual environment as credible, and not just computerized images. The third area attribute is "not being aware of the artificiality of social interaction", which refers to the extent to which the user experiences interacting with a human being rather than with the computer simulation. The fourth area attribute of social presence is "not being aware of the social mediation". This is defined as the extent to which the user is unaware of the process by which the social interaction is mediated. Attention is focused on the verbal and non-verbal communication and not the interface through which it is realized.

2.1.3. Key aspects of self-presence

The first area attribute of self-presence is "sense of self being in the virtual environment", which describes the extent to which users experience an intuitive sense of their self actually being in the virtual environment. The second area attribute is "sense of bodily connectivity". This refers to the extent to which users experience a connection between their real and virtual body, thereby experiencing their real body and virtual embodiment as one and the same thing. The third area attribute is "sense of bodily extension", which describes the extent to which users experience their body being extended through a medium into the virtual world. The fourth area attribute of self-presence is "emotional connectivity", which refers to the extent to which users have certain emotional experiences as a result of corresponding events happening to their virtual embodiment (e.g. when sad events happen to their virtual embodiment in the virtual world, users experience sadness). In Ratan and Hasler's (2009) terminology the first three area attributes would be characterized as proto self-presence, while emotional connectivity would be referred to as core self-presence. In their work, a third aspect termed extended self-presence is also described, but this was not included in the MPS as it measures possible antecedents of self-presence rather than the actual experience of self-presence (e.g. "to what extent have you customized your avatar to make it look the way it does?").

2.2. Item selection and adaptation

The item selection process was based on the previous work of identifying common themes and thereby extracting the 13 key aspects (area attributes) of presence described above. The literature sources used to identify items, the area attributes they covered, and the number of items used are described in Appendix 1. Of the 16 studies initially selected, items were only obtained from six of these. This does not mean that items from the remaining scales were not a part of the extraction process described in Section 2.1. Due to many of the studies containing either exactly the same items or differently worded items with the same underlying theme, items were selected from the least number of studies as possible without missing any of the 13 key aspects of presence or any potentially important wording differences within a given key aspect. Multiple items were included for each area attribute when they had a different focus. However, only one item was selected in instances where several items assessed the same content using different terminology. If necessary, items were modified in order to fit with a five-point Likert scale (1 = completely disagree, 2 = disagree, 3 = neither disagree nor agree, 4 = agree, 5 = strongly agree); and a small minority of items were further modified in order to better fit within a variety of VR contexts (e.g. "avatar" was changed to "virtual embodiment" to make the MPS useful in simulations that do not let users create their own avatar). With regards to the number of items initially selected, 21 were selected to capture physical presence; 7 were selected to capture social presence; and 12 were selected to capture self-presence (see Appendix 2 for a full list of items).

2.3. Confirmatory factor analysis and item response theory

In this study confirmatory factor analysis (CFA) and item response theory (IRT) were used to investigate the psychometric properties of the initial MPS. CFA is optimal for investigating the construct validity of a scale when there is a strong theoretical hypothesis about the structure of the scale; and it is often regarded as a stronger source of evidence compared to more exploratory approaches in scale validation (e.g., Fabrigar, Wegener, MacCallum, & Strahan, 1999). In this case, CFA is used to investigate if the structure of the items selected from previous literature (see Appendix 2) measures the three sub-dimensions of physical, self, and social-presence as described by the theoretical framework suggested by Lee (2004). Once a general structure is identified, IRT is optimal for investigating the quality of each item within each sub-dimension. IRT is a family of measurement models which has been denominated the "measurement paradigm of the 21st century" (e.g. Hays, Morales, & Reise, 2000; Ware, 2003) because it provides detailed information about the validity of a measurement instrument based on the investigation of whether the instrument lives up to a set of assumptions. These assumptions include unidimensionality (that the items in a scale measure only one latent trait), local independence (that there is no redundancy between items), item fit (the items all measure the latent trait in a consistent way), measurement invariance or differential item functioning (DIF; item estimation is independent of the sub-groups of individuals completing the measure) (Bond & Fox, 2001).

Therefore, the use of CFA and IRT complement each other well because CFA can be used to identify a general structure of a measurement instrument; and once that structure has been identified

IRT can provide more detailed information about how well the specific items function within each sub-dimension. These methodologies are becoming increasingly used in a wide variety of fields within science (Griffith et al., 2009; Reise, Widaman, & Pugh, 1993).

2.4. Statistical analyses

Descriptive statistics were calculated using IBM SPSS software (version 23.0.0; SPSS Inc., Chicago, IL). CFA analysis to investigate the dimensionality of the MPS were conducted in Mplus version 7 (L. K. Muthén & Muthén, 2012) using polychoric correlations. Reported goodness-of-fit indices include the comparative fit index (CFI), the Tucker-Lewis index (TLI), and the root mean square error of approximation (RMSEA). An acceptable fit is indicated by CFI and TLI $\geq$0.90, and RMSEA $\leq$0.06 (Hu & Bentler, 1999). The estimation method used is Muthén's three-step procedure (B. Muthén, 1984).

The Partial Credit Model (PCM; Masters, 1982) within the framework of IRT was used to investigate the psychometric properties of each sub-scale within the MPS. The PCM is optimal for polytomous data such as the five-point Likert scale used in this study. Analyses were conducted with RUMM2030 (Andrich, Sheridan, & Luo, 2010). The evaluation criteria applied were unidimensionality, local dependence, item fit, ordered response thresholds, measurement invariance (DIF) and reliability as evaluated by the person separation index (PSI) and Cronbach's alpha.

The evaluation criteria have been explained elsewhere and are only described briefly here (for more information see Pallant & Tennant, 2007; Tennant & Conaghan, 2007; Makransky, Rogers, & Creed, 2015). Unidimensionality was evaluated according to a formal test proposed by Smith (2002). This test uses the first residual factor in a principal components analysis (of residuals) to determine two groups of items: those with positive and those with negative residuals. Each set of items is then used to calculate an independent trait estimate for each person in the sample. When items form a unidimensional scale, it is expected that the person estimates from the 2 item subsets should be similar. An independent samples *t*-test is used to determine whether there is a significant difference between the two person estimates. This is repeated for each person with the expectation that the percentage of tests lying outside the range of -1.96 to 1.96 should not exceed 5% (Makransky et al., 2015).

The individual items fit of each item according to the expectations of the model was evaluated based on fit residuals larger than $\pm$ 2.5 (Pallant & Tennant, 2007). Local dependence (which can indicate redundancy among items in a scale) was assessed by investigating if residual correlations between the items in the scale were larger than a critical value. A critical value of 0.2 is widely used in the literature (Chen & Thissen, 1997; Makransky & Bilenberg, 2014). The suitability of the five-point Likert-type format response categories was investigated by assessing if there were any reverse thresholds (e.g. Pedersen, Mathiasen, Christensen, & Makransky, 2016). Finally, measurement invariance in the form of DIF across the demographic variable of gender for each of the items in the MPS was investigated. For example, DIF occurs when different subgroups within the sample (e.g., men versus women) have different scores on a specific item despite equal levels of the latent trait (e.g., physical presence). Items with significant Chi-square statistics at the 0.05 level (two-sided and with a Bonferroni correction applied separately within each DIF-variable) are reported as exhibiting DIF in this study.

3. Study 1

The objectives of Study 1 were to: 1) investigate the validity and reliability of the MPS by assessing if the items fit the hypothesized theoretical structure described by Lee (2004), and 2) to select the items within each dimension that would result in a short manageable scale, while simultaneously retaining the content validity (see Appendix 2) and construct validity (fit to the PCM) of the instrument.

3.1. Sample and procedure

The sample consisted of 161 (65% female) students from the University of Copenhagen who participated in an undergraduate genetics course in the spring 2016 semester. The sample participated in a learning session that consisted of a 30-min pre-test that included demographic characteristics and questions to assess students' knowledge of genetics; a 2-h session for the medical genetics virtual laboratory simulation; and a 30-min post-test which included the presence items from the initial MPS presented in Appendix 2 and the same knowledge questions about genetics presented in the pre-test (the knowledge questions were used by the teachers of the course as a measure of progress, but are not used in this study).

The laboratory simulation was a desktop virtual reality version of a medical genetics simulation developed by the simulation development company Labster. It was designed to facilitate learning within the field of genetics at a university level by allowing the user to virtually work through the procedures in a lab by using and interacting with the relevant lab equipment and by teaching the essential content through an inquiry-based learning approach (Bonde et al., 2004; Makransky et al., 2016). In this simulation, students are introduced to a young pregnant couple, where the fetus may suffer from a syndrome caused by a chromosomal abnormality. The students are able to make a genome wide analysis of the fetal DNA and karyotype in the virtual laboratory, and practice communicating their conclusions to the couple using a simulated genetic counseling approach (for more information on the simulation see Makransky et al., 2016; Labster, 2016a).

Students were informed that the data could be used for research, and care was taken not to expose the students to any risk or burden (Helsinki Declaration article 17). While the simulation was a mandatory part of the curriculum, the presence scale was voluntary, and only data from students who gave permission to use their results was used. Responses were anonymous (Helsinki Declaration article 24) and no stipend was provided. The study protocol was submitted to the Regional Committees on Health Research Ethics for Southern Denmark, which indicated that no written consent was required by Danish law.

3.2. Results

3.2.1. Investigating the dimensionality of the initial MPS with CFA

The first set of analyses was conducted with CFA to determine the number of dimensions that best described the structure of the 40 items selected for the MPS. More specifically, the results of a unidimensional model where all of the items measure one single unidimensional scale are compared, with those of a three dimensional model based on Lee's (2004) theoretical model of physical, social, and self-presence.

There were two items that had a very bad fit to the model, which made it difficult to interpret the results (PHYS_15: "I was aware of the device {computer, VR-goggles} through which the virtual environment was displayed"; and PHYS_20: "I experienced delay between my actions and expected outcomes in the virtual environment"). These items were eliminated before conducting additional analyses. Fit of the remaining 38 items is shown in the top section of Table 2. The fit to a unidimensional scale indicated suboptimal fit (RMSEA: 0.114, CFI: 0.841, TLI: 0.832). The fit to the

Table 2
Confirmatory factor analysis results.

	RMSEA	CFI	TLI
Initial MPS (38 items)			
1 dimensional model	0.114	0.841	0.832
3 dimensions: physical, social, and self-presence	0.094	0.893	0.886
Revised MPS (15 items)			
1 dimensional model	0.139	0.918	0.904
3 dimensions: physical, social, and self-presence	0.094	0.964	0.956

three dimensional model with physical, social, and self-presence was better (RMSEA: 0.094, CFI: 0.893, TLI: 0.886). Table 3 shows the values of estimate, standard errors, z-value, and p-values of 38 items in original MPS from the three dimensional CFA.

3.2.2. Investigating the validity of the three sub-dimensions with the PCM

The validity of the items within each dimension (physical, social, and self-presence) was assessed according to the criteria described above for the PCM. There were 21 items hypothesized to be related physical presence, seven items to related to social presence, and 12 items related to self-presence (from Appendix 2). Each group of items were subject to the following tests: general fit of the PCM; unidimensionality test, to investigate if the items measure only one latent trait; local dependence, to ensure that there is no dependence or redundancy between the items in the scale; investigation of reverse thresholds to ensure that the five-point Likert response format functioned well for all items; DIF or measurement invariance across gender; and reliability in the form of PS and Cronbach's alpha. Since the objective was to develop a short scale that had content and construct validity, as well as reliability, items were eliminated based on the criteria for the PCM. One assumption that is tested when evaluating the fit of a scale to the PCM is sufficiency (Rasch, 1960), this means that any subset of items from the scale would also meet the requirements of the PCM and thus maintain the construct validity of the scale. Therefore, even when fit to the PCM was obtained, items were eliminated in order to have as short a scale as possible. This was only done if the elimination of items did not cause a substantial decrease in reliability or content validity. The final scale with 15 items is presented in Table 4.

Table 3
The CFA estimates, standard errors, Z-, and P-values for the 38 items in the original MPS.

	Estimate	S.E.	Est./S.E.	P-value (2 tailed)
Physical Presence				
PHYS_1	0.255	0.080	3.177	0.001
PHYS_2	0.451	0.067	6.699	0.000
PHYS_3	0.766	0.041	18.578	0.000
PHYS_4	0.774	0.041	18.702	0.000
PHYS_5	0.708	0.042	16.931	0.000
PHYS_6	0.839	0.029	28.749	0.000
PHYS_7	0.799	0.034	23.665	0.000
PHYS_8	0.702	0.042	16.734	0.000
PHYS_9	0.372	0.072	5.170	0.000
PHYS_10	0.579	0.062	9.406	0.000
PHYS_11	0.717	0.045	15.875	0.000
PHYS_12	0.423	0.070	6.065	0.000
PHYS_13	0.334	0.073	4.555	0.000
PHYS_14	0.558	0.059	9.386	0.000
PHYS_16	−0.270	0.080	−3.396	0.001
PHYS_17	0.193	0.086	2.236	0.025
PHYS_18	0.422	0.071	5.987	0.000
PHYS_19	0.479	0.061	7.839	0.000
PHYS_21	0.155	0.072	2.151	0.031
Social Presence				
SOC_1	0.693	0.047	14.803	0.000
SOC_2	0.678	0.048	14.066	0.000
SOC_3	0.721	0.041	17.463	0.000
SOC_4	0.793	0.033	24.002	0.000
SOC_5	0.717	0.046	15.576	0.000
SOC_6	0.676	0.044	15.422	0.000
SOC_7	0.806	0.034	23.470	0.000
Self Presence				
SELF_1	0.758	0.033	22.950	0.000
SELF_2	0.806	0.030	27.160	0.000
SELF_3	0.827	0.026	31.262	0.000
SELF_4	0.845	0.021	40.641	0.000
SELF_5	0.882	0.020	44.410	0.000
SELF_6	0.756	0.033	23.169	0.000
SELF_7	0.909	0.017	52.891	0.000
SELF_8	0.801	0.030	26.858	0.000
SELF_9	0.854	0.019	44.502	0.000
SELF_10	0.882	0.017	50.689	0.000
SELF_11	0.817	0.024	34.096	0.000
SELF_12	0.857	0.020	42.615	0.000

The results for physical presence were consistent with the results from the CFA in that the same two items PHYS_15 and PHYS_20 had a very bad fit to the model, which made it difficult to interpret the results. Therefore, these items were eliminated in an initial step before conducting additional analyses. The remaining 19 items did not fit the PCM χ^2 *(38)* $= 15.04$, $p < 0.001$. Acceptable fit was obtained with an eight item scale that included four of the five area attributes of physical presence outlined in Table 1. The results showed that the items that were intended to measure the area attribute of control/act in the virtual environment did not measure the same latent construct as the other items in the measure. Furthermore, the number of items could be decreased to 5 without compromising the reliability, the content, or the construct validity of the scale, so three additional items were eliminated for area attributes that had multiple items.

The final scale with five physical presence items met most of the requirements of the PCM as shown in Table 5. That is, the scale had good general fit to the PCM χ^2 *(10)* $= 14.38$, $p = 0.16$. Furthermore, all of the five items fit the model; there were no items with local dependence, reverse thresholds, or DIF across gender. The test for unidimensionality indicated that there were 5.70% significant t-tests, which is slightly above the critical value of 5%, and indicates that there could be an issue regarding the dimensionality of the scale. Furthermore, four of the five area attributes were retained. Finally, the PSI and Cronbach's alpha reliability indices were good, with values of 0.85 and 0.84 respectively.

The results for the social presence scale showed that there was acceptable general fit to the model χ^2 *(14)* $= 16.98$, $p = 0.26$. However, items SOC_4 ("I perceived the people in the virtual environment as being only computerized images, not real people") and SOC_6 ("When I think about my experience in the virtual environment, I remember it as more like interacting with a computer than working with another person") did not function optimally. The two items were consequently deleted, resulting in a five item scale that retained the four area attributes hypothesized to be a part of the social presence construct (see Table 4).

The final scale with five social presence items met most of the requirements of the PCM as shown in Table 5. That is, the scale had good general fit to the PCM χ^2 *(10)* $= 10.82$, $p = 0.37$, all of the five items fit the model, there was no local dependence, or reverse thresholds. Item 2 did exhibit non-uniform DIF across genders, and the test for unidimensionality indicated that there were 5.70% significant t-tests, which is slightly above the critical value of 5%, indicating that there could be an issue regarding the dimensionality of the scale. Finally, the PSI and Cronbach's alpha reliability indices were acceptable, with values of 0.79 and 0.83 respectively.

The results for the self-presence showed that the 12 items did not function well as a scale. The first major issue was that there was strong local dependence between the items measuring core self-

Table 4
Final selection of items for the MPS and designation of area attribute.

Label	Item	Area attribute
Physical Presence		
PHYS_2	The virtual environment seemed real to me.	PR
PHYS_3	I had a sense of acting in the virtual environment, rather than operating something from outside.	NAPM
PHYS_4	My experience in the virtual environment seemed consistent with my experiences in the real world.	PR
PHYS_5	While I was in the virtual environment, I had a sense of "being there".	SBVE
PHYS_10	I was completely captivated by the virtual world.	NPARE
Social Presence		
SOC_1	I felt like I was in the presence of another person in the virtual environment.	SC
SOC_2	I felt that the people in the virtual environment were aware of my presence.	HR
SOC_3	The people in the virtual environment appeared to be sentient (conscious and alive) to me.	HR
SOC_5	During the simulation there were times where the computer interface seemed to disappear, and I felt like I was working directly with another person.	NASM
SOC_7	I had a sense that I was interacting with other people in the virtual environment, rather than a computer simulation.	NAASI
Self-presence		
SELF_2	I felt like my virtual embodiment was an extension of my real body within the virtual environment.	SBE
SELF_3	When something happened to my virtual embodiment, it felt like it was happening to my real body.	SBC
SELF_4	I felt like my real arm was projected into the virtual environment through my virtual embodiment.	SBE
SELF_6	I felt like my real hand was inside of the virtual environment.	SBC
SELF_7	During the simulation, I felt like my virtual embodiment and my real body became one and the same.	SBC

Note. Physical realism (PR), not paying attention to real environment (NARE), sense of being in the virtual environment (SBVE), not aware of the physical mediation (NAPM), sense of coexistence (SC), human realism (HR) not aware of artificiality of social interaction (NAASI), not aware of the social mediation (NASM), sense of bodily connectivity (SBC), sense of bodily extension (SBE).

Table 5
PCM results of the items selected for the final MPS.

	Chi square fit	PS reliability	Cronbach's alpha	Item fit	Thresholds	LD	DIF gender	Uni-Dimensionality
Physical presence	0.16	0.85	0.84	All OK	All OK	None	No	5.59%
Social presence	0.37	0.79	0.83	All OK	All OK	None	SOC_2	5.70%%
Self-presence	0.21	0.85	0.93	All OK	All OK	None	SELF_4	4.43%

presence and proto self-presence indicating that the items measured two different latent traits. Another problem with the scale was that item SELF_1 ("I felt physically close to objects and other characters in the virtual environment") did not measure the same latent trait as the other items. A closer investigation of this item indicated that the item was highly correlated with the latent trait of physical presence, suggesting either that the area attribute it was intended to measure (self being in the virtual environment), or the item itself, was more related to physical presence than self-presence. In order to investigate if this area attribute could be retained in self-presence, supplemental analyses were conducted to investigate whether the items intended to measure the area attribute of sense of being in the virtual environment from the physical presence scale, were actually better suited for self-presence. These were: PHYS_5 ("While I was in the virtual environment, I had a sense of 'being there'"), PHYS_6 ("I felt present in the virtual environment") and PHYS_7 ("I felt that the virtual environment surrounded me"). The results from the CFA and PCM clearly indicated that these items were better measures of physical presence than of self-presence.

Therefore, the empirical results indicated that the items which were intended to measure the area attributes of sense of self being in the virtual environment and emotional connectivity, did not form a unidimensional scale with the items intended to measure sense of bodily connectivity and sense of bodily extension. A large number of iterative analyses were then conducted to investigate if it was possible to include some of the items from the two area attributes without any positive results. The best five of the remaining six items that measured the area attributes of bodily connectivity and sense of bodily extension were then retained to form the final self-presence scale (see Table 4).

The final scale with 5 self-presence items met most of the requirements of the PCM as shown in Table 5. That is, the scale had good general fit to the PCM $\chi^2\ (10) = 13.25$, $p = 0.21$, all of the five items fit the model, the scale was unidimensional, there was no local dependence between the items, or reverse thresholds. Item 4 did exhibit non-uniform DIF across genders. Finally, the PSI and Cronbach's alpha reliability indices were good with values of 0.85 and 0.93 respectively.

3.2.3. *Investigating the dimensionality of the final MPS with CFA*

A new CFA with the revised MPS with 15 items (5 per sub-dimension) was then conducted to assess if the structure of the MPS was retained in its new form. The results that are shown in the bottom of Table 2 indicate that the hypothesized three dimensional model had acceptable and better fit (RMSEA: 0.094, CFI: 0.964, TLI: 0.956) than the unidimensional model (RMSEA: 0.139, CFI: 0.918, TLI: 0.904). The final three dimensional CFA is shown in the left panel of Fig. 1. The largest difference between the loadings was for the factor loading of PHYS_2 ("The virtual environment seemed real to me") which was only 0.424 in Study 1, but 0.860 in Study 2. The results could be due to differences in the sample characteristics, or the representational fidelity of the two simulations.

3.3. *Conclusions of study 1*

The results from Study 1 indicate that the items used in this study measure the three dimensional theoretical model of presence: physical, social, and self-presence as described in Lee (2004). Furthermore, IRT analyses indicated that it was possible to limit the number of items to 15 in the MPS (five items per sub-dimension) while maintaining the construct validity of the measure. A final CFA indicated acceptable fit of the revised MPS to the three dimensional theoretical model described by Lee (2004). However, the deletion of a large number of items to obtain validity puts into question the generalizability of the results in this study. Therefore, a

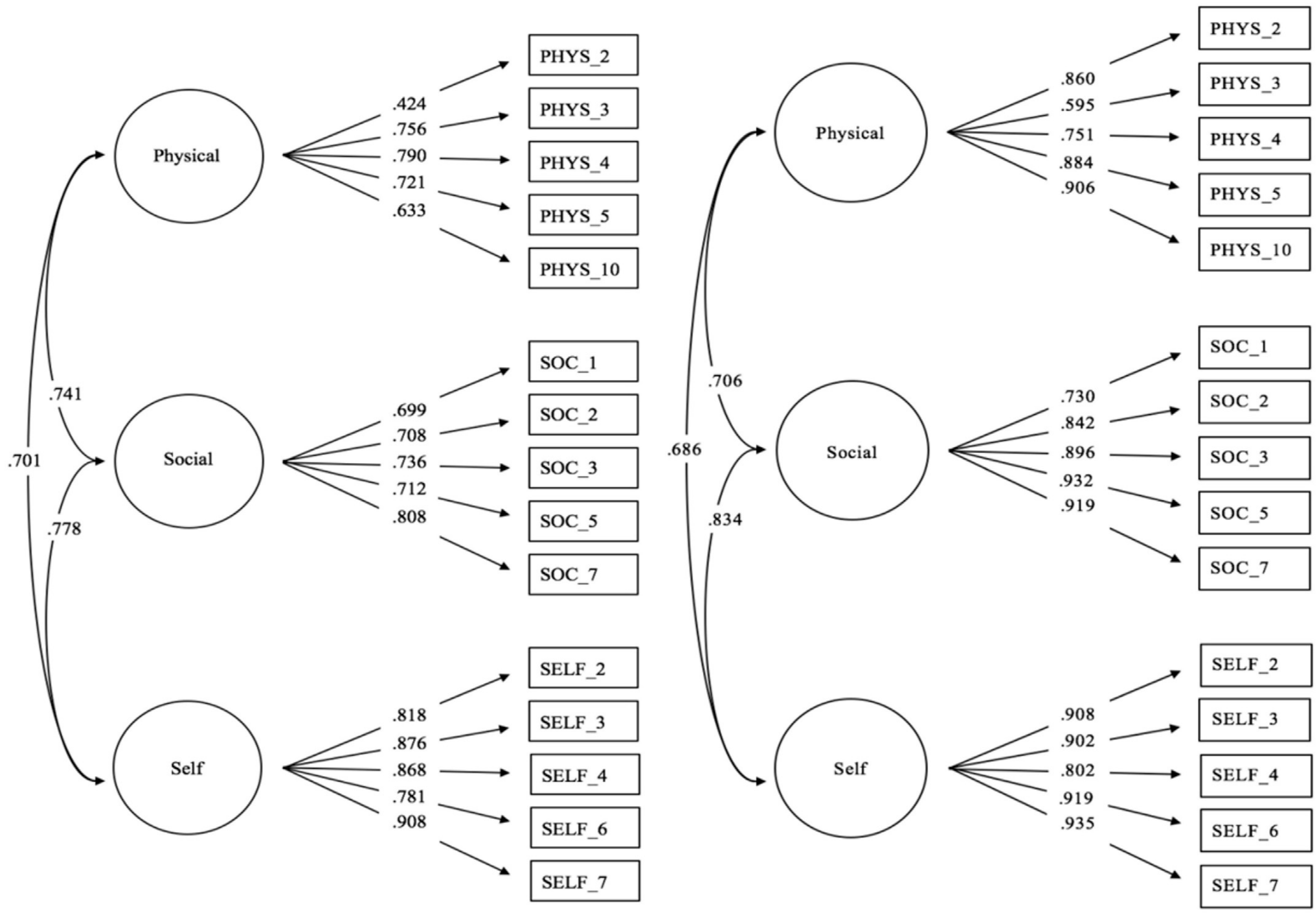

Fig. 1. CFA model with the final MPS in Study 1 and in the cross validation Study 2 samples.

follow-up cross validation was required to assess the generalizability of the findings in Study 1.

4. Study 2

The objective of Study 2 was to investigate whether the findings from Study 1 that indicated that the MPS is a valid and reliable multidimensional measure of presence would generalize to a new sample in a new context.

4.1. Sample and procedure

The sample consisted of 118 (71% female) undergraduate students from the University of Glasgow who participated in an undergraduate biology course in the Fall 2016 semester. The presence items from the revised MPS (Table 4) were administered as part of a post-test immediately after students had taken part in a 1-h virtual laboratory simulation session. The revised MPS was used along with a number of other questions that were used by the teachers to assess the simulation.

The simulation was a desktop virtual reality version of a bacteria isolation simulation developed by the simulation development company Labster. In the simulation students were set the task of isolating a bacterium that had caused an incidence of food poisoning. They were introduced to the principles of using selective and differential culture media in microbiology, and given repeated opportunities to 'practice' streaking out bacteria on agar plates, incubate them, and culture isolated colonies free from contamination. This is a key technique in microbiology lab practice (for more information on the simulation see Makransky, Thisgaard, & Gadegaard, 2016, and Labster, 2016b).

Students were informed that the data could be used for research, and care was taken not to expose the students to any risk or burden (Helsinki Declaration article 17). The study was approved by the College of Medical, Veterinary & Life Sciences Ethics Committee at the University of Glasgow (October 2016).

4.2. Results

The results of the CFA confirm the results from Study 1 (see Table 6). That is, the hypothesized three dimensional model had acceptable and better fit (RMSEA: 0.114, CFI: 0.973, TLI: 0.967) than the unidimensional model (RMSEA: 0.184, CFI: 0.928, TLI: 0.916). The three dimensional CFA from Study 2 is shown in the right panel of Fig. 1.

Table 7 shows the results of the fit of each sub-dimension to the PCM for Study 2. The results support the validity of each sub-

Table 6
Confirmatory factor analysis results.

Cross validation (15 items)	RMSEA	CFI	TLI
Unidimensional	0.184	0.928	0.916
3 dimensions: physical, social, and self-presence	0.114	0.973	0.967

Table 7
PCM results of the items selected for the final MPS.

	Chi square fit	PS reliability	Cronbach's alpha	Item fit	Thresholds	LD	DIF gender	Uni-dimensionality
Physical presence	0.06	0.85	0.86	All OK	PHYS_2	None	None	4.24%
Social presence	0.21	0.86	0.90	All OK	All OK	None	None	5.08%
Self-presence	0.82	0.87	0.94	SELF_2	All OK	SELF_4 SELF_6	None	3.39%

dimension. That is, the items in the physical presence sub-dimension met almost all of the requirements of the PCM. The scale had good general fit to the PCM χ^2 *(5)* $= 10.56$, $p = 0.06$. All of the five items also fit the model; and there was no local dependence between the items, which is evidence that there is no redundancy between the items in the scale. Furthermore, the five-point Likert response format functioned well for all items with the exception of PHYS_2, which had reverse thresholds. The test for unidimensionality indicated that the items measured a single unidimensional scale, and there were no items with DIF for gender. Finally, the PSI and Cronbach's alpha reliability indices were good with values of 0.85 and 0.86 respectively.

Similarly, there was almost perfect fit for the social presence scale. The scale had good general fit to the PCM χ^2 *(5)* $= 7.17$, $p = 0.21$. All of the five items also fit the model; and there was no local dependence between the items. Furthermore, the five-point Likert response format functioned well for all items, and there were no items with DIF for gender. The only problem was that there were 5.08% significant tests, which is just slightly over the critical value of 5% and thus an indication that the items might measure more than one dimension. Finally, the PSI and Cronbach's alpha reliability indices were good with values of 0.86 and 0.90 respectively.

The self-presence scale also had good general fit to the PCM χ^2 *(5)* $= 2.21$, $p = 0.82$. Furthermore, the five-point Likert response format functioned well for all items. The test for unidimensionality also indicated that the items measured a single unidimensional scale, and there were no items with DIF for gender. However, there were some measurement issues in the scale that had not been identified in Study 1. Item SELF_2 had a fit residual of 2.70, which indicates that this item did not measure the latent trait of self-presence as well as the other items in the sub-dimension. There was also a high fit residual of 0.34 between item SELF_4 (“I felt like my real arm was projected into the virtual environment through my virtual embodiment”) and SELF_6 (“I felt like my real hand was inside of the virtual environment”) which could be evidence of redundancy between the items. Finally, the PSI and Cronbach's alpha reliability indices were good, with values of 0.87 and 0.94 respectively.

4.3. Study 2 conclusion

The results from Study 2 suggest that the MPS is a valid and reliable multidimensional measure of presence. The CFA indicates that the MPS measures the theoretical model of presence suggested by Lee (2004) with a different sample in a new context, thereby providing evidence of the generalizability of the findings from Study 1. Furthermore, IRT analyses indicated that each of the sub-dimensions in the MPS had acceptable general fit to the PCM. There were some small problems regarding the fit of the sub-dimensions including reverse thresholds for PHYS_2 in the physical presence sub-dimension, a slight indication of lack of unidimensionality for the social presence sub-dimensions, and a lack of fit for SELF_2 and local dependence between SELF_4 and SELF_6 in the self-presence sub-dimension. These results, and a comparison of the results from the two studies, are discussed in more detail in the next section.

5. Discussion

Presence is one of the most important psychological constructs for understanding human-computer interaction and although there are many scales that proport to measure presence in virtual environments, there is currently no single measure of presence based on Lee's (2004) unifying definition of the construct which includes physical, social, and self-presence sub-dimensions. The objective of this paper was to describe the development of the Multimodal Presence Scale (MPS) and the validation of the scale with confirmatory factor analysis (CFA) and item response theory (IRT).

The results of Study 1 supported the construct validity of the MPS and suggested that the MPS measures the three dimensional theoretical model of presence: physical, social, and self-presence as described in Lee (2004). The results of Study 2 supported the validity of the 15 items that were selected in Study 1 (5 for each sub-dimension), and suggested that the MPS is a valid and reliable multidimensional measure of presence.

Although the CFA indicated acceptable fit for the three dimensional model in both studies based on the CFI and TLI fit indices, the results for the RMSEA were higher than expected. Moreover, the results showed that each of the three sub-dimensions in the MPS had acceptable fit for the Partial Credit Model (PCM) within the framework of IRT in both studies. Although there was some evidence of misfit to the PCM in each of the studies, the results were not consistent across studies, which suggest that the aberrations could be due to Type 1 error (because the data had good fit to the model in both studies). However, more research is needed to investigate these issues further.

The only consistent source of misfit was that there was slight evidence of multidimensionality for the social presence sub-dimension. The principle components analysis identified two potential sub-dimensions within social presence. The three items intended to measure the area attributes of sense of coexistence and human realism loaded positively on the first component whereas the items intended to measure the area attributes of “not aware of the artificiality of social interaction” and “not aware of the social mediation” loaded negatively. Although the result was identified in both studies, the critical value of 5% was within the confidence interval of the observed values, meaning that there was no clear evidence of multidimensionality. Therefore, further studies are needed to investigate the unidimensionality of the sub-dimension.

The importance of validating the MPS with CFA and IRT can be demonstrated with the physical sub-dimension. The final MPS contains the same four area attributes as the IPQ (see Appendix 1); the only previous study to have utilized CFA. This supports the validity of this analysis for identifying a general structure of a construct. The IPQ was not however validated further with IRT, which would have supplied additional information about how well specific items function within a given sub-dimension. Through this analysis the results of the present study showed that the physical sub-dimension of the MPS could be reduced to five items (as opposed to 14 items in the IPQ) without compromising the reliability, the content, or the construct validity of the scale.

The five items from Witmer and Singer (1998) hypothesized to measure control/act in the virtual environment did not measure the latent trait of physical presence as defined by the remaining content of the scale. A closer look at the content of these items suggests that they could measure immediacy of control. Immediacy of control is defined as the ability to change the view position or direction, giving the impression of smooth movement through a virtual environment; and the ability to pick up, examine, and manipulate objects within the virtual environment (Dalgarno, Hedberg, & Harper, 2002). The consequences of the user's action should be appropriately obvious and apparent to the user in order to afford expected continuities (McGreevy, 1992), and is thus related to presence. Immediacy of control has been found to predict presence (e.g., Lee, Wong, & Fung, 2010); therefore it could be that these features are an antecedent of presence, but not a part of the construct.

The two area attributes of sense of self being in the virtual environment and emotional connectivity were eliminated from the self-presence sub-dimension. The results also indicated that the item intented to measure sense of self being in the virtual environment loaded on the physical presence sub-dimension. The items intended to measure emotional connectivity clearly measured a separate latent trait. Therefore, self-presence as measured in the MPS is a quite narrow construct with representation of two area attributes: sense of bodily connectivity and sense of bodily extension. One potential explanation for these results is that the virtual laboratories simulations used in our study were quite limited in terms of (1) self-representation in the virtual environment and (2) the ability to induce emotional reactions. The viewing in the virtual world was from a personal perspective, so a participant's self-image was only viewed as an arm. Thus, if the participant looked down, he would not see a virtual representation of his entire body. Most research where self-presence is measured includes the use of avatars where one's self-image is more fully represented (e.g. Ratan & Hasler, 2009; Ratan, Cruz, & Vorderer, 2007). Furthermore, the simulation's limited ability to induce emotional reactions might be due to the fact, that they were rather dull with regards to emotional content. Accordingly, future research is needed where the MPS is used in other virtual environments which use a richer self-representation and contain more emotionally arousing content.

5.1. Practical implications and study limitations

The results of this study suggest that the MPS provides a short yet valid measure of presence based on Lee's (2004) definition of the term. The fact that the scale is short means that it could easily be used in most applied settings with a wide range of applications where the measurement of presence is needed. A unifying operationalization and measurement of presence in VR will also make it possible to compare and generalize results across studies in different fields.

One possible limitation of this study was, that the item selection process was theoretically guided by Lee's (2004) unifying theory of presence, and this filtering of items based on a theoretical understanding might introduce bias. A more open analysis of items with exploratory factor analysis (EFA) might have suggested a latent structure of presence different than that presented by Lee (2004). However, the objective of this study was not to validate Lee's model, but to develop and validate a measure of presence based on this theoretical model. Instead of a strictly data-driven analysis such as EFA, CFA provides the option of investigating whether the data fits the structure of an a priori hypothesized model in a given context.

Another limitation of this study was that the MPS was validated within the context of VR learning simulations. Therefore, the items were adapted to fit this context. The items would have to be adapted if they are to be used in a different context. Also different results might be expected across different VR environments. Future research is thus needed to investigate whether the scale would function equally well in a different context. Future research should also investigate the generalizability of these findings for different samples in different cultures and across different languages. There are great advantages of using CFA and IRT to validate psychometric instruments, because they provide a detailed understanding of the validity of the measures. Furthermore, they make it possible to investigate the measurement invariance of a scale across different cultures or contexts (Makransky & Glas, 2013; Palic, Kappel, & Makransky, 2016). Therefore, we suggest that future research take this approach.

Acknowledgements

We would like to thank the genetics professors Anne Nørremølle, Asli Silahtaroglu, and Iben Bache from the University of Copenhagen who collected the data used in Study 1. We would also like to thank the biology professors Helen Gadegaard and Nicola Veitch from the Unviersity of Glasgow who collected the data for Study 2.

This work has been funded by Innovation Fund Denmark.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.chb.2017.02.066.

References

Andrich, D., Sheridan, B., & Luo, G. (2010). *Rasch models for measurement: RUMM2030*. Perth, Australia: RUMM Laboratory.
Bonde, M. T., Makransky, G., Wandall, J., Larsen, M. V., Morsing, M., Jarmer, H., et al. (2004). Improving biotech education through gamified laboratory simulations. *Nature Biotechnology, 32*(7), 694–697.
Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the social sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.
Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. New York: The Guilford Press.
Chen, W.-H., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics, 22*(3), 265–289.
Dalgarno, B., Hedberg, J., & Harper, B. (2002). The contribution of 3D environments to conceptual understanding. In O. J. McKerrow (Ed.), *Winds of change in the sea of Learning: Proceedings of the 19th annual conference of the australasian society for computers in learning i tertiary education* (Vol. 1, pp. 149–158). Auckland, New Zealand: UNITEC, Institute of Technology.
Embertson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272–299. http://dx.doi.org/10.1037/1082-989X.4.3.272.
Griffith, J. W., Sumner, J. A., Debeer, E., Raes, F., Hermans, D., Mineka, S., … Craske, M. G. (2009). An item response theory/confirmatory factor analysis of the autobiographical memory test. *Memory, 17*(6), 609–623. http://dx.doi.org/10.1080/09658210902939348.
Hays, R. D., Morales, L. S., & Reise, S. P. (2000). Item response theory and health outcome measurement in the 21st century. *Medical Care, 38*(9 Suppl), 28–42.
Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. http://dx.doi.org/10.1080/10705519909540118.
Klimmt, C., & Vorderer, P. (2003). Media psychology "is not yet there": Introducing theories on media entertainment to the presence debate. *Presence: Teleoperators and Virtual Environments, 12*(4), 346–359. http://dx.doi.org/10.1162/105474603322391596.
Labster. (2016a). *Labster - cytogenetics lab*. Retrieved December 16th 2016 from https://www.youtube.com/watch?v=_VsabhW1LkA.
Labster. (2016b). *Labster - bacteria isolation lab*. Retrived December 16th 2016 from https://www.youtube.com/watch?v=zZYUob44efE.
Lee, K. M. (2004). Presence, explicated. *Communication Theory, 14*(1), 27–50.
Lee, E. A.-L., Wong, K. W., & Fung, C. C. (2010). How does desktop virtual reality

enhance learning outcomes? A structural equation modeling approach. *Computers & Education, 55*(4), 1424–1442. http://dx.doi.org/10.1016/j.compedu.2010.06.006.

Makransky, G., & Bilenberg, N. (2014). Psychometric properties of the parent and teacher ADHD rating scale (ADHD-RS): Measurement invariance across gender, age, and informant. *Assessment, 21*(6), 694–705. http://dx.doi.org/10.1177/1073191114535242.

Makransky, G., Bonde, M. T., Wulff, J. S. G., Wandall, J., Hood, M., Creed, P. A., … Nørremølle, A. (2016). Simulation based virtual learning environment in medical genetics counseling: An example of bridging the gap between theory and practice in medical education. *BMC Medical Education, 16*(1), 98. http://dx.doi.org/10.1186/s12909-016-0620-6.

Makransky, G., & Glas, C. A. W. (2013). Modelling differential item functioning with group-specific item parameters: A computerized adaptive testing application. *Measurement, 46*, 3228–3237.

Makransky, G., Rogers, M. E., & Creed, P. A. (2015). Analysis of the construct validity and measurement invariance of the career decision self-efficacy scale: A rasch model approach. *Journal of Career Assessment, 23*(4), 645–660. http://dx.doi.org/10.1177/1069072714553555.

Makransky, G., Thisgaard, M., & Gadegaard, H. (2016). Virtual simulations as preparation for lab Exercises: Assessing learning of key laboratory skills in microbiology and improvement of essential non-cognitive skills. *PLoS One, 11*(6).

Masters, G. N. (1982). A rasch model for partial credit scoring. *Psychometrika, 47*(2), 149–174. http://dx.doi.org/10.1007/BF02296272.

McGreevy, M. W. (1992). The presence of field geologists in Mars-like terrain. *Presence: Teleoperators and Virtual Environments, 1*(4), 375–403. http://dx.doi.org/10.1162/pres.1992.1.4.375.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika, 49*(1), 115–132. http://dx.doi.org/10.1007/BF02294210.

Muthén, L. K., & Muthén, B. O. (2012). *Mplus (version 7)*. Los Angeles, CA, USA.

Palic, S., Kappel, M., & Makransky, G. (2016). Rasch validation and cross-validation of the health of nation outcome scales (HoNOS) for monitoring of traumatized refugees in western psychiatric care. *Assessment*. http://dx.doi.org/10.1177/1073191115594690.

Pallant, J. F., & Tennant, A. (2007). An introduction to the Rasch measurement model: An example using the hospital anxiety and depression scale (HADS). *British Journal of Clinical Psychology, 46*, 1–18. http://dx.doi.org/10.1348/014466506X96931.

Pedersen, S. S., Mathiasen, K., Christensen, K. B., & Makransky, G. (2016). Psychometric analysis of the Patient Health Questionnaire in Danish patients with an implantable cardioverter defibrillator (The DEFIB-WOMEN study). *Journal of Psychosomatic Research, 90*, 105–112. http://dx.doi.org/10.1016/j.jpsychores.2016.09.010.

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

Ratan, R. A., Cruz, M. S., & Vorderer, P. (2007). Multitasking, presence, and self-presence on the Wii. In *Proceedings of the 10th annual international workshop on presence*.

Ratan, R. A., & Hasler, B. (2009). Self-presence standardized: Introducing the self-presence questionnaire (SPQ). In *Proceedings of the 12th annual international workshop on presence*.

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin, 114*(3), 552–566. http://dx.doi.org/10.1037/0033-2909.114.3.552.

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

Smith, E. V. (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principle component analysis of residuals. *Journal of Applied Measurement, 3*, 205–231.

Steuer, J. (1992). Defining virtual Reality: Dimensions determining telepresence. *Journal of Communication, 42*(4), 73–93. http://dx.doi.org/10.1111/j.1460-2466.1992.tb00812.x.

Tennant, A., & Conaghan, P. G. (2007). The Rasch measurement model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Care and Research, 57*(8), 1358–1362. http://dx.doi.org/10.1002/art.23108.

Ware, J. E., Jr. (2003). Conceptualization and measurement of health-related quality of life: Comments on an evolving field. *Archives of Physical Medicine and Rehabilitation, 84*(4 Suppl 2), 43–51.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments, 7*(3), 225–240. http://dx.doi.org/10.1162/105474698565686.