

Measuring presence with verbal versus pictorial scales: a comparison between online- and ex post-ratings

Bartholomäus Wissmath · David Weibel · Fred W. Mast

Received: 31 January 2009 / Accepted: 25 June 2009 / Published online: 10 July 2009
© Springer-Verlag London Limited 2009

Abstract In this study, we compare subjective online- and post-immersion measures. Although its relevance appears obvious from a theoretical and applied research perspective, this question has not yet been addressed in previous studies. In addition, we also compare verbally and pictorially anchored scales. These factors were measured in different contents using a $2 \times 2 \times 2$ design. We manipulated time of measure (online vs. ex post), type of measure (verbal vs. visual), and content (language vs. language-free). Participants ($N = 162$) evaluated two video clips in terms of presence. No differences between averaged online- and post-immersion measures were found and online judgments did not interfere with the sensation of presence. In line with findings from other areas of research, the use of pictorially anchored items has major advantages. Our results suggest that those items require less mental workload and assess the sensation of presence more directly than verbally anchored items. We discuss the theoretical implications of our findings.

Keywords Presence · Online measurement · Post-rating · Verbal measures · Pictorial measures · SAM

B. Wissmath (✉) · D. Weibel · F. W. Mast
Department of Psychology, University of Bern,
Muesmattstrasse 45, Bern 9, Switzerland
e-mail: wissmath@psy.unibe.ch
URL: <http://www.kog.psy.unibe.ch>

D. Weibel
e-mail: david.weibel@psy.unibe.ch

F. W. Mast
e-mail: fred.mast@psy.unibe.ch

B. Wissmath · D. Weibel
Swiss University Institute of Distance Education,
Überlandstrasse 12, Postfach 265, 3900 Brig, Switzerland

1 Introduction

In his overview of the current and upcoming challenges in presence research, Lombard (2008) outlined the importance of refining and standardizing definitions and terminology of (tele)presence. Along with the technological development, a plethora of conceptualizations have been proposed by scholars of various disciplines. According to Lombard and Ditton (1997), they include social richness, realism, transportation, immersion, social actor within medium, and medium within a social actor. These conceptualizations can be further grouped into the two main categories: physical presence and social presence. Thus, there is still no general conceptualization of presence and the community is still challenged to refine and standardize presence definitions.

The development of a comprehensive and coherent theory of presence implies in the first place the development of corresponding standard, flexible, valid, and reliable measures (Lombard 2008). However, a commonly accepted paradigm for the assessment of presence does not yet exist. This conclusion has been drawn several times since Minsky (1980) coined the term telepresence (e.g., Hendrix and Barfield 1996; van Baren and IJsselstein 2004). From our point of view, however, it is not necessary to agree on a universal definition prior to developing valid assessment tools. The history of the intelligence research illustrates well that a universally accepted definition is not a necessary precondition for the development of valid assessment techniques and high impact research (cf. Sternberg, 2000).

In this article, we will briefly review the milestones in the development of a theory of presence and presence measures. Then, we will introduce the features of the key assessment methods. This has been done before but those attempts ended up enlisting existing measures and

speculating about possible advantages and drawbacks of each method. Here, we will provide an empirical analysis of two important features: type and time of measurement. Both issues have been considered important and caused various arguments in the past—the latter even more than the former (cf. van Baren and IJsselsteijn 2004; Slater 2004).

2 Theories of presence

Presence has become a core dimension when it comes to describing and understanding how humans interact with digital and virtual environments. According to Steuer (1992), (tele)presence describes the extent to which one feels present in the mediated environment, rather than in the immediate physical environment. Mediated contents are perceived as real and one's self-awareness is immersed into this other world (Draper et al. 1998). Sadowsky and Stanney (2002) describe presence as “a sense of belief that one has left the real world and is now ‘present’ in the virtual environment” (p. 791), while Slater and Wilbur (1997) define it as a “state of consciousness, the (psychological) sense of being in the virtual environment” (p. 604). According to Lombard and Ditton (1997), presence is a perceptual illusion of non-mediation. The sensation of presence in virtual environments implies the departure from the physical environment and the arrival in the mediated environment (Sadowsky and Stanney 2002). Thus, inspired by Lombard and Ditton's (1997) and Steuer's (1992) theoretical considerations, Kim and Biocca (1997) found the two dimensions *arrival* in the mediated environment and *departure* from the non-mediated physical environment to form the sensation of presence. Yet, another description was proposed by Heeter (1992) who describes presence as the feeling of a virtual environment being responsive to human actions. According to Heeter, this feeling occurs, above all, in the context of 3-D virtual reality (VR) environments.

IJsselsteijn et al. (2000) reviewed various investigations and concluded that four factors contribute to the sensation of presence. The first factor depends on the *extent and fidelity of sensory information* provided by the environment. This includes resolution of the visual display or the audio channel (e.g., mono vs. surround speakers). The second factor—*matching between sensors and display*—refers to sensory-motor contingencies: The closer the system maps the user's actions in terms of perceptible effects, the higher the user will experience the level of presence. The third factor—*content*—is broadly defined and includes the objects, actors, and events presented in the mediated environment. The fourth factor consists of *user characteristics* such as cognitive resources, interest, or experience

with a particular medium. From our point of view, all factors besides the second one apply to interactive as well as to non-interactive media as the second factor only matters in environments where users can take actions (i.e., interactive environments).

Television and video clips, which will be used in this study, are clearly to be considered as non-interactive, but they still have the potential to elicit sensations of presence (Bracken 2005; Kim and Biocca 1997; Lombard et al. 2000). Wirth et al. (2007) reviewed the presence literature and concluded that media factors are by far not the only determinants of presence and that less immersive media than VRs like television broadcasts or even books clearly bear the potential to evoke even strong sensations of spatial presence.

Additionally, Wirth et al. (2007) developed a conceptual two-level process model of spatial presence which integrates psychological concepts such as attention allocation. The authors conceive the sensation of spatial presence as a binary experience. Thereby, a fundamental precondition for the experience of spatial presence is the *attention allocation* toward the medium and the establishment of a mental model of the mediated environment (i.e., *spatial situational model*; SSM). Once these preconditions are met, the individual must confirm the “medium-as-PERF-hypothesis” (primary egocentric reference frame). The idea is that the subjective frame of reference is captured and controlled by the mediated environment. Additional factors determining the sensation of spatial presence are *involvement* (the active and intensive processing of the mediated world), *possible actions* within the mediated environment and *suspension of disbelief* (avoiding features contradicting the medium-as-PERF-hypothesis) (Wirth et al. 2007). The model further includes user factors like domain-specific interest or visual imagery.

In the following section, we will demonstrate that the existing presence measures can be grouped along important criteria and that the corresponding pros and cons of each need to be systematically examined. This is what Lombard (2008) considers necessary, and this study contributes to the development of presence measures.

3 The measurement of presence

Though the ideal instrument to assess presence is not known so far, the desired features of such an instrument have been described more than a decade ago (Hendrix and Barfield 1996): relevance, validity, reliability, sensitivity, non-intrusiveness, robustness, and convenience. In their compendium, van Baren and IJsselsteijn (2004) provide an overview of presence measures. In a first step, all instruments can be grouped into behavioral or physiological

measures on the one hand, and subjective ratings as well as subjective reports on the other hand.

3.1 Behavioral and physiological measures

Various behavioral and physiological presence assessment techniques have been developed over the last two decades (cf. Freeman et al. 2000). These indicators reflect responses to specific stimuli. Examples for behavioral indicators are postural responses (e.g., postural adjustments when immersed in a racing game), eye movements, or interactions with virtual objects as if they were real. As these measures are usually assessed implicitly (i.e., without asking participants to provide any presence judgments) and during the actual exposure they are not prone to self-report and recollection biases. The same advantage also applies to physiological indicators such as measures of arousal (e.g., heart rate, blood pressure, electrodermal response, and respiration rate), measures of emotional responses (e.g., EMG and ocular response), or patterns of brain activity (e.g., electroencephalography and fMRI).

However, a drawback of these data is that their collection can be intrusive (van Baren and IJsselsteijn 2004). Hence, the assessment of these indicators can eventually prevent participants from the sensation of presence. In addition, the measures are sometimes unreliable and often hard to interpret because the relation between these measures and the subjective level of presence is not necessarily strong (van Baren and IJsselsteijn 2004). Moreover, physiological indicators are hardly specific to one particular concept. This raises questions concerning convergent and discriminant validity (cf. Campell and Fiske 1959) and could be a core reason why a generally accepted set of physiological indicators for the sensation of presence does not exist so far.

Another major problem is context dependence: even when a specific behavioral or physiological measure is a valid and reliable indicator for presence in a specific environment, it might fail in another context. Accelerated heart rate could be a valid indicator for presence during a virtual ghost train ride but not in an online Yoga class. Thus, postural responses might be a suitable indicator for presence in the context of a virtual roller coaster ride but much less in an environment lacking any visual flow. Therefore, even the most enthusiastic proponents of behavioral and physiological indicators (cf. Slater, 2004) admit that their value is limited and most likely is their use most beneficial in combination with subjective data.

3.2 Subjective measures

In the study outlined below, we focus on subjective measures because the sensation of presence is in the first place

of a subjective experience (Sheridan 1992). Furthermore, almost every empirical study on presence includes subjective data. According to van Baren and IJsselsteijn (2004), subjective indicators can be further categorized into presence questionnaires, continuous assessment, qualitative measures, psychophysical ratings (e.g., cross-modality matching), and subjective corroborative measures (e.g., time estimation).

Presence questionnaires are usually administered after the exposure. An indication for the wide use of post-rating measures can be found in the presence measurement compendium (van Baren and IJsselsteijn 2004), which lists 29 subjective measures (e.g., Witmer and Singer 1998; Slater 1999). Among those, only one single instrument is intended for continuous assessment, while the 28 other measures are post-rating instruments. According to Insko (2003), key advantages of these post-rating questionnaires are the ease to administer, high face validity, and the lack of measurement-related interferences during exposure. Additional benefits of questionnaires are the opportunity to conduct factor analyses which allow the identification of the underlying dimensions of presence, low cost, mobility, sensitivity, and ease to analyze and interpret (van Baren and IJsselsteijn 2004). These advantages made subjective post-rating scales the most common and used presence measure.

There is no measure without any side effects or drawbacks. For Sadowsky and Stanney (2002), a major concern is to assure validity as participants must understand the concept of presence. Slater (2004) argues that presence questionnaires could be invalid as the phenomenon to be measured could be brought into existence merely by asking questions about it.

According to Insko (2003), further disadvantages associated with post-immersion questionnaires are anchoring effects, inaccurate recall, and inability to assess temporal variations in the subjective sense of presence. Van Baren and IJsselsteijn (2004) underline possible recency effects in the post-rating judgments. To overcome these problems associated with the traditional presence questionnaire approach, Slater (2004) suggests the multivariate adoption of physiological measures but acknowledges the importance of subjective ratings and qualitative reports (Slater and Garau 2007). In our study, we will address two of the major objections brought forward toward subjective ratings: the limitations of subjective ex post-ratings and response biases associated with verbal measures.

3.3 Continuous subjective assessment

To overcome some of the abovementioned limitations of presence questionnaires, continuous assessment has been proposed (cf. IJsselsteijn et al. 1998, 2000). This allows the

assessment of variations in the subjective experience of presence, which are likely to occur not only through changes in the stimulus but also through the participant himself (e.g., onset time or saturation of presence). Thus, continuous assessment is a possibility to overcome limitations associated with post-rating measures (IJsselsteijn et al. 2000). A method originally developed to assess picture quality of TV images was adapted to continuously assess the experience of presence during exposure. Participants had to provide online judgments of presence by means of a hand-held slider. A task, the authors consider to require little attention and effort to operate. For non-interactive stimuli a considerable temporal variation depending on the sensory input was found (IJsselsteijn et al. 1998). IJsselsteijn et al. (2000) consider continuous assessment of presence to be mainly applicable to non-interactive media as continuous measurement devices may interfere with operating in an interactive environment.

A drawback of this procedure may be that participants are required to divide their attention between the physical and the mediated environment. Attention allocation toward the display in turn is a central component of spatial presence (Wirth et al. 2007). Thus, online ratings could restrain participants from experiencing arrival in the mediated environment as well as departure from the actual physical setting. Both of these factors were found to be central components of presence (Kim and Biocca 1997). Thus, the continuous subjective assessment of presence could bear similar side effects as intrusive physiological measures.

In addition, one could expect that a person who is fully immersed in the displayed environment not only forgets about the real world but also forgets to shift the slider to “fully immersed”. Not only the reliability but also the validity of this method should be scrutinized as during exposure, participants could confuse presence with other perhaps more familiar judgments such as liking or enjoyment.

The advantages and drawbacks of online judgments and post-ratings have been described above. To our knowledge, a direct and systematic comparison of online- and post-exposure methods to assess the subjective sense of presence does not exist so far.

3.4 Verbal versus non-verbal subjective assessment

Findings from other fields such as emotion research (cf. Lang, 1985) suggest that verbally based measures are prone to biases. In presence research, Sadowsky and Stanney (2002) and Slater stressed possible flaws associated with these approaches. Thus, various non-verbal methods to assess presence have been suggested in the past (Insko 2003; IJsselsteijn et al. 2000). The main approaches are psychophysical or subjective corroborative indicators. The former

include free-modulus magnitude estimation (Snow and Williges 1998), paired comparison (Welch et al. 1996), the VR Turing test (Schloerb 1995), and cross-modality-matching (Welch 1997). In cross-modality-matching, for example, a participant usually has to express the degree of presence experienced in a different modality such as the brightness of a lamp (the brighter the lamp the more presence experienced). Although psychophysical approaches were found to be sensitive, inexpensive, unobtrusive, and easy to use, only a limited number of presence studies have actually used such measures (cf. van Baren and IJsselsteijn 2004; IJsselsteijn et al. 2000).

Other than psychophysical approaches, subjective corroborative measures aim to capture psychological processes, which are conceptually related to presence. Among those, time estimation, attention allocation, recognition, recollection, or spatial cognition (van Baren and IJsselsteijn 2004) have been suggested. These indicators may serve as unobtrusive measures but they assess at best concepts, which are, in one way or the other, closely related to presence but they are still different from presence. An example illustrating the possible lack of validity could be recalled. Thus, an individual might be able to exactly recall the features of a virtual environment because she or he was not immersed and thus had attentional resources available to exploring the environment.

Although there exist even further non-verbal methods to assess presence, the vast majority of subjective presence instruments rely on verbal measures. The advantages of verbally anchored subjective measures such as high face validity or applicability to almost every mediated environment might explain the dominance of these verbal indicators. In most cases, participants rate statements or questions assessing the experience of presence on Likert-type scales or bipolar items.

Long before concerns toward verbally based measures were raised in research articles on presence, visually oriented scales were developed to overcome several limitations of the verbal measures. To our knowledge, pictorial items were first developed to assess emotional responses. The Self-Assessment-Manikin (SAM) test uses just three simple judgments (Lang 1985; Bradley and Lang 1994). Those three pictorial items represent the dimensions valence, arousal, and dominance. Bradley and Lang systematically compared the SAM measure with semantic differential scales consisting of far more items than the SAM. Thereby, high correlations between those measures and their corresponding scales were revealed. This result proved to be stable as it could be replicated in various contexts and thus suggests SAM to be a reliable and valid substitute for semantic differential scales. Bradley and Lang conclude that SAM ratings track personal responses to affective stimuli sometimes even better than semantic

differential scales. The authors conclude that SAM ratings are more valid because they more directly assess the emotional state in the respondent than semantic differential scales. Whereas SAM ratings seem to reflect the subjective experience more closely, the wording of semantic differential scales could mislead the respondents to judge the characteristics of the stimulus rather than the actual psychological effect the stimulus has on them (Bradley and Lang 1994). A vast body of evidence based on cardiac and electrodermal responses as well as facial EMG measures to assess emotional expressions shows close correlations to the SAM ratings (Greenwald et al. 1989; Lang et al. 1993). However, it is noteworthy that each concept is represented by just one single SAM item. This precludes any attempt to calculate inter-item reliabilities. The reliability of the SAM measure could only be assessed by the simultaneous use of other scales representing the same construct.

The SAM is easy to use and to understand even for children and people who speak another language. It is equally suited for paper and pencil as well as computer-based responses (Bradley and Lang 1994). In contrast to verbally anchored measures, visually oriented scales are supposed to be culture free (Lang 1985; Bradley and Lang 1994). Another advantage in comparison to verbal measures is that participants are able to respond quicker and usually appreciate those items more than filling out verbal questionnaires (Lang 1985). It has also been suggested that SAM measures create less mental workload (Jex 1988) than verbally anchored questionnaires. Information processing resources are required in both, examining stimuli as well as answering questionnaire items. Thereby, the working memory is central in information processing. It is a limited-capacity system to store and manipulate information (Baddeley 2000). In his model, Baddeley considers two distinct components; the phonological loop and visuospatial sketch pad. A third component, the central executive, allocates the attention flexibly to one of the two former components. Hence, this model suggests that modality matters and capacity is limited. An investigation addressing the impact of sensory modality (auditory vs. visual) on sustained attention and psychological stress offers further support for this assumption (Szalma et al. 2004). In that study, visual signals resulted in less mental workload and less psychological stress than auditory signals. The findings mentioned above should be of particular

importance in designs with repeated measurement as well as in the evaluation of multiple stimuli (Bradley and Lang 1994). These features could also be highly relevant when the measures are collected not *ex post* but during exposure. In this case, the participants have to divide their attention between a mediated environment and the rating scale. This leads to increased perceptual load, as more items need to be perceived at the same time (Lavie and Cox 1997; Lavie 2005). Therefore, in order not to interfere the sensation of presence, measures should create as little perceptual load as possible.

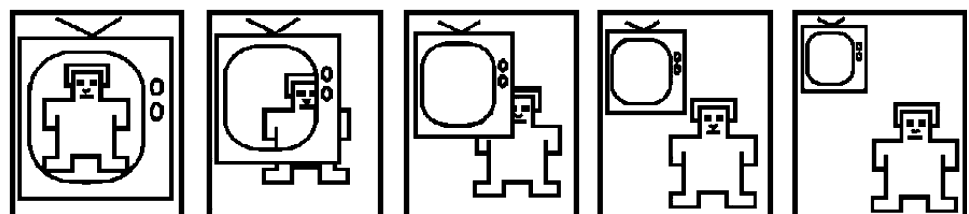
The numerous positive features of visual instruments might have suggested the development of a presence SAM. Lang (2000) and Schneider et al. (2004) developed a single presence SAM item representing the sensation of presence in a mediated environment (cf. Fig. 1). The pictorially anchored presence assessment technique starts with the following instruction (Lombard 2005, p. 4):

Please use the figures below to indicate your feelings or emotional response to the media experience. The pictures go from a person who feels he or she is **INSIDE THE PICTURE, A PART OF THE STORY, A PART OF THE ACTION** on the left end, to a person who feels he or she is **OUTSIDE THE PICTURE, REMOVED OR SEPARATED FROM THE STORY, NOT PART OF THE ACTION** on the right end. Please put an 'X' through the picture, or in the space between any two pictures, that best represents how you felt during the media experience.

4 Hypotheses

Even the proponents of continuous assessment of presence admit that this additional task requires attention and some effort to operate (IJsselstein et al. 2000). Inevitably, participants have to divide their attention between the physical and the mediated environment even if the handling of the rating device has been trained extensively prior to the task. Attention allocation toward the display is a central component of spatial presence (Wirth et al. 2007). In addition, the rating task could lead to a more critical evaluation of the mediated environment which could in turn result in less suspension of disbelief. Being fully present in a displayed

Fig. 1 Depiction of the SAM item



environment requires one to fade out the immediate physical environment (Kim and Biocca 1997; Wirth et al. 2007). This could prevent participants from providing online ratings. These considerations offer the following hypothesis:

H1: the online assessment of presence draws the allocation of attention away from the medium and thus leads to lower levels of presence.

When measuring presence, not only the time but also the modality of the assessment is a controversial issue. The validity, reliability, and efficiency of verbally anchored measures have been questioned (cf. Bradley and Lang 1994; Sadowsky and Stanney 2002; Slater 2004; van Baren and IJsselsteijn 2004). In validation studies, SAM ratings were found to be more valid than semantic differential scales (Bradley and Lang 1994). The SAM measures usually require less attention and effort than verbally anchored measures. When applied online, this should increase the probability to experience spatial presence as less attention is allocated away from the mediated environment (Wirth et al. 2007). For ex post-ratings, SAM items should also lead to higher ratings of presence because SAM items as well as presence are often associated with positive emotions (Lang 1985; Wirth et al. 2007). In contrast, the negative impact of verbally anchored measures might induce a more critical attitude and the suspension of disbelief might be retrospectively impaired. Therefore, we expect the following:

H2: compared to verbal measures, visually oriented items will allow higher levels of presence.

However, we do not expect the advantages of visually anchored scales to equally pay off in all virtual environments and situations. Lavie (2005) suggests attention capacity to be specific in terms of modality. Correspondingly, there are different components involved in visual and auditory information processing (Baddeley 2000). Mediated environments differ in terms of sensory modalities and thus create different patterns of mental workload. Therefore, verbal and non-verbal measures could be unequally appropriate depending on the modality features of the environment (Jex 1988; Szalma et al. 2004). When measuring presence online in visual environments, verbal measures could be more accurate as there is less interference within the same modality. For environments in which language-based information is of major importance such as audio books, verbal measures are more likely to interfere with information processing than a visual measure. These considerations suggest the following:

H3: verbal measures of presence interfere with the sensation of presence in language-based environments,

whereas visual items interfere with the illusion of non-mediation in visual environments.

Our last hypothesis concerns characteristics of the mediated environment. We include this assumption here because it allows a factorial design to test the third hypothesis. IJsselsteijn et al. (2000) reviewed a vast body of research and concluded that among others a content factor (including the objects, actors and events presented by mediated environment) determines the sensation of presence. Thus, language-based and language-free environments should differ in terms of their potential to elicit presence. Thus, we present the following hypothesis:

H4: the content of the medium will influence the levels of presence.

Taken together, our hypotheses address various unsettled issues in context of presence measurement. In the first place, we intend to explore whether online assessment interferes with the sensation of presence. Second, we try to examine the potential of visually oriented scales in this particular field. Third, we aim to investigate whether verbal and visual items are equally suited for different environments. We think that those issues have not only practical but also theoretical implications.

5 Method

The overall design was a $2 \times 2 \times 2$ within-subjects-design. Factors were time of measurement (levels: online vs. ex post), type of measure (levels: verbal vs. visual), and audio content (levels: language vs. language-free). Participants ($N = 162$) watched two video clips. Thereby, participants evaluated one clip online and one clip ex post. In addition, each participant provided one verbal and one visual judgment. A hierarchical design allowed us to implement the three factors without having each participant to evaluate all eight conditions, respectively, all factor combinations. In addition, we counterbalanced the conditions which ruled out confounding variables such as carry-over effects.

5.1 Stimuli

The two video clips used for this experiment were two 4-min lasting sequences of television broadcasts. The language-free clip showed the planet earth (cf. Fig. 2, left side). The language-based clip dealt with the perils of global warming. A journalist interviewed an environmental campaigner before and while they made a trip in an off-road vehicle (cf. Fig. 2, right side).

Fig. 2 Screenshots of the stimuli



5.2 Independent variables and measures

The first factor was *time of measurement*. It was manipulated by having participants in one condition to give three online ratings on a single-item measure every 60 s, starting from 30 s after the clip had started. The average of these three measures was included within the analysis. In the other condition, participants evaluated their sense of presence on a single item after viewing the clip (ex post).

The second factor was *type of measure*. In one condition, participants evaluated their feelings of presence on the following verbally coded single item: “To what extent do you feel located in the world displayed by the clip?” (5-point ranging from 1, “not at all” to 5, “very much”). In the other condition, the pictorial SAM item described above was administered (cf. Fig. 1).

The third factor was the *content*. In the language-free condition, film sequences from the space shuttle (cf. Fig. 2) dubbed with instrumental music were presented. In the language condition, the interviewer and the interviewee discussed the perils of global warming, the future of the planet, and human behavior when facing global dangers. Although presenting a visually identical clip twice with different soundtracks could lead to an increased internal validity, we chose two different clips in order to avoid demand characteristics and response biases. Moreover, our central research interest lies in the interaction between content and type of measure rather than the main effect of content.

5.3 Participants

One hundred and sixty-two undergraduate students volunteered to participate in this study. All participants were recruited in an introductory lecture course in psychology. There was no monetary compensation or extra credit. The majority was female (82.7%) which reflects the overall student population in this particular department.

5.4 Procedure

Participants were randomly assigned to groups of 8–12 persons and to one of the eight experimental conditions. The experiments were conducted in standard lecture rooms. The alignment of the chairs was similar to the setting in a movie theater. The lateral distance between the subjects was 60 cm. The average distance between the screen and the audience members was 7 m. The stimuli were presented by means of high-definition video projectors. Equal projection size (1.69 m × 3.00 m) and audio levels for all conditions were ensured. Each participant saw both clips. The questionnaires were provided in paper–pencil form and were handed over before the experiment began. The experimenter told the participants that there were no wrong answers and that they could end participation at any time. In addition, the experimenter asked the participants to remain absolutely quiet during the experiment and not to interact with other participants in any way. A research assistant observed the participants to ensure that those instructions were followed. Before starting the clip, which had to be rated online, the experimenter asked the participants to carefully read the instructions. The post-rating measures were mentioned only after the presentation of the clip. In those conditions, a blank sheet of paper covered the post-rating items in order to avoid possible distractions.

During the experiment, each participant rated the sense of presence for one clip by means of the verbal and the other clip by means of the visual measure. Each participant rated one clip online (three judgments while the clip was running, 60 s time lag between measures, the answer was triggered by a handclapping of the experimenter) and one clip post-exposure. In order to prevent sequence-effects, the order of the within-factor levels was counterbalanced. The whole experiment took about 10 min.

5.5 Analysis strategy

All hypotheses have been analyzed by means of a three-factorial linear mixed model. Linear mixed modeling

(LMM; cf. McCulloch and Searle 2000) is a further generalization of general linear models (GLM). That is, LMM correctly models correlated errors, whereas procedures in the GLM family usually do not. LMM handles data when observations are not independent. In addition, LMM encompasses all models in the variance components (VARCOMP) procedure. This allows implementation of hierarchical designs. Thereby, it is not necessary to realize all possible combinations of the implemented factors. These features make LMM the ideal analysis strategy for our study.

6 Results

In this section, we will report the results of the hierarchical three-factorial linear mixed modeling and an ANOVA.

H1: the online assessment of presence draws the allocation of attention away from the medium and thus leads to lower levels of presence.

The first hypothesis predicting the online assessment ($M = 2.27$; $SD = 0.94$) of presence leading to lower levels of presence than the post-rating ($M = 2.26$; $SD = 1.08$) could not be supported, $F(1, 158) = 0.01$, $P = 0.91$. Regardless of the measure applied, the participants indicated medium presence ratings. In other words, there is no significant difference between the averaged online measures and the ex post measures.

To assess possible changes in presence during exposure, we compared the three measures over both conditions (iconographic as well as verbal measures). Therefore, we calculated a repeated measure ANOVA. As sphericity could not be assumed (Mauchly's $W = 0.90$; $df = 2$; $P < 0.01$), we corrected for Huynh-Feldt's *Epsilon*. The factor time of measure turned out to be important as the presence scores rose during exposure, $F(1.85, 298) = 6.89$, $P < 0.01$. Figure 3 depicts the results. No significant difference between verbal and iconographic online measures resulted, $F(1.87, 298) = 0.29$, $P = 0.73$. The first online rating tended to be lower than the ex post-rating, whereas the last online rating tended to be higher than the ex post-rating. Most noteworthy, the second online rating ($M = 2.28$; $SD = 1.09$) was at the same level as the ex post-ratings ($M = 2.26$; $SD = 1.08$). This is confirmed by the respective t test, $t(161) = 0.30$, $P = 0.86$.

H2: compared to verbal-based measures, visually oriented scales will allow higher levels of presence.

The second hypothesis was strongly supported. As predicted, visually oriented scales ($M = 2.45$; $SD = 0.99$) result in higher levels of presence compared to verbally based measures ($M = 2.10$; $SD = 0.98$), $F(1, 158) = 9.60$,

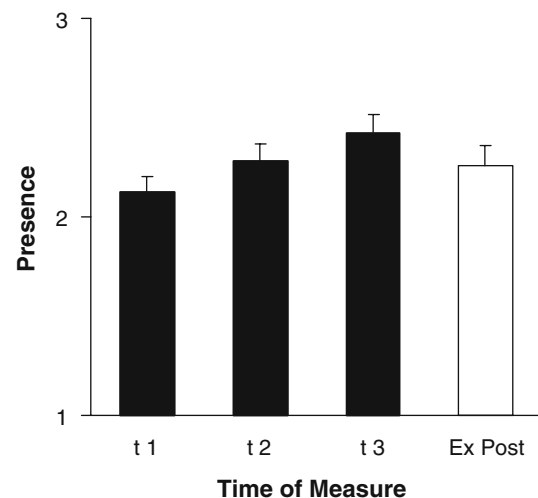


Fig. 3 Mean presence scores (+SE) for online and ex post measures

$P < 0.01$. The descriptives of the results described above are summed up in Table 1.

H3: verbal measures of presence interfere with the sensation of presence in language-based environments, whereas visually oriented scales interfere with the illusion of non-mediation in visual environments.

The third hypothesis could not be supported. However, there is an interaction between type of measure and content, $F(1, 158) = 7.74$, $P = 0.01$. Verbal measures interfere with presence experienced in visual environments. In contrast, visual measures seem to be similarly suited for both types of contents (cf. Fig. 4).

H4: the content of the medium will influence the levels of presence.

The fourth hypothesis was supported. The content of the medium influences the levels of presence as the language-free condition ($M = 2.09$; $SD = 0.92$) produced lower

Table 1 Means and standard deviations (time of measure \times type of measure)

Time of measure	Type of measure					
	Visual		Verbal		Total	
	Mean	SD	Mean	SD	Mean	SD
Online						
t1	2.20	1.02	2.03	1.06	2.12	1.04
t2	2.41	1.12	2.13	1.07	2.28	1.09
t3	2.54	1.29	2.28	1.08	2.42	1.20
Online total	2.38	1.15	2.15	1.07	2.27	0.94
Ex post	2.51	1.01	2.04	1.02	2.26	1.08
Total	2.45	0.99	2.10	0.98	2.27	1.00

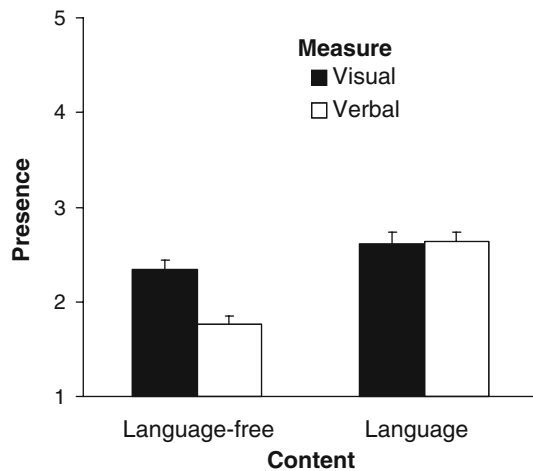


Fig. 4 Mean presence scores (+SE) for visual and verbal measures in language and language-free conditions

presence than the language-based clip ($M = 2.63$; $SD = 0.98$), $F(1, 158) = 36.92$, $P < 0.000$.

Finally, for exploratory reasons, the interactions of time of measure \times content as well as time of measure \times type of measure \times content were tested. Both were non-significant with $F(1, 158) = 3.03$, $P > 0.05$, and $F(1, 158) = 3.50$, $P > 0.05$, respectively.

7 Discussion

One of the most noteworthy findings of this investigation is that averaged online judgments of presence and post-rating presence measures lead to comparable results. The question whether overall online and post-rating judgments of presence are interchangeable has been raised various times (Ijsselsteijn et al. 1998, 2000; van Baren and Ijsselsteijn 2004). So far, scholars who investigated the experience of presence by means of subjective post-rating measures could not be sure that these widely used post-experience judgments are reliable and valid indicators of the underlying psychological process. In our case, the subjective sensation of presence continuously increased as the exposure continued. Thereby, the mean second value in the online assessment as well as the average overall online rating matched the ex post value. Thus, participants seem to provide highly accurate judgments of their overall presence experience. Hence, post-exposure measures seem not to be prone to biases like inaccurate recall and recency effects as van Baren and Ijsselsteijn (2004) expected. This finding suggests that post-rating scales are by all means useful and not less precise than online measures.

In addition, subjective online ratings seem not to draw a meaningful portion of the attention away from the displayed environment and thus do not diminish the sensation of

presence. Furthermore, it seems unlikely that the online assessment results in a more critical attitude toward the mediated environment or in an impairment of the suspension of disbelief. Online ratings provide the possibility to observe presence not only afterward but also during the exposure to the mediated environment. Here, we add further evidence to the body of research suggesting variations in the subjective sensation of presence over time (cf. Ijsselsteijn et al. 2000).

The next central finding is that visual items lead to higher ratings of presence. This could be due to the fact that they are easier to understand and it takes less time to respond when compared to verbally anchored items (Bradley and Lang 1994). Another explanation could be that participants appreciate these items more (Lang 1985) and therefore experience more presence. It is also possible that SAM measures are more valid than verbal measures. In line with findings from the assessment of emotional states (Bradley and Lang 1994), SAM could reflect the subjective experience more closely than verbally anchored items. Bradley and Lang assume the wording of semantic differential scales to mislead the respondents to judge the characteristics of the stimulus rather than their actual sensation. Taken together, our results suggest the SAM item to be a valid indicator for the sensation of presence.

Not surprisingly, the content of the displayed clip influenced the presence ratings. As the clips differed in many aspects associated with the content such as topic, display of humans, editing, location and the like, there is no way to relate the differences in the presence assessment to specific features of the clips. Still we consider the effect of this factor to be meaningful not only because it adds evidence to the growing body of research demonstrating that content factors matter (Ijsselsteijn et al. 2000), but also increases our confidence to believe in the interchangeability of online- and post-rating measures when averaged overall judgments are to be compared. In combination with the variance obtained in all measures, this effect provides a strong point that our participants experienced different levels of presence in the course of our experiment. Thus, it cannot be argued that low overall variance caused averaged online- and post-ratings to be on the same level.

We expected verbal and non-verbal measures to differently affect the sensation of presence in language containing and language-free environments. The corresponding interaction was significant. However, against our prediction, verbal measures seem to interfere with the sensation of presence only if the stimulus is visual. This could again be an issue of validity (Bradley and Lang 1994). As the non-language condition generally elicited less sensations of presence than the language condition, we could assume that the verbal measure might have provoked the participants to express their boredom or lack of enjoyment rather than their

sense of presence. Another conclusion again in favor of visually anchored measures is that the mental workload to respond to them is small enough, not impairing the allocation of attention toward the mediated environment.

To finally ensure the validity of the findings discussed above, we made sure that our findings concerning time of measure equally apply for different contents. Thereby, the corresponding interaction clearly failed to reach significance. It is important to mention that our design includes a large sample, repeated measures, and linear mixed modeling. These three features contribute to high power, which in turn provides strong support for all of our findings, especially when we refer to the interchangeability of measures suggested in the discussion of the first hypothesis.

In this investigation, a single overall item proved to be a valid and reliable indicator of presence. However, from our point of view the item should be further validated by direct comparisons with other presence measures. Thereby, both subjective as well as behavioral and/or physiological presence indicators should be used to further strengthen our confidence in this particular approach. Additionally, we suggest the application in non-interactive environments such as television broadcasts or in IMAX theatres as well as interactive environments such as online-games or VR-applications. During the assessment of presence in an interactive environment the rating tool should not be intrusive (i.e., physically limit the possibilities to explore and interact). One possible technique could be cross-modality-matching if participants explore a CAVE by walking around, they could hold a small rubber ball in their hand. Thereby, the more spatially present they feel the more they compress the ball which measures the pressure. In addition, the participants should fully automatize the usage of the online rating device prior to the actual measurement.

8 Limitations

The present study has a number of limitations and their identification should help to refine future research efforts. For this study, a student population was tested under experimental conditions. Clearly, a more representative sample (i.e., in terms of gender, age, and education) as well as the replication of these findings for other measures would increase confidence in the results. In addition, our findings are based on non-interactive environments and overall presence judgments. Because previous research found considerable variations in presence during the exposure to video clips (cf. IJsselstein et al. 1998), we have not manipulated presence directly. However, a direct presence manipulation could further increase the variance in the ratings and thus allow for further insight. Moreover, having implemented single-item measures, the impact of

time and type of measurement on the sub-dimensions of presence has not been addressed here. Before directly generalizing our findings to more interactive environments, longer or shorter exposure times, and other presence measures, additional research is required. We also have to point out that our operationalization of online measures is not identical to continuous assessment techniques. Research in the area of divided attention (cf. Lavie 2005) suggests impaired validity and reduced sensations of presence when mental workload is too high. Therefore, to strengthen confidence in the interchangeability of averaged online and ex post measures, future research should compare online and ex post measures in more complex and more interactive-mediated environments.

9 Conclusions

This study contributes to findings suggesting advantages of visually anchored measures in terms of efficacy and validity. When assessing an overall estimation of presence, the advantages of visually anchored measures seem to pay off especially with visual environments. The second noteworthy finding is that the sensations of presence continuously rise during exposure. This is further evidence for temporal variations in the subjective sensation of presence.

The third important finding in this piece of research is the interchangeability of averaged online and ex post measures of presence. This finding suggests the further use of post-rating scales as participants are to be able to ex post provide accurate overall estimations of the presence experienced. However, when temporal variations of presence are of particular interest, our study suggests that overall online ratings do not interfere with the sense of presence.

References

- Baddeley AD (2000) Short-term and working memory. In: Tulving E, Craik FMI (eds) *The Oxford handbook of memory*. Oxford University Press, New York, pp 77–92
- Bracken CC (2005) Presence and image quality: the case of high definition television. *Media Psychol* 7:191–205
- Bradley MM, Lang PJ (1994) Measuring emotion: the self-assessment manikin and the semantic differential. *J Behav Ther Exp Psy* 25:49–59
- Campbell DT, Fiske DW (1959) Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol Bull* 56:81–105
- Draper JV, Kaber DB, Usher JM (1998) Telepresence. *Hum Factors* 40(3):354–375
- Freeman J, Avons S, Meddis R, Pearson D, IJsselstein W (2000) Using behavioural realism to estimate presence: a study of the utility of postural responses to motion stimuli. *Presence Teleop Virt* 9:149–164

- Greenwald MK, Cook EW, Lang PJ (1989) Affective judgement and psychophysiological response: dimensional covariation in the evaluation of pictorial stimuli. *J Psychophysiol* 3:51–64
- Heeter C (1992) Being there: the subjective experience of presence. *Presence Teleop Virt* 1:262–271
- Hendrix C, Barfield W (1996) Presence within virtual environments as a function of visual display parameters. *Presence Teleop Virt* 5:274–289
- IJsselsteijn WA, De Ridder H, Hamberg R, Bouwhuis D, Freeman J (1998) Perceived depth and the feeling of presence in 3DTV. *Displays* 18:207–214
- IJsselsteijn WA, De Ridder H, Freeman J, Avons SE (2000) Presence: concept, determinants and measurement. *Proc SPIE* 3959:520–529
- Insko B (2003) Measuring presence: subjective, behavioral and physiological methods. In: Riva G, Davide F, IJsselsteijn WA (eds) *Being there: concepts, effects, and measurement of user presence in synthetic environments*. Ios Press, Amsterdam, pp 211–222
- Jex H (1988) Measuring mental workload: problems, progress, and promises. In: Hancock PA, Meshkati, N (eds) *Human mental workload*. Elsevier, North-Holland, pp 5–39
- Kim T, Biocca F (1997) Telepresence via television: two dimensions of telepresence may have different connections on memory and persuasion. *J Comput Mediat Comm* 3(2). Available at <http://jcmc.indiana.edu/vol3/issue2/kim.html>. Accessed 14 Aug 2005
- Lang P (1985) *The cognitive psychophysiology of emotion: anxiety and the anxiety disorders*. Lawrence Erlbaum, Hillsdale
- Lang A (2000) Self assessment Manikin (SAM) measure. Available at http://www.temple.edu/ispr/frame_measure_t.htm. Accessed 11 Oct 2002
- Lang PJ, Greenwald MK, Bradley MM, Hamm AO (1993) Looking at pictures: affective, facial, visceral and behavioural reactions. *Psychophysiology* 30:261–273
- Lavie N (2005) Distracted and confused? Selective attention under load. *Trends Cogn Sci* 9:75–82
- Lavie N, Cox S (1997) On the efficiency of attentional selection: efficient visual search results in inefficient rejection of distraction. *Psy Sci* 8:395–398
- Lombard M (2005) Temple presence inventory (TPI) questionnaire items and factor structures. Available at <http://astro.temple.edu/~lombard/research/P2qstnre.doc>. Accessed 08 Sep 2008
- Lombard M (2008) Presence and telepresence scholarship: challenges ahead. Keynote presentation at *When Media Environments Become Real (WMEBR)*. University of Bern, Switzerland
- Lombard M, Ditton TB (1997) At the heart of it all: The concept of presence. *J Comput-Mediat Comm* 3(2). Available at <http://jcmc.indiana.edu/vol3/issue2/lombard.html>. Accessed 30 Jul 2000
- Lombard M, Reich RD, Grabe ME, Bracken CC, Ditton TB (2000) Presence and television: the role of screen size. *Hum Comm Res* 26(1):75–98
- McCulloch CE, Searle SR (2000) *Generalized, linear, and mixed models*. Wiley, New York
- Minsky M (1980) Telepresence. *Omni* 2(9):45–51
- Sadowsky W, Stanney K (2002) Measuring and managing presence in virtual environments. In: Stanney KM (ed) *Handbook of virtual environments technology*. Lawrence Erlbaum Associates, Hillsdale
- Schloerb DW (1995) A quantitative measure of telepresence. *Presence Teleop Virt* 4:64–80
- Schneider EF, Lang A, Shin M, Bradley SD (2004) Death with a story: how story impacts emotional, motivational, and physiological responses to first-person shooter video games. *Hum Comm Res* 30:361–375
- Sheridan T (1992) Musings on telepresence and virtual presence. *Presence Teleop Virt* 1:120–126
- Slater M (1999) Measuring presence: a response to the Witmer and Singer questionnaire. *Presence Teleop Virt* 8:560–566
- Slater M (2004) How colorful was your day? Why questionnaires cannot assess presence in virtual environments. *Presence Teleop Virt* 13:484–493
- Slater M, Garau M (2007) The use of questionnaire data in presence studies: do not seriously Likert. *Presence Teleop Virt* 16:447–456
- Slater M, Wilbur S (1997) A framework for immersive virtual environments (FIVE): speculations on the role of presence in virtual environments. *Presence Teleop Virt* 6:603–616
- Snow MP, Williges RC (1998) Empirical models based on free-modulus magnitude estimation of perceived presence in virtual environments. *Hum Factors* 40:386–402
- Sternberg RJ (2000) Implicit theories of intelligence as exemplar stories of success: why intelligence test validity is in the eye of the beholder. *Psychol Public Pol L* 6:159–167
- Steuer J (1992) Defining virtual reality: dimensions determining telepresence. *J Comm* 42:72–92
- Szalma JL, Warm JS, Matthews G, Dember WN, Weiler EM, Meier A, Eggemeier FT (2004) Effects of sensory modality and task duration on performance, workload, and stress in sustained attention. *Hum Factors* 46:219–233
- van Baren J, IJsselsteijn WA (2004) Measuring presence: a guide to current measurement approaches. IST-FET OMNIPRES project, deliverable 5. Available at: <http://www.informatik.umu.se/~jworth/PresenceMeasurement.pdf>. Accessed 11 Nov 2008
- Welch RB (1997) The presence of aftereffects. In: Salvendy G, Smith MJ, Koubek RJ (eds) *Designs of computing systems: cognitive considerations*. Elsevier, Amsterdam, pp 271–276
- Welch RB, Blackmon TT, Liu A, Mellers BA, Stark LW (1996) The effects of pictorial realism, delay of visual feedback, and observer interactivity on the subjective sense of presence. *Presence Teleop Virt* 5:263–273
- Wirth W, Hartmann T, Böcking S, Vorderer P, Klimmt C, Schramm H et al (2007) A process model of the formation of spatial presence experiences. *Media Psychol* 9:493–525
- Witmer B, Singer M (1998) Measuring presence in virtual environments: a presence questionnaire. *Presence Teleop Virt* 7:225–240