A Comparison of Two Presence Measures Based on Experimental Results

Espen Nystad, Angelia Sebok
OECD Halden Reactor Project
{espen.nystad@hrp.no, angelia.sebok@hrp.no}

Abstract

This paper describes a study performed to evaluate two commonly used measures of presence. The study involved 24 participants using a VR-based procedural training application with different technology types. A variety of performance measures and background information were collected during the study. These variables were factors that are assumed to be related with sense of presence. The presence questionnaires of Witmer & Singer and Slater, Usoh & Steed were used to assess sense of presence in the virtual environment. Half of the subjects received the W&S questionnaire and half received the SUS questionnaire. Following the data collection, a variety of correlations were calculated and evaluated to provide insights into the two measures of presence. The questionnaire of Slater, Usoh and Steed gave results that were most consistent with the presence construct defined by the authors.

1 Introduction

The main advantage of Virtual Environments (VEs) is that they offer the user the opportunity to be present in another world. This means that training can be offered in an almost-real environment, safe from the hazards associated with the real world. For safety-critical industries, e.g., the nuclear industry, this is a tremendous benefit.

This feeling of being present in a simulated environment is referred to as sense of presence (SOP). This paper discusses two commonly-used presence metrics. A study conducted at the Virtual Reality Centre of the OECD Halden Reactor Project provides the basis for this discussion and offers further insights into these presence metrics. The study reported here was a part of a larger study investigating the use of different technology types for training in VEs [8].

When conducting research on presence, it is important to know what is really being measured. There is still no agreement on how to measure SOP. To reach an agreement on which measure to use, one must find a measure that demonstrably assesses the concept defined as presence. For a measure of presence to be valid, the responses of a subject group to the measure should be correlated with factors that are assumed to determine presence. For instance, if the level of realism in the VE is assumed to affect presence, then different levels of realism should result in different responses to the presence measure. This is called construct validity. Two ways of determining construct validity are by 1) considering how well the measure can reveal differences between groups, and 2) by observing correlations with important factors [2]. This study will identify potential factors that should be associated with an enhanced sense of presence, and see how the two presence measures relate to these factors. The presence measures studied are the Witmer and Singer Presence Questionnaire (PQ) [10, 17] and the Slater, Usoh Steed inventory (SUS) [12,13]. The experimental results will be presented in terms of these expectations and their implications will be discussed.

2 The presence concept

To be able to measure presence, it is necessary to have a clear concept of what presence is. Singer and Witmer [10] define presence as: 'the subjective experience of being in one place when one is physically in another.' and in virtual environments as 'experiencing the computer generated environment rather than the actual physical locale'. Steuer [15] similarly defines telepresence as 'the extent to which one feels present in the mediated environment, rather than the immediate physical environment', and Sheridan defines virtual presence as 'feeling like you are present in the environment generated by the computer' [9].

Presence has been identified as a key component in the definition of virtual reality [15,20]. Steuer [15] argues for using presence as *the* defining trait of VR to make the definition independent of the technological implementation of the VR system.

In the present study we distinguish between presence and immersion. Immersion is seen as a characteristic of the VR technology and describes the degree to which the technology engrosses the user in the VE; the degree to which the VE surrounds the user and the degree to which the technology gives a realistic or high-fidelity representation of the VE. By realistic / high-fidelity representation is implied a representation that is close to the way we perceive the real world, e.g. with stereo (rather than mono), real time shadows (rather than no shadows) or high resolution (rather than low). This definition of immersion is in line with Slater [11].

3 Measuring presence

A common way to measure presence is by using questionnaires because they are fairly simple to administer and analyse. The SUS and PQ questionnaires have been chosen for two reasons. Both are frequently used measures of presence, and both consist of a relatively small number of questions, which make them convenient to use. The PQ consists of 32 questions. Of these, 19 have been through a process of empirical testing by reliability and cluster analyses. These 19 questions were used in the present study. Three subscales have been determined from cluster analysis. They address issues related to the level of involvement and control, the naturalness of the interaction, and the quality of the interface. The questions were derived from a literature review of factors assumed to influence presence. The factors were categorised as control, sensory, realism and distraction factors. The PQ questions are in the form of a Lickert-scale with seven response choices (e.g., ranging from “not at all” to “completely”).

Whereas the PQ measure is constructed from factors assumed to influence presence, the SUS measure is more directly related to the feeling of being in the virtual environment. There are different versions of this questionnaire; one with three questions and one with six questions. In the present study the one with six questions was used. The questionnaire includes three aspects: the sense of being in the VE, the extent to which the VE becomes more dominant than the real world, and the extent to which the VE is remembered as a place the user visited rather than images the user has seen. The SUS questions are in the form of a Lickert-scale with seven response choices

Slater [11] has criticized the PQ measure for confounding the assessment of the user's personal characteristics and assessment of the immersiveness of the system. The questionnaire asks about the user's subjective response to various system factors. The answers will therefore vary according to individual differences of the users, not only according to variations in system characteristics. Slater is also critical of the way the questionnaire is based on assumptions about which factors should influence presence. He argues that the questionnaire "cannot be used to study the factors, according to W&S, that influence presence. The presence score is constructed out of those factors. It is their sum." [11].

Both the PQ and the SUS measures have been used extensively to study the effects of various factors on SOP. Youngblut and Perrin [19] did a review of roughly thirty studies where the PQ or SUS measures were used. They looked at factors that were related with each of the questionnaires, and reported for which of these factors there were consistent results across two or more studies. They found thatPQ was positively related with display field of view (FOV), while for head tracking, task-related experience and gender, the results consistently showed no relationship with the PQ. The SUS questionnaire was positively related with immersive tendencies. Youngblut and Perrin concluded that the data did not give an indication about the validity of the measures, or whether they measured the same construct.

Some previous studies have compared the two presence measures. Youngblut and Huie [18] studied learning of procedures using either immersive or desktop VE technologies. There was no difference between the immersive and desktop VEs in SOP experienced during training for any of the measures. Performance in the transfer of training test was positively correlated with the SUS measure, but not with the PQ. Youngblut and Perrin [19] used the two questionnaires in a study of VR maintenance training. They found a partial negative correlation between PQ and the Immersive Tendencies Questionnaire (ITQ, see section 4.3.2), and a partial positive correlation between SUS and ITQ. Both measures were negatively correlated with number of errors on a transfer of training task.

Usoh et al. [16] used the two questionnaires to compare a real office environment with a virtual one. The PQ was not able to distinguish between the two environments, while the SUS showed a marginal difference.

4 Factors associated with presence

A number of factors are believed to influence SOP. In this section, we describe the factors that were examined in this study and how they are assumed to relate to SOP. The factors are divided into system factors, performance and personal factors.. The system factors are immersion/realism and usability, the personal factors are familiarity with the environment and immersive tendencies. A valid measure of presence should be sensitive to variations in these factors.

4.1 System factors

4.1.1 Immersion In general, SOP is expected to be greater in more immersive, realistic systems. Sheridan suggest that SOP is determined by the extent of sensory information, i.e. the fidelity or richness of the VE [9]. Similarly, Steuer describe sensory depth (the resolution or quality of the sensory information) as one factor influencing SOP [15], and Zeltzer claim that SOP "provides a rough, lumped measure of the number and fidelity of available input and output channels" [20]. Witmer and Singer claim that SOP is related to scene realism, with FOV and dimensionality being factors that govern scene realism. They also suggest that a display that isolates the user from the physical environment may give a higher SOP [17]. Display systems with stereoscopic view can be expected to have associated higher SOP than monoscopic view displays because they provide more sensory information and realism. Large-screen, wide FOV displays are expected to be associated with higher SOP than standard desktop displays. HMDs with orientation tracking should make the user feel more immersed in the VE and more isolated from the physical surroundings, and therefore give a higher SOP than displays without orientation tracking or displays that isolate the user to a lesser degree.

4.1.2 Usability System usability is expected to be positively correlated with SOP. Particularly, SOP is assumed to be related to the control capabilities, e.g. the ability to modify viewpoint [9], response time [15], and the extent to which mapping between controls and effects in the VE are natural and predictable [15]. VEs in which users can easily implement control actions lead to users feeling present. If the control mechanisms for moving about and interacting with the VE are intuitive and easy to use, they become more or less transparent and the user can focus on the environment itself. Conversely, if users must struggle to figure out how to take control actions, must wait for feedback, are unable to navigate easily, they will not be convinced that they are in the environment. One way to measure usability is by questionnaires. In this study Brooks' usability test was used [1].

4.2 Performance

SOP is generally assumed to be associated with improved performance [9, 15, 20]. According to Stanney et al. [14], the claim that there should be a relation between presence and performance has face validity and is supported by perceptual and cognitive theories. While Draper, Kaber, and Usher [3] state that the evidence linking presence and performance is scarce, they identify a number of sources that claim that presence is associated with enhanced performance. However, the nature of the relationship is not clear. Draper, Kaber and Usher note that many of the factors commonly thought to determine presence, are also determinants of performance. Although this makes it difficult to discover if there is a causal relationship between presence and performance, it supports the assumption that there is a positive correlation between the two.

4.3 Personal factors

4.3.1 Familiarity with environment User familiarity with the corresponding real-world environment is expected to be associated with SOP in a VE. Witmer and Singer predict that SOP should increase as the VE becomes more meaningful to the user [17]. If users are familiar with the corresponding real-world environment, and if the VR model is sufficiently realistic, familiarity might be expected to be positively correlated with SOP. Users recognize the environment and it becomes meaningful to them. Users who have never seen the actual physical environment might be less likely to be convinced by a VR model.

4.3.2 Immersive tendencies Steuer predict that individual factors will influence SOP [15]. Slater, Usoh and Steed assumed that individual factors influencing SOP are the representation system used in a given context (visual, auditory or kinaesthetic) and the dominant point of view of information processing (internal or external) [12]. Witmer and Singer have constructed a questionnaire the Immersive Tendencies Questionnaire, based on the assumption that some people have a higher tendency to feel involved in a VE than others [10,17]. The ITQ tries to assess this tendency. If presence is in fact related to an individual's ability to be drawn in, then immersive tendencies can be expected to be positively correlated with SOP.

4.4 Summary of expected results

Presence should be positively correlated with:

1. Level of immersion
2. Usability
3. Performance
4. Familiarity with the environment.
5. Immersive tendencies

5 Method

5.1 Participants

24 employees of the OECD Halden Reactor Project volunteered for the study. These consisted of 22 males and 2 females, with ages ranging from 25 to 61 years. Gender was not assumed to influence presence. Therefore no attempt was made to balance gender in the experimental design.

Participants were from three different work groups. These groups represented subjects with high , medium and low familiarity with the real-world environment depicted in the VE and the maintenance task performed in the study..

5.2 Experimental design

The experimental task was to learn a multiple-step maintenance procedure in a nuclear reactor setting. The procedure was divided into three parts, and all subjects learned all three parts of the procedure. Four different technology types with different levels of immersion were evaluated: desktop display, desktop display with stereo view, large screen with stereo view, and HMD with orientation tracking. See Table 1 for details of each technology type. Each subject used three of the four technology types, and each technology type was used by the same number of subjects. The order of technology types was balanced across subjects. SOP was assessed for each part of the procedure. Half of the subjects were given the PQ measure and the other half the SUS measure. All subjects answered the presence questionnaires after each of the three procedure training sessions. There were 36 responses for each of the two questionnaires. Since all participants did not use the same technology types, the design was treated as a between-subjects design instead of a repeated-measures design when analysing the data.

Technology type	*Approx. Physical FOV*	*Reso-lution*	*Depth effect*	*Head orient. tracking*
Desktop mono	30º horiz. 20º vert.	1024x768	Mono	No
Desktop stereo	30º horiz. 20º vert.	1024x768	Stereo	No
Large screen	90º horiz. 45º vert.	2048x768	Stereo	No
HMD	55º horiz. 45º vert.	800x600	Mono	Yes

Table 1 Technology types evaluated in the experiment

5.3 Independent variables

Subject group and technology type were the two independent variables in the study. The three subject groups varied in the level of familiarity with the environment and task. Four different technology types was used (Table 1). Given our definition of immersion and the discussion of immersion factors in section 4.1.1, the technology types can be ordered on a scale of increasing immersion primarily by increasing FOV. Where the FOV is identical, as is the case with the two desktop displays, the stereo display is more immersive than the mono display. The HMD has a FOV that lies between the desktop and large screen displays, and it has head-tracking. So it is more immersive than the desktop mono display. On the other hand, the resolution is lower. It is therefore difficult to say how this compares with the large screen and desktop stereo displays.

5.4 Dependent variables

The main variable of interest was SOP, measured with the PQ and SUS questionnaires. Several measures were collected to reveal potential relationship with SOP. These include objective measures and questionnaires.

5.3.1 Usability A 10-item questionnaire, Brooks' usability questionnaire, was used to evaluate system usability [1]. It included statements about system ease-of-use, learnability and complexity. The items used a five-point Lickert scale, and the subject was asked to indicate the amount of agreement with these statements, from “Strongly disagree” to “Strongly agree”.

5.3.2 Performance Two types of performance were assessed in the study. Firstly, the number of incorrect actions taken by the participant during the training sessions were recorded in the data logs as errors. Secondly, transfer of training was assessed by having the subjects recall the procedure steps and the tools used while looking at pictures from the real reactor hall. The number of tools incorrectly remembered or forgotten were counted. The number of procedure steps omitted or incorrectly remembered were also counted.

The fewer errors the subject made, the better the performance. It is therefore expected that errors are negatively correlated with presence.

5.3.3 Familiarity with environment Participants were asked how many times they had been in the real reactor hall which was represented in the VE. The answers were recorded in the following categories: 1-5 times, 6-25 times, 26-100 times, 101-300 times and more than 300 times.

5.3.4 Immersive tendencies A questionnaire was given to all participants to assess their immersive tendencies, or their ability to get drawn into external events. The Witmer and Singer ITQ was used [10,17]. It has 18 questions with a seven-point Likert scale. The inventory consist of the subscales focus, involvement and games.

5.5 VR application

Three procedural training programs were developed to assess procedural learning. These training programs are based on a control-station change-out procedure that is used at the Halden boiling-water reactor. The procedures are small portions (8-12 actions) of the entire control-station change-out procedure.

A training application, VRTexp, was developed specifically for this experiment based on an existing training toolkit [7]. VRTexp was used for showing the procedures in a VR model of the reactor hall, and provided a geometric FOV of 45 degrees (figure 1). This application was built using an extended version of XJ3D, which uses Java 3D to visualize VRML models and allows the application to be run on a variety of VR technology types, including stereoscopic displays.

Users navigated and looked around in the VE with a mouse. In the HMD condition, users were able to change direction of gaze by moving the head instead of using the mouse.

5.6 Experimental Procedure

Each participant attended the study for four sessions held on separate days. Each session lasted from 30 minutes to 1 hour, depending on the material to be covered and the subject’s personal working speed. The order of session presentation was identical for all subjects. The first procedure was learned in the first session. A retention test was given during the second session. The second and third procedures were learned in the third and fourth sessions.

For all VR sessions, participants arrived at the VR lab and were shown the equipment they would be using for that day’s session. They received instructions on the use of the equipment, navigating in the VR model, and interacting with objects in the VE. Participants worked with a practice scenario for 10 minutes to familiarize themselves with selecting objects and navigating in the VE. This familiarization phase was offered for all technology types, to ensure that subjects had equal exposure to VR models in all technology types before beginning the training.

Figure 1 Screenshot from the VR application

Following the familiarization phase, subjects began the procedure training. Subjects experienced 5 repetitions of each procedure. The first two repetitions were passive viewings; in the final three, the subjects were actively performing the procedures.The experimenter first gave the subjects verbal instructions on how the procedure should be performed. The subject then watched two virtual "videos" (i.e., animation sequences of the actions to be performed) of the procedure. In the first viewing, the subject controlled the rate at which the steps were presented. During the second viewing, the subject watched a video of the entire procedure. This played at a pre-determined rate that the subject could not influence. In both videos, text was presented at the bottom of the display to describe the step that was about to occur or actually occurring.

Following these first two viewings, the subject was then given a short (1-2 minutes) set of instructions for actively performing the procedure. Then the subject began doing the procedure him/herself. Here, the text to describe the steps was not presented. Subjects did have access to this text (by using a Help command). Further, if the subject was unable to remember the step, even with the text prompt, s/he could instruct the application to execute the task.

When a subject performed a procedure in the VE, s/he did so by selecting (i.e., clicking on) objects - generally the tool that would be used to perform the task. These were associated with an animation sequence. The subject selected a tool and the appropriate action occurred in the VE. If the subject selected an incorrect object, feedback would be given in the text field: "Please select another object."

After each session, the subjects filled out the presence and usability questionnaires. After the first session, they also filled out a background questionnaire including environment familiarity. After the second session they filled out the ITQ.

5.7 Data handling and analyses

Upon completion of the data collection, analyses were conducted. Data were obtained from computer logs and questionnaires. The responses to each presence measure were summed across questions to get the total presence score or subscale scores. The scores were then standardised.

In the data analyses, Spearman's rank order correlation was used to calculate correlations between the presence measures and each of the hypothesized influencing factors. ANOVA analyses were used to look for differences in presence between the three subject groups and between the four technology types. A significance level of 0.05 was used.

6 Results

The following tables show correlations between reported sense of presence and various factors believed to be related to presence. For the SUS measure, the correlations with the total (sum) score is presented in the tables. For the PQ measure, correlation with the total score and each of the subscales involved/control (IC), natural (N) and interface quality (IQ) are presented.

6.1 System factors

6.1.1 Presence and level of immersion Analysis of variance was performed for the SUS and PQ measures respectively to see if there were any differences in presence for the four technology types. Neither of the measures showed any significant differences.

6.1.2 Presence and usability Usability, as measured by Brooks' usability test [1], was significantly correlated with the PQ total score and the involved/control and interface quality subscales (Table 2). It was not correlated with the SUS measure.

6.2 Performance

Correlation between presence and errors during the active training is shown in Table 3. There was a significant negative correlation with the SUS measure only in the third active repetition. There was no significant correlation between the PQ measure and errors during training.

The correlation between presence and retention errors on the memory test one day after learning the procedure are shown in Table 4. Separate correlations were calculated for tool errors and procedure errors. Tool errors were negatively correlated with the SUS measure and positively correlated with the PQ interface quality scale.

SUS	**PQ**			
	IC	N	IQ	Total
not sig.	R=0.39, p=0.02	not sig.	R=0.69, p<0.0001	R=0.46, p=0.005

Table 2 Correlations between presence and usability

SUS	PQ			
	IC	N	IQ	Total
Session 3: R = -0.38, p=0.02	not sig.	not sig.	not sig.	not sig.

Table 3 Correlations between presence and errors during training

	SUS	PQ			
		IC	N	IQ	Total
Equipm. errors	R=-0.87, p=0.0002	not sig.	not sig.	R=0.58, p=0.05	not sig.
Procedure errors	not sig.	not sig.	not sig.	not sig.	not sig.

Table 4 Correlations between presence and retention errors

SUS	PQ			
	IC	N	IQ	Total
R=0.71, p<0.0001	not sig.	not sig.	not sig.	not sig.

Table 5 Correlations between presence and familiarity with the environment

	SUS	PQ			
		IC	N	IQ	Total
Focus	not sig.	not sig.	R=-0.38, p=0.02	not sig.	not sig.
Involv	R=0.67, p<0.0001	not sig.	not sig.	not sig.	not sig.
Game	not sig.	not sig.	not sig.	not sig.	not sig.
Total	R=0.40, p=0.02	not sig.	R=-0.40, p=0.01	not sig.	not sig.

Table 6 Correlations between presence and immersive tendencies

6.3 Personal factors

6.3.1 Presence and familiarity with environment The number of times the subject had been in the real reactor hall was positively correlated with SUS, but not with PQ (Table 5).

The three subject groups were also investigated to see if there was a difference in presence between them. The three groups differed in their familiarity with the environment and also in familiarity with the procedure learned in the study. There was a significant interaction effect between subject group and presence measure, ($F(2, 48)=16.13$, $p< 0.0001$), see Figure 2. Tukey HSD post hoc-test showed a significant difference between the presence measures for both the high familiarity (p=0.01) and low familiarity subjects (p=0.001). The high familiarity subjects rated presence lower with the PQ measure than the SUS measure, while the opposite was true for the low familiarity subjects.

6.3.2 Presence and immersive tendencies The SUS measure was positively correlated with the total immersive tendencies score, and specifically with the involvement subscale. The PQ natural subscale was negatively correlated with the ITQ focus subscale and the ITQ total score (Table 6).

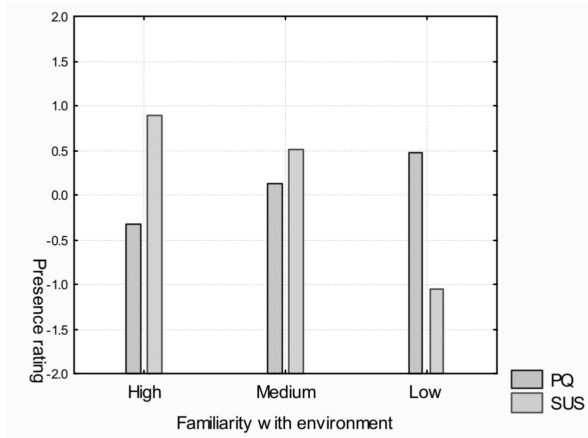

Figure 2 Interaction between presence measure and familiarity with environment

6.4 Summary of presence findings

Table 7 summarizes the results of the study described in the previous sections. The expected relationship between presence and each of the factors is also given in the table.

Factor		*Expected relation-ship*	*SUS*	*PQ*
System factors	***Level of immersion***	*positive*	none	none
	Usability	*positive*	none	positive **
Perfor-mance	***Performance (Errors during training)***	*negative*	negative *	none
	Performance (Retention errors)	*negative*	negative **	positive (partly)
Personal factors	***Familiarity with environment***	*positive*	positive **	none
	Immersive tendencies	*positive*	positive *	negative (partly)

*: p<0.05

**: p<0.01

partly: correlation with one or more of the subscales of the PQ; but not with the total score.

Table 7 Summary of results

7 Discussion

In this study, we investigated a number of factors thought to be associated with sense of presence. We have evaluated the relationship between these factors and both the SUS and PQ presence scores. Each of the individual results will be discussed and the overall picture will be summarized.

7.1 System factors

The level of immersion varied in the form of screen size, stereoscopic or monoscopic presentation and head-tracking. There was no difference in reported sense of presence among the four technology types. This was true for both the SUS and PQ presence measures. The result is in accordance with the lack of difference between immersive and desktop technologies found previously for both measures [18] and the lack of a relationship with head-tracking found for the PQ [19]. Immersion is believed to be an important factor to lead the user to feel present in the VE. The measures may be lacking in sensitivity for this factor. Another reason may be that the technologies are too different. Usoh et al. [16] suggest that comparisons across different virtual environments, e.g. comparing desktop to immersive technologies may not be valid using subjective presence measures. However, other presence questionnaires have been found to be sensitive to immersion factors [4, 6].

The other VE-related factor thought to be associated with sense of presence is rated system usability. If the system is perceived as easy to use, it should contribute to (or at least not interfere with) the sense of being in the environment. The results indicate that only the PQ measure was correlated with subjective ratings of usability. The correlation was positive. The PQ questionnaire subscales that were positively correlated with usability were Interface quality and Involved / control. Since these two subscales include questions which deal directly with the quality of the interaction, they are expected to correlate positively with usability. The SUS questionnaire, on the other hand, does not specifically address quality-of-interaction issues and did not correlate with rated system usability.

7.2 Performance

SUS was negatively correlated with errors during learning and retention errors, as expected. This is consistent with the relationship found between presence and transfer of training by Youngblut and Perrin [19], and Youngblut and Huie [18]. It is not possible to say something definite about the relationship between presence and performance in this study, but it may be that familiarity with the environment and task is an underlying factor influencing both presence and performance. The PQ interface quality scale was partly positively correlated with retention errors. This is opposite to the expected relationship. The explanation for this may be connected with the negative relation between PQ and familiarity indicated in figure 1, although there is not a statistical significant relationship. If familiarity with the environment is in fact influencing both presence and performance, then the negative correlation between presence as measured by the PQ and performance could be mainly due to variation in the familiarity factor. In other words, performance may be more related to familiarity than to the PQ measure. PQ looks to be negatively related with familiarity and, as a consequence of this, positively related with performance errors.

7.3 Personal factors

The results showed a strong correlation between the SUS measure and familiarity with the environment. This finding is consistent with the relationship found between chess experience and sense of presence by Hoffman et al.[5]. That study used two questions from the SUS measure and two other similar questions. Both results show that presence as measured by the SUS questionnaire is related to the meaningfulness of the VE, and that presence can be enhanced when the content of the VE matches previous experience. The prior experience with the environment probably makes the user associate the virtual environment with the real one, and memories of the real environment add to the experience to make it more like the real thing. Meaningfulness of experience is one of the factors that Witmer and Singer predict will influence presence [17]. But no statistical significant relationship was found in our study between the PQ measure and familiarity with the environment. In this respect, the two questionnaires seem to measure different aspects of the presence construct. A lack of realism in interaction and selection of objects may underlie this difference. The interaction in performing the procedure tasks may have felt less natural for the familiar subjects, since they knew how the tasks are performed in reality. In the PQ measure, this could result in a low score on interaction quality and naturalness. The unfamiliar subjects may have been less critical to the way the task was performed, and rated these scales higher. The questions in the SUS measure are more about the feeling of being in the VE than about the tasks performed. So the familiar subjects may have felt more 'there' than the unfamiliar subjects. This points to a possibly important distinction between the two measures. The PQ seems to be oriented around the tasks performed in the VE, while the SUS is more oriented around the place itself.

Another personal factor believed to be associated with sense of presence is immersive tendencies. The SUS measure was correlated with the Immersive tendencies Questionnaire, and this relationship was accounted for mainly by the involvement scale of the ITQ. This scale refers to the ability to get deeply involved in an activity or a stimulus , e.g. books, television or movies. The PQ measure was partly negatively correlated with the ITQ, for the natural subscale. The expected relationship was not found. This may be because, as Slater claims [11], the PQ confounds individual differences and assessment of variations in system characteristics.

7.4 Summary and conclusion

For the five factors where a significant positive or negative relation was expected (counting the two performance measures as one factor), the PQ measure gave expected results for one factor: usability. The SUS measure gave expected results for three factors: performance, familiarity with the environment, and immersive tendencies. Both measures were, contrary to expectation, unrelated with level of immersion. There were no factors for which the measures gave similar, expected results.

Our findings indicate that the SUS and PQ questionnaires are tapping into two different aspects of presence. It seems that responses to the SUS measure are more dependent on personal factors than factors related to technology and the VR system. The SUS questionnaire indicates a sensitivity to user familiarity with the environment, which provides a strong indication of its real-world usefulness. SUS is also sensitive to performance. Some authors see the link between presence and performance as the main reason for studying the presence phenomenon [3]. This is a good argument for using the SUS measure. However, the possibilities for and quality of interaction is commonly seen as a determinant of presence, and a good presence measure should be sensitive for these factors. The SUS did not show a significant correlation with these factors as indicated by the usability questionnaire. This may be because the SUS questions ask about the feeling of being in the place, and not specifically about interaction with the environment.

In this study, the PQ was sensitive to usability. Youngblut and Perrin [18] reported that the PQ was consistently related with FOV. The PQ therefore seems to be more related to technology or interaction factors.

The SUS questionnaire fits better with the definition of presence – a sense of being in the virtual environment – and the expectations we had identified based on that concept. The PQ seems to assess a too narrow slice of the presence concept. This study suggests that a good deal of research is still needed to develop a valid measure of presence.

8 References

[1] Brooks, J. (1998). SUS — A "quick and dirty" usability scale. (World Wide Web document). http://www.cee.hw.ac.uk/~ph/sus.html. Updated October 23, 1998. Digital Equipment Corporation.

[2] Cronbach, Lee J. & Meehl, Paul E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281-302.

[3] Draper, J. V., Kaber, D. B., & Usher, J. M. (1998). Telepresence. *Human Factors, 40(3),* 354-375.

[4] Hendrix, C., & Barfield, W. (1996). Presence within virtual environments as a function of visual display parameters. Presence: Teleoperators and Virtual Environments, 5, 274 - 289.

[5] Hoffman, H. G., Prothero, J., Wells, M. J., & Groen, J. (1998). Virtual chess: Meaning enhances users' sense of presence in virtual environments. *International Journal of Human-Computer Interaction, 10(3)*, 251-263.

[6] Lessiter, J., Freeman, J., Keogh, E., Davidoff, J. (2001). A cross-media presence questionnaire: The ITC-Sense of Presence Inventory. Presence: Teleoperators & Virtual Environments, 10, 282-298.

[7] Louka, M.N. (1999). A Software Toolkit for Implementing Low-Cost Virtual Reality Training Systems. HWR-581. Halden, Norway: OECD Halden Reactor Project.

[8] Sebok, A. & Nystad, E. (2004). Training in virtual reality: A comparison of technology types. HWR-734. Halden, Norway: OECD Halden Reactor Project.

[9] Sheridan, T. B. (1992). Musings on telepresence and virtual presence. *Presence, 1(1),* 120-125.

[10] Singer, M. J., & Witmer, B. G. (1996). *Presence Measures for Virtual Environments: Background & Development. (ARI draft research note and questionnaires).* Alexandria, VA: Simulator Systems Research Unit, U. S. Army Research Institute for the Behavioral and Social Sciences.

[11] Slater, M. (1999). Measuring presence: A response to the Witmer and Singer presence questionnaire. Presence 8(5), 560-565.

[12] Slater, M., Usoh, M., Steed, A. (1994). Depth of presence in virtual environments. *Presence 3, 130-144.*

[13] Slater, M., Usoh, M., Steed, A. (1995). Taking Steps: The influence of a Walking Technique on Presence in Virtual Reality. ACM Transactions on Computer Human Interaction, 2, 201-219.

[14] Stanney, K. et al. (1998). Aftereffects and sense of presence in virtual environments: Formulation of a research and development agenda. *International Journal of Human-Computer Interaction, 10(2),* 135-187.

[15] Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence. *Journal of communication, 42(4),* 73-93.

[16] Usoh, M., Catena, E., Arman, S., & Slater, M. (2000). Using presence questionnaires in reality. *Presence, 9(5),* 497-503.

[17] Witmer, B.G., & Singer, M .J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence 7(3)*, 225-240.

[18] Youngblut, C., & Huie, (2003). The relationship between presence and performance in virtual environments: Results of a VERTS study. In *Proceedings of the IEEE Virtual Reality (VR'03).*

[19] Youngblut, C., & Perrin, B. M. (2002). Investigating the relationship between presence and performance in virtual environments. Paper presented at IMAGE 2002 Conference, Arizona.

[20] Zeltzer, D. (1992). Autonomy, Interaction, Presence. *Presence 1(1),* 127-132.