Title

Presence is Key: Unlocking Performance Benefits of Immersive Virtual Reality

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 40(0)

Authors

de Back, Tycho T
van Hoef, Rens
Tinga, Angelica M
et al.

Publication Date

2018

Presence is Key: Unlocking Performance Benefits of Immersive Virtual Reality

Tycho T. de Back[1] (T.T.deBack@uvt.nl)
Rens van Hoef[2] (R.vanHoef@lancaster.ac.uk)
Angelica M. Tinga[1] (A.M.Tinga@uvt.nl)
Max M. Louwerse[1] (M.M.Louwerse@uvt.nl)
[1]Tilburg University, Department of Communication and Information Sciences, Tilburg, The Netherlands
[2]Lancaster University, Department of Psychology, Lancaster, United Kingdom

Abstract

High immersion in virtual reality is often hypothesized to improve learning and memory. This immersion benefit is frequently attributed to presence, the user's feeling of being present inside the computer-generated environment. Obtaining learning gains due to high immersion may however be difficult, as is evidenced by the null results of multiple studies in this area. In the current study we investigated the role of presence in low- and high-immersion virtual reality settings. No differences in performance in object location and spatial memory were found between low- and high-immersive conditions. Yet, when considering self-report measures of presence, performance improvements in the high immersive condition did become apparent. The finding of the importance to consider the role of presence in virtual reality highlights the complexity of immersion effects in simulated environments.

Keywords: performance; virtual reality; immersion; presence

Introduction

Immersive Virtual reality (VR) is starting to gain popularity in the cognitive sciences, likely due to its unique ability to induce highly realistic experiences. The same is true for VR applications in various societal sectors, including the gaming and movie industries, but also the automotive sector, social media, art, architecture, education and others. An important part of the realism of VR is brought forth by presence, which is its ability to give the user a sense of "being there", the feeling of existing inside the virtual environment (Heeter, 1992; Steuer, 1992). Whereas the sense of being present is subjective, presence itself is the result of immersion. Immersion, however, is objective and can be defined as the technological fidelity of VR as produced by hard- and software (Bowman & McMahan, 2007; Slater, 1999).

A common distinction in VR research is that between low immersive VR with a small viewing area as provided by conventional desktop monitors, and high immersive VR with a large viewing area which surrounds the user, as is afforded by head-mounted displays (HMD's) and Cave automatic environments (CAVE's) (Bowman & McMahan, 2007). High immersion is assumed to yield improvements in learning (Meehan, Insko, Whitton, & Frederick P. Brooks, 2002). Importantly however, studies investigating this hypothesized beneficial effect of immersion on learning have yielded decidedly mixed results, with benefits present in some studies and absent in others.

In a study on problem-solving, Kozhevnikov, Gurlitt, and Kozhevnikov (2013) found improvements on problem-solving skills for an immersive HMD condition compared to a non-immersive desktop monitor condition. Similarly, in a study by Limniou, Roberts, and Papadopoulos (2008), students obtained a better understanding of chemical reactions after exposure to molecule animations in a CAVE compared to animations displayed on a desktop monitor.

Conversely, in two experiments on the acquisition and transfer of knowledge of plants in desktop and HMD conditions, Moreno and Mayer (2002) found no evidence for an effect of increased immersion on higher learning outcomes. Likewise, in a study on the recall of object location and shape, Mania, Troscianko, Hawkes, and Chalmers (2003) did not find performance differences between desktop monitor and HMD conditions.

In light of the mixed results reported in the literature, it can be questioned whether immersion has a strong positive effect on learning. As immersion does not appear to consistently yield learning gains, this presents challenges for applications of virtual reality and learning which rely heavily on immersion. Finding an answer to the question why the effects of immersion are not consistent could aid in addressing these challenges. The answer may be found in the concept of immersion itself, which may be modulated by other factors. A prime candidate for such a factor is presence, as it is one of the foremost benefits of immersion and has been linked to positive learning outcomes (Mikropoulos, 2006; Witmer & Singer, 1998). Despite this, few studies have investigated the possible modulatory effect of presence on immersive learning.

The central question of the current study is whether presence measures can be employed to find performance differences between low- and high-immersive conditions. To this aim, an experiment was conducted that compared learning performance in a low immersive monitor condition and a high immersive HMD condition, while taking the modulating factor of presence into account.

Experiment

Object recognition and spatial tasks were selected for the experiment of the current study as they actively exploit the rich spatial information immersive VR provides (Bowman & McMahan, 2007), and are thus a natural fit for comparing performance between low- and high-immersive conditions.

In addition to presence, a possible modulatory effect of flow was examined. Flow is a state of optimal experience in which the perceived demands and skills of a task are matched, and has been found to be positively correlated with presence (Csikszentmihalyi, 2000; Hegarty, Richardson, Montello, Lovelace, & Subbiah, 2002; Rheinberg, 2008). Enhanced levels of natural interaction enabled by immersive VR may lead to increases in flow (Liao, 2006). Flow may be disrupted when natural interaction is suboptimal by affecting the balance between perceived demands and skills.

Lastly, sense of direction on task performance was assessed. Sense of direction is a measure of spatial ability relevant for route- and landmark learning, map drawing and other spatial tasks (Hegarty et al., 2002). Its use is therefore appropriate in light of the object recognition and spatial tasks of the current study.

To measure presence, flow, and sense of direction self-report questionnaires were employed. These questionnaires and their respective subscales are often presumed to be independent and are therefore assumed to explain distinct parts of the variance in the data, as is the case for the presence questionnaire used for the current study (Lessiter, Freeman, Keogh, & Davidoff, 2001).

After examining the effect of low versus high immersion, the presence, flow and sense of direction measures will be used to examine their influence on performance in these two immersion conditions. Additionally, the independency of the measures and their subscales will be tested.

Method

Participants

Thirty-three students of Tilburg University (16 females, age: M = 24.61, SD = 3.34, range = 18-32) participated in this experiment either for a monetary reward (15 Euro) or 1.5 course credits. Participants did not have a current or past condition of migraine or epilepsy, had normal or corrected-to-normal vision, did not wear glasses during the experiment, and were not pregnant. Finally, participants were excluded if they had visited Japan, as the spatial navigation task involved Japanese locations.

Apparatus and Materials

Virtual environments were presented with a HMD and a conventional desktop monitor (HMD: Oculus Rift consumer edition; horizontal field of view approximately 110 degrees; resolution 2160 x 1200 pixels; 90 Hz refresh rate. Monitor: Dell P2214H; 21.5-inch; resolution 1920 x 1080 pixels; 60 Hz refresh rate). The distance to the monitor was approximately 50 cm. For navigation, a Microsoft Xbox wireless controller was used which allowed rotations by 45-degree increments. Visual stimuli were presented with HMD and desktop monitor versions of Google Street View (software: VISO Places, VISO VIRTUAL).

An object recognition and spatial navigation task was performed using Google Street View 360-degree panoramas of Asakusa, a district in Taito, Tokyo, Japan. A practice session used an unrelated area in Burlington, Oklahoma, United States. Two routes, A and B, of similar complexity were used. Route A contained 16 decision points (go left, right, straight) and was 670 m in length. Route B contained 15 decision points and was 708 m in length. Examples of visual stimuli located along the two routes which were to be memorized are shown in Fig. 1. An object recognition and response time test consisted of three object types, namely specific (object at decision point specifically drawn attention to by the experimenter during the learning phase); non-specific (located along the route but not specifically indicated by the experimenter) and unseen (unseen objects).

Self-report measures of presence, flow and spatial ability were collected via questionnaires. Presence was measured using the ITC-Sense of Presence Inventory, one of the few presence questionnaires compatible with both low- and high-immersive media types (Lessiter et al., 2001). The questionnaire consists of 44 items, divided into subscales measuring spatial presence (sense of being present in the displayed content), engagement (level of psychological involvement), ecological validity/naturalness (perceived realism and naturalness of the displayed content) and negative effects (unwanted side effects such as dizziness resulting from media consumption). The internal reliability ranged between .76 – .94 (Lessiter et al., 2001).

Flow was measured using the Flow Short Scale. This questionnaire measures flow experience, and additionally perceived importance (of the task at hand), demand (imposed by the task), skills (available skills to perform the task) and perceived fit of demands and skills. (Rheinberg, Vollmeyer, & Engeser, 2003). The internal reliability of the scales was approximately .90.

Spatial ability was assessed with the Santa Barbara Sense of Direction Scale, which has an internal reliability of .88 (Hegarty et al., 2002). All questions were administered using Qualtrics survey software (Qualtrics, Provo, UT).

Dependent variables

To assess performance differences between monitor and HMD conditions, we measured participant memory of two virtual routes and objects present along these routes by recording route navigation errors as well as correct object recognition numbers and accompanying response times. Map drawing quality was determined by comparing drawn

Figure 1: Examples of visual stimuli located along Google Street View routes A (top) and B (bottom).

and ideal maps for street length accuracy and correct left-right turns. Lastly, we measured both the number of correctly placed objects along the route as well as the deviance from the correct object placement order.

Procedure

After receiving information about the tasks to be performed, all participants carried out both the low immersive monitor and high immersive HMD conditions in a randomly assigned order. Irrespective of condition, in the practice phase participants first familiarized themselves with a gamepad used for navigation. This was done on a monitor, so as to facilitate communication with the experimenter. Participants in the HMD condition next continued the practice phase inside the VR headset in order to practice the use of head rotations for interacting naturally with the virtual environment. The practice phase lasted between one and two minutes and was ended when participants demonstrated proficiency with the interaction methods.

In the learning phase, participants in both conditions performed an object memory and navigation task in one of two randomly assigned routes. At the start of the learning phase, the experimenter used a predefined script to verbally guide the participants along a predetermined route, which contained a fixed number of decision points (go left, right, continue forward). At each decision point, the experimenter stated the decision point number, asked to look at one specific object or person visible in the scene, and requested to look from left to right. This ensured that all main elements of the scene had been observed by the participants at least once. After reaching the destination, the participants were put back at the start of the route, concluding the learning phase.

The testing phase consisted of four parts and started out with a navigation test. In the first part, the participants were asked to travel along the same route as quickly as possible while minimizing the number of navigation errors at the decision points. When the correct route was deviated from, the experimenter halted the participant to ensure he/she would not get lost, recorded the error and guided the participant back to the last correct position. This process was repeated until the destination was reached. In the second and third parts of the testing phase, participants completed questionnaires on flow, presence and sense of direction and performed a yes/no recognition task on objects visible along the route. Next, for the fourth and last part, participants used a graph paper to draw a map of the route they had just virtually travelled, and placed pictures of objects which were visible along the route at the correct locations within five minutes. For consistency, the start and end points of the route were predesignated on the paper. The procedure was then repeated for the second condition, and was followed up with a short exit questionnaire. After the participants were debriefed, the session was completed. The total session duration was approximately 90 minutes.

Data Analysis

Linear Mixed Effects (LME) models were used on both accuracy and response time variables. Condition (monitor, HMD) was dummy coded and used as a fixed factor, as well as the individual self-report variables of flow, presence and sense of direction. Route (routes A and B) and route order (A-B or B-A) were used as random factors. Subject was not included as random factor in the model as this would remove the variance of interest.

Results and Discussion

Due to a technical issue, the scores and response times of the object recognition test of one of the two routes contained missing values for 8.8% of the data. Missing values were equally distributed among the tested objects and the conditions, with route and route order being counterbalanced, so that the technical issue is not expected to have affected the variables of interest. Values of ±2.5 standard deviations from the (individual) means were removed from further analyses. This amounted to 3% of the data of both the navigation error numbers and the response times of the recognition test.

The reliability of the presence subscales was good to excellent, with Cronbach's α ranging between .72

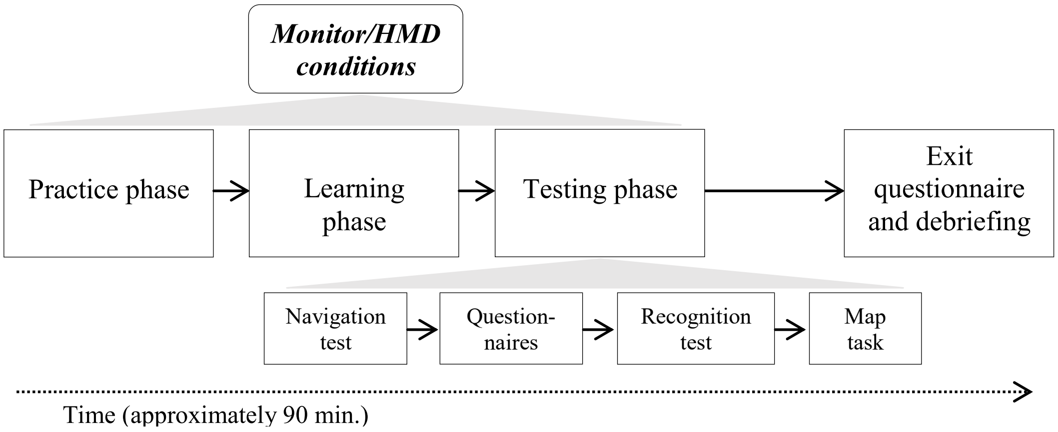

Figure 2: Overview of experimental procedure.

(ecological validity/naturalness) and .94 (spatial presence). The Cronbach's α of flow experience of .80 and sense of direction of .86 were also good. Flow subscale perceived importance had a Cronbach's α of .36 and was removed from further analyses as it was below the acceptable limit.

Immersion

An analysis of the main effect of immersion revealed no significant performance differences between monitor and HMD conditions for either of the outcome variables. This is consistent with some of the findings in the literature (Mania et al., 2003; Moreno & Mayer, 2002), and illustrates that clear-cut performance benefits due to high immersion can be difficult to obtain.

Moderator Variables

Immersion effects may be modulated by presence and other subjective measures. In order to examine this, it was first assessed whether the high immersive HMD condition effectively induced presence in the first place. In accordance with the scoring instructions of the presence questionnaire, the subscales were analyzed individually. Presence scores were shown to be significantly higher in the HMD condition for all four subscales, confirming that the high immersive condition indeed successfully induced presence. For the flow questionnaire this pattern was confirmed for flow experience, but not for the worry, skills and demands scales.

Having determined that the HMD condition induced presence, we proceeded to investigate a possible modulating effect of the subjective measures. For each of the outcome variables we assessed whether the interaction between condition and the questionnaire scale scores was significant. The analyses indicated significant interactions between condition and all four individual presence scales, however not with the scales of the flow and sense of direction questionnaires. In the current study, the effect of immersion was thus solely modulated by presence. With the exception of the outcome variable object order (correct placement order of objects on the route map), all significant interactions coincided with improved estimated marginal means (EMM's) in the HMD condition. This demonstrated that presence can be effectively used for finding differences between low- and high-immersive conditions, and that higher presence scores in the HMD condition of the current study were associated with better task performance. A summary of the results is presented in Table 1.

Principal Component Analysis

Self-report questionnaires and their subscales are often presumed to be independent (Lessiter et al., 2001). This will be examined by conducting a principal component analysis (PCA) on all scales. If the scales are independent, they are expected to fall under different components. If they do not, this may be an incentive for further investigation.

To assess whether a PCA could be conducted, factorability was tested using the Kaiser-Meyer-Olkin measure of sampling adequacy (KMO) and Barlett's test. KMO was within the bounds of acceptability, KMO = .65 (Field, 2009), as were the KMO values of all individual moderator variables, KMO > .50. Barlett's test indicated a sufficient correlation of the moderators, $p < .001$. PCA using varimax rotation was performed using a cut-off of an eigenvalue of ≥

Table 1: Result summary: coefficient estimates β, standard errors SE(β), test statistic t, significance level p and estimated marginal means EMM for all significant condition by subjective measure interactions.

Outcome variable	Object type	Scale	Coef. β	SE(β)	t	p	EMM HMD	EMM Monitor	Condition with improved EMM
Recognition test score	Specific	Negative effects	-.086	.034	-2.546	.014	.922	.873	HMD
RT of correct responses on recognition test	Non-specific	Engagement	.597	.273	2.185	.033	1.731	1.585	Monitor
RT of correct responses on recognition test	Unseen	Spatial presence	.674	.293	.306	.025	1.808	1.891	HMD
RT of correct responses on recognition test	Unseen	Ecological validity/ Naturalness	.660	.286	2.312	.024	1.805	1.824	HMD
Map: Object order	NA	Engagement	.175	.083	2.100	.040	.275	.195	HMD
Map: Object order distance	NA	Negative effects	.801	.369	2.173	.034	2.049	2.618	HMD

Note. Coef. = coefficient estimate; RT = response time; HMD = head-mounted display.

1 for component extraction and 0.4 for component loadings. This resulted in three components together accounting for 71.7% of the variance, with loadings as shown in Table 2.

Component 1 was described as Presence, with three of the four scales of the presence questionnaire being represented, namely spatial presence, engagement and ecological validity/naturalness. Component 2 was described as Competence, as it contained flow experience and skills, both scales of the flow questionnaire. This component also contained the single scale of the sense of direction questionnaire, as well as the fourth presence scale negative effects. Component 3 was described as Difficulty and is comprised of demands, skills and perceived fit of demands and skills, each scales of the flow questionnaire. Of components competence and difficulty, skills loaded most strongly on competence and fits most closely here given the nature of the other loadings of this component. The negative effects scale had a strong negative loading on competence whereas the other loadings on this component were positive. We therefore removed negative effects from this component.

The PCA analysis regrouped ten questionnaire scales into only three components, despite the fact that the questionnaires were considered to be independent.

In the previous analysis we tested whether the presumably independent theoretical constructs in the questionnaires yielded an effect for immersion. Given the PCA findings it is worthwhile to determine whether this effect holds for the questionnaires as determined by the PCA components.

Three variables presence, competence and difficulty were created by averaging the scores of the scales contained in each of the components. The new component variables were used to analyze their modulating effect on immersion, as was done previously for the individual scales. The results indicated a significant interaction between condition and the presence variable for response times of correct answers on the recognition test for unseen objects, β = .866, SE(β) = .354, t = 2.450, p = .017. Importantly, the sign of the interaction coefficient showed that as the average of the spatial presence, engagement and ecological validity/naturalness scale scores improved, persons in the HMD condition were more likely to have lower response times on this combination of outcome measure and object type. This was reflected in the EMM of the HMD condition, which was lower and therefore improved compared to that of the monitor condition, HMD EMM = 1.840, monitor EMM = 1.896. The interactions between condition and the competence and demands variables were not significant.

In conclusion, after grouping the supposedly independent scales into a small number of components, a benefit of immersion on task performance re-emerged by controlling for the presence component variable. This replicates the result of the analyses performed for the individual scales, and fortifies the finding that of the measures of presence, flow and sense of direction, solely presence was effective for detecting performance benefits linked to high immersion.

The PCA showed several scales of the same questionnaire to be grouped under the same component, in spite of their assumed independence. This finding calls for further research into the possibilities for reducing the length of these often extensive questionnaires.

General Discussion

The current study investigated the role of subjective measures in an immersion experiment. Presence, the sense of existing inside a virtual environment, is a characteristic benefit of immersive technology. It is also an established fact that the same level of immersion may induce different degrees of presence, as it is subjective and differs between persons and their mental states. This is of importance, as presence has been linked to learning gains (Mikropoulos, 2006; Witmer & Singer, 1998). Given the fact that presence and its susceptibility to interpersonal differences are widely recognized, it is surprising that studies on immersion and task performance have rarely taken the effect of presence into account. This is even more striking as previous studies

Table 2: Three components resulting from the PCA analysis and rotated factor loadings of the scales higher than .40.

Questionnaire	Scale	Presence	Competence	Difficulty
Presence	Spatial presence	.84		
	Engagement	.91		
	Ecological validity/naturalness	.90		
	Negative effects		-.80	
Flow	Flow experience		.74	
	Demands			.84
	Skills		.69	-.41
	Perceived fit of demands and skills			.82
Sense of direction	Sense of direction		.54	
Eigenvalues		2.70	2.65	1.10
Variance explained (%)		30.02	29.42	12.27

have had difficulty in finding performance differences due to immersion, which may in part have been caused by not controlling for critical factors affecting performance, of which presence is a prime example. The current study addresses this issue and shows that presence can be employed to find performance differences due to immersion.

Further studies are required to elucidate why significant performance differences between monitor and HMD conditions were found for object recognition and object order tasks, but were absent for route navigation and map drawing. One path for improvement would be to use real walking in the HMD condition instead of using a gamepad, as proprioceptive and vestibular information has been shown to improve navigation and the quality of the mental representation of locations (Ruddle, Volkova, & Bülthoff, 2011). Additionally, regarding map drawing, a solution may be to reduce task difficulty, which was high, irrespective of condition. This may be concluded from the small percentage of completed maps, which was 40% for both conditions.

Immersion is brought forth by a range of technical factors, and may provide various benefits differentially affecting behavior and performance. These benefits include presence, increased spatial awareness and others (Bowman & McMahan, 2007). An interesting future avenue would therefore be to parametrize the use of immersive factors assumed to benefit specific tasks, and to assess the role of relevant subjective measures, as done here for presence.

The current study demonstrated the importance of measuring presence when examining the effect of immersion on performance. When presence is disregarded, important performance benefits may remain undetected.

Acknowledgements

The current research is partially funded by the PACCAR foundation and DAF Trucks and on a grant to the last author by the European Union, OP Zuid, the Ministry of Economic Affairs, the Province of Noord-Brabant and the municipalities of Tilburg and Gilze Rijen. The usual exculpations apply.

References

Bowman, D. A., & McMahan, R. P. (2007). Virtual Reality: How Much Immersion Is Enough? *Computer, 40*(7), 36-43.

Csikszentmihalyi, M. (2000). *Beyond boredom and anxiety*. San Francisco: Jossey-Bass.

Field, A. (2009). *Discovering statistics using SPSS*: Sage publications.

Heeter, C. (1992). Being there: The subjective experience of presence. *Presence: Teleoperators and Virtual Environments*, *1*(2), 262-271.

Hegarty, M., Richardson, A. E., Montello, D. R., Lovelace, K., & Subbiah, I. (2002). Development of a self-report measure of environmental spatial ability. *Intelligence, 30*(5), 425-447.

Kozhevnikov, M., Gurlitt, J., & Kozhevnikov, M. (2013). Learning Relative Motion Concepts in Immersive and Non-immersive Virtual Environments. *Journal of Science Education and Technology, 22*(6), 952-962.

Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J. (2001). A Cross-Media Presence Questionnaire: The ITC-Sense of Presence Inventory. *Presence: Teleoperators and Virtual Environments, 10*(3), 282-297.

Liao, L. F. (2006). A Flow Theory Perspective on Learner Motivation and Behavior in Distance Education. *Distance Education, 27*(1), 45-62.

Limniou, M., Roberts, D., & Papadopoulos, N. (2008). Full immersive virtual environment CAVETM in chemistry education. *Computers & Education, 51*(2), 584-593.

Mania, K., Troscianko, T., Hawkes, R., & Chalmers, A. (2003). Fidelity metrics for virtual environment simulations based on spatial memory awareness states. *Presence-Teleoperators and Virtual Environments, 12*(3), 296-310.

Meehan, M., Insko, B., Whitton, M., & Brooks Jr., F. P. (2002). Physiological measures of presence in stressful virtual environments. *ACM Trans. Graph.*, *21*(3), 645-652.

Mikropoulos, T. A. (2006). Presence: a unique characteristic in educational virtual environments. *Virtual Reality, 10*(3), 197-206.

Moreno, R., & Mayer, R. E. (2002). Learning science in virtual reality multimedia environments: Role of methods and media. *Journal of Educational Psychology, 94*(3), 598-610.

Rheinberg, F. (2008). Intrinsic motivation and flow. In J. Heckhausen & H. Heckhausen (Eds.), *Motivation and action*. Cambridge, UK: Cambridge University Press.

Rheinberg, F., Vollmeyer, R., & Engeser, S. (2003). *Die Erfassung des Flow-Erlebens* [The measurement of flow experiences]. In J. Stiensmeier-Pelster & F. Rheinberg (Eds.), *Diagnostik von Motivation und Selbstkonzept*. Göttingen: Hogrefe.

Ruddle, R. A., Volkova, E., & Bülthoff, H. H. (2011). Walking improves your cognitive map in environments that are large-scale and large in extent. *ACM Trans. Comput.-Hum. Interact., 18*(2), 1-20.

Slater, M. (1999). Measuring presence: A response to the Witmer and Singer presence questionnaire. *Presence: Teleoperators and Virtual Environments, 8*(5), 560-565.

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

VISO Places (Version 117) [Computer software]. (2016). Retrieved from http://places.visovirtual.com/

Weibel, D., Wissmath, B., Habegger, S., Steiner, Y., & Groner, R. (2008). Playing online games against computer- vs. human-controlled opponents: Effects on presence, flow, and enjoyment. *Computers in Human Behavior, 24*(5), 2274-2291.

Witmer, B. G., & Singer, M. J. (1998). Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments, 7*(3), 225-240.