The role, measurement and delivery of immersion in training environments

A/Prof. Debbie RICHARDS , Dr. Mark DRAS, Mr. John PORTE and Ms. Meredith TAYLOR
Computing Department, Macquarie University, North Ryde, NSW, 2109
{richards, madras, jporte}@ics.mq.edu.au

Abstract. Immersion is believed to be a relevant factor in providing an effective training simulation. In this study we consider one influencing factor, the media used. We sought to test whether differences existed in perceived levels of immersion when using a single screen, three screen or a semi-cylindrical projection system. We were also interested in what role the level of perceived immersion played in learning. We also briefly consider the phenomenon of immersion and its measurement.

1. INTRODUCTION

Border security has become a major concern to governments and the general population. Learning to detect suspicious behaviour in passengers is a key part of protecting borders; however, such knowledge is difficult to acquire via traditional or formal education. To allow the acquisition of tacit or practical knowledge via problem solving, the trainee needs to explore the environment learning through a process of trial and error. Role plays are sometimes used for such purposes, but an even less threatening environment can be offered by computer-based simulations. Furthermore, virtual environment technologies are able to produce systems that allow increased control of the environment together with increased ecological validity (degree of relevance to the real world). For high risk and security situations of the type that we are exploring, a virtual environment is also safer and more reliable.

An effective training environment that transfers tacit and experiential type knowledge (knowledge that can't just be read in a book or told to someone) requires that the trainees *experience* relevant scenarios. Engagement has been found to be a key factor in learning. Based on the literature it is hypothesized that immersion will increase the level of engagement. While a sense of immersion, such as being there or losing oneself, can be induced in a variety ways (such as an intriguing plot in a novel), we were interested to test whether different output devices provided different levels of immersion, as is popularly claimed, and whether the (perceived) level of immersion correlates to the learning achieved.

2. METHODOLOGY

The three alternative output devices used included: 1) a single desktop screen; 2) three desktop screens that work together to display the output and 3) an immersive semi-cylindrical projection system known as a CAVE (CAVE Automatic Virtual Environment). Figure 1 shows device 2 in the foreground with the same scene being shown on the CAVE in the background.

Figure 1. Virtual Reality (VR) Lab showing the three screen configuration in the foreground and CAVE in the background.

The participants consisted of undergraduate and postgraduate students recruited from campus advertisements. Participants attended our Virtual Reality Laboratory and watched three training demonstrations, each on a different media, each with a different scenario.

Media 1: Single screen.
Media 2: Three screen configuration.
Media 3: CAVE.
Scenario 1: Food
Scenario 2: Drugs
Scenario 3: Passport

The three scenarios contained similar context, type of content and degree of difficulty to understand. Each scenario ran for approximately the same length of time. Each participant saw every scenario and each presentation method, but they only saw one presentation method for each scenario.

To allow between and within subject comparison we created 3 scenarios and 3 media. There are nine (9) combinations of scenarios and media.

Media 1- Scenario 1, Media 2 – Scenario 2, Media 3 – Scenario 3
Media 2 – Scenario 1, Media 3 – Scenario 2, Media 1 – Scenario 3
Media 3 – Scenario 1, Media 1 – Scenario 2, Media 2 – Scenario 3
A latin-squares design was used to allow for order effects. For each of these scenario-device combinations there are 6 orders resulting in 18 overall combinations. Two participants were assigned to each combination, resulting in 36 participants in total.

A number of instruments were used to measure what participants experienced during the study. The experimental procedure is summarized in Figure 2.

Witmer and Singer's (1998) immersive tendencies questionnaire was used to test the individual's tendency to becoming immersed – this could be in a book, a game, a movie, etc.

To measure the participant's perceived sense of immersion or presence within the environment we devised a presence questionnaire (Appendix A) based on Csikszentmihalyi and Csikszentmihalyi (1988), other relevant studies, e.g. VR Snowplow Simulator Training by IOWA (IOWA Dept of Transport, 2007) and Witmer and Singer's (1998) presence questionnaire. The LightStone Biometric device was used to measure the physical levels of immersion which was intended to allow comparison with the self-reported data from the presence questionnaire.

A pretest to measure what the participant already knew about the Custom's Officer domain was performed. This data was correlated to the results to questionnaires performed after each task which sought to determine what the participant noticed, remembered, understood and whether learning had taken place which could be transferred to a similar situation. The same number of questions with similar difficulty levels were used so that we could measure and compare the responses.

Motion sickness has been associated with VR simulators. While the movement within our environment was minimal and there were no roller coasters, car rides or similar motion-based experiences, it is common practice to measure whether the environment has produced any sensation of motion sickness, since the CAVE results in the participant potentially experiencing a situation in which their various senses may send conflicting information. For example, your eyes might believe you are moving, but your stomach does not feel the accompanying appropriate body movements. Motion sickness may reduce the efficacy of the training as it may inhibit concentration and in extreme cases require the training to terminate prematurely. Kennedy et al's (1993) Simulator sickness questionnaire was used for this purpose.

1. Introduction - sign consent forms, told sequence of events and location - 5 mins
2. Immersive tendencies questionnaire 10 mins
3. Pretest for knowledge of the custom's domain
4. Attach Lightstone Device – perform Task 1 - 6 mins
5. Presence questionnaire 3 mins
6. Task 1 questionnaire 5 mins
7. Attach Lightstone Device – perform Task 2 6 mins
8. Presence questionnaire 3 mins
9. Task 2 questionnaire 5 mins
10. Attach Lightstone Device – perform Task 3 6 mins
11. Presence questionnaire 3 mins
12. Task 3 questionnaire 5 mins
13. Simulator sickness questionnaire 5 mins (only after the CAVE environment)
14. Realism and Training Questionnaire 7 mins

Figure 2. Experimental Procedure

. The order in which the motion sickness questionnaire was administered changed depending on when the CAVE was being experienced. This in turn depended on the group to which the participant had been assigned.

3. RESULTS

Participants included 20 males and 16 females aged in their early 20s. English was a second language for up to half of the participants. Since our goal is to build a training system we were interested to determine how much of the information provided was noticed and remembered. To test this, participants were required to answer some questions about the scenario they had just watched. Missed answers were given a score of 0 if the participant had answered most questions but only missed one or two. There was only one case in which a participant missed answering one side of the questionnaire sheet. For this participant each unanswered question was scored with the average of the other participants' scores for this question.

We ran a 2-way ANOVA with replication with factors viewing medium and scenario type on the post-scenario questionnaire scores to see if these factors had any effect on the scores (Table 1).

We found that the screen type had no significant effect on the scores and that there was no significant interaction between the screen type and the scenario. We did find, however, that the scenario did have a significant effect on the score. Post-hoc Tukey tests performed on the participants' scores for each scenario show that participants scored worse on the 'passport' scenario on average than they did on the 'drugs' or 'food' scenarios (HSD[1] = 0.06653; food - passport = 0.09365; drugs - passport = 0.06896; food - drugs = 0.02469). Comments left by the participants about the passport scenario indicate that there was some confusion over whether the passenger in the scenario was male or female.

[1] Tukey's HSD (Honestly Significant Differences) Test

Table 1. Two-way ANOVA run on post-scenario questionnaire scores by scenario and viewing media

Anova: Two-Factor With Replication

SUMMARY	Passport	food	drugs	Total
1screen				
Count	12	12	12	36
Sum	8.79	9.39	9.03	27.217
Average	0.73	0.78	0.75	0.768
Variance	0.0053	0.0087	0.0237	0.01195
3screens				
Count	12	12	12	36
Sum	8.2105	9.3056	8.733	26.249
Average	0.6842	0.7754	0.7278	0.72915
Variance	0.01057	0.01577	0.01411	0.01415
CAVE				
Count	12	12	12	36
Sum	7.68421	9.3611	9.4	26.445
Average	0.64035	0.78009	0.78333	0.73459
Variance	0.01704	0.01157	0.01950	0.01968
Total				
Count	36	36	36	
Sum	24.684	28.0556	27.1667	
Average	0.68567	0.77932	0.75463	
Variance	0.01179	0.01133	0.01823	

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	0.014	2	0.007	0.516	0.598	3.09
Columns	0.170	2	0.085	6.095	0.003	3.09
Interaction	0.055	4	0.014	0.996	0.413	2.46
Within	1.377	99	0.014			
Total	1.617	107				

Participants were also given a presence questionnaire (Witmer and Singer, 1998) after they viewed each scenario to measure how immersed and involved in the scenario the viewing medium made them feel.

Table 2. ANOVA calculated on presence scores for single screen, 3 screens, and CAVE viewing systems

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
1screen	36	527	14.63889	62.86587
3screen	36	520	14.44444	56.76825
cave	36	672	18.66667	46.74286

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	409.1	2	204.5	3.7	0.028	3.083
Within Groups	5823.2	105	55.46			
Total	6232.3	107				

An ANOVA performed on the data indicated a significant ($p < 0.05$) difference between the average presence score calculated for each screen (Table 2). The presence score for the CAVE system was shown to be significantly higher than the scores for both the one screen and three screen systems (Table 2). However, we were not able to find any significant correlation between the level of presence a participant experienced when viewing a scenario and their score on the post-scenario questionnaire ($R^2 =$ 0.000; Figure 3 and Table 3).

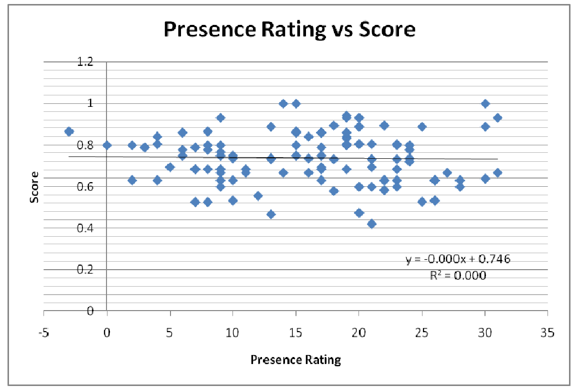

Figure 3. Scatter plot of presence rating by post-scenario questionnaire score

Table 3. Regression statistics for presence rating vs post-scenario questionnaire score

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.024379
R Square	0.000594
Adjusted R Square	-0.00883
Standard Error	0.12347
Observations	108

Before viewing the scenarios, participants were given an immersion questionnaire (also from (Witmer and Singer, 1998) to test whether they had any immersive tendencies. Immersion is stated in Witmer and Singer (1998) as:

…a psychological state characterized by perceiving oneself to be enveloped by, included in, and interacting with an environment that provides a continuous stream of stimuli and experiences.

It is also claimed in Witmer and Singer (1998) that a virtual environment that produces a strong sense of immersion will also produce a higher level of presence. The Immersion Tendencies Questionnaire (ITQ) is designed to measure how easily a person is able to become immersed in an activity or environment. We wished to test if there was a correlation between a participant's ITQ score and their presence score for each of the viewing media.

Table 4. Correlation coefficients for immersion rating versus presence score for each of the 3 viewing media

	cave	*1screen*	*3screen*
Immersion	0.268812	0.022281	0.210924

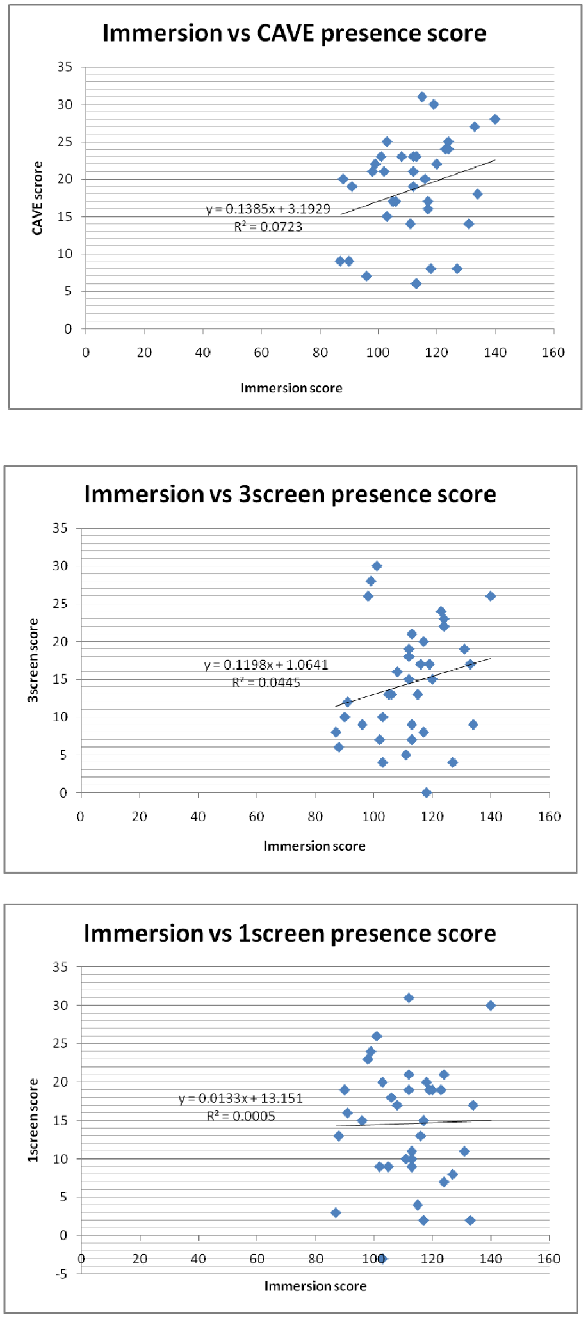

Figure 4. Scatter plots of immersion by presence score for each of the 3 viewing media

As can be observed from the scatter plots in Figure 4 and the correlation coefficients in Table 4, there appears to be little if any relation between the self-reported level of immersiveness in the participants' personalities and the presence scores provided by the participants for each of the viewing media. This could be explained by the fact that the participants were required to assess their own immersive tendencies. As can be observed in Table 5, none of the participants rated themselves higher than 113/198 or 74% immersiveness.

Table 5. Descriptive statistics for immersion rating

	Immersion
Mean	111.6944
Standard Error	2.211018
Median	112.5
Mode	113
Standard Deviation	13.26611
Sample Variance	175.9897
Kurtosis	-0.41365
Skewness	0.014725
Range	53
Minimum	87
Maximum	140
Sum	4021
Count	36

Descriptive statistics for the presence ratings for the 3 viewing media (Table 6) show a much wider distribution with the range being larger than the mean in all cases.

Participants were asked which viewing system they preferred across six categories. The categories were Training (which system they found best for training), Enjoyment (the system they found the most enjoyable), Realism (the system they found the most realistic), Immersion (the system they found the most immersive), Learning (the system that helped them learn the most), and which viewing system they preferred in general. The results are shown in Figure 5. As can be observed, the CAVE system scored significantly higher in almost all categories. It received an almost identical score to the single screen in the category of Learning (CAVE = 12; single screen = 11). Comments written by the participants in order to explain their scores indicated that some found that the single screen made it easier to concentrate on the scenario as this was the type of screen they were most used to seeing.

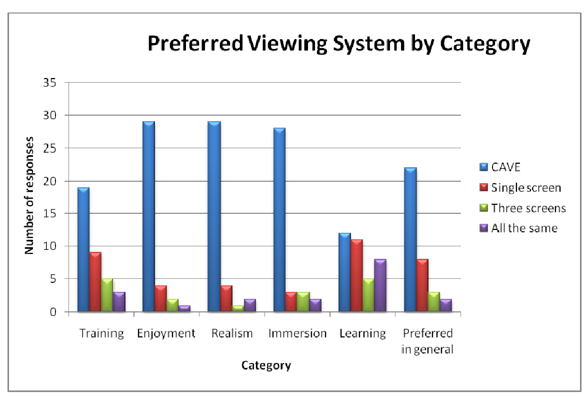

Figure 5. Preferred viewing system by category

Table 6: Descriptive statistics for presence rating for each of the three viewing media

1 screen Presence rating	
Mean	14.63889
Standard Error	1.321467
Median	15.5
Mode	19
Standard Deviation	7.9288
Sample Variance	62.86587
Kurtosis	-0.25299
Skewness	-0.10364
Range	34
Minimum	-3
Maximum	31
Sum	527
Count	36

CAVE Presence rating	
Mean	18.66667
Standard Error	1.139479
Median	20
Mode	23
Standard Deviation	6.836875
Sample Variance	46.74286
Kurtosis	-0.58653
Skewness	-0.39708
Range	25
Minimum	6
Maximum	31
Sum	672
Count	36

3 screen presence rating	
Mean	14.44444
Standard Error	1.255745
Median	14
Mode	13
Standard Deviation	7.534471
Sample Variance	56.76825
Kurtosis	-0.72496
Skewness	0.228572
Range	30
Minimum	0
Maximum	30
Sum	520
Count	36

4. DISCUSSION AND CONCLUSION

As part of our design we also collected biometric data. Our motivation was to avoid criticisms such as those voiced by Slater and Garau (2007) which points out the limitations of asking people via a survey instrument whether they feel "immersed" or a sense of presence/"being there". In particular, they criticize the use of Likert scale data. We also have reservations regarding the responses given by our participants as evidenced in the lack of correlation between their tendency to become immersed and their reporting of levels of immersion. As an alternative, Cox et al. 2006 proposed the use of eye movement data (the number of fixations, fixation duration and saccade length at 10 second intervals) to provide an objective way of measuring immersion. Yang, Marsh and Shahabi (2005) have developed a computer-based tool which captures "immersidata" from input devices to record commands and keystrokes synchronised with videotape recording of the player. We have performed some initial analysis of the biometric data collected via the Lightstone device but are currently still trying to find patterns and determine how best to interpret that data and relate it to the participants' reported sense of immersion.

Our study seeks to add to our understanding of the role of immersion on learning, specifically perception and memory and knowledge transfer (via a number of questions in our post survey). Understanding the role of immersion is difficult given that the concept of immersion is not well-defined or understood (Tijs, 2006), despite its popular use in the simulation and gaming industry. Jennett et al. (2008) define immersion as comprising: 1) flow - clear goals, high degree of concentration, a loss of the feeling of self-consciousness (sense of serenity), distorted sense of time, direct and immediate feedback, balance between ability level and challenge, sense of personal control, intrinsically rewarding (Csikszentmihalyi, 1990), 2) cognitive absorption - comprising temporal dissociation, attention focus, heightened enjoyment, control and curiosity (Agarwal and Karahana, 2000) and 3) presence - defined by Slater et al. (1994) as a psychological sense of being in a virtual environment affected by the factors of control, sensory, distraction, and realism. Adams (2004) has defined three different types of immersion: tactical (physical and immediate), strategic (cerebral and goal driven) and narrative (lost in the plot). While a particular simulation/game may predominately offer one of these types, depending on the individual and the experience they seek, multiple types of immersion could be provided by the same simulation.

Our presence questionnaire sought to determine the level of immersion and we did find the CAVE to provide a significantly greater sense/impression of immersion. However, we could not find any correlation between this perception and what was learnt/remembered by participants. This finding correlates to our past study into realism (Richards and Barles, 2005) and interactivity (Richards, 2006) which also found that while our participants preferred realism and interactivity, it did not improve (or degrade) their learning. In a study looking at the use of VR technology for acquiring spatial knowledge, Patrick et al. (2008), found "that the low-cost projection screen might be as effective as a headmounted display [HMD] for educational or training exercises involving spatial cognition". (p. 484). (Of course, they note the high cost of the projector system). Furthermore, in comparison to the HMD they found the screen to be quicker to set up, less invasive and uninviting, lower likelihood of motion sickness and offering the opportunity for multiple concurrent participants. In fact in our study we had two subjects conduct the CAVE task concurrently. Based on the results in our study and observation of (hundreds of) participants involved in various CAVE experimental studies in the past three years we note that participants consistently comment/rate the CAVE highly (we assume) due to the novelty of the environment. Participants will report favourably even in (uncommon) situations where, for example, we discover after the demonstration that the 3D stereo was inversed and couldn't have been providing depth information.

While not explored in this study, we conjecture that 1) learning through immersion could be more beneficial to particular types of lessons, like the spatial experiment mentioned below, or interactive lessons; 2) the learning gap between the three different types of media may be increased by manipulating factors such as the quality of the simulation, including audio quality, camera movements, appropriate animations, lighting, and so on. Perhaps the take home message is that learning can occur under many conditions which is largely unaffected by the media. If, however, the sense of enjoyment, engagement and/or immersion are seen to be enhanced by a particular media (e.g CAVE, HMD) or method (e.g. interactivity), then perhaps individuals will be motivated to participate in the learning/training experience without the need to provide carrots ($15 for participation in this study) or sticks (loss of marks or employment opportunities).

REFERENCES

1. Adams, E. (2004) "Postmodernism and the Three Types of Immersion". Accessed: 2-12-08. http://www.gamasutra.com/features/20040709/adams_01.shtml.
2. Agarwal, R. & Karahana, E. (2000). "Time flies when you're having fun: Cognitive Absorption and beliefs about information technology usage". *MIS Quarterly*, 24 (4), 665-694.
3. Cox, A.L., Cairns, P.A., Berthouze, N & Jennett, C. (2006) "The Use of Eyetracking for Measuring Immersion". *CogSci* 2006.
4. Csikszentmihalyi, M. (1990). Flow: The Psychology of Optimal Experience. New York: Harper and Row.
5. Csikszentmihalyi, M. and Csikszentmihalyi, I. (1988) Optimal Experience: Psychological Studies of Flow in Consciousness, Cambridge: Cambridge University Press.
6. Iowa Department of Transportation. (2007). Evaluation of Virtual Reality Snowplow Simulator Training: Final Report, IOWA State University and IOWA Dept of Transport accessed 3-12-08 http://www.ctre.iastate.edu/reports/snowplow_simulator_final.pdf
7. Jennett, C., Cox, A.L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., Walton, A. (2008). "Measuring and Defining the Experience of Immersion in Games". *Int. Journal of Human Computer Studies*, 66(9), 641-661
8. Kennedy, R.S., Lane, N.E. Berbaum, K.S. and Lilienthal, M. G. (1993) "Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness". *International Journal of Aviation Psychology*, 3(3):203-220.
9. Patrick, E., Cosgrove, D., Slavkovic, A., Rode, J. A., Verratti, T., and Chiselko, G. (2000). Using a large projection screen as an alternative to head-mounted displays for virtual environments. In Proc.of the SIGCHI Conference on Human Factors in Computing Systems (The Hague, The Netherlands, April 01 - 06, 2000). CHI '00. ACM, New York, NY, 478-485.
10. Richards, D. and Barles, J. (2005) Actors vs. Animation for Adult Learning ? Interactive Entertainment, IE'2005, Sydney Nov. 23-25, 2005.
11. Richards, D. (2006) Is Interactivity Actually Important? Proc.3rd Aust. Conf. on Interactive Entertainment (IE'2006), 4-6 Dec. 2006, Perth, 59-66.
12. Slater, M., Linakis, V., Usoh, M. & Kooper, R. (1996) Immersion, presence, and performance in virtual environments: An experiment using tri-dimensional chess. Paper presented at the 1996 Virtual Reality and Software and Technology Conference, Hong Kong.
13. Tijs, T.J.W. (2006) Immersion in Games: An Eyetracking Study. Master's thesis. Department of Computer and Systems Science, Royal Institute of Technology, Stockholm.
14. Witmer, B. G., and M. J. Singer. (1998). "Measuring presence in virtual environments: A presence questionnaire". *Presence* 7:225–240.
15. Yang, K., Marsh, T. & Shahabi, C. (2005) Continuous archival and analysis of user data in virtual and immersive game environments. Proceedings of the 2nd ACM workshop on Continuous archival and retrieval of personal experiences, Singapore. pp. 13-22.

APPENDIX A: PRESENCE QUESTIONNAIRE

Post experiment questionnaire

1. Please estimate in minutes/seconds how long the demonstration took.__min___sec
2. I was thinking of something other than the demonstration
3. I was distracted by something other than the demonstration
4. How completely were all of your senses engaged?
5. How much did the visual aspects of the environment involve you?
6. How much did the auditory aspects of the environment involve you?
7. How aware were you of events occurring in the real world around you?
8. How aware were you of your display?
9. How inconsistent or disconnected was the information coming from your various senses?
10. How much did your experiences in the virtual environment seem consistent with your realworld experiences?
11. How compelling was your sense of moving around inside the virtual environment?
12. To what degree did you feel confused or disoriented at the beginning of breaks or at the end of the experimental session?
13. How involved were you in the virtual environment experience?
14. How quickly did you adjust to the virtual environment experience?
15. How much did the visual display quality interfere or distract you from the training experience?
16. Were you involved in the experimental task to the extent that you lost track of time?